Geodesic Exponential Kernels: When Curvature and Linearity Conflict

Aasa Feragen
DIKU, University of Copenhagen
Denmark
aasa@diku.dk

François Lauze
DIKU, University of Copenhagen
Denmark
francois@diku.dk

Søren Hauberg
DTU Compute
Denmark
sohau@dtu.dk

Abstract

We consider kernel methods on general geodesic metric spaces and provide both negative and positive results. First we show that the common Gaussian kernel can only be generalized to a positive definite kernel on a geodesic metric space if the space is flat. As a result, for data on a Riemannian manifold, the geodesic Gaussian kernel is only positive definite if the Riemannian manifold is Euclidean. This implies that any attempt to design geodesic Gaussian kernels on curved Riemannian manifolds is futile. However, we show that for spaces with conditionally negative definite distances the geodesic Laplacian kernel can be generalized while retaining positive definiteness. This implies that geodesic Laplacian kernels can be generalized to some curved spaces, including spheres and hyperbolic spaces. Our theoretical results are verified empirically.

1. Introduction

Standard statistics and machine learning tools require input data residing in a Euclidean space. However, many types of data are more faithfully represented in general non-linear metric spaces (e.g. Riemannian manifolds). This is, for instance, the case when analyzing shapes [10, 16, 26, 33, 41, 56, 58, 66], DTI images [25, 46, 49, 64], motion models [14, 61], symmetric positive definite matrices [12, 50, 62], illumination-invariance [13], human poses [32, 47], tree structured data [20, 22, 23], metrics [28, 31], probability distributions [2]; for general manifold learning metrics [60] or in general for data invariant to a group action [42]. The underlying metric space captures domain specific knowledge, e.g. non-linear constraints, which is available a priori. The intrinsic geodesic metric encodes this knowledge, often leading to improved statistical models.

A seemingly straightforward approach to statistics in metric spaces is to use kernel methods [54], designing kernels \( k(x, y) \) which only rely on geodesic distances \( d(x, y) \) between observations [15]:

\[
k(x, y) = \exp(-\lambda d(x, y)^q), \quad \lambda, q > 0. \tag{1}
\]

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.

Positive definiteness of a kernel \( k \) is critical for the use of kernel methods such as support vector machines or kernel PCA, as it ensures the existence of a reproducing kernel Hilbert space where these methods act [54]. In this paper, we analyze exponential kernels on geodesic metric spaces and show the following results, summarized in Table 1.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>Geodesic exp. ((q &gt; 2))</th>
<th>Riemannian manifolds</th>
<th>Geodesic exp. ((q = 1))</th>
<th>Metric spaces</th>
<th>Extends to general</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian ((q = 2))</td>
<td>No (only if flat)</td>
<td>No (only if Euclidean)</td>
<td>Yes, iff metric is CND</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Laplacian ((q = 1))</td>
<td>Not known</td>
<td>Yes, iff metric is CND</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Geodesic exp. ((q &gt; 2))</td>
<td>No</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

- The geodesic Gaussian kernel is positive definite (PD) for all \( \lambda > 0 \) only if the underlying metric space is flat (Theorem 1). In particular, when the metric space is a Riemannian manifold, the geodesic Gaussian kernel is PD for all \( \lambda > 0 \) if and only if the manifold is Euclidean (Theorem 2). This negative result implies that Gaussian kernels cannot be generalized to any non-trivial Riemannian manifolds of interest.
- The geodesic Laplacian kernel is PD if and only if the metric is conditionally negative definite (Theorem 4). This condition is not generally true for metric spaces, but it holds for a number of spaces of interest. In particular, the geodesic Laplacian kernel is PD on spheres, hyperbolic spaces, and Euclidean spaces (Table 2).
- For any Riemannian manifold \((M, g)\), the kernel \((1)\) will never be PD for all \( \lambda > 0 \) if \( q > 2 \) (Theorem 3).

Generalization of geodesic kernels to metric spaces is motivated by the general lack of powerful machine learning techniques in these spaces. In that regard, our first results are disappointing as they imply that generalizing Gaussian kernels to metric spaces is not a viable direction forward. Intuitively, this is not surprising as kernel methods embed the data in a linear space, which cannot be expected to capture the curvature of a general metric space. Our second result is therefore a positive surprise: it allows the Laplacian kernel to be applied in some metric spaces, although this has strong implications for their geometric properties. This gives hope
that other kernels can be generalized, though our third result indicates that the geodesic exponential kernels \((1)\) have limited applicability on Riemannian manifolds.

The paper is organized as follows. We state our main results and discuss their consequences in Sec. 2, postponing proofs until Sec. 3, which includes a formal discussion of the preliminaries. This section can be skipped in a first reading.

Many efficient machine learning algorithms are available in Euclidean spaces; their generalization to metric spaces is an open problem. Kernel methods form an immensely popular class of algorithms including support vector machines and kernel PCA [54]. These algorithms rely on the specification of a kernel \(k(x, y)\), which embeds data points \(x, y\) in a linear Hilbert space and returns their inner product. Kernel methods are very flexible, as they only require the computation of inner products (through the kernel). However, the kernel is only an inner product if it is PD, so kernel methods are only well-defined for kernels which are PD [54].

