KMITL Sci. Tech. J. Vol. 17 No. 1 Jan.-Jun. 2017

Geodesic Distance Kernels

Uraiwan Somboon[1*], Praiboon Pantaragphong[1] and Sorin V. Sabau[2]

[1]Department of Mathematics, Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[2]School of Biological Science, Department of Biology, Tokai University,
Sapporo Campus, Sapporo, Japan

Abstract

In this paper, the authors deals with non-symmetric kernels induced by weighted quasi-metrics on Hilbert spaces and they study their fundamental properties. These are new and original. Such kind of metrics is obtained from Finsler metrics for example. We show that the use of such kernels may provide a solution to the conflict between positive definiteness of the kernel and the curvature of the underlying space.

Keywords: Finsler metrics, Hilbert spaces, non-symmetric kernels, weighted quasi-metric spaces

1. Introduction

Kernel methods are fundamental tools for statistical analysis and machine learning [1, 2]. Considering the data set as a subset $\mathbb{R}^n$ one can use an embedding into the higher dimensional Hilbert space where the problem becomes linear and hence easy to solve. Even though the data is usually regarded as lying on the Euclidean space, in many cases, one needs to work with data sitting in more general type of space, in particular data lying on spaces that are not necessarily flat. Examples include data analysis for computer vision, where rotation metrics belong to the Lie group SO (3), normalized histograms from the unit n-sphere S^n and other type of data that belong to smooth manifolds, as Riemannian and Finsler manifolds.

In the present paper, we will show how to extend the theory of kernels using the geodesic distance from Euclidean and Riemannian setting to the much general case of a weighted quasi-distance. It is known that this kind of distance naturally appears in the case of a special Finsler manifold called Randers space [3].

In particular, we are interested in answering to the following question. "Is it possible to use (non-symmetric) kernel methods in order to analyze the data on a curve manifolds as S^n ?" It is clear that the answer to this question is NO for the symmetric kernels case.

However, if we consider geodesic kernel based on weighted quasi distance, then the bundle representation [4] allows us to extend the space where the data lives. By such a decompactification of the base manifold it is clearly possible to use the geometry of non-symmetric kernels for analyzing data on curved manifolds, as S^n .

*Corresponding author: Tel.: +66 91 7300313
E-mail: uraiwan.somboon@gmail.com

Here is the structure of the paper. In section 2 (subsection 2.1), we review the main results about symmetric kernels defined on topological spaces and recall that PD and CND properties of the symmetric kernels imply that the base topological spaces are actually Hilbert spaces. Moreover, we explain in subsection 2.2 why the underlying metric space of a PD symmetric kernel must be flat and hence why one cannot use PD symmetric kernels when working with data extracted from spaces that are not flat.

In section 3, we recall the main geometrical results of weighted quasi-metrics (see [3] and [4]). In special we call the attention to the bundle representation of weighted quasi-metric, a fundamental notion for the generalizations following.

Finally, in section 4, we define for the first time non-symmetric kernels induced by weighted quasi-metrics, study their fundamental properties and show that these can allow using data from more general spaces than the flat ones. All the content of this section is new and makes the core of the present paper. Further details and some concrete algorithms will be given in a forthcoming paper.

2. Symmetric Kernels

2.1 General theory

Let X be a topological space.

Definition2.1 A continuous function $\Phi : X \times X \to \mathbb{C}$ or $\mathbb{R}$, $\Phi(x,x) = 0$, $\Phi(x,y) = \Phi(y,x)$ is called a *symmetric kernel* on X.

We recall the following definition.

Definition 2.2 ([5]) The kernel K on the topological space X is called *positive definite* (PD) if for any $n \in \mathbb{N}$, and elements $x_1,...,x_n \in X$ and any scalars $c_1,...,c_n \in \mathbb{R}$ we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K\left(x_i, x_j\right) \geq 0 \quad . \tag{2.1}$$

Example 2.3 Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $f : X \to \mathcal{H}$ acontinuous function.

Then the kernel $\Phi(x,y) = \langle f(x), f(y)\rangle, \forall x, y \in X$ is positive definite.

Remark 2.4 Obviously $\Phi(x,y) = \langle x, y\rangle$ is positive definite kernel.

Definition 2.5 ([5]) The kernel K on the topological space X is called *conditionally negative definite* (CND) if it satisfies

1. $K(x,x) = 0$ for all $x \in X$.
2. $K(x,y) = K(y,x)$ for all $x, y \in X$.
3. Forany $n \in \mathbb{N}$, and elements $x_1,...,x_n \in X$ and any real numbers $c_1,...,c_n$ with $c_1 + c_2 + \cdots + c_n = 0$ we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K\left(x_i, x_j\right) \leq 0 \; . \tag{2.2}$$

For symmetric kernels the following results are fundamental.

Theorem 2.6 ([5]) **(The GNS construction for PD kernels)** *If $\Phi : X \times X \to \mathbb{R}$ is a PD kernel on a topological space X then there exist*

- *a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$*
- *a continuous function $f : X \to \mathcal{H}$*

such that $\Phi(x,y) = \langle f(x), f(y)\rangle$

The following result is important for applications and will be extensively used in the presentpaper.

Theorem 2.7 ([5]) **(The GNS construction for CND kernels)** *If* $\Psi : X \times X \to \mathbb{R}$ *is a CND kernel on a topological space X, then there exist*

- *a real Hilbert space* $(\mathcal{H}, \langle \cdot, \cdot \rangle)$
- *a continuous function* $f : X \to \mathcal{H}$

such that $\psi(x, y) = \| f(x) - f(y) \|^2$.

