A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion

Jian Sun   Maks Ovsjanikov   Leonidas Guibas

Stanford University

Abstract

We propose a novel point signature based on the properties of the heat diffusion process on a shape. Our signature, called the Heat Kernel Signature (or HKS), is obtained by restricting the well-known heat kernel to the temporal domain. Remarkably we show that under certain mild assumptions, HKS captures all of the information contained in the heat kernel, and characterizes the shape up to isometry. This means that the restriction to the temporal domain, on the one hand, makes HKS much more concise and easily commensurable, while on the other hand, it preserves all of the information about the intrinsic geometry of the shape. In addition, HKS inherits many useful properties from the heat kernel, which means, in particular, that it is stable under perturbations of the shape. Our signature also provides a natural and efficiently computable multi-scale way to capture information about neighborhoods of a given point, which can be extremely useful in many applications. To demonstrate the practical relevance of our signature, we present several methods for non-rigid multi-scale matching based on the HKS and use it to detect repeated structure within the same shape and across a collection of shapes.

1. Introduction

A geometric shape is often given by its bounding surface, whose discrete representation in the computer is a mesh, or sometimes a point set. Although such representations are convenient in applications such as rendering and visualization, they are not suitable, at least in a direct way, for many others including full or partial shape comparison, structure (e.g. symmetry) detection, partial matching, shape classification and retrieval, to name just a few. In these applications, shapes or parts of a shape are considered to be similar if there exist rigid or isometric transformations between them. Thus, it is desirable to derive shape signatures that are invariant under such transformations to facilitate comparison and differentiation between shapes or parts of a shape. The properties of the signatures are of great importance to the success of efficient and effective comparison and by extension the various applications that depend on them.

A large amount of research has been done in developing signatures to facilitate various tasks in computer vision, geometry processing and data analysis. However, most of the signatures are presented without a rigorous explanation of their properties and their effectiveness is often illustrated only in specific applications through examples. In this paper, we propose a novel point signature, and argue from both *theoretical* and *practical* points of view that it has the following desirable properties:
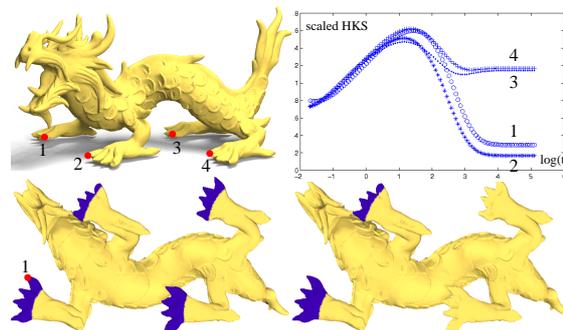
Figure 1: *top left: dragon model; top right: scaled HKS at points* 1*,* 2*,* 3 *and* 4*. all four signatures are close at small t's while big t's separate the points on the front claws from those on back; bottom left: the points (blue), whose signature is close to the signature of point* 1 *based on the smaller half of the t's; bottom right: based on the entire range of t's.*

- It organizes information about the intrinsic geometry of a shape in an efficient, multi-scale way.
- It is stable under perturbations of the shape.
- It is concise and commensurable, but remains informative.
- It can be estimated faithfully and efficiently.

Besides, our signature is invariant under isometric deformations and can therefore be used in applications that involve deformable shapes.

Our signature captures information about the neighborhood of a point on a shape by recording the dissipation of heat from the point onto the rest of the shape over time. Because heat diffuses to progressively larger neighborhoods, the time parameter provides a natural notion of scale to describe the shape around a point. This means, in particular, that the detailed, highly local shape features are observed through the behavior of heat diffusion over short time, while the summaries of the shape in large neighborhoods are observed through the behavior of heat diffusion over longer time. This property of heat diffusion enables us to perform multi-scale matching between points by comparing their signatures at different time intervals, as shown in Figure 1.

It is well-known that heat diffusion on the surface of a shape is fully described by the heat kernel, associated with the Laplace-Beltrami operator. Rather than using the heat kernel itself as a point signature, we define our Heat Kernel Signature (or HKS) by considering its restriction to the temporal domain. HKS inherits many nice properties from the heat kernel, such as being intrinsic and stable against perturbations of the shape. More remarkably, we prove that under certain mild assumptions, the set of Heat Kernel Signatures at all points on the shape, fully characterizes the shape up to isometry. This means that our Heat Kernel Signatures are concise and easily commensurable since they are defined over the common temporal domain, but at the same time, preserve almost all of the information contained in the heat kernel. We also show that HKS can be computed faithfully and efficiently from the discrete representation of the shape.

The Heat Kernel Signature has the potential to benefit many applications, including robust discovery of correspondences, shape registration, and partial matching, especially in the context of deformable shapes. In this paper, we present several applications ranging from multi-scale point comparison to intrinsic symmetries and detection of repeated structure across a collection of shapes.

1.1. Related work

A large amount of work has been done on designing various local point signatures in the context of shape analysis. One common strategy to derive point signatures is to summarize the shape distribution in a neighborhood of a point under investigation [Joh97, CJ97, BMP00]. Spin images [Joh97] and shape context [BMP00] are two widely used point signatures that fall into this category. Both spin images and shape context are invariant under rigid transformations. Hilaga et al. [HSKK01] and Gal et al. [GSCO07] extend the shape context to the non-rigid setting, by using geodesic distances. However signatures based on geodesic distances are very sensitive to topological noise, since addition or removal of a small connection creating a handle can change geodesic distances dramatically over a large portion of the shape. Moreover, signatures for different neighborhoods are computed and stored independently, which can be both time and memory consuming.

Li and Guskov [LG05] define multi-scale local surface signatures by first obtaining a series of increasingly smoothed versions of a given shape and then constructing point signatures for features found at each smoothed version of the shape. Another approach to obtain a multi-scale signature is by convolving a function of a certain geometric property, such as the indicator of the interior of the domain, with a series of kernel functions like a Gaussian of increasing width. Integral invariant signatures proposed by Manay et al. [MYHS04] for domains in 2D, that have been used for global shape matching in [GMGP05] and analyzed extensively by Pottmann et al. [PWHY09], fall into this category. Our Heat Kernel Signature can also be considered within this class of signatures, but is significantly different in that no explicit integration is needed, which reduces the computation and storage costs.

Another related point signature is the Global Point Signature (GPS) proposed by Rustamov [Rus07]. For a fixed point x, $GPS(x)$ is a vector whose components are scaled eigenfunctions of the Laplace-Beltrami operator evaluated at x. GPS is invariant under isometric deformations of the shape, yet does not use geodesic distances explicitly. Ovsjanikov et al. [OSG08] develop an algorithm to detect global intrinsic symmetries based on GPS by observing that the intrinsic symmetries of a shape become extrinsic or Euclidean in the signature space. Unfortunately, GPS is a global signature and cannot be used to detect partial symmetries or to perform partial matching. Unlike the GPS, our signature allows to perform multi-scale comparison between neighborhoods of points on the same shape, or even across different shapes. In Section 3 we detail the differences between our signature and the GPS, and in Section 4 we show that our signature is more stable against noise in practice.

Our Heat Kernel Signature is also closely related to diffusion maps and diffusion distances proposed by Lafon [Laf04] for data representation and dimensionality reduction, and later used by de Goes et al. [dGGV08] for shape segmentation and by Bronstein et al. [BBK*09] for shape matching. In Section 3, we reveal the relation of our signature to the eccentricity based on diffusion distance.