Many popular choices of kernels for Euclidean data rely only on the Euclidean distance between data points; for instance the widely used Gaussian kernel (given by \((1)\) with \(q = 2\)). Kernels which only rely on distances form an obvious target for generalizing kernel methods to metric spaces, where distance is often the only quantity available.

2. Main results and their consequences

Before formally proving our main theorems, we state the results and provide hints as to why they hold. We start with a brief review of metric geometry and the notion of a flat space, both of which are fundamental to the results.

In a general metric space \((X, d)\) with distance metric \(d\), the length \(l(\gamma)\) of a path \(\gamma: [0, L] \rightarrow X\) from \(x\) to \(y\) is defined as the smallest upper bound of any finite approximation of the path (see Fig. 1):

\[
l(\gamma) = \sup_{0 = t_0 < t_1 < \ldots < t_n = 1, n \in N} \sum_{i=1}^{n} d(t_{i-1}, t_i).
\]

A path \(\gamma: [0, L] \rightarrow X\) is called a geodesic [9] from \(x\) to \(y\) if \(\gamma(0) = x, \gamma(L) = y\) and \(d(\gamma(t), \gamma(t')) = |t - t'|\) for all \(t, t' \in [0, L]\). In particular, \(l(\gamma) = d(x, y) = L\) for a geodesic \(\gamma\). In a Euclidean space, geodesics are straight lines. A geodesic from \(x\) to \(y\) will always be the shortest possible path from \(x\) to \(y\), but geodesics with respect to a given metric do not always exist, even if shortest paths do. An example is given later in Fig. 3.

A metric space \((X, d)\) is called a geodesic space if every pair \(x, y \in X\) can be connected by a geodesic. Informally, a geodesic metric space is merely a space in which distances can be computed as lengths of geodesics, and data points can be interpolated via geodesics.

Riemannian manifolds are a commonly used class of metric spaces. Here distances are defined locally through a smoothly changing inner product in the tangent space. Intuitively, a Riemannian manifold can be thought of as a smooth surface (e.g. a sphere) with geodesics corresponding to shortest paths on the surface. A geodesic distance metric corresponding to the Riemannian structure is defined explicitly as the length of the geodesic joining two points. Whenever a Riemannian manifold is complete, it is a geodesic space. This is the case for most manifolds of interest.

(Remote) Figure 1. Path length in a metric space is defined as the supremum of lengths of finite approximations of the path.

2.1. Main results

In Theorem 1 of this paper we prove that geodesic Gaussian kernels on metric spaces are PD for all \(\lambda > 0\) only if the metric space is flat. Informally, a metric space is flat if it (for all practical purposes) is Euclidean. More formally:

Definition 1. A geodesic metric space \((X, d)\) is flat in the sense of Alexandrov if any geodesic triangle in \(X\) can be isometrically embedded in a Euclidean space.

Here, an embedding \(f: X \rightarrow X'\) from a metric space \((X, d)\) to another metric space \((X', d')\) is isometric if \(d'(f(x), f(y)) = d(x, y)\) for all \(x, y \in X'\). A geodesic triangle \(abc\) in \(X\) consists of three points \(a, b, c\) joined by geodesic paths \(\gamma_{ab}, \gamma_{bc}\) and \(\gamma_{ac}\). The concept of flatness essentially requires that all geodesic triangles are identical to Euclidean triangles; see Fig. 2.

With this, we state our first main theorem:

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. In fact, the motivation for generalizing kernel methods is to cope with data residing in non-flat metric spaces.

As a 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 Euclidean.**

**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 have several consequences. The first and main consequence is that defining geodesic Gaussian kernels on Riemannian manifolds or other geodesic metric spaces has limited applicability as most spaces of interest are not flat. In particular, on Riemannian manifolds the kernels will generally only be PD if the original data space is Euclidean. In this case, nothing is gained by treating the data space as a Riemannian manifold, as it is perfectly described by the well-known Euclidean geometry, where many problems can be solved in closed form. In Sec. 4 we re-interpret recent work which does, indeed, take place in Riemannian manifolds that turn out to be Euclidean.

Second, this result is not surprising: Curvature cannot be captured by a flat space, and Schönberg’s classical theorem (see Sec. 3.1) indicates a strong connection between PD Gaussian kernels and linearity of the employed distance measure. This is made explicit by Theorems 1 and 2.

While this paper was in print, a result similar to Theorem 2 appeared in [39]. However, the authors do not note that as a consequence, Gaussian RBF kernels that use the geodesic distance only apply to Riemannian manifolds that are Euclidean spaces, where they coincide with the standard Gaussian kernels [54]. In order to apply Gaussian kernels to non-Euclidean spaces they are forced to replace the Riemannian structure by a Euclidean chordal metric.

The obvious next question is the extent to which these negative results depend on the choice \( q = 2 \) in (1), which results in a Gaussian kernel. A recent result by Istas [35] implies that for Riemannian manifolds, passing to a higher power \( q > 2 \) will never lead 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 (1) is not PD.