The GNS construction of CND kernels allows to prove the following important result.

Theorem 2.8 ([5]) **(Schoenberg)** *If* X *is a topologicalspace and* $\Psi : X \times X \to \mathbb{R}$ *is a continuouskernel on* X *such that*

(i) $\Psi(x, x) = 0,$ *and*

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

then the followingtwoproperties are equivalent

(A) Ψ *is CND kernel*

(B) *the exponentialkernel* $K(x, y) := exp(-\lambda \cdot \Psi(x, y))$ *is PD for all* $\lambda \geq 0$.

2.2 Geodesic distance induced kernels

In particular, if (M, ρ) is a metric space, then it is customary to consider kernels induced by ρ given in the form: $K(x, y) = \exp\left(-\lambda \cdot \rho^q(x, y)\right), \quad \lambda, q > 0$**.**The Schoenberg theorem and GNS construction for CND kernels imply**.**

Theorem 2.9 *If the Gaussiankernel* $K(x, y) = \exp\left(-\lambda \cdot \rho^2(x, y)\right)$ *of the metric space* (M, ρ) *is PD, then there exist*

- *a real Hilbert space* $(\mathcal{H}, \langle \cdot, \cdot \rangle)$
- *a continuous function* $f : X \to \mathcal{H}$

suchthat $\rho(x, y) = \| f(x) - f(y) \|_{\mathcal{H}} = d_{\mathcal{H}}\left(f(x), f(y) \right),$

where $d_{\mathcal{H}}$ *is the induced distance of* $(\mathcal{H}, \langle \cdot, \cdot \rangle)$.

Remark 2.10 In other words, if the Gaussian geodesic kernel is PD, then the metric space (M, ρ) can be isometrically embedded in the Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ constructed in the GNS construction for CND kernels (see above).

The following result is well-known [1, 2].

Theorem 2.11 *Let* (X, d) *be a geodesic metric space and assume that the Gaussian kernel* $K(x, y) = \exp\left(-\lambda \cdot d^2(x, y)\right)$ *is PD on* X *for all* $\lambda > 0.$ *Then* (X, d) *is flat in the sense of Alexandrov, i.e. any geodesic triangle in* (*X*, *d*) *can be isometrically embedded in an Euclidean space.*

Theorem 2.12 *Let* (M, g) *be a complete, smooth Riemannian manifold with Riemannian distance function d on M. Let us assume that the Riemannian distance induced Gaussian kernel* $K(x, y) = \exp\left(-\lambda \cdot d^2(x, y)\right)$ *is PD on* X *for all* $\lambda > 0$ *. Then the Riemannian manifold* M *is isometric to an Euclidean space.*

From here it follows that the geodesic Gaussian kernel can be PD only if the underlying spaceis flat. In particular, if the distance is induced by a Riemannian metric, then the Gaussian kernelis PD if and only if the Riemannian space is flat, i.e. an Euclidean space.

The geometrical reason behind this unexpected result actually comes from the injectivity of the isometric embedding. Indeed, we recall from [6] that, if (X, d_x) and (Y, d_y) are metric spaces, then a map $\varphi : X \to Y$ is called *isometry onto its image* if it preserves distance, that is

$$d_Y(\varphi(x), \varphi(y)) = d_X(x, y) \text{ for any } x, y \in X . \tag{2.3}$$

Remark that the definition above automatically implies that φ must be injective. Here is a simple proof of this fact.

Recall that φ is injective by definition if, for any $x_1, x_2 \in X$, $\varphi(x_1) = \varphi(x_2)$, then $x_1 = x_2$.

If we assume $\varphi(x_1) = \varphi(x_2)$, then formula (2.3) implies

$$d_Y(\varphi(x_1), \varphi(x_2)) = 0 = d_X(x_1, x_2)$$

And since d_X is a metric, it follows $x_1 = x_2$, i.e. the isometry φ must be an injection.

A map φ betweentwolengthspacesiscalled an *arcwise isometry* if

$$\mathcal{L}_X(\gamma) = \mathcal{L}_Y(\varphi(\gamma)) \text{ for any path } \gamma.$$

An injective arcwise isometry iscalled an *isometric embedding.*

Remark 2.13

1. An isometric embedding is not the same notion as isometry onto its image. For instance, a simple curve $\gamma : [0,1] \to \mathbb{R}^2$, $\gamma(t) = (\cos t, \sin t)$ is an isometric embedding, but not an isometry onto its image.

2. The isometric embeddings of Riemannian spaces are studied in Differential Geometry and they are actually arcwise isometric embeddings.

Example 2.14 Let us consider the unit sphere

$$S^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\} \textit{ in } \mathbb{R}^3.$$

Then

1. $S^2 \to \mathbb{R}^3$ can be isometrically embedded in $\mathbb{R}^3$ as Riemannian manifold.

2. On the other hand, there exists no embedding $\phi : S^2 \to \mathbb{R}^3$ $(\textit{or } \mathbb{R}^n)$ that would be an isometry onto its image.

Indeed, consider the coordinate system (θ, φ) on S^2 given by

$$x = \sin\theta \cdot \cos\varphi, \; y = \sin\theta \cdot \sin\varphi, \; z = \cos\theta,$$

where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$.

We have the Riemannian isometric embedding

$$\phi : S^2 \to \mathbb{R}^3, (\theta, \varphi) \mapsto \phi(\theta, \varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$$

and observe that this map cannot be injective. One can see for instance that

$$\phi(0, \varphi) = (0, 0, 1) \text{ for any } \varphi \in [0, 2\pi].$$

In general, if the manifold X , where data belongs, is compact, it is impossible to find an isometry between X and $\mathbb{R}^n$, therefore is impossible to obtain a PD Gaussian kernel from the geodesic distance on a compact manifold.