Finally, point signatures are often amalgamated into concise global signatures that can be used for shape indexing and shape retrieval. For example, the well-known shape distribution [OFCD02] can be viewed as the sum of the shape contexts over all sample points. In [GSCO07], Gal et al. employ a similar strategy where the histogram of intrinsic eccentricity and local diameter over all points are used to perform pose-oblivious shape indexing. Thus, the properties of point signatures are inherited by the global shape signatures.

2. Heat Operator and Heat Kernel

In this section we introduce the basic facts about heat diffusion on Riemannian manifolds that are necessary to define

our Heat Kernel Signature. Let M be a compact Riemannian manifold possibly with boundary. The heat diffusion process over M is governed by the heat equation

$$\Delta_M u(x,t) = -\frac{\partial u(x,t)}{\partial t}, \tag{1}$$

where Δ_M is the Laplace-Beltrami operator of M. If M has boundaries, we additionally require u to satisfy the Dirichlet boundary condition $u(x,t) = 0$ for all $x \in \partial M$ and all t. Given an initial heat distribution $f : M \to \mathbb{R}$, let $H_t(f)$ denote the heat distribution at time t, namely $H_t(f)$ satisfies the heat equation for all t, and $\lim_{t \to 0} H_t(f) = f$. H_t is called the *heat operator*. Both Δ_M and H_t are operators that map one real-valued function defined on M to another such function. It is easy to verify that they satisfy the following relation $H_t = e^{-t\Delta_M}$. Thus both operators share the same eigenfunctions and if λ is an eigenvalue of Δ_M, then $e^{-\lambda t}$ is an eigenvalue of H_t corresponding to the same eigenfunction.

It is well-known (see e.g. [Hsu02]) that for any M, there exists a function $k_t(x,y) : \mathbb{R}^+ \times M \times M \to \mathbb{R}$ such that

$$H_t f(x) = \int_M k_t(x,y)f(y)dy, \tag{2}$$

where dy is the volume form at $y \in M$. The minimum function $k_t(x,y)$ that satisfies Eq. 2), is called the *heat kernel*, and can be thought of as the amount of heat that is transferred from x to y in time t given a unit heat source at x. In other words $k_t(x,\cdot) = H_t(\delta_x)$ where δ_x is the Dirac delta function at x: $\delta_x(z) = 0$ for any $z \neq x$, and $\int_M \delta_x(z)dz = 1$. For compact M, the heat kernel has the following eigen-decomposition:

$$k_t(x,y) = \sum_{i=0}^{\infty} e^{-\lambda_i t}\phi_i(x)\phi_i(y), \tag{3}$$

where λ_i and ϕ_i are the ith eigenvalue and the ith eigenfunction of the Laplace-Beltrami operator, respectively.

The heat kernel $k_t(x,y)$ has many nice properties. For instance, it is symmetric: $k_t(x,y) = k_t(y,x)$ and satisfies the semigroup identity: $k_{t+s}(x,y) = \int_M k_t(x,z)k_s(y,z)dz$. The other properties more relevant to this work are that heat kernel is an *isometric invariant*, it is *informative*, *multi-scale*, and *stable*. Below we describe each of these properties.

The heat kernel can also be interpreted as the transition density function of Brownian motion on the manifold. This means that for any Borel subset $C \subseteq M$, $\int_C k_t(x,y)dy = \mathbb{P}(X_t^x \in C)$, the probability that the Brownian motion starting at a point x will be in C at time t. Brownian motion $X_t^x : \mathbb{R}^+ \to M$ is the most basic continuous time Markov process on M and many of the following properties have nice interpretations in the probabilistic setting. However, due to lack of space, we do not pursue this direction, and refer the interested reader to an excellent book by E. Hsu [Hsu02].

Intrinsic. By intrinsic, we mean the heat kernel is an isometric invariant, as stated in the following proposition.

Proposition 1 (Intrinsic Property) *If* $T : M \to N$ *is an isometry between two Riemannian manifolds M and N, then* $k_t^M(x,y) = k_t^N(T(x),T(y))$ *for any $x,y \in M$ and any $t > 0$.*

The invariance of the heat kernel under isometric deformations is a direct consequence of the invariance of the Laplace-Beltrami operator, which implies that the heat equation only involves intrinsic properties of the manifold. One practical implication of this property is that the heat kernel can be used to analyze shapes undergoing isometric deformations, which is useful, for example, in matching articulated shapes, such as humans or animals in different poses.

Informative. In addition to its isometric invariance, the heat kernel contains all of the information about the intrinsic geometry of the shape and hence fully characterizes shapes up to isometry, as stated in the following proposition.

Proposition 2 (Informative Property) *Let* $T : M \to N$ *be a surjective map between two Riemannian manifolds. If* $k_t^N(T(x),T(y)) = k_t^M(x,y)$ *for any $x,y \in M$ and any $t > 0$, then T is an isometry.*

This proposition is a simple consequence of the following equation (see e.g. [Gri06]). For any x,y on a manifold,

$$\lim_{t \to 0} t \log k_t(x,y) = -\frac{1}{4}d^2(x,y)$$

where $d(x,y)$ is the geodesic distance between x and y on M. As we will see in Section 3, however, a lot of the information in the heat kernel is redundant, and we can obtain a concise and informative signature by only considering its restriction to the temporal domain.

Multi-Scale. The multi-scale property of the heat kernel, in the context of point signatures, means that for small values of t, the function $k_t(x,\cdot)$ is mainly determined by small neighborhoods of x, and these neighborhoods grow bigger as t increases. This implies, in particular, that for small t, $k_t(x,\cdot)$ only reflects local properties of the shape around x, while for large values of t, $k_t(x,\cdot)$ captures the global structure of M from the point of view of x.

To formalize the above intuition, let us consider the heat diffusion process on a subset of M. Let D be a smooth, relatively compact domain on M and H_t^D be the heat operator associated with D with the Dirichlet boundary conditions as described above. Obviously, the heat kernel $k_t^D(x,y)$ is a local quantity that only depends on the domain D. The following proposition shows that $k_t^D(x,y)$ is a good approximation of $k_t(x,y)$ either in the case that D is arbitrary small as long as t is small enough or in the case that t is arbitrary large as long as D is big enough.

