Geodesic Exponential Kernels: When Curvature and Linearity Conflict

Aasa Feragen[1], François Lauze[1], Søren Hauberg[2]
[1]DIKU, University of Copenhagen. [2]Cognitive Systems, DTU Compute.

Enforcing known constraints in data models often provides more faithful data representations and stronger results than simple Euclidean approaches. Constraints lead to nonlinear data spaces such as Riemannian manifolds or more general metric spaces. Data types typically modeled this way include shapes, DTI tensors, symmetric positive definite matrices, human poses, probability distributions and graphs. However, data analysis methods in nonlinear spaces often suffer from lack of computational efficiency, accuracy or both. Motivated by this, a recent trend in computer vision has been to define kernel methods in nonlinear data spaces. A popular way to generate kernels on nonlinear data spaces is through exponential kernels $k(x,y)$ which only rely on geodesic distances $d(x,y)$ between observations:

$$k(x,y) = \exp\left(-\lambda (d(x,y))^q\right), \qquad \lambda, q > 0. \tag{1}$$

For $q = 2$ this gives a geodesic generalization of the *Gaussian kernel*, and $q = 1$ gives the geodesic *Laplacian kernel*. While this idea has an appealing similarity to familiar Euclidean kernel methods, we show that it is highly limited if the metric space is curved. In Theorem 1 of this paper we prove that **geodesic Gaussian kernels on metric spaces are positive definite (PD) for all $\lambda > 0$ only if the metric space is flat**.

Theorem 1 *Let (X,d) be a geodesic metric space, and assume that $k(x,y) = \exp(-\lambda d^2(x,y))$ is a PD geodesic Gaussian kernel on X for all $\lambda > 0$. Then (X,d) is flat in the sense of Alexandrov.*

This is a negative result, in the sense that most metric spaces of interest are not flat. As a straightforward consequence of Theorem 1, we show that **geodesic Gaussian kernels on Riemannian manifolds are PD for all $\lambda > 0$ only if the Riemannian manifold is a Euclidean space**:

Theorem 2 *Let M be a complete, smooth Riemannian manifold with its associated geodesic distance metric d. Assume, moreover, that $k(x,y) = \exp(-\lambda d^2(x,y))$ is a PD geodesic Gaussian kernel for all $\lambda > 0$. Then the Riemannian manifold M is isometric to a Euclidean space.*

These two theorems raise several points. The first and main point is that defining geodesic Gaussian kernels on Riemannian manifolds or other geodesic metric spaces has limited applicability as most relevant spaces are curved. In particular, on Riemannian manifolds the kernels will generally only be PD if the original data space is Euclidean, in which case the geodesic Gaussian kernel is just the standard Gaussian kernel.

Second, *this result is not surprising:* Curvature cannot be captured by a flat space, and the classical Schönberg theorem indicates a strong connection between PD Gaussian kernels and linearity of the employed distance measure. This is made explicit by Theorems 1 and 2.

Third, do these results depend on the choice $q = 2$ in Equation (1)? For Riemannian manifolds, a higher power $q > 2$ *never* leads to a PD kernel for all $\lambda > 0$:

Theorem 3 *Let M be a Riemannian manifold with its associated geodesic distance metric d, and let $q > 2$. Then there is some $\lambda > 0$ so that the kernel in Equation (1) is not PD.*

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

Kernel	Extends to general Metric spaces	Extends to general Riemannian manifolds
Gaussian ($q = 2$)	No (only if flat)	No (only if Euclidean)
Laplacian ($q = 1$)	Yes, iff metric is CND	Yes, iff metric is CND
Geodesic exp. ($q > 2$)	Not known	No

Table 1: Overview of results: For a geodesic metric, when is the geodesic exponential kernel in Equation (1) positive definite for all $\lambda > 0$?

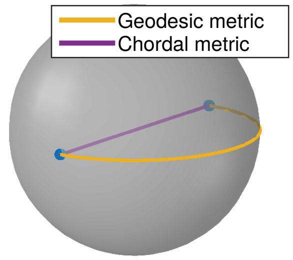

Figure 1: The sphere $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ is an example of a popular manifold data space parametrizing e.g. SIFT features or probability distributions. The Euclidean chordal metric on $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ is measured directly in $\mathbb{R}^{d+1}$ and therefore relies on the embedding, while the geodesic metric is measured along $\mathbb{S}^d$.

The existence of a $\lambda > 0$ such that the kernel is not PD may seem innocent, but this means that the kernel bandwidth parameter cannot be learned.

In contrast, the choice $q = 1$ in Equation (1), giving a geodesic Laplacian kernel, leads to a more positive result: The geodesic Laplacian kernel will be positive definite if and only if the distance d is *conditionally negative definite* (CND). This provides a PD kernel framework which, for several popular Riemannian data manifolds, takes advantage of the geodesic distance.

Theorem 4 *i) The geodesic distance d in a geodesic metric space (X,d) is CND if and only if the corresponding geodesic Laplacian kernel is PD for all $\lambda > 0$.*

ii) In this case, the square root metric $d_{\sqrt{}}(x,y) = \sqrt{d(x,y)}$ is also a distance metric, and $(X, d_{\sqrt{}})$ can be isometrically embedded as a metric space into a Hilbert space H.

iii) The square root metric $d_{\sqrt{}}$ is not a geodesic metric, and $d_{\sqrt{}}$ corresponds to the chordal metric in H, not the intrinsic metric on the image of X in H.

In Theorem 4, for $\phi : X \to H$, the *chordal metric* $\|\phi(x) - \phi(y)\|_H$ measures distances directly in H rather than intrinsically in the image $\phi(X) \subset H$, see also Fig. 1.

We study PD'ness of Laplacian kernels in several popular data spaces; see Table 2. Part *ii)* of Theorem 4 illustrates that any geodesic metric space whose geodesic Laplacian kernel is always PD must necessarily have strong linear properties: Its square root metric is isometrically embeddable in a Hilbert space. This illustrates an intuitively simple point: A PD kernel has no choice but to linearize the data space through the reproducing kernel Hilbert space. Therefore, its ability to capture the original data space geometry is deeply connected to the linear properties of the original metric.

Space	Distance metric	PD Gaussian kernel?	PD Laplacian kernel?
$\mathbb{R}^n$	Euclidean metric	✓	✓
$\mathbb{R}^n$, $n > 2$	l_q-norm $\|\cdot\|_q$, $q > 2$	÷	÷
Sphere $\mathbb{S}^n$	classical intrinsic	÷	✓
Real projective space $\mathbb{P}^n(\mathbb{R})$	classical intrinsic	÷	÷
Grassmannian	classical intrinsic	÷	÷
Sym_d^+	Frobenius	✓	✓
Sym_d^+	Log-Euclidean	✓	✓
Sym_d^+	Affine invariant	÷	÷
Sym_d^+	Fisher information metric	÷	÷
Hyperbolic space $\mathbb{H}^n$	classical intrinsic	÷	✓
1-dimensional normal distributions	Fisher information metric	÷	✓
Metric trees	tree metric	÷	✓
Geometric graphs (e.g. kNN)	shortest path distance	÷	÷
Strings	string edit distance	÷	÷
Trees, graphs	tree/graph edit distance	÷	÷

Table 2: For a set of popular metric and manifold data spaces and metrics, we record whether the metric is a geodesic metric, and whether its corresponding Gaussian and Laplacian kernels are PD.