GEODESIC DISTANCE NON-SYMMETRIC KERNELS

URAIWAN SOMBOON

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE (APPLIED MATHEMATICS)

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2017

KMITL-2017-SC-M-001-007

Faculty of Science
King Mongkut's Institute of Technology Ladkrabang
Thesis Certification

Thesis Title “GEODESIC DISTANCE NON-SYMMETRIC KERNELS”
Student Name Miss Uraiwan Somboon
Student ID 58605028
Degree Master of Science (Applied Mathematics)
Department Mathematics
Thesis Advisor Assoc.Prof.Praiboon Pantaragphong

Thesis Committee	Signatures
Dr.Sukrawan Mavecha Chairperson	Sukrawan Mavecha
Dr.Thawatchai Khumprapussorn Examiner	Thawatchai
Assoc.Prof.Dr.Jessada Tariboon External Examiner	Jessada Tariboon
Assoc.Prof.Praiboon Pantaragphong Thesis Advisor	Praiboon P.

Examination Date 25th May 2017 **Time** 09.30 - 12.30 a.m.
Place Faculty of Science room 304

Approved by Faculty of Science

(Assoc.Prof.Dr. DusaneeThanaboripat)
Dean

Date 28 June 2017

หัวข้อวิทยานิพนธ์	เคอร์เนลที่ไม่สมมาตรของระยะทางจีออเดสิก
ชื่อนักศึกษา	นางสาวอุไรวรรณ สมบูรณ์
รหัสประจำตัว	58605028
ปริญญา	วิทยาศาสตรมหาบัณฑิต (คณิตศาสตร์ประยุกต์)
ภาควิชา	คณิตศาสตร์
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
พ.ศ.	2560
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ. ไพรบูลย์ พันธรักษ์พงษ์

บทคัดย่อ

ในงานวิจัยนี้ผู้วิจัยได้จัดการและศึกษาสมบัติพื้นฐานของเคอร์เนลที่ไม่สมมาตรที่เกิดจากปริภูมิเมตริกกึ่งเสมือนแบบมีน้ำหนักบนปริภูมิฮิลเบิร์ตซึ่งในงานวิจัยนี้ ผู้วิจัยแสดงให้เห็นว่าการใช้เคอร์เนลที่ไม่สมมาตรอาจจะให้ผลเฉลยที่ขัดแย้งระหว่างสมบัติที่เป็นบวกแน่นอนของเคอร์เนลและค่าความโค้งของปริภูมิพื้นฐาน

คำสำคัญ : เคอร์เนลที่ไม่สมมาตร ปริภูมิเมตริกกึ่งเสมือนแบบมีน้ำหนัก ปริภูมฮิลเบิร์ต เมตริกฟินส์เลอร์ สมบัติที่เป็นบวกแน่นอนของเคอร์เนล

Thesis Title	Geodesic Distance Non-Symmetric Kernels
Student Name	Miss Uraiwan Somboon
Student ID	58605028
Degree	Master of Science (Applied Mathematics)
Department	Mathematics
Faculty	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Year	2017
Thesis Advisor	Assoc.Prof.Praiboon Pantaragphong

Abstract

In this research, we deal with non-symmetric kernels induced by weighted quasi-metrics on Hilbert spaces and we study their fundamental properties. 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 : non-symmetric kernel, weighted quasi-metric space, Hilbert space, positive definiteness of kernel

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor Associate Professor Praiboon Pantaragphong for the continuous support of my M.Sc. study and related research, for his patience, motivation, and immense knowledge. The door to his office was always open whenever I ran into a trouble spot or had a question about my research or writing. Furthermore I would also like to express my sincere gratitude to Professor Sorin V. Sabau from Tokai university for introducing me to the topic as well for the useful comments, remarks and engagement through the learning process of this master thesis. Their guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my M.Sc. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr.Sukrawan Mavecha, Dr.Thawatchai Khumprapussorn, and Associate Professor Dr.Jessada Tariboon from King Mongkut's University of Technology North Bangkok, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I am sincerely grateful to financial supports from the Graduate Study of the Faculty of Science, King Mongkut's Institute of Technology Ladkrabang Fund.