Remark 2.15 It is interesting to see that actually we can embed S^2 as metric space into an infinite dimensional Hilbert space [6].

3. Weighted Quasi-Metric

We recall that, if M is a non-empty set and d areal-valued function
$d: M \times M \to \mathbb{R}$ that satisfies

1. Positiveness: $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$
2. Symmetry: $d(x, y) = d(y, x)$.
3. Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$

for any $x, y, z \in M$, then (M, d) is called *a metric space*.

More generally, a metricspace (M, d) that do not satisfy the symmetry condition $d(x, y) = d(y, x)$ is called *aquasi-metric space.*

Moreover, an important class of quasi-metric space are the so-called *weighted quasi-metric space*. A *weighted quasi-metric* ρ is a function $\rho: M \times M \to \mathbb{R}$. This is quasi-metric and such that there exists a *weight function* $w: M \to [0, \infty)$ satisfying

Weightability: $\rho(x, y) + w(x) = \rho(y, x) + w(y)$ for any $x, y \in M$.

We can define the symmetrization of ρ, where $\rho: M \times M \to \mathbb{R}$, is a quasi-metric. Indeed the function $d: M \times M \to [0, \infty)$ given by

$$d(x, y) := \frac{1}{2}\left[\rho(x, y) + \rho(y, x)\right], \tag{3.1}$$

Is called *the symmetrization* of ρ for any $x, y \in M$.

Lemma 3.1 *Observe that* (M, d) *is metric space.*

Proposition 3.2 *Using the symmetrization* d *of weight of quasi-metric* ρ *with the weight function* $w: M \to [0, \infty)$ *we have* $\rho(x, y) = d(x, y) + \frac{1}{2}\left[w(y) - w(x)\right]$ *for any* $x, y \in M$.

Moreover, we have $\frac{1}{2}\left|w(x) - w(y)\right| \leq d(x, y)$ *for any* $x, y \in M$.

Proposition 3.3 *If* (M, d, w) *is a weighted quasi-metric space, then the perimeter length of any geodesic triangle on* M *does not depend on the orientation, i.e.*

$$\rho(x, y) + \rho(y, z) + \rho(z, x) = \rho(x, z) + \rho(z, y) + \rho(y, x) \text{ for any } x, y \in M. \tag{3.2}$$

If (X, q, w) and (Y, p, u) are two weighted quasi-metric spaces, the mapping $\varphi: X \to Y$ with the properties

$$p(\varphi(x), \varphi(y)) \leq q(x, y), \quad \forall x, y \in X \tag{3.3}$$

$$u(\varphi(x)) \leq w(x), \quad \forall x \in X \tag{3.4}$$

Is called a *morphism* of weighted quasi-metric spaces.

In the case we have equality in relation (3.3), then the morphism ϕ is called an *isometric morphism*. In this case w and $u \circ \varphi$ differ by a constant only.

Moreover, an *isomorphism* of the weighted quasi-metric spaces (X, q, w) and (Y, p, u) is a bijective function $\varphi: X \to Y$ that preserves both the quasi-metric and the weight function.

Finally, an *embedding* of (X, q, w) into (G, Q, W) is an isomorphism of (X, q, w) onto a subspaceof (G, Q, W). Here, a *subspace* (Y, p, u) *of a weighted quasi-metric space* (G, Q, W) is a subset $Y \subset G$, the function p and u are the restriction of Q and W to $Y \times Y$ and Y, respectively

Example 3.4 (The product of a metricspacewith a half ray) Consider a metric space (S,d) and the half ray $I := [0,\ \infty)$. Then the product space $G := S \times I$ inherits a natural structure of (generalized) weighted quasi-metric space (G, Q, W), where

$$\begin{aligned} &Q : G \times G \to [0,\infty),\ Q(u,\ v) := d(x,\ y) + \eta - \xi, \\ &W : G \to [0,\infty),\ W(u) := 2\xi, \quad \forall u = (x,\ \xi),\ v = (y,\ \eta) \in S \times I. \end{aligned} \tag{3.5}$$

Remark 3.5 The generalized weighted quasi-metricspace $(S \times I, Q, W)$ constructed in example 3.4 is sometimes called *the bundle over* (S,d) (see [4]).

Example 3.6 (The graph of a function) We consider the case of the graph of anon-negative valued function $f : S \to [0,\infty)$ defined on a metric space (S,d). Indeed, if we denote the graph of f by $G_f := \{(x,\ f(x)) : x \in S\}$ then (G_f, Q, W) is a naturally induced weighted quasi-metric space structure defined by

$$\begin{aligned} &Q : G_f \times G_f \to [0,\infty),\ Q(u,v) := d(x,\ y) + f(y) - f(x), \\ &W : G_f \to [0,\infty),\ W(u) := 2f(x),\ \forall u = (x,\ f(x)),\ v = (y,\ f(y)) \in G_f. \end{aligned} \tag{3.6}$$

Based on these, one has

Theorem 3.7 ([3]) *Every weighted quasi-metric space (X,q,w) is embeddable in a bundle over a suitable metric space (S,d).*

Theorem 3.8 ([3])

1. *Let (S,d) be a metric space and $f : S \to [0,\infty)$ a 1-Lipschitz function. Then the graph of f is a weighted quasi-metric space (G_f, Q, W).*

2. *Conversely, every weighted quasi-metric space (X,q,w) can be constructed in this way.*

4. Non-Symmetric Kernels

All kernels studied up to here were symmetric kernels, and basically, induced by Riemannian distances, which are also symmetric.