Proposition 3 (Multi-Scale Property [Gri06]) *(i)   For any smooth and relatively compact domain $D \subseteq M$,* $\lim_{t \to 0} k_t^D(x,y) = k_t(x,y)$. *(ii) For any $t \in \mathbb{R}^+$, and any $x,y \in D_1$, the Dirichlet heat kernel $k_t^{D_1}(x,y) \leq k_t^{D_2}(x,y)$ if $D_1 \subseteq D_2$. Moreover, if $\{D_n\}$ is an expanding and exhausting sequence, i.e., $\cup_{n=1}^{\infty} D_n = M$ and $D_{n-1} \subset D_n$, then* $\lim_{n \to \infty} k_t^{D_n}(x,y) = k_t(x,y)$ *for any t.*
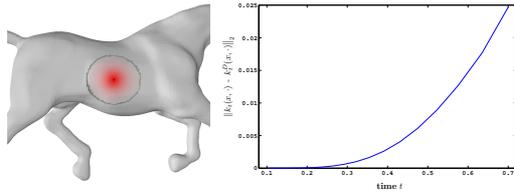
(a) $k_t^D(x,\cdot)$ for $t = 0.3$ (b) $\|k_t^D(x,\cdot) - k_t(x,\cdot)\|_2$ for various t

Figure 2: *(a) Heat kernel $k_t^D(x,\cdot)$ colored from grey to bright red determined by a neighborhood D, whose boundary is shown in dark grey. (b) The L_2 difference $\|k_t(x,\cdot) - k_t^D(x,\cdot)\|_2$ between the heat kernel computed using D, and the one computed on the entire shape for various t.*

To illustrate this multi-scale property, in Figure 2(b) we plot the L_2 difference $\|k_t(x,\cdot) - k_t^D(x,\cdot)\|_2$, where x is a point on the horse model and D is the region whose boundary is shown in dark grey in Figure 2(a). Figure 2(a) also shows the heat kernel $k_t^D(x,\cdot)$ for $t = 0.3$, around which the L_2 error begins to grow. Note that for this value of t, $k_t^D(x,y)$ at points y close to the boundary of D is no longer negligible. The heat kernels shown in Figure 2 are estimated from the mesh based on the method described in Section 4.

Stable. One of the particularly useful properties of the heat kernel, inherited by the signature that we define in Section 3, is that it is stable under perturbations of the underlying manifold. Many deformable models in practice are not isometric, and a successful point signature should not be sensitive to small perturbations. An intuitive reason for the stability of the heat kernel comes from its interpretation as the transition probability of the Brownian motion on the manifold. Intuitively, this means that $k_t(x,y)$ is a weighted average over *all* paths between x and y possible in time t, which should not be greatly affected by local perturbations of the surface. More concretely, if X_t^x is the Brownian motion on a domain M, and we perturb a subset $P \subset M$ (for example by introducing a tunnel between a small set of points in M), then only the paths passing through P will be affected. This intuition can be formalized by conditioning on the event that X_t^x passes through P.

Pointwise convergence of the heat kernel $\mathbb{R}^d$ under perturbations of the diffusion process was shown by Chen et al. [CZHZ98], who prove the stability of the heat kernel under changes of the Laplace operator. They show that if L and $\tilde{L}$ are two uniformly elliptic operators in $\mathbb{R}^d$ given as $L = \nabla \cdot (A(x)\nabla)$ and $\tilde{L} = \nabla \cdot (\tilde{A}(x)\nabla)$, with the corresponding heat kernels $k_t(x,y)$ and $\tilde{k}_t(x,y)$, then $\left|k_t(x,y) - \tilde{k}_t(x,y)\right|$ can be bounded by a function of the difference between A and $\tilde{A}$. In the case of the Laplace operator, A is the identity matrix $I_{d \times d}$ and $\tilde{A}(x) \in \mathbb{R}^{d \times d}$ models the perturbations of the diffusion process. The heat equation corresponding to $\tilde{L}$ becomes: $\frac{\partial u}{\partial t} = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(\tilde{A}_{ij}(x)\frac{\partial p}{\partial x_j}\right)$. In particular, their result shows that the heat kernels converge pointwise whenever the corresponding diffusion matrices A and $\tilde{A}$ converge. These results can also be extended to general Riemannian

manifolds [†]. In Section 4, we also show the stability of the heat kernel in the discrete setting.

3. Heat Kernel Signature (HKS)

In Section 2, we showed that the heat kernel is intrinsic and stable against noise. Moreover, the heat kernel provides a natural multi-scale characterization of the neighborhoods of a given point. These properties make the heat kernel a very lucrative candidate for a point signature. Specifically, one can take the family of functions $\{k_t(x,\cdot)\}_{t>0}$ parametrized by t as a signature for each point $x \in M$. Unfortunately, the complexity of this signature would be extremely high, since for a single point signature $\{k_t(x,\cdot)\}_{t>0}$ is a function defined on $\mathbb{R}^+ \times M$, the product of temporal and spatial domains. Even worse, it would be difficult to compare the signatures of two different points. To make sense of the comparison, one would have to align two signature functions, say $k_t(x,\cdot)$ and $k_t(x',\cdot)$, which would require a map from the neighborhood of x to that of x'.

As noted above, the heat kernel contains a large amount of redundant information. This is because the heat diffusion process is governed by the heat equation $\Delta_M u(x,t) = -\frac{\partial u(x,t)}{\partial t}$, which implies that the change of the signature function in the spatial domain is manifested by its change over time. Thus our strategy to overcome the above difficulties is to consider the restriction of the heat kernel to a subset of $\mathbb{R}^+ \times M$, while keeping as much information as possible.

Given a point x on the manifold M, we define its *Heat Kernel Signature*, $HKS(x)$ to be a function over the temporal domain:

$$HKS(x) : \mathbb{R}^+ \to \mathbb{R}, HKS(x,t) = k_t(x,x). \quad (4)$$

One of the main results of this paper is the following *Informative Theorem*, which says that, despite restricting the signature to $\mathbb{R}^+ \times \{x\}$ and dropping the entire spatial domain, under mild assumptions, $\{k_t(x,x)\}_{t>0}$ keeps all of the information of $\{k_t(x,\cdot)\}_{t>0}$.

Theorem 1 (Informative Theorem) *If the eigenvalues of the Laplace-Beltrami operators of two compact manifolds M and N are not repeated, and T is a homeomorphism from M to N, then T is isometric if and only if $k_t^M(x,x) = k_t^N(T(x),T(x))$ for any $x \in M$ and any $t > 0$.*

Intuitively, if the eigenvalues are not repeated, one can deduce both the eigenvalues and the squared values of the eigenfunctions at a point x, by analyzing the behavior of its HKS, $k_t(x,x) = \sum_{i=0}^{\infty} e^{-\lambda_i t}\phi_i(x)^2$ as a function of t. Then using a nice property of the eigenfunctions, which says that the negative and the positive nodal domains are interleaved,

---

[†] Personal communication with Dr. W. Zheng (a coauthor of [CZHZ98])
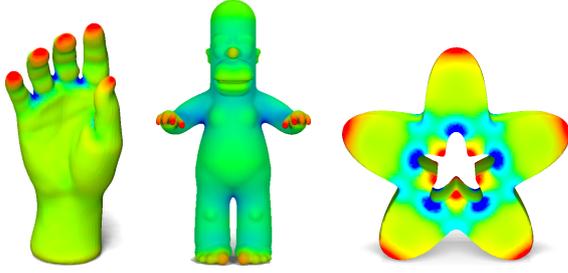
Figure 3: *Heat kernel function $k_t(x,x)$ for a small fixed t on the hand, Homer, and trim-star models. The function values increase as the color goes from blue to green and to red, with the mapping consistent across the shapes. Note that high and low values of $k_t(x,x)$ correspond to areas with positive and negative Gaussian curvatures respectively.*

it is easy to show that the eigenfunctions are uniquely determined up to a sign, and so is $k_t(x,y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(y)$, which proves the theorem by Proposition 2, see Appendix for a formal proof of this Theorem.