The existence of a \( \lambda > 0 \) such that the kernel is not PD may seem innocent; however, as a consequence, the kernel bandwidth parameter cannot be learned.

In contrast, the choice \( q = 1 \) in (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). CND metrics have linear embeddability properties analogous to those of PD kernels; see Sec. 3.1 for formal definitions and properties. 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 \( d_{\sqrt{\cdot}}(x,y) = \sqrt{d(x,y)} \) is also a distance metric, and \( (X,d_{\sqrt{\cdot}}) \) can be isometrically embedded as a metric space into a Hilbert space \( H \).

iii) The square root metric \( d_{\sqrt{\cdot}} \) is not a geodesic metric, and \( d_{\sqrt{\cdot}} \) 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. 3.

In Sec. 4 we discuss several popular data spaces for which geodesic Laplacian kernels are PD (see Table 2); examples include spheres, hyperbolic spaces and more. Nevertheless, we see from part ii) of Theorem 4 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. Therefore, its ability to capture the original data space geometry is deeply connected to the linear properties of the original metric.

### 3. Proofs of main results

In this section we prove the main results of the paper; this section may be skipped in a first reading of the paper. In the first two subsections we review and discuss classical geometric results on kernels, manifolds and curvature, which we will use to prove the main results.

#### 3.1. Kernels

A modern and comprehensive treatment of the classical results on PD and CND kernels referred to here, can be found in [5, Appendix C].

---

\(^2\)Another curious connection between kernels and curvature is found in [11], which shows that Gaussian and polynomial kernels on \( \mathbb{R}^n \) and \( \mathbb{R}^2 \), respectively, have flat feature space images \( \phi(\mathbb{R}^n) \) and \( \phi(\mathbb{R}^2) \).
Definition 2. A positive definite (PD) kernel on a topological space $X$ is a continuous function $k: X \times X \rightarrow \mathbb{R}$ such that for any $n \in \mathbb{N}$, any elements $x_1, \ldots, x_n \in X$ and any numbers $c_1, \ldots, c_n \in \mathbb{R}$, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j k(x_i, x_j) \geq 0.$$ 

Definition 3. A conditionally negative definite (CND) kernel on a topological space $X$ is a continuous function $\psi: X \times X \rightarrow \mathbb{R}$ which satisfies

i) $\psi(x, x) = 0$ for all $x \in X$

ii) $\psi(x, y) = \psi(y, x)$ for all $x, y \in X$

iii) for any $n \in \mathbb{N}$, any elements $x_1, \ldots, x_n \in X$ and any real numbers $c_1, \ldots, c_n$ with $\sum_{i=1}^{n} c_i = 0$, we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \psi(x_i, x_j) \leq 0.$$

Example 1. If $d: H \times H \rightarrow \mathbb{R}$ is the metric induced by the norm on a Hilbert space $H$, then the map $d^2: H \times H \rightarrow \mathbb{R}$ given by $d^2(x, y) = (d(x, y))^2$ is a CND kernel [5].

The following two theorems are key to understanding the connection between distance metrics and their corresponding exponential kernels.

Theorem 5 (Due to Schönberg [55], Theorem C.3.2 in [5]). If $X$ is a topological space and $\psi: X \times X \rightarrow \mathbb{R}$ is a continuous kernel on $X$ with $\psi(x, x) = 0$ and $\psi(y, x) = \psi(x, y)$ for all $y, x \in X$, then the following are equivalent:

- $\psi$ is a CND kernel
- the kernel $k(x, y) = e^{-\lambda \psi(x, y)}$ is PD for all $\lambda \geq 0$. 

Theorem 6 (Part of Theorem C.2.3 in [5]). If $\psi: X \times X \rightarrow \mathbb{R}$ is a CND kernel on a topological space $X$, then there is a real Hilbert space $H$ and a continuous mapping $f: X \rightarrow H$ such that $\psi(x, y) = \|f(x) - f(y)\|_H^2$ for all $x, y \in X$. 

From the above, it is straightforward to deduce:

Corollary 1. If the geodesic Gaussian kernel is PD, then there is a mapping $f: X \rightarrow H$ into some Hilbert space $H$ such that

$$d(x, y) = \|f(x) - f(y)\|_H$$

for each $x, y \in X$. Note that this mapping $f$ is not necessarily related to the feature mapping $\phi: X \rightarrow V$ such that $k(x, y) = \langle \phi(x), \phi(y) \rangle_V$.