We will consider in this section another type of kernels, namely non-symmetric kernels. In special, we are interested in non-symmetric kernels induced by weighted quasi-metrics. If (M, ρ, w) is a weighted quasi-metric, we can consider the kernels

- $\mathcal{G}(x,\ y) := \exp(-\lambda \cdot \rho^2(x,\ y))$, the Gaussian kernel and
- $\mathcal{L}(x,\ y) := \exp(-\lambda \cdot \rho(x,\ y))$, the Laplacian kernel.

The definition of PD and CND kernels are same as for the symmetric case.

We start with two examples.

Example 4.1 Let $(\mathcal{H}, \langle\cdot,\ \cdot\rangle)$ be a Hilbert space and $f : M \to \mathcal{H}$ a continuous function. Then the function

$$K(x,y) := \langle f(x),\ f(y)\rangle + \frac{1}{2}\left[w(y) - w(x)\right],\quad \forall x,y \in M \tag{4.1}$$

is a non-symmetric PD kernel, where $w : M \to [0,\ \infty)$ is a 1-Lipschitz continuous function.

Example 4.2 With same notation as in example (4.1), the function

$$K(x,y) := \left[d_{\mathcal{H}}(x,y) + \frac{1}{2}\big(w(y) - w(x)\big) \right]^2 \qquad (4.2)$$

is a non-symmetric CND kernel, where $d_{\mathcal{H}}(x,y) = \sqrt{\langle x-y,\, x-y \rangle}$ is the induced distance function of the Hilbert space $(\mathcal{H}, \langle \cdot,\ \cdot \rangle)$.

In order to prove this example, we need the following.

Lemma 4.3 *If* $w: \mathcal{H} \to \mathbb{R}$ *is a continuous function and* $c_1 + c_2 + \cdots + c_n = 0,$ *the*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left[w(x_i) - w(x_j) \right]^2 = -2 \cdot \left[\sum_{i=1}^{n} c_i w(x_i) \right]^2 \le 0 \textit{ for all } x_1, \ldots, x_n \in \mathcal{H}. \qquad (4.3)$$

Proof. For $n = 2$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left[w(x_i) - w(x_j) \right]^2 = c_1^2 \left[w(x_1) - w(x_1) \right]^2 + c_1 c_2 \left[w(x_1) - w(x_2) \right]^2 +$$

$$c_1 c_2 \left[w(x_2) - w(x_1) \right]^2 + c_2^2 \left[w(x_2) - w(x_2) \right]^2$$

$$= 2 c_1 c_2 \left[w(x_1) - w(x_2) \right]^2.$$

Since $c_1 + c_2 = 0,\ c_1 c_2 = -c_1^2$. Then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left[w(x_i) - w(x_j) \right]^2 = -2 \cdot c_1^2 \left[w(x_1) - w(x_2) \right]^2 \le 0.$$

For $n = 3$:

$$\sum_{i=1}^{3}\sum_{j=1}^{3} c_i c_j \left[w(x_i) - w(x_j) \right]^2 = 2\Big[\left(c_1 c_2 + c_1 c_3 \right) w\left(x_1\right)^2 + \left(c_1 c_2 + c_2 c_3 \right) w\left(x_2\right)^2 + \left(c_1 c_3 + c_2 c_3 \right) w\left(x_1\right)^2$$

$$-2 c_1 c_2 w(x_1) w(x_2) - 2 c_1 c_3 w(x_1) w(x_3) - 2 c_2 c_3 w(x_2) w(x_3) \Big].$$

Since $c_1 + c_2 + c_3 = 0,$ we have

$$c_1 c_2 + c_1 c_3 = -c_1^2,\ c_1 c_2 + c_2 c_3 = -c_2^2,\ c_1 c_3 + c_2 c_3 = -c_3^2.$$

Then

$$\sum_{i=1}^{3}\sum_{j=1}^{3} c_i c_j \left[w(x_i) - w(x_j) \right]^2 = -2\Big[c_1^2 w\left(x_1\right)^2 + c_2^2 w\left(x_2\right)^2 + c_3^2 w\left(x_3\right)^2$$

$$+ 2 c_1 c_2 w(x_1) w(x_2) + 2 c_1 c_3 w(x_1) w(x_3) + 2 c_2 c_3 w(x_2) w(x_3) \Big]$$

$$= -2\left[c_1 w(x_1) + c_2 w(x_2) + c_3 w(x_3) \right]^2 \le 0.$$

Induction step: let k be positive integers and suppose (4.3) is CND for $n = k$, we will show that (4.3) is CND for $n = k+1$.