Informative Theorem assures that the set of Heat Kernel Signatures is almost as informative as the family of functions $\{k_t(x,\cdot)\}_{t>0}$ which is much more complex. Most notably, the Heat Kernel Signatures at different points are defined over a common temporal domain, which makes them easily commensurable. Moreover, HKS inherits many of the properties of the heat kernel, illustrated in Section 2, including being intrinsic, multi-scale and robust. By mimicking the heat diffusion process on a shape, the geometric information about the neighborhoods of a point x at different scales is compactly encoded in the HKS $\{k_t(x,x)\}_{t>0}$. Thus, our signature encodes geometric information about the shape as a set of functions over the temporal domain, which are not only isometrically invariant but also inexpensive to store and easy to compare.

One hypothesis of the Informative Theorem is that the Laplace-Beltrami operator does not have repeating eigenvalues. In general, the Theorem does not hold if this hypothesis is violated. A simple counterexample is a sphere, where $k_t(x,x)$ is the same for all x. However, it is clearly possible to construct non-isometric maps between two spheres of the same radius. Similarly, it is possible to show that there are non-isometric maps between congruent tori, that preserve the Heat Kernel Signatures at all points. These examples, however, are degenerate, in that they possess infinitesimal isometric self-mappings. It would be interesting to see if the theorem holds for shapes with a discrete set of intrinsic symmetries, such as the Star model shown in Figure 3, which we leave for future work.

Relation to curvature. In addition to the Informative Theorem above, which is rather global in nature, the following expansion shows that for small t, the Heat Kernel Signature of a point x is directly related to the scalar curvature $s(x)$ (Gaussian curvature on a surface) at x (e.g. [MP49]):

$$k_t(x,x) = (4\pi t)^{-d/2} \sum_{i=0}^{\infty} a_i t^i,$$

where $a_0 = 1$ and $a_1 = \frac{1}{6}s(x)$. This expansion corresponds to the well-known property of the heat diffusion process, which states that heat tends to diffuse slower at points with positive curvature, and faster at points with negative curvature. Figure 3 plots the values of $k_t(x,x)$ for a fixed small t on three shapes, where the colors are consistent across the shapes. Note that the values of this function are large in highly curved areas, and small in negatively curved areas. Note that even for the trim-star, which has sharp edges, $k_t(x,x)$ provides a meaningful notion of curvature at all points. For this reason, the function $k_t(x,x)$ can be interpreted as the intrinsic curvature at x at scale t. We use this intuition to detect salient features on the shape in Section 5.1.

Relation to diffusion distance. The Heat Kernel Signature is also closely related to diffusion maps and diffusion distances proposed by Lafon [Laf04] for data representation and dimensionality reduction. The diffusion distance between $x, y \in M$ at time scale t is defined as

$$\begin{aligned} d_t^2(x,y) &= k_t(x,x) + k_t(y,y) - 2k_t(x,y) \\ &= \sum_i e^{-\lambda_i t} (\phi_i(x) - \phi_i(y))^2. \end{aligned}$$

The eccentricity of x in terms of diffusion distance, denoted $ecc_t(x)$, is defined as the average squared diffusion distance over the entire manifold (called ADD in [dGGV08]):

$$ecc_t(x) = \frac{1}{A_M} \int_M d_t^2(x,y) dy = k_t(x,x) + H_M(t) - \frac{2}{A_M},$$

where A_M is the surface area of M, and $H_M(t) = \sum_i e^{-\lambda_i t}$ is the heat trace of M. Since both $H_M(t)$ and $\frac{2}{A_M}$ are independent of x, if we consider both $ecc_t(x)$ and $k_t(x,x)$ as functions over M, their level sets, in particular extrema points, coincide. Thus, for small t, we expect the extremal points of $ecc_t(x)$ to be located at the highly curved areas.

Relation to GPS As mentioned in the introduction, the HKS is also related to the Global Point Signature (GPS) [Rus07], defined as: $GPS(x) = \left(\frac{\phi_1(x)}{\sqrt{\lambda_1}}, \frac{\phi_2(x)}{\sqrt{\lambda_2}}, \cdots, \frac{\phi_i(x)}{\sqrt{\lambda_i}}, \cdots \right)$. Note that GPS is not easily comparable for points across different shapes because the eigenfunctions are only defined up to a change in sign (or, more generally, orthonormal transformations in the eigenspace). Furthermore, when eigenvalues are close to each other, perturbation theory shows that the corresponding eigenfunctions may switch [GVL96]. Our signature can be viewed as a weighted sum of the squares of eigenfunctions and thus is not sensitive to the order of the eigenfunctions or their sign as shown in Proposition 4 below. Most importantly, the HKS of points on different shapes are commensurable, which allows to use our signature naturally for multi-scale matching.

4. Discrete Setting

In most applications, the underlying manifold is not known. Instead, we are often given an approximation of the the shape

represented by a mesh. In this section, we describe how to estimate HKS from a mesh and show its stability against noise in the discrete setting with an explicit error bound.

In the d-dimensional Euclidean space $\mathbb{R}^d$, the heat kernel has an explicit expression which is simply a Gaussian $k_t(x,y) = (4\pi t)^{-d/2} e^{-\frac{\|x-y\|^2}{4t}}$. For a general compact manifold, the heat kernel does not have an explicit expression, but one can compute the heat kernel via the corresponding Laplace-Beltrami operator. Many schemes have been proposed to estimate the Laplace-Beltrami operator from discrete meshes [PP93, MDSB02, RWP06, BSW08]. Perhaps, the most commonly used method in the computer graphics community is the so-called cotangent scheme. However, it has been shown that the cotangent scheme does not converge in general [Xu04, War05]. Recently, Belkin et al. [BSW08] proposed a convergent scheme called mesh Laplace operator. Unlike the FEM methods, which include the cotangent scheme, the convergence of their operator does not require the triangles in the mesh to be well-shaped. For this reason we use the mesh Laplace operator to estimate the Laplace-Beltrami operator.

Given a mesh with n vertices, the mesh Laplace operator L, is a sparse matrix of size $n \times n$, which can be written as $L = A^{-1}W$ where A is a positive diagonal matrix, whose element $A(i,i)$ represents the area associated with vertex i, and W is a symmetric semi-definite matrix. Under these conditions the generalized eigenproblem $W\phi = \lambda A\phi$ has a full set of real eigenvalues and eigenvectors, which means that $L = \Phi\Lambda\Phi^T A$, where Λ is a diagonal matrix of eigenvalues, and Φ is a matrix whose columns correspond to the right eigenvectors of L.

To interpret the heat operator and the heat kernel in the discrete setting, we re-write the heat equation on a mesh that approximates the underlying manifold. If u_t is a time-dependent function defined on the vertices, such that $u_t(x)$ is the amount of heat at vertex x at time t, then the equivalent to Equation 1 is: $Lu_t = \frac{\partial u_t}{\partial t}$. It is well known that solutions to this equation have the form: $u_t = e^{-tL}u_0$ where u_0 is an arbitrary vector representing the initial distribution of heat, and e^{-tL} is a matrix exponential:

$$e^{-tL} = \sum_{i=0}^{\infty} \frac{(-tL)^i}{i!}. \tag{5}$$

When $L = \Phi\Lambda\Phi^T A$, as above, $e^{-tL} = \Phi e^{-t\Lambda}\Phi^T A = K_t A$ Similarly to the continuous case, we interpret the matrix e^{-tL} as the heat operator, and the elements (x,y) of the matrix K_t as the values of the heat kernel $k_t(x,y)$. Then, for a given initial vector f, the heat distribution at time t is given as $u_t = e^{-tL}f$, or $u_t(x) = \sum_y k_t(x,y)f(y)A(y)$, which is the discrete version of Eq. (2). Note that because HKS is a restriction of the heat kernel instead of the heat operator, it is insensitive to the meshing and thus commensurable for points with different area weights.