3.2. Curvature

While curvature is usually studied using differential geometry, we shall access curvature via a more general approach that applies to general geodesic metric spaces. This notion of curvature, originating with Alexandrov and Gromov, operates by comparing the metric space to spaces whose geometry we understand well, referred to as model spaces. The model spaces $M_\kappa$ are spheres (of positive curvature $\kappa > 0$), the Euclidean plane (flat, curvature $\kappa = 0$) and hyperbolic space (negative curvature $\kappa < 0$). Since metric spaces can be pathological, curvature is approached by bounding the curvature of the space at a given point from above or below. The bounds are attained by comparing geodesic triangles in the metric space with triangles in the model spaces, as expressed in the $CAT(\kappa)$ condition:

Definition 4. Let $(X, d)$ be a geodesic metric space $X$. Let $abc$ be a geodesic triangle of perimeter $< 2D_\kappa$, where $D_\kappa$ is the diameter of $M_\kappa$, that is, $D_\kappa = \infty$ for $\kappa \leq 0$, and $D_\kappa = \frac{\pi}{\sqrt{-\kappa}}$ for $\kappa > 0$. There exists a triangle $\bar{abc}$ in the model space $M_\kappa$ with vertices $\bar{a}, \bar{b}$ and $\bar{c}$ and with geodesic edges $\bar{\gamma}_{\bar{a}\bar{b}}, \bar{\gamma}_{\bar{b}\bar{c}}$ and $\bar{\gamma}_{\bar{a}\bar{c}}$, whose lengths are the same as the lengths of the edges $\gamma_{ab}, \gamma_{bc}$ and $\gamma_{ac}$ in $abc$. This is an $M_\kappa$-comparison triangle for $abc$ (see Fig. 4).

For any point $x$ sitting on the segment $\gamma_{bc}$, there is a corresponding point $\bar{x}$ on the segment $\bar{\gamma}_{\bar{b}\bar{c}}$ in the comparison triangle, such that $d_{M_\kappa}(\bar{x}, \bar{b}) = d(x, b)$. If we have

$$d(x, a) \leq d_{M_\kappa}(\bar{x}, \bar{a})$$

for every such $x$, and similarly for any $x$ on $\gamma_{ab}$ or $\gamma_{ac}$, then the geodesic triangle $abc$ satisfies the $CAT(\kappa)$ condition.

The metric space $X$ is a $CAT(\kappa)$ space if any geodesic triangle $abc$ in $X$ of perimeter $< 2D_\kappa$ satisfies the $CAT(\kappa)$ condition given in eq. 2. Geometrically, this means that triangles in $X$ are thinner than triangles in $M_\kappa$. The metric space $X$ has curvature $\leq \kappa$ in the sense of Alexandrov if it is locally $CAT(\kappa)$.

While curvature in the $CAT(\kappa)$ sense allows the study of curvature through the relatively simple means of geodesic distances alone, it is a weaker concept of curvature than the standard sectional curvature used in Riemannian geometry. Nevertheless, the two concepts are related, as captured by the following theorem due to Cartan and Alexandrov:

Theorem 7 (Theorem II.1A.6 [9]). A smooth Riemannian manifold $M$ is of curvature $\leq \kappa$ in the sense of Alexandrov if and only if the sectional curvature of $M$ is $\leq \kappa$.

The proof of the main theorem will, moreover, rely on the following theorem characterizing manifolds of constant zero sectional curvature:

Theorem 8 (Part of Theorem 11.12 [45]). Let $M$ be a complete, simply connected $n$-dimensional Riemannian manifold with constant sectional curvature $C = 0$. Then $M$ is isometric to $\mathbb{R}^n$. 

We are now ready to prove our main theorems.
3.3. Geodesic Gaussian kernels on metric spaces: Proof of Theorem 1

As in the statement of Theorem 1, assume that the metric space \((X, d)\) is a geodesic space as defined in Sec. 2, and that \(k(x, y) = e^{-λd(x, y)}\) is a PD geodesic Gaussian kernel on \(X\) for all \(λ > 0\). An important consequence of Theorem 6 is that the map \(f: X \rightarrow H\) must take geodesic segments to geodesic segments, which in \(H\) are straight line segments.

Lemma 1. If \(γ: [0, L] \rightarrow X\) is a geodesic of length \(L\) from \(a = γ(0)\) to \(b = γ(L)\) in \(X\), then \(f(γ([0, L]))\) is the straight line from \(f(a)\) to \(f(b)\) in \(H\), and

\[
f(γ(t)) = f(a) + \frac{t}{L} (f(b) − f(a)) \tag{3}
\]

for all \(t ∈ [0, L]\).

Proof. Since \(γ: [0, L] \rightarrow X\) is a geodesic, it contains every point \(γ(t)\) for all \(t ∈ [0, L]\), and since \(γ\) is a geodesic of length \(L\), we have \(d(γ(0), γ(t)) = t\) for all \(t ∈ [0, L]\), so

\[
\|f(γ(0)) − f(γ(t))\| = d(γ(0), γ(t)) = t.
\]

This is only possible if \(f ∘ γ\) is the straight line from \(f(a)\) to \(f(b)\) in \(H\). Equation (3) follows directly, as it is the geodesic parametrization of a straight line from \(f(a)\) to \(f(b)\).

This enables us to prove Theorem 1:

Proof of Theorem 1. Let \(a, b, c ∈ X\) be three points in \(X\) and form a geodesic triangle spanned by their joining geodesics \(γ_{ab}, γ_{bc}\) and \(γ_{ca}\). Then the points \(f(a), f(b)\) and \(f(c)\) in \(H\) are connected by straight line geodesics \(f ∘ γ_{ab}, f ∘ γ_{bc}\) and \(f ∘ γ_{ca}\) by Lemma 1. These points and geodesics in \(H\) span a 2-dimensional linear subspace of \(H\) in which they form a Euclidean comparison triangle.