Thus

$$\sum_{i=1}^{k+1}\sum_{j=1}^{k+1} c_i c_j \left[w(x_i) - w(x_j) \right]^2 = 2\Big[\left(c_1 c_2 + c_1 c_3 + \cdots + c_1 c_{k+1} \right) w\left(x_1\right)^2 + \left(c_1 c_2 + c_2 c_3 + \cdots + c_2 c_{k+1} \right) w\left(x_2\right)^2$$

$$+ \cdots + \left(c_1 c_{k+1} + c_2 c_{k+1} + \cdots + c_k c_{k+1} \right) w\left(x_{k+1}\right)^2 - 2 c_1 c_2 w(x_1) w(x_2)$$

$$-2 c_1 c_3 w(x_1) w(x_3) - \cdots - 2 c_k c_{k+1} w(x_k) w(x_{k+1}) \Big].$$

Since $c_1 + c_2 + \cdots + c_{k+1} = 0,$ we have

$c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1} = -c_1^2,\ c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1} = -c_2^2, ...,\ c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1} = -c_{k+1}^2.$

Then $\sum_{i=1}^{3}\sum_{j=1}^{3} c_ic_j\left[w(x_i) - w(x_j)\right]^2 = -2\left[c_1w(x_1) + c_2w(x_2) + \cdots + c_{k+1}w(x_{k+1})\right]^2 \le 0.$

Proposition 4.4

(i) *If K_1 and K_2 are CND, then $s{\cdot}K_1 + t{\cdot}K_2$ is also CND, for all $t,\ s > 0$, i.e. the set of CND kernels on M is a convex cone.*

(ii) *If $\{K_t\}_t$ is a family of CND kernels converging point-wise on $M \times M$ to a continuous kernel $K : M \times M \to \mathbb{R}$, then K is also CND, i.e. the set of CND kernels on M is closed.*

(iii) *If K is a PD kernel on M, then $H(x,y) = K(x,x) - K(x,y)$* is CND.

Proof. (i) For any $x_1,...,\ x_n$ and $c_1,...,\ c_n$ such that $c_1 + \cdots + c_n = 0$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK_1\left(x_i, x_j\right) \le 0,\quad \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK_2\left(x_i, x_j\right) \le 0.$$

Then,

$$\begin{aligned} 0 &\ge s\cdot\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK_1\left(x_i, x_j\right) + t\cdot\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK_2\left(x_i, x_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j\cdot\left(s\cdot K_1 + t\cdot K_2\right)\left(x_i, x_j\right). \end{aligned}$$

(ii) By hypothesis, for any $(x,y) \in M \times M$, we have for any $\varepsilon > 0$, there exists $T > 0$ such that $\left|K_t(x,\ y) - K(x,\ y)\right| < \varepsilon$, for $t > T$. For any $x_1,...,\ x_n \in M$, we have

$$-\varepsilon < K(x_i,\ x_j) - K_t(x_i,\ x_j) < \varepsilon,\quad -\varepsilon < K_t(x_i,\ x_j) - K(x_i,\ x_j) < \varepsilon$$

and hence, for any $c_1,...,c_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} c_i = 0$, we obtain

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j\cdot\varepsilon < \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK_t\left(x_i, x_j\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK\left(x_j, x_i\right) < \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j\cdot\varepsilon$$

and using hypothesis that $\sum_{i=1}^{n} c_i = 0$, it follows

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK_t\left(x_i, x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK\left(x_j, x_i\right) \le 0.$$

(ii) For any $c_1,...,\ c_n \in \mathbb{R}$, $\sum_{i=1}^{n} c_i = 0$, and $x_1,...,\ x_n \in M$, we compute

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jH\left(x_i, x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK\left(x_i, x_i\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK\left(x_i, x_j\right) = -\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK\left(x_i, x_j\right) \le 0$$

due to the fact that K is PD.

Remark 4.5 Let $K : M \times M \to \mathbb{R}$ be an arbitrary non-symmetric kernel and let us denote

$$H : M \times M \to R,\ H(x,y) = \frac{1}{2}\left[K(x,y) + K(y,x)\right]$$

the average symmetrized kernel.

Then, an elementary computation shows that for any $x_1,...,x_n \in M$ and any $c_1,...,c_n \in \mathbb{R}$, we have $\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jK\left(x_i, x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_jH\left(x_j, x_i\right).$

That is,

(i) K is PD if and only if H is PD.
(ii) K is CND if and only if H is CND.

We obtain the following important result.

Lemma 4.6 *(Fundamental Lemma)*

(i) $\rho(x,y)$ *is PD if and only if* $d(x,y)$ *is PD.*

(ii) $\rho(x,y)$ *is CND if and only if* $d(x,y)$ *is CND.*

The proof is trivial if we take into account that $d(x,y)=\frac{1}{2}\left[\rho(x,y)+\rho(y,x)\right]$.

Example 4.7 Let $(\mathcal{H},\langle\cdot,\ \cdot\rangle)$ be a Hilbert space. We define the following weighted quasi-metric space; $(\mathcal{H}\times[0,\infty),\bar{Q},\bar{W})$, where

$$\bar{Q}:(\mathcal{H}\times[0,\infty))\times(\mathcal{H}\times[0,\infty))\to[0,\infty)\ ,\ (u,v)\mapsto\bar{Q}(u,v):=\langle x,y\rangle+\eta-\xi,$$

Where $u=(x,\xi)$, $v=(y,\eta)$ are points in $\mathcal{H}\times[0,\infty)$ .Obviously $\bar{Q}$ is a quasi-metric on $\mathcal{H}\times[0,\infty)$.

Moreover, the mapping $\bar{W}:\mathcal{H}\times[0,\infty)\to[0,\infty),\ u\mapsto\bar{W}(u)=2\cdot\xi$ is a weight, where $u=(x,\xi)\in\mathcal{H}\times[0,\infty)$.

One can now easily see that $(\mathcal{H}\times[0,\infty),\bar{Q},\bar{W})$ is a weighted quasi-metric space. We will call it the associated weighted quasi-metric space to a given Hilbert space $(\mathcal{H},\langle\cdot,\ \cdot\rangle)$.

We obtain an important result.