Figure 4: *Absolute values of the 4th eigenfunction of the Laplace-Beltrami operator of the dragon model (left) and of the 8th eigenfunction of the armadillo model (right), both corresponding to non-repeating eigenvalues. The absolute values of these functions increase as the color changes from blue to red. Note that these values are not preserved under the bilateral symmetries.*

For each pair of vertices x, y, the corresponding entry $K_t(x,y) = \sum_{i=1}^{n} e^{-\lambda_i t}\phi_i(x)\phi_i(y)$, which coincides with the eigen-decomposition of the heat kernel in the smooth case (see Eq. (3)). This expression allows us to estimate the heat kernel through the eigenvalues and the eigenvectors of the Laplace operator of the mesh. In practice, we use the sparse eigensolver implemented in Matlab to compute the eigenvalues and the eigenvectors of L.

Stability In Section 2 we have shown that heat kernel is stable in the smooth case. In the following proposition, we show the estimated heat kernel based on the matrix exponential is also stable against noise, see Appendix for its proof.

Proposition 4 *Assume $L = A^{-1}W$ and $\tilde{L} = \tilde{A}^{-1}\tilde{W}$ with $\tilde{A}^{-1} = A + E$ and $\tilde{W} = W + F$ where $\|E\| < \varepsilon$ and $\|F\| < \delta$. Then if K_t and $\tilde{K}_t$ are the induced heat kernels from L and $\tilde{L}$ respectively, we have $\|K_t - \tilde{K}_t\| = O(\sqrt{\varepsilon} + \delta)$. Here $\|\cdot\|$ is any matrix norm induced from vector norm or its equivalence.*

Note that unlike the stability of the eigenvector computations used in [Rus07] and [OSG08], which depends on the spectral gap, the stability of the HKS only depends on the perturbation itself, which is easier to control. Figure 4 illustrates the absolute values of the eigenvectors ϕ on two shapes that have a bilateral symmetry T, but $|\phi(T(x))| \neq |\phi(x)|$, as required by [OSG08], although both eigenfunctions correspond to non-repeating eigenvalues. Our signature is more stable and can be used to detect the symmetries in these cases, as shown in Section 5.1. Furthermore it provides a multi-scale method to distinguish between purely local and global effects of shape distortion.

5. Multi-Scale Matching

We have shown that HKS $\{k_t(x,x)\}_{t>0}$ of a point $x \in M$ encodes the information about its neighborhoods in a multi-scale way. This enables us to perform multi-scale matching by comparing HKS's corresponding to different time intervals. Specifically, in order to see if two points x and x' are matched at the given scales specified by the time interval $[t_1, t_2]$, we can compute the difference between two HKS's:

$(\int_{t_1}^{t_2} |k_t(x,x) - k_t(x',x')|^2 dt)^{1/2}$ and consider x and x' to be matched at a given range of scales if the difference is zero. However, in practice, we are interested in matching approximately and thus must be careful on how the difference approaches 0. Since $k_t(x,x) = \sum_i e^{-\lambda_i t} \phi_i^2(x)$ decays exponentially as t increases, we need a more appropriate strategy of computing the difference between two signatures.

We introduce two heuristics, which work very well in our experiments. First, observe that the difference $|k_t(x,x) - k_t(x',x')|$ decreases exponentially as t increases, which makes the difference at large scales negligible compared to those at small scales. To address this issue, we scale each $k_t(x,x)$ by $\int_M k_t(x,x)dx$, which normalizes the integral of the scaled $k_t(x,x)$ over M to be one for all t's. This normalization ensures that the differences between two signatures at different time scales contribute approximately equally. Note that $\int_M k_t(x,x)dx$ is also called the heat trace at time t and can be computed as $\sum_i e^{-\lambda_i t}$.

In addition, we observe that for a fixed point x, the variation of the function $k_t(x,x)$ is large for small t's but decays as t increases. Intuitively, this is because $k_t(x,x)$ is determined by the average behavior of heat dissipation in the neighborhood determined by the scale t. The average behavior in a small neighborhood is easily affected as the scale changes, while average behavior in a large neighborhood is much more stable. Thus in order to efficiently represent HKS, we scale the temporal domain logarithmically, which gives a more faithful approximation to HKS at small scales, when the signature changes more rapidly.

| Model	#vertices	MATRIX	EIGEN	HKS
Dancer	9.97k	52	117	0.38
Armadillo	15k	137	183	0.56
Octopus	66.2k	3236	987	2.38
Dragon	100k	3512	2015	3.65

Table 1: *Timing for computing HKS (on a machine with 2.4GHz CPU and 16G RAM). The 3rd, 4th and 5th column show the timing (in seconds) for constructing Laplace matrix (based on geodesic distances), solving eigenproblem and computing HKS at all points.*

With these observations in mind, we define the difference between two Heat Kernel Signatures at the scale $[t_1, t_2]$ as

$$d_{[t_1,t_2]}(x,x') = \left(\int_{t_1}^{t_2} \left(\frac{|k_t(x,x) - k_t(x',x')|}{\int_M k_t(x,x)dx} \right)^2 d\log t \right)^{1/2}. \quad (6)$$

In practice, we sample HKS uniformly over the logarithmic scaled temporal domain and obtain a vector to represent the HKS for each point. Then we estimate the integral in Eq. (6) by computing the L_2-norm of the difference between two corresponding vectors.

Table 1 shows the timing to compute the Heat Kernel Signature for some models shown in the paper. For each model, we compute 300 eigenvalues and eigenvectors and compute the HKS by uniformly sampling 100 points in the
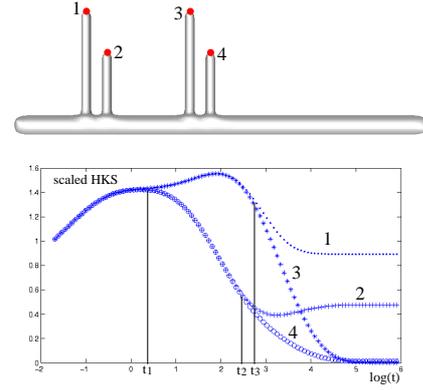
Figure 5: *Top: a synthetic model; bottom: the scaled HKS corresponding to the four marked points.*

logarithmically scale over the time interval $[t_{min}\ t_{max}]$ with $t_{min} = 4\ln 10/\lambda_{300}$, $t_{max} = 4\ln 10/\lambda_2$. The HKS remains almost unchanged for $t > t_{max}$ as it is mainly determined by the eigenvector ϕ_2. To estimate the HKS with $t < t_{min}$ faithfully, one needs to compute more eigenvalues and eigenfunctions. However, there is a limit on how small t can be for which the HKS can be estimated faithfully, which is determined by the resolution of the mesh.