Without loss of generality, pick any two points \(x\) and \(y\) on the geodesic triangle and measure the distance \(d(x, y)\). The corresponding distance in the comparison triangle is \(\|f(x) − f(y)\|\), and by the definition of \(f\) we know that \(d(x, y) = \|f(x) − f(y)\|\), so the geodesic triangle is isometrically embedded into the comparison triangle. Hence, \(X\) is flat in the sense of Alexandrov.

Corollary 2. The metric space \(X\) is contractible, and hence simply connected.

Proof. By Theorem 1, \(X\) must necessarily be a \(\text{CAT}(0)\), and contractible by [9, Corollary II.1.5].

3.4. Geodesic Gaussian kernels on Riemannian manifolds: Proof of Theorem 2

We prove that for a complete, smooth Riemannian manifold \(M\) with associated geodesic distance metric \(d\), if the geodesic Gaussian kernel \(k(x, y) = e^{-λd(x, y)}\) is PD for all \(λ > 0\), then \(M\) is isometric to a Euclidean space.

Proof of Theorem 2. We start out by showing that the sectional curvature of \(M\) is 0 everywhere.

By Theorem 1, \(M\) is a \(\text{CAT}(0)\) space, so in particular it has curvature \(≤ 0\) in the sense of Alexandrov. Therefore, by Theorem 7, the sectional curvature of \(M\) is \(≤ 0\).

To prove the claim, we need to show that \(M\) does not have any points with negative sectional curvature. To this end, assume that there is some point \(p \in M\) such that the sectional curvature of \(M\) at \(p\) is \(κ < 0\). Then, since sectional curvature on smooth Riemannian manifolds is continuous, there exists some neighborhood \(U\) of \(p\) and some \(κ' < 0\) such that the sectional curvature in \(U\) is \(≤ κ' < 0\). But then, by Theorem 7, \(U\) also has curvature \(≤ κ'\) in the sense of Alexandrov, which cannot hold due to Theorem 1. It follows that the sectional curvature of \(M\) at \(p\) cannot be \(κ < 0\); hence, the sectional curvature of \(M\) must be everywhere 0.

Since \(M\) is simply connected by Corollary 2, we apply Theorem 8 to conclude that \(M\) must be isometric to \(\mathbb{R}^m\).

3.5. The case \(q > 2\)

Proof of Theorem 3. This is a direct consequence of [35, Theorem 2.12].

3.6. Geodesic Laplacian kernels: Proof of Theorem 4

Another consequence of Schönherr’s Theorem 5 is that the geodesic Laplacian kernel defined by (1) with \(q = 1\) is PD if and only if the distance \(d\) is CND. This provides a PD kernel framework which, for several popular Riemannian data manifolds, utilizes the geodesic distance.

Proof of Theorem 4. i) By Theorem 5, \(d\) is CND if and only if the Laplacian kernel \(k(x, y) = e^{-λd(x, y)}\) is PD for all \(λ > 0\).

ii) By Theorem 6, there exists a real Hilbert space \(H\) and a continuous map \(f: X \rightarrow H\) such that

\[
d(x, y) = \|f(x) − f(y)\|_H^2\ \text{for all } x, y ∈ X. \tag{4}
\]

That is, \(d_\sqrt{γ}(x, y) = \|f(x) − f(y)\|_H\) for all \(x, y ∈ X\). The map \(f\) must be injective, because if \(f(x) = f(y)\) for \(x \neq y\) then by (4), \(0 = \|f(x) − f(y)\|_H = d(x, y) > 0\), which is false. Therefore, \(d_\sqrt{γ}\) coincides with the restriction to \(f(X)\) of the metric on \(H\) induced by \(\|\cdot\|\). Since the restriction of a metric to a subset is a metric, \(d_\sqrt{γ}\) is a metric, and by definition, \(f\) is an isometric embedding of \((X, d_\sqrt{γ})\) into \(H\).

iii) Since \(f\) is an isometric embedding as metric spaces, \(d_\sqrt{γ}\) must correspond to the chordal metric in \(H\).

Assume that \(d_\sqrt{γ}\) is a geodesic metric on \(X\), then by Lemma 1, \(f\) maps geodesics in \((X, d_\sqrt{γ})\) to straight line segments in \(H\). Focusing on a single geodesic segment \(γ: [0, L] \rightarrow X\), we obtain

\[
d_\sqrt{γ}(γ(t), γ(t')) = \|f ∘ γ(t) − f ∘ γ(t')\| = |t − t'|
\]

for all \(t, t' ∈ [0, L]\). Since \(d = d^2_\sqrt{γ}\) is a metric by assumption, the square \(d_\sqrt{γ}(t, t') = |t − t'|^2 =

3036
\(d(\gamma(t), \gamma(t'))\) is a metric on \([0, L]\). But this is not true, as the triangle inequality fails to hold.

Therefore, \(d_\sqrt{-}\) cannot be a geodesic metric on \(X\).