Theorem 4.8 *If the weighted quasi-metric* ρ *is PD then there exists a continuous function*

$$\varphi:\ (G_f,Q,W)\to(\mathcal{H}\times[0,\infty),\bar{Q},\bar{W})$$

from the representation (G_f,Q,W) *of* (M,ρ) *to the weighted quasi-metric space* $(\mathcal{H}\times[0,\infty),\bar{Q},\bar{W})$ *which is an isomorphism of bundles, where* $(\mathcal{H}\times[0,\infty),\bar{Q},\bar{W})$ *is the associated weighted quasi-metric of the Hilbert space* $\mathcal{H}$ *obtained by the GNS construction for PD kernels.*

Proof. Since ρ is PD it follows that d is also PD. Hence, from the GNS construction theorem of PD kernels it follows that it exists a Hilbert space $(\mathcal{H},\langle\cdot,\ \cdot\rangle)$ and a continuous mapping $\psi:M\to\mathcal{H}$ such that $d(x,y)\ =\langle\psi(x),\ \psi(y)\rangle,\ \ \forall x,\ y\in M$.

We will extend this mapping to G_f by defining

$$\varphi:G_f\to\mathcal{H}\times[0,\infty),\ \ u\mapsto\varphi(u)\ =\varphi(x,\ f(x))\ =\ (\psi(x),\ f(x)).$$

We will show now that this is an isomorphism of (G_f,Q,W) with $\left(\mathcal{H}\times[0,\ \infty)\subset\mathbb{R}^n\left(\mathbb{C}^n\right)\times\mathbb{R},\bar{Q},\bar{W}\right)$, where

$$Q\left((X,\ a),(Y,\ b)\right)=\langle X,\ Y\rangle+b-a,\ \text{ and } \bar{W}(x,a)=a\ .$$

We have, $Q(u,v)=Q\left((x,\ f(x)),(y,\ f(y))\right)=d(x,y)+\ f(y)-\ f(x)$

$=\langle\psi(x),\ \psi(y)\rangle+\ f(y)-\ f(x)\ =\bar{Q}(\varphi(u),\varphi(v))$

and similarly $W(u)=W\left(\left(x,\ f(x)\right)\right)=\ f(x)\ \ =W\left(\psi(x),\ f(x)\right)=\bar{W}\left(\varphi(u)\right)$.

We have the following general result.

Proposition 4.9 *Let* $\psi:M\times M\to\mathbb{R}$ *be a (non-symmetric) kernel on* M *. If* $e^{-\lambda\cdot\psi}$ *is PD, then* ψ *is CND.*

Let us consider Laplacian kernels induced by weighted quasi-metrics. We have

Theorem 4.10

(i) *The weighted quasi-metric space* (M, ρ, w) *is CND if and only if Laplacian kernel* $e^{-\lambda \cdot \rho(x,y)}$ *is PD for all* $\lambda > 0$.

(ii) *In this case we have the bundles isomorphism* $\varphi : (G_f, Q, W) \to (\mathcal{H} \times [0,\infty), \bar{Q}, \bar{W})$ *Where* $Q(u,v) = \sqrt{d(x,y)} + f(y) - f(x)$, *and* $(\mathcal{H} \times [0,\infty), \bar{Q}, \bar{W})$ *is the associated weighted quasi-metric of the Hilbert space* $\mathcal{H}$ *obtained by the GNS construction for CND kernels.*

Proof. (i) We assume $\mathcal{L}(x,y) = e^{-\lambda \cdot \rho(x,y)}$ is PD. Then we observe that $\Phi = 1 - e^{-\lambda \cdot \rho} = 1 - \mathcal{L}$ is CND.

Indeed, for any $c_1, ..., c_n$ such that $c_1 + \cdots + c_n = 0$ we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left[1 - e^{-\lambda \cdot \rho(x_i,x_j)}\right] = -\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \left[e^{-\lambda \cdot \rho(x_i,x_j)}\right] \le 0$$

by hypothesis. Using now the fact that the set of CND kernels on M is a closed set we have that $\lim_{t\to 0}\frac{1-\mathcal{L}}{t} = \lim_{t\to 0}\frac{1-e^{-t\cdot\rho(x,y)}}{t}$ is also CND kernel. Recalling from L'Hospital's theorem $\lim_{t\to 0}\frac{1-e^{-t\cdot\rho(x,y)}}{t} = \rho(x,y)$, it follows that $\rho(x,y)$ is CND kernel.

Conversely, we assume now that (M, ρ) is CND. Then it follows that (M, d) is CND and hence there exists $\varphi : M \to (\mathcal{H}, \langle \cdot, \cdot \rangle)$ such that $d(x,y) = \langle \varphi(x), \varphi(y) \rangle$.

A simple computation shows that $e^{-\lambda \cdot \rho(x,y)} = e^{-\lambda \cdot d(x,y)} \cdot e^{-\frac{\lambda}{2}[w(y)-w(x)]}$.