5.1. Results

We first illustrate the multi-scale property of HKS on a synthetic model shown in Figure 5 by considering the signatures of the four marked points. Since all four points have isometric neighborhoods at small scales, their HKS's are the same for small t's ($< t_1$). Point 1 and point 3 have isometric neighborhoods at middle scales and thus their HKS's coincide even for middle t's ($[t_1,t_3]$), as do the HKS's of point 2 and point 4 ($[t_1,t_2]$). The signatures of points 3 and 4 are similar for large t's since they are both located in the middle of the base bar and their neighborhoods at large scale are close. Similar phenomena can be observed for the signatures of the four marked points on different claws on the dragon model in Figure 1.

In Figure 6, we color plot the distance function $d_{[t_1,t_2]}(x,\cdot)$ between the HKS of the marked point x and the signatures of other points on the model. As we can see, at small scales, all four feet are similar to each other. On the other hand, if large values of t, and consequently large neighborhoods are taken into account, the difference function can separate the front feet from the back feet, since the head of the dragon is quite different from its tail.

Figure 7 shows the application of HKS to multi-scale self-matching for several other models. For a given point x on each model, we report other points, whose signatures are within a threshold of HKS of x at different scales. Again, points with similar small neighborhoods are reported at small scales, while only those with similar global structure are reported at larger scales. To further illustrate the resilience of our method to noise, we introduce a small tunnel

Figure 6: *Color plot of the difference between the HKS defined by the range of scales $[t_1, t_2]$ of the point marked by the purple sphere and the signatures of other points on the shape. The difference increases as the color changes from red to green to blue. left: both t_1 and t_2 are small; right: t_1 is small, while t_2 is large.*

to the armadillo model. Note that despite this modification, relevant points are reported for both small and large scales.

One further advantage of our signature, is that unlike the GPS signature that is only defined up to orthonormal transformations within each eigenspace, the HKS signature is defined canonically, and allows to compare points across several shapes. Figure 8 shows the points on 4 independent poses of a horse model whose signatures at small and large scales are close to the signature of a point picked on one of the models.

As noted earlier, HKS can also be used to identify and differentiate between salient features on the shape. To find the feature points, we use the local maxima of the function $k_t(x, x)$ for a large t. These local maxima seem to capture the extremities of long protrusions on the surface, which allows us to find, for example, the hands and feet on the human model. We declare point x as a feature if $k_t(x, x) > k_t(x_i, x_i)$ for all x_i in the two ring neighborhood of x. Figure 9(a) shows the features picked on the 3 poses of the Armadillo model. Note that although this model has eigenfunctions that fail to be symmetric, as shown in Figure 4, the features picked by the method above are consistent across the two sides. For each of the poses of the Armadillo, our method identified 18 feature points. Figure 9(b) shows the classical MDS embedding of all 54 feature points in $\mathbb{R}^2$ based on the distances between their signature vectors for $t = [0.1, 4]$. Note that at this scale, the feature points on the hands and the feet of the Armadillo are hard to distinguish. Figure 9(c) shows the same embedding when HKS vectors are compared for $t = [0.1, 80]$. By increasing the scale, and incorporating more information we are able to distinguish the features on the hands from those on the feet. Note also that the lack of perfect symmetry is manifested by the fact that signatures of individual points from the left and the right sides of each model have non zero distance between them.

Because we only identify a few salient features on each shape, the complexity of this method is quite low. To illustrate this, we apply it to find repeated structure in the dataset of 175 models of the girl dancing swing provided by Vlasic et al. [VBMP08]. The top of Figure 10 shows the feature points found on 4 models in the dataset, while the bottom shows the classical MDS embedding of the feature points
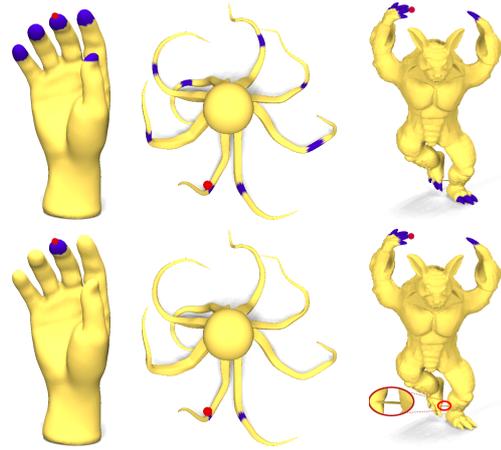
Figure 7: *The points in purple are matched with the marked point. Top: matching based on small t's; bottom: matching based on all t's.*

found on all 175 models based on distances between their signatures. In this example, we use a medium range of t's, although the reported embedding seems to persist across many scales. Note that not only are the features persistent across a large set of models, but also the heat-kernel signature allows to distinguish between different classes of features, such as the hands, the feet, and the head.

6. Conclusion and Future Work

We have described a concise and provably informative intrinsic point signature based on the heat diffusion process, and demonstrated its effectiveness in multi-scale shape matching. Our main observation is that HKS, though being the restriction of the heat kernel to the time domain, often preserves all the shape information up to isometric transformations. We also show that HKS can be computed faithfully and efficiently, and is stable against perturbations.

In the future, we would like to improve the efficiency of computing the HKS. Because computing the eigen-decomposition is rather costly an alternative approach of estimating the heat kernel $K_t = e^{-tL}A^{-1}$ is using a partial sum of the infinite series in Eq. (5). This method would be especially attractive for for small values of t, since only a few terms would be needed to obtain an accurate estimation of e^{-tL}. However, computing matrix exponential is a well-known difficult problem, and it would be interesting to see if the structure of L allows to perform this computation in a stable and efficient way for all values of t.

We would also like to obtain a quantative relation between the time parameter t and the size of the neighborhood around x that is characterized by the HKS of x at scale t, which would be help to choose scales automatically for partial matching. Finally, we would like to see if we can characterize entire regions or shapes using the Heat Kernel Signatures, and to explore other applications of the HKS.
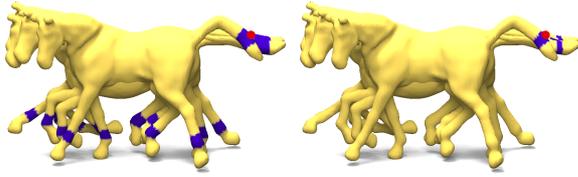
Figure 8: *Four different poses of a horse. left: matching based on half of t's; right: matching based on all t's*

References

[BBK*09] BRONSTEIN A. M., BRONSTEIN M. M., KIMMEL R., MAHMOUDI M., SAPIRO G.: A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *Submitted to the Intl. Journal of Computer Vision (IJCV)* (2009).

[BMP00] BELONGIE S., MALIK J., PUZICHA J.: Shape context: A new descriptor for shape matching and object recognition. In *In NIPS* (2000), pp. 831–837.

[BSW08] BELKIN M., SUN J., WANG Y.: Discrete Laplace operator on meshed surfaces. In *Proceedings of SOCG* (2008), pp. 278–287.

[Che76] CHENG S.-Y.: Eigenfunctions and nodal sets. *Commentarii Mathematici Helvetici 51*, 1 (1976), 43–55.

[CJ97] CHUA C. S., JARVIS R.: Point signatures: A new representation for 3d object recognition. *Int. J. Comput. Vision 25*, 1 (1997), 63–85.

[CZHZ98] CHEN Z.-Q., Z.QIAN, HU Y., ZHENG W.: Stability and approximations of symmetric diffusion semigroups and kernels. *Journal of Functional Analysis 152*, 1 (1998), 255–280.