As noted in Table 2 below, for a number of popular Riemannian manifolds, the geodesic distance metric is CND, meaning that geodesic Laplacian kernels are PD.

**Remark 5.** For a CND distance metric \(d: \mathcal{X} \times \mathcal{X} \to \mathbb{R}\), a PD kernel \(k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}\) can also be constructed through the formula \(k(x, x') = d(x, x')^2 - d(x, x_0) - d(x_0, x')^2\) \([7, 53]\), where \(x_0 \in \mathcal{X}\) is any point. For other distance-based kernels, e.g. the rational-quadratic kernel, little is known.

### 4. Implications for popular manifolds and related work

Many popular data spaces appearing in computer vision are not flat, meaning that their geodesic distances are not CND and their geodesic Gaussian kernels will not be PD. Table 2 lists known results on CND status of some popular data spaces. In particular, the classical intrinsic metrics on \(\mathbb{R}^n\), \(\mathbb{H}^n\) and \(\mathbb{S}^n\) are all CND\(^3\). As the Fisher information metric on 1-dimensional normal distributions defines the hyperbolic geometry \(\mathbb{H}^2\) \([2]\), it will give a CND geodesic metric. For projective space, on the other hand, \([51]\) provides an example showing that the classical intrinsic metric is not CND. As Grassmannians are generalizations of projective spaces, their geodesic metrics are therefore also not generally CND.

Symmetric, positive definite \((d \times d)\) matrices form another important data manifold, denoted \(\text{Sym}_+^d\). While the popular Frobenius and Log-Euclidean \([3]\) metrics on \(\text{Sym}_+^d\) are actually Euclidean, little is known theoretically about whether the geodesic distance metrics of non-Euclidean Riemannian metrics on \(\text{Sym}_+^d\) are CND. In Sec. 5 we show empirically that neither the affine-invariant metric \([49]\) nor the Fisher information metric on the corresponding fixed-mean multivariate normal distributions \([2, 4]\) induce a CND geodesic metric. Note how the qualitatively similar affine-invariant and Log-Euclidean metrics differ in whether they generate PD exponential kernels.

Non-manifold data spaces are also popular, e.g. the edit distance on strings was shown not to be CND by Cortes et al. \([17]\). As tree- and graph edit distances generalize string edit distance, the same holds for these. For this reason, PSD graph kernels are often similarity-based \([8, 21]\), not distance-based. The metric along a metric tree, on the other hand, is CND. In Sec. 5 we show empirically that this does not generalize to the shortest path metric on a geometric graph, such as the \(k\text{NN}\) or \(\epsilon\)-neighborhood graphs often used in manifold learning \([1, 6, 52, 59]\).

#### 4.1. Relation to previous work

Several PD kernels on manifolds have appeared in the literature, some of them even Gaussian kernels based on distance metrics on manifolds such as spheres or Grassmanian manifolds, which we generally consider as curved manifolds. The reader might wonder how this is possible given the above presented results. The explanation is that the distances used in these kernels are not geodesic distances and, in many cases, have little or nothing to do with the Riemannian structure of the manifold. We discuss a few examples.

**Example 2.** In \([37]\), a PD kernel is defined on \(\text{Sym}_+^d\) by using a geodesic Gaussian kernel with the log-Euclidean metric \([49]\). The log-Euclidean metric is defined by pulling the (Euclidean) Frobenius metric on \(\text{Sym}_d\) back to \(\text{Sym}_+^d\) via the diffeomorphic matrix logarithm. Equivalently, data in \(\text{Sym}_+^d\) is mapped into the Euclidean \(\text{Sym}_d\) via the diffeomorphic log map, and data is analyzed there. The geodesic Gaussian kernel is PD because the Riemannian manifold is actually a Euclidean space. In such cases, the Riemannian framework only adds an unnecessary layer of complexity.

**Example 3.** In \([38]\), radial kernels are defined on spheres \(\mathbb{S}^n\) by restricting kernels on \(\mathbb{R}^{n+1}\) to \(\mathbb{S}^n\), giving radial kernels with respect to the chordal metric on \(\mathbb{S}^n\). Due to the symmetry of \(\mathbb{S}^n\), any kernel which is radial with respect to the chordal metric, will also be radial with respect to the geodesic metric on \(\mathbb{S}^n\). This result is next used to define PD radial kernels on the Grassmannian manifold \(\mathcal{G}_n^r\) and on the Kendall shape space \(\mathcal{S}P^n\). However, these kernels are not radial with respect to the usual Riemannian metrics on these spaces, but with respect to the projection distance and the full Procrustes distance, respectively, both of which are not geodesic distances with respect to any Riemannian metric on \(\mathcal{G}_n^r\) and \(\mathcal{S}P^n\), respectively\(^4\). These kernels, thus, have little to do with the Riemannian geometry of \(\mathcal{G}_n^r\) and \(\mathcal{S}P^n\).