Remark that since $(M, d_{\mathcal{H}})$ is CND it follows that $e^{-\lambda \cdot d(x,y)} = e^{-\lambda \cdot d^2_{\mathcal{H}}(\varphi(x),\varphi(y))}$ is PD by Schoenberg Lemma for symmetric kernels. For any $x_1, ..., x_n$ and any $c_1, ..., c_n$ we compute

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j e^{-\lambda \cdot \rho(x_i,x_j)} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j e^{-\lambda \cdot d(x_i,x_j)} \cdot e^{-\frac{\lambda}{2}\left[w(x_j)-w(x_i)\right]}.$$

For $n = 2$ we have,

$$c_1^2 e^{-\lambda\cdot\rho(x_1,x_1)} + c_1 c_2 e^{-\lambda\cdot\rho(x_1,x_2)} + c_2 c_1 e^{-\lambda\cdot\rho(x_2,x_1)} + c_2^2 e^{-\lambda\cdot\rho(x_2,x_2)}$$

$$= c_1^2 e^{-\lambda\cdot d(x_1,x_1)} + c_1 c_2 e^{-\lambda\cdot d(x_1,x_2)} + c_2^2 e^{-\lambda\cdot d(x_2,x_2)} + c_1 c_2 e^{-\lambda\cdot d(x_1,x_2)} \left\{ e^{-\frac{\lambda}{2}[w(x_2)-w(x_1)]} + e^{-\frac{\lambda}{2}[w(x_1)-w(x_2)]} - 1 \right\}.$$

Since d is CND, then $e^{-\lambda \cdot d(x,y)}$ is PD, i.e.

$$c_1^2 e^{-\lambda\cdot d(x_1,x_1)} + c_1 c_2 e^{-\lambda\cdot d(x_1,x_2)} + c_2 c_1 e^{-\lambda\cdot d(x_2,x_1)} + c_2^2 e^{-\lambda\cdot d(x_2,x_2)} \ge 0 . \tag{4.4}$$

On the other hand, observe that

$$e^{-\frac{\lambda}{2}[w(x_2)-w(x_1)]} + e^{-\frac{\lambda}{2}[w(x_1)-w(x_2)]} - 1 > 0. \tag{4.5}$$

Indeed, if we denote $e^{-\frac{\lambda}{2}[w(x_2)-w(x_1)]} = Y$ then

$$e^{-\frac{\lambda}{2}[w(x_2)-w(x_1)]} + e^{-\frac{\lambda}{2}[w(x_1)-w(x_2)]} - 1 = Y + \frac{1}{Y} - 1 = \frac{Y^2 - Y + 1}{Y} > 0, \tag{4.6}$$

Because $Y^2 - Y + 1 > 0$, $Y > 0$ for any Y and from (4.4), (4.5) it results that $e^{-\lambda\cdot\rho(x,y)}$ is PD.

(ii) Recall that (M, ρ) is CND and hence (M, d) is CND. Then, by using GNS construction for CND symmetric kernels, there exists $\varphi : (M, d) \to (\mathcal{H}, \langle \cdot, \cdot \rangle)$ such that

$$d(x,y) = d^2_{\mathcal{H}}(\varphi(x), \varphi(y)) = \langle \varphi(x) - \varphi(y),\ \varphi(x) - \varphi(y) \rangle \tag{4.7}$$

and hence, we have an isometric embedding (see [6]) $\varphi:\left(M,\sqrt{d}\right)\to\left(\mathcal{H},\ d_{\mathcal{H}}\right)$.

It follows that we can construct $Q(x,y)=\sqrt{d(x,\ y)}+f(y)-f(x)$ and the conclusion follows similarly with theorem 4.8.

We turn now to Gaussian kernels $\mathcal{G}(x,\ y)\ =e^{-\lambda\cdot\rho(x,y)},\ \lambda>0.$

Theorem 4.11 *If* $\left(M,d^2\right)$ *is CND then,*

(i) $\left(M,\rho^2\right)$ *is CND, for* $\rho(x,\ y)=d(x,\ y)+f(y)-f(x)$ and

(ii) *there exists a continuous function* $\varphi:\left(G_f,Q,W\right)\to\left(\mathcal{H}\times[0,\ \infty),\bar{Q},\bar{W}\right)$ *that is an isomorphism of bundles.*

Proof. (i) Using $\rho(x,y)=d(x,y)+f(y)-f(x)$ we have

$$\rho^2(x,y)\ =d^2(x,y)+2\cdot d(x,\ y)\cdot\left[f(y)-f(x)\right]+\left[f(y)-f(x)\right]^2.$$

And for any $x_1,...,x_n$ and any $c_1,...,c_n$ such that $c_1+\cdots+c_n=0$ we have

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\rho^2(x_i,\ x_j)&=\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\,\frac{\rho^2(x_i,\ x_j)+\rho^2(x_j,\ x_i)}{2}\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jd^2(x_i,\ x_j)+\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\left[f(x_i)-f(x_j)\right]^2\\&=\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jd^2(x_i,\ x_j)-2\cdot\left[\sum_{i=1}^{n}c_if(x_i)\right]^2,\end{aligned}$$

due to lemma 4.3. Since d^2 is CND it results $\frac{1}{2}\left[\rho^2(x,y)+\rho^2(y,x)\right]$ is CND and due to remark 4.5, ρ^2 is CND.

(ii) Since $\left(M,d^2\right)$ is CND, by the GNS construction for symmetric kernels it follows that there exists $\varphi:M\to\left(\mathcal{H},\langle\cdot,\ \cdot\rangle\right)$ continuous function such that

$$d^2(x,y)\ =\left\|\varphi(x)-\varphi(y)\right\|_{\mathcal{H}}^2=d_{\mathcal{H}}^2(\varphi(x),\ \varphi(y)),\tag{4.8}$$

that is $d(x,y)=d_H\left(\varphi(x),\varphi(y)\right)$.

Then, we can extend this function to $\varphi:\left(G_f,Q,W\right)\to\left(\mathcal{H}\times[0,\ \infty),\bar{Q},\bar{W}\right)$

by $\varphi(u)=\varphi(x,\ w(x))=\ (\varphi(x),2f(x))$. By computing, we get

$$Q(u,v)=d(x,y)+f(y)-f(x)=d_H\left(\varphi(x),\ \varphi(y)\right)+w(y)-w(x)=\bar{Q}(\varphi(u),\ \varphi(v)),$$

and $W(u)=W\left(x,\ w(x)\right)=\ 2\cdot f(x)\ =\bar{W}\left(\varphi(x)\right)$.