[dGGV08] DE GOES F., GOLDENSTEIN S., VELHO L.: A hierarchical segmentation of articulated bodies. *Computer Graphics Forum 27*, 5 (July 2008), 1349–1356.

[GMGP05] GELFAND N., MITRA N. J., GUIBAS L. J., POTTMANN H.: Robust global registration. In *Symposium on Geometry Processing* (2005), pp. 197–206.

[Gri06] GRIGOR'YAN A.: Heat kernels on weighted manifolds and applications. *Cont. Math 398* (2006), 93–191.

[GSCO07] GAL R., SHAMIR A., COHEN-OR D.: Pose-oblivious shape signature. *IEEE Transactions on Visualization and Computer Graphics 13*, 2 (2007), 261–271.

[GVL96] GOLUB G. H., VAN LOAN C. F.: *Matrix computations (3rd ed.).* Johns Hopkins University Press, 1996.

[Hör85] HÖRMANDER L.: *The Analysis of Linear Partial Differential Operators, IV: Fourier Integral Operators.* Grundlehren. Math. Wiss. 275, Springer, Berlin, 1985.

[HSKK01] HILAGA M., SHINAGAWA Y., KOHMURA T., KUNII T. L.: Topology matching for fully automatic similarity estimation of 3d shapes. In *Proc. SIGGRAPH* (2001), pp. 203–212.

[Hsu02] HSU E. P.: *Stochastic Analysis on Manifolds.* American Mathematical Society, 2002.
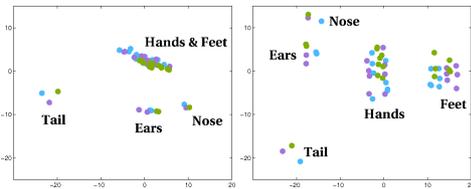
[Joh97] JOHNSON A.: *Spin-Images: A Representation for 3-D Surface Matching.* PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.

[Laf04] LAFON S.: *Diffusion Maps and Geometric Harmonics.* PhD thesis, Yale University, 2004.

[LG05] LI X., GUSKOV I.: Multi-scale features for approximate alignment of point-based surfaces. In *Symposium on Geometry processing* (2005), pp. 217–226.

[MDSB02] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: Discrete differential geometry operators for triangulated 2-manifolds. In *Proc. VisMath'02* (Berlin, Germany, 2002).

[MP49] MINAKSHISUNDARAM S., PLEIJEL A.: Some properties of the eigenfunctions of the Laplace operator on riemannian manifolds. *Can. J. Math* (1949).

[MYHS04] MANAY S., YEZZI A. J., HONG B. W., SOATTO S.: Integral invariant signatures. In *Proceedings of ECCV* (2004), pp. 87–99.

[OFCD02] OSADA R., FUNKHOUSER T., CHAZELLE B., DOBKIN D.: Shape distributions. *ACM Transactions on Graphics 21*, 4 (Oct. 2002), 807–832.

[OSG08] OVSJANIKOV M., SUN J., GUIBAS L. J.: Global intrinsic symmetries of shapes. *Comput. Graph. Forum 27*, 5 (2008), 1341–1348.

[PP93] PINKALL U., POLTHIER K.: Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics 2*, 1 (1993), 15–36.

[PWHY09] POTTMANN H., WALLNER J., HUANG Q.-X., YANG Y.-L.: Integral invariants for robust geometry processing. *Computer Aided Geometric Design 26*, 1 (2009), 37 – 60.

[Rus07] RUSTAMOV R. M.: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In *Symposium on Geometry Processing* (2007), pp. 225–233.
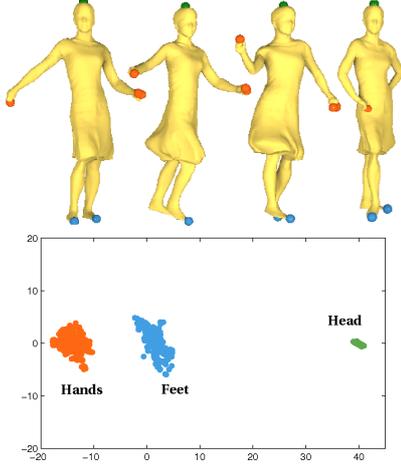
[RWP06] REUTER M., WOLTER F.-E., PEINECKE N.: Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. *Computer-Aided Design 38*, 4 (2006), 342–366.

[VBMP08] VLASIC D., BARAN I., MATUSIK W., POPOVIĆ J.: Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph. 27*, 3 (2008), 1–9.

[VL77] VAN LOAN C.: The sensitivity of the matrix exponential. *SIAM J. Numer. Anal. 14*, 6 (1977).

[War05] WARDETZKY M.: Convergence of the cotangent formula: An overview. In *Discrete Differential Geometry.* Birkhäuser Basel, 2005, pp. 89–112.

[Xu04] XU G.: Discrete Laplace-Beltrami operators and their convergence. *Comput. Aided Geom. Des. 21*, 8 (2004), 767–784.

(a) Feature points found by our method

(b) $t = [0.1, 4]$ (c) $t = [0.1, 80]$

Figure 9: *(a) Feature points on three poses of the Armadillo. (b-c) Classical MDS embedding of the feature points based on their HKS at different scales. The color of each point corresponds to the pose from which it is taken.*

Figure 10: *Top: Feature points on 4 models in the dataset by Vlasic et al. [VBMP08]. Bottom: the MDS embedding of the feature points from all 175 models based on their Heat Kernel Signatures at medium scales.*

7. Appendix

Proof of Theorem 1 We prove the theorem in three steps.

Step 1: We claim that M and N have the same spectrum and $|\phi_i^M(x)| = |\phi_i^N(T(x))|$ for any eigenfunction ϕ_i and any $x \in M$. We prove the above claim by contradiction. In the following, we sort the eigenvalues in the increasing order. The claim can fail first at the kth eigenvalue for some k, namely $\lambda_k^M \neq \lambda_k^N$ but $\lambda_i^M = \lambda_i^N$ and $|\phi_i^M(x)| = |\phi_i^N(T(x))|$ for any $i < k$ and any $x \in M$, or fail first at the kth eigenfunction for some k, namely there exists a point x such that $|\phi_k^M(x)| \neq |\phi_k^N(T(x))|$ but $\lambda_i^M = \lambda_i^N = \lambda_i$ for any $i \leq k$ and $|\phi_i^M(x)| = |\phi_i^N(T(x))|$ for any $i < k$ and any $x \in M$. In the former case, WLOG, assume $\lambda_k^M < \lambda_k^N$. There must exist a point $x \in M$ such that $\phi_k^M(x)^2 = \varepsilon > 0$ for some ε. From Eq. (3), we have

$$k_t^M(x,x) - k_t^N(T(x), T(x))$$
$$> e^{-\lambda_k^M t} \phi_k^M(x)^2 - \sum_{i=k}^{\infty} e^{-\lambda_i^N t} \phi_i^N(T(x))^2$$
$$= e^{-\lambda_k^M t} \big(\varepsilon - \sum_{i=k}^{\infty} e^{-(\lambda_i^N - \lambda_k^M)t} \phi_i^N(T(x))^2 \big). \qquad (7)$$