**Example 4.** In \([18]\) it is noted that since the feature map \(\phi\) corresponding to a Euclidean Gaussian kernel maps data onto a hypersphere \(\mathbb{S}\) in the reproducing kernel Hilbert space \(V\) \([54]\), it might improve classification to consider the geodesic distance on \(\mathbb{S}\) rather than the chordal distance from \(V\). This is, however, done by projecting each \(\phi(x) \in V\) onto the tangent space \(T_{\phi(x)}\mathbb{S}\) at a fixed base point \(\phi(\hat{x})\), where the linear kernel in \(V\) is employed. This explains why the resulting kernel \(k_\mathcal{E}\) is PD: the kernel linearizes the sphere and, thereby, discards the spherical geometry.

**Example 5.** In \([34]\) and \([40]\), geodesic Laplacian kernels are defined on spheres; as shown above, these are PD.

**Example 6.** In \([36]\), a kernel is defined on a general sample space \(\mathcal{X}\) by selecting a generating probability distribution \(P_\theta\) on \(\mathcal{X}\) and defining a Fisher kernel on \(\mathcal{X}\). Denote by \(M_\theta\) the Riemannian manifold defined by a parametrized family of probability distributions \(P_\theta, \theta \in \Theta, \mathcal{X}\) endowed with the Fisher information metric. The kernel \(k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}\)

\footnotesize\(^4\)Assume that either of these metrics were a Riemannian geodesic distance metric. The family of PD radial kernels defined in \([38]\) on both \(\mathcal{G}_n^r\) and \(\mathcal{S}P^n\) include Gaussian kernels with the projection distance and the full Procrustes distance, respectively. By our previous results, 1) these were geodesic distances with respect to some Riemannian metric, this Riemannian metric would define a Euclidean structure on \(\mathcal{G}_n^r\) and \(\mathcal{S}P^n\), respectively. This is impossible, since these manifolds are both compact.

\footnotesize\(^3\)As a curious side note, this implies that \(\sqrt{\|x - y\|}\) is a metric on \(\mathbb{R}^n\).
is defined by mapping samples in \( X \) to the tangent space \( T_x M_\theta \) and applying the Riemannian metric at \( P_\theta \in M_\theta \). This is PD because the kernel is an inner product on data mapped into a Euclidean tangent space. Again, the statistical manifold is linearized and the resulting kernel does not fully respect its geometry.

In several of these examples the data space is linearized by mapping to a tangent space or into a linear ambient space, which always gives a PD kernel. It should, however, be stressed that the resulting kernels neither respect the distances nor the constraints encoded in the original Riemannian structure. Thus, the linearization will inevitably remove the information that the kernel was aiming to encode.

In general, whenever a data space is embedded into a Euclidean/Hilbert space, and the chordal metric is used in (1), the exponential kernel on a dataset coincides with an exponential kernel on the dataset embedded in the Euclidean/Hilbert ambient space. This therefore gives a PD kernel, and by the Whitney embedding theorem [45], universal kernels can thus be defined on any manifold. These kernels will, however, disregard any constraints encoded by the geodesic distance.

It is tempting to refer to the Nash theorem [48], which states that any Riemannian manifold can be isometrically embedded into a Euclidean space. Here, however, “isometric embedding” refers to a Riemannian isometry, which preserves the Riemannian metric (the smoothly changing inner product) — not to be confused with a distance metric! Therefore, in a Riemannian isometric embedding \( f: X \rightarrow \mathbb{R}^n \) we typically have \( d(x, y) \neq \|f(x) - f(y)\| \). A kernel based on chordal distances in a Nash embedding will, thus, not generally be related to the geodesic distance.

Note, moreover, that the Nash theorem does not guarantee a unique embedding; in fact there are viable embeddings generating a wide range of distance metrics inherited from the ambient Euclidean space. Therefore, an exponential kernel based on the chordal metric will typically have little to do with the intrinsic Riemannian structure of the manifold.

There exist PD kernels that take full advantage of Riemannian geometry without relying on geodesic distances:

**Example 7.** Gong et al. [29] design a PD kernel for domain adaptation using the geometry of the Grassmann manifold: Let \( S_1 \) and \( S_2 \) be two low-dimensional subspaces of \( \mathbb{R}^n \) estimated with PCA on two related data sets. This gives two points \( x_1, x_2 \) on the Grassmann manifold. A test point can be projected into all possible subspaces along the Grassmann geodesic connecting \( x_1 \) and \( x_2 \), giving an infinite dimensional feature vector in a Hilbert space. Gong et al. [29] show how to compute inner products in this Hilbert space in closed-form, thereby providing a PD kernel which takes geometry into account without relying on geodesic distances.

### 5. Experiments

We now validate our theoretical results empirically. First, we generate 500 randomly drawn symmetric PD matrices of size \( 3 \times 3 \). We compute the Gram matrix of both the Gaussian and Laplacian kernels under both the affine-invariant metric [49] and the Fisher information metric on the corresponding fixed-mean multivariate normal distributions [2,4]. Fig. 5a shows the eigenspectrum of the four different Gram matrices. All four kernels have negative eigenvalues, which imply that none of them are positive definite. This empirically proves that neither the affine-invariant metric nor the Fisher information metric induce CND geodesic distance metrics in general, although we know this to hold for the Fisher information metric on \( \text{Sym}_d^+ = \mathbb{R}_+^d \).