Remark 4.12 The natural question left is if ρ^2 is CND implies d^2 is CND. One can see that this is not true in general. Indeed, let us assume that ρ^2 is CND. Then, a computation similar to the one in proof of theorem 4.11 gives

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\rho^2(x_i,\ x_j)=&\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jd^2(x_i,\ x_j)\ +\ 2\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jd(x_i,\ x_j)\cdot\left[f(x_j)-f(x_i)\right]\\&+\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\left[f(x_j)-f(x_i)\right]^2.\end{aligned}$$

It can be proved by induction that $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d(x_i, x_j)\cdot\left[f(x_j)-f(x_i)\right]=0.$

Then, we get $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \rho^2(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) - 2\cdot\left[\sum_{i=1}^{n} c_i f(x_i)\right]^2.$

and hence
$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j)=2\cdot\left[\sum_{i=1}^{n} c_i f(x_i)\right]^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \rho^2(x_i, x_j). \tag{4.9}$$

One can easily see now that the first term in the sum on right hand side is positive and the second term is negative. That is ρ^2 being CND do not imply d^2 is CND in general.

Therefore, we conclude that ρ^2 is CND do not imply d^2 is CND and hence the space is not isometric embedding to the flat case.

We recall that $\varphi: X \to Y$ is an isomorphism, an isometry of the weighted quasi-metrics (X, q, w) and (Y, p, u) if $p(\varphi(x), \varphi(y)) = q(x, y)$ and $u(\varphi(x)) = w(x)$ for any $x, y \in X$ (see section 3 and [3]).

We also recall that for a weighted quasi-metric p, we have $p(x, y) = p(y, x) = 0 \Rightarrow x = y$, for any $x, y \in X$. Indeed it is easy to see that

$$p(x, y) = d(x, y) + \frac{1}{2}\left[w(y) - w(x)\right],\ \ p(y, x) = d(y, x) + \frac{1}{2}\left[w(x) - w(y)\right]$$

leads to $p(x, y) = p(y, x) \Leftrightarrow w(y) - w(x) = 0$, i.e. $p(x, y) = p(y, x) = d(x, y)$ and since d is a usual distance function it follows $x = y$. Let us assume $\varphi(x_1) = \varphi(x_2)$.
From the definition of isometry we have

$$p\left(\varphi(x_1), \varphi(x_2)\right) = q(x_1, x_2),\ u\left(\varphi(x_1)\right) = w(x_1),\ u\left(\varphi(x_2)\right) = w(x_2),$$

and hence $p\left(\varphi(x_1), \varphi(x_2)\right) = 0 = q(x_1, x_2),\ p\left(\varphi(x_2), \varphi(x_1)\right) = 0 = q(x_2, x_1)$
$\Rightarrow q(x_1, x_2) = q(x_2, x_1) = 0$ and therefore, $x_1 = x_2$. This shows that an isomorphism (i.e. isometry) of weighted quasi-metrics must also be injective.

The intuition behind can conclude that the replacement of the metric space (X, d) with a weighted quasi-metric space (X, ρ, w) leads to the de-compactification of X. Indeed, even though X is a compact manifold, a weighted quasi-metric on X is actually equivalent with the weighted quasi-metric space (G_f, Q, W), where $G_f = \{(x, f(x)) : x \in X\}$ is the graph of the Lipschitz function $f: X \to [0, \infty)$, $f(x) = \frac{1}{2}w(x)$. In other words, we replace the compact manifold X with the non-compact total space $X \times [0, \infty)$ of the bundle over X.

Example 4.13 Consider $X = S^2$ the unit 2-dimensional sphere in $\mathbb{R}^2$ or $\mathbb{R}^n$ with distance function d, and let $w: S^2 \to \mathbb{R}$ be the function $w(\theta, \varphi) = \tan\theta$ for the parametrization $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$. Clearly (S, ρ, w) obtained from d and w is a (generalized) weighted quasi-metric built on the total space $S^2 \times \mathbb{R}$.

5. Acknowledgements

We express our gratitude to Prof. H. Shimada for many useful suggestions that have improved the material.

References

[1] Feragen, A., Lauze, F. and Hauberg, S., **2015**. Geodesic Exponential Kernels: When Curvature and linearity Conflict. *Proceedings of Computer Vision and Pattern Recognition 2015*, Boston, June 7-12, 3032-3042.

[2] Jayasumana, S., Hartley, R., Salzmann, M., Li, H. and Harandi, M., **2015**. Kernel Methods on Riemannian Manifold with Gaussian RBF Kernels. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 37(12), 2464-2477.

[3] Sabau, S.V., Shibuya, K. and Shimada, H., **2014**. Metric structures associated to Finsler metrics. *Publicationes Mathematicae Debrecen*, 84(1-2), 89-103.

[4] Vitolo, P., **1999**. The Representation of Weighted Quasi-Metric Spaces. *Rendiconti dell'Istituto di Matematica dell'Università di Trieste,* XXXI, 95-100.

[5] Valette, A., Bekka, B. and Harpe, P.D., **2008**. *Kazhdan's Property (T)*. New Mathematical Monographs.

[6] Burago, D., Burago, Y. and Ivanov, S., **2001**. *A course on Metric Geometry*. American Mathematical Society.