By the local Weyl law [Hör85], we have $|\phi_i^N(T(x))| = O((\lambda_i^N)^{(d-1)/4})$ where d is the dimension of N. In addition, the sequence $\{\lambda_i^N\}_{i=0}^{\infty}$ is increasing and hence $\lambda_i^N - \lambda_k^M > 0$ for any $i \geq k$. As the exponential decay can cancel the increasing of any polynomial, we have

$$\lim_{t \to \infty} \sum_{i=k}^{\infty} e^{-(\lambda_i^N - \lambda_k^M)t} \phi_i^N(T(x))^2 = 0$$

By choosing a big enough t, we have $k_t^M(x,x) - k_t^N(T(x), T(x)) > 0$ from Eq. (7), which contradicts the hypothesis. In the latter case, WLOG, assume $\varepsilon = \phi_k^M(x)^2 - \phi_k^N(T(x))^2 > 0$. We have

$$k_t^M(x,x) - k_t^N(T(x), T(x))$$
$$> e^{-\lambda_k t} (\phi_k^M(x)^2 - \phi_k^N(T(x))^2) - \sum_{i=k+1}^{\infty} e^{-\lambda_i^N t} \phi_i^N(T(x))^2$$
$$= e^{-\lambda_k t} \big(\varepsilon - \sum_{i=k+1}^{\infty} e^{-(\lambda_i^N - \lambda_k)t} \phi_i^N(T(x))^2 \big). \qquad (8)$$

Since the sequence $\{\lambda_i^N\}_{i=0}^{\infty}$ is strictly increasing, similarly for a big enough t, we have $k_t^M(x,x) - k_t^N(T(x), T(x)) > 0$ from Eq. (8), which contradicts the hypothesis.

Step 2: We show that either $\phi_i^M = \phi_i^N \circ T$ or $\phi_i^M = -\phi_i^N \circ T$ for any i. The argument is based on the properties of the nodal domains of the eigenfunction ϕ. A nodal domain is a connected component of $M \setminus \phi^{-1}(0)$. The sign of ϕ is consistent within a nodal domain, that is either all positive or all negative. For a fixed eigenfunction, the number of the nodal domains is finite. Since $|\phi_i^M(x)| = |\phi_i^N(T(x))|$ and T is continuous, the image of a nodal domain under T can not cross two nodal domains, that is a nodal domain can only be mapped to another nodal domain. A special property of the nodal domains [Che76] is that a positive nodal domain is only neighbored by negative ones, and vice versa. Pick a fixed point x_0 in a nodal domain. If $\phi_i^M(x_0) = \phi_i^N(T(x_0))$, we claim that $\phi_i^M(x) = \phi_i^N(T(x))$ for any point x on the manifold. Certainly the claim holds for the points inside the nodal domain D containing x_0. Due to the continuity of T, the neighboring nodal domains of D must be mapped to those next to the one containing $T(x_0)$. Because of the alternating property of the signs of neighboring nodal domains, the claims also hold for those neighboring ones. We can continue on expanding nodal domains like this until they are exhausted, which proves the claim. Thus $\phi_i^M = \phi_i^M \circ T$. Similarly, $\phi_i^M(x_0) = -\phi_i^N(T(x_0))$ leads to $\phi_i^M = -\phi_i^N \circ T$.

Step 3: We have for any $x, y \in M$ and $t > 0$ $k_t^M(x,y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i^M(x) \phi_i^M(y) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i^N(T(x)) \phi_i^N(T(y)) = k_t^N(T(x), T(y))$, which proves the theorem by Proposition 2.

Proof of Proposition 4 The difficulty of the proof comes from the fact that $e^{X+Y} \neq e^X e^Y$ in general. Let $E_1 = \tilde{A}^{\frac{1}{2}} - A^{\frac{1}{2}}$ and $E_2 = \tilde{A}^{-\frac{1}{2}} - A^{-\frac{1}{2}}$. It is easy to verify that both $\|E_1\| < \sqrt{\varepsilon}$ and $\|E_2\| < \sqrt{\varepsilon}$. Now let $L_1 = A^{\frac{1}{2}} L A^{-\frac{1}{2}}$, then L_1 is symmetric and $e^{-tL} = A^{\frac{1}{2}} e^{-tL_1} A^{-\frac{1}{2}}$. Similarly $\tilde{L}_1 = \tilde{A}^{\frac{1}{2}} \tilde{L} \tilde{A}^{-\frac{1}{2}}$ is symmetric and $e^{-t\tilde{L}} = \tilde{A}^{\frac{1}{2}} e^{-t\tilde{L}_1} \tilde{A}^{-\frac{1}{2}}$. We also have:

$$\|L_1 - \tilde{L}_1\| = \|A^{-\frac{1}{2}} W A^{-\frac{1}{2}} - \tilde{A}^{-\frac{1}{2}} \tilde{W} \tilde{A}^{-\frac{1}{2}}\|$$
$$= \|A^{-\frac{1}{2}} W A^{-\frac{1}{2}} - (A^{-\frac{1}{2}} + E_2)(W+F)(A^{-\frac{1}{2}} + E_2)\|$$
$$\leq \|E_2(W+F)(A^{-\frac{1}{2}} + E_2) + A^{-\frac{1}{2}} F(A^{-\frac{1}{2}} + E_2)\| = O(\sqrt{\varepsilon} + \delta)$$

As both $\tilde{L}_1$ and $\tilde{L}$ are symmetric, from [VL77], we have:

$$\frac{\|e^{-tL_1} - e^{-t\tilde{L}_1}\|}{\|e^{-tL_1}\|} \leq t\|L_1 - \tilde{L}_1\| e^{-t\|L_1 - \tilde{L}_1\|} = O(\sqrt{\varepsilon} + \delta).$$

It is easy to show that e^{-tL} and e^{-tL_1} have the same eigenvalues and thus $\|e^{-tL}\| = \|e^{-tL_1}\|$, which leads to

$$\frac{\|e^{-tL} - e^{-t\tilde{L}}\|}{\|e^{-tL}\|} = \frac{\|A^{\frac{1}{2}} e^{-tL_1} A^{-\frac{1}{2}} - \tilde{A}^{\frac{1}{2}} e^{-t\tilde{L}_1} \tilde{A}^{-\frac{1}{2}}\|}{\|e^{-tL_1}\|}$$
$$\leq \frac{\|\tilde{A}^{\frac{1}{2}} (e^{-tL_1} - e^{-t\tilde{L}_1}) \tilde{A}^{-\frac{1}{2}}\| + \|A^{\frac{1}{2}} e^{-tL_1} E_2 + E_1 e^{-t\tilde{L}_1} \tilde{A}^{-\frac{1}{2}}\|}{\|e^{-tL_1}\|}$$
$$\leq t\|L_1 - \tilde{L}_1\| e^{-t\|L_1 - \tilde{L}_1\|} + \|E_2\| \|A^{\frac{1}{2}}\| + \|A^{-\frac{1}{2}} + E_2\| \|E_1\|$$
$$= O(\sqrt{\varepsilon} + \delta)$$

Therefore we have:

$$\|K_t - \tilde{K}_t\| = \|e^{-tL} A^{-1} - e^{-t\tilde{L}} \tilde{A}^{-1}\|$$
$$\leq \|(e^{-tL} - e^{-t\tilde{L}}) A^{-1}\| + \|e^{-t\tilde{L}} E\| = O(\sqrt{\varepsilon} + \delta).$$