Next, we consider kernels on the unit sphere. We generate data from salient points in the 1934 painting *Etude de femmes* by Le Corbusier. At each salient point a HOG [19] descriptor is computed; as these descriptors are normalized they are points on the unit sphere. Fig. 5b shows the eigenspectrum of the Gram matrix of the geodesic Gaussian and Laplacian kernels. While the geodesic Gaussian kernel has negative eigenvalues, the geodesic Laplacian does not. This verifies our theoretical results from Sec. 4 and Table 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{R}^n ) [35, 55]</td>
<td>Euclidean metric</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \mathbb{R}^n, n &gt; 2 ) [35]</td>
<td>( l_q )-norm ( | \cdot |_q, q &gt; 2 )</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sphere ( S^n ) [35]</td>
<td>classical intrinsic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Real projective space ( \mathbb{P}^n(\mathbb{R}) ) [51]</td>
<td>classical intrinsic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Grassmannian</td>
<td>classical intrinsic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \text{Sym}_d^+ )</td>
<td>Frobenius</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \text{Sym}_d^{++} )</td>
<td>Log-Euclidean</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>( \text{Sym}_d^{+++} )</td>
<td>Affine invariant</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Hyperbolic space ( \mathbb{H}^n ) [35]</td>
<td>Fisher information metric</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1-dimensional normal distributions</td>
<td>Fisher information metric</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Metric trees [63], [35, Thm 2.15]</td>
<td>tree metric</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Geometric graphs (e.g., kNN)</td>
<td>shortest path distance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Strings [17]</td>
<td>string edit distance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Trees, graphs</td>
<td>tree/graph edit distance</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 2. For a set of popular metric and manifold data spaces and metrics, we record whether the metric is a geodesic metric, whether it is a Euclidean metric, whether it is a CND metric, and whether its corresponding Gaussian and Laplacian kernels are PD.
We also consider data on the Grassmann manifold. First, we consider one-dimensional subspaces as spanned by samples from a 50-dimensional isotropic normal distribution. We again consider both the Gaussian and the Laplacian kernel; here both under the usual intrinsic metric, but also under the extrinsic metric [30]. Fig. 5c shows the eigenspectra of the different Gram matrices. Only the Gaussian kernel under the intrinsic metric appears to have negative eigenvalues, while the remaining have strictly positive eigenvalues.

Next, we consider 15-dimensional subspaces of $\mathbb{R}^{100}$ drawn from a uniform distribution on the corresponding Grassmannian. We only consider kernels under the intrinsic metric, and the eigenspectra are shown in Fig. 5d. The Gaussian kernel has negative eigenvalues, while the Laplacian kernel does not. Note that this does not prove that the Laplacian kernel is PD on the Grassmannian; in fact, we know theoretically from [51] that it is generally not.

Finally, we consider shortest-path distances on nearest neighbor graphs as commonly used in manifold learning. We take 124 one-digits from the MNIST data set [44], project them into their two leading principal components, form a $\epsilon$-neighborhood graph, and compute shortest path distances. We then compute the eigenspectra of both the Gaussian and Laplacian kernel; Fig. 5e show these spectra. Both kernels have negative eigenvalues, which empirically show that the shortest-path graph distance is not CND.

**6. Discussion and outlook**

We have shown that exponential kernels based on geodesic distances in a metric space or Riemannian manifold will only be positive definite if the geodesic metric satisfies strong linearization properties:

- for Gaussian kernels, the metric space must be flat (or Euclidean).
- for Laplacian kernels, the metric must be conditionally negative definite. This implies that the square root metric can be embedded in a Hilbert space.

With the exception of select metric spaces, these results show that geodesic exponential kernels are not well-suited for data analysis in curved spaces.

This does, however, not imply that kernel methods can never be extended to metric spaces. Gong et al. [29] provide an elegant kernel based on the geometry of the Grassmann manifold, which is well-suited for domain adaptation. This kernel is not a geodesic exponential kernel, yet it strongly incorporates the geodesic structure of the Grassmannian. As an alternative, the Euclidean Gaussian kernel is a diffusion kernel. Such kernels are positive definite on Riemannian manifolds [43], and might provide a suitable kernel. However, these kernels generally do not have closed-form expressions, which may hinder their applicability.

Most existing machine learning tools assume a linear data space. Kernel methods only encode non-linearity via a non-linear transformation between a data space and a linear feature space. Our results illustrate that such methods are limited for analysis of data from non-linear spaces. Emerging generalizations of learning tools such as regression [24, 33, 57] or transfer learning [27, 65] to nonlinear data spaces are encouraging. We believe that learning tools that operate directly in the non-linear data space, without a linearization step, is the way forward.
Acknowledgements

We thank Nicolas Charon, Oswin Krause, Michael Schober and Bernhard Schölkopf for insightful discussions. A.F. is funded in part by the Danish Council for Independent Research (DFF), Technology and Production Sciences. S.H. is funded by the DFF, Natural Sciences.

References