Finally, I must express my very profound gratitude to my parents for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Uraiwan Somboon

Table of Contents

List of Tables

List of Figures

Abbreviations

Abbreviations	Explanation
CND	Conditionally Negative Definite
PD	Positive Definite
SVM	Support Vector Machine
PCA	Principal Component Analysis

Chapter 1

Introduction

1.1 Research Motivation

Kernels methods are fundamental tools for statistical data analysis and machine learning [5, 6]. Considering the data set as a subset $\mathbb{R}^n$ one can use an embedding into a 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 spaces, in particular data lying on spaces that are not necessarily flat. Examples include data analysis for computer vision, where rotation matrices belong to the Lie group SO(3), normalized histograms from the unit n-sphere $\mathbb{S}^n$ and other type of data that belong to smooth manifolds, as Riemannian and Finsler manifolds.

Such manifolds lack a vector space structure and hence, popular algorithms or methods, like support vector machine (SVM), principal component analysis (PCA) or other clustering methods cannot be put in practical use [6].

The positive definiteness of a kernel is essential for the use of kernel methods as support vector machines or kernel PCA. Indeed, the bandwidth parameters can be learned from data only if the kernel is positive definite (PD).

There are several approaches to this problem. The simplest and most widely used method is to ignore the curvature of the manifold including the data, that is to work only with data extracted from an Euclidean space $\mathbb{R}^n$. This approach is generally good from the methodological point of view, but gives results of poor accuracy and undesirable effects for data that should be regarded to sit on curved manifolds, like $\mathbb{S}^n$.

Another approach is to consider only data defined on Riemannian manifolds that can be regarded as inner product spaces, for instance Sym_d^+, the space of $d \times d$ symmetric positive definite matrices, or the Grassmann manifold $\mathcal{G}_n^r$, the space of r dimensional linear subspace of an n-dimensional Euclidean space, provided $r < n$.[6].

In this research we will 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 [7].

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 manifold as $\mathbb{S}^n$?" It is clear that the answer to this question is NO for the symmetric kernels case.

1.2 Objectives of the study

1) To define non-symmetric kernels and investigate their properties.

2) To study in particular non-symmetric geodesic kernels induced by weighted-quasi distance functions.

3) To show that the positive definiteness of a weighted quasi-metric geodesic kernel do not impose that the data space must be flat.

1.3 Scope of the study

We investigate the algebraic, geometrical and analytical properties of symmetric and non-symmetric kernels. In particular, the kernels investigated in the research are Gaussian kernels induced by classical distance functions and weighted-quasi distances. The general theory makes use of complex and real Hilbert spaces. In the applications using geodesic kernels data can live on any topological manifold, including Euclidean space, spheres, or some matrices Lie groups.

1.4 Benefits of the Study

1) Develop a new mathematical theory of weighted-quasi metric spaces.

2) Show that using weighted quasi-metric distances we can use kernel methods for data analysis even in the case when data space is not flat.

3) Provide mathematical work frame for developing algorithms and programs for SVM and other kernel based clustering methods. These clustering methods and algorithms can be applied to any kind of data sets from Physics, Biology, Economics, etc.

1.5 Research methodology

1) Study advanced topics in complex matrix theory.

2) Study advanced topics in linear algebra.

3) Study the geometry of metric spaces.

4) Study the algebraic and geometrical meaning of positive definite (PD) and conditionally negative definite (CND) kernels.

5) Define geodesic distance kernels using weighted quasi distance.

6) Prove that PD of the geodesic distance kernel induced by a weighted quasi distance do not necessarily impose the flatness of the data space.

7) Conclude that we have solved the problem that motivates this research, make suggestions for further works and write the thesis.

Table 1.1: The Research Schedule

Activity	Time frame (month of year)																					
	2015					2016												2017				
	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5
Step 1																						
Step 2																						
Step 3																						
Step 4																						
Step 5																						
Step 6																						
Step 7																						

Chapter 2

Preliminaries

The purpose of this chapter is to provide basic concepts and tools as weighted quasi-metric spaces, Hilbert spaces and symmetric kernels used in the research.

2.1 Weighted quasi-metric spaces

In this section, we recall the main geometrical results of weighted quasi-metrics (see [7] and [9]). In special we call the attention to the bundle representation of weighted quasi-metric, a fundamental notion for the generalization following.

Definition 2.1. Let X be a non-empty set and d a real-valued function $d : X \times X \to \mathbb{R}$ that satisfies

(i) Positiveness: $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$.

(ii) Symmetry: $d(x,y) = d(y,x)$.

(iii) Triangle inequality: $d(x,y) \leq d(x,z) + d(z,y)$

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

Example 2.2. (i) The function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$d(x,y) = \left[\sum_{i=1}^{n}(x_i - y_i)^2\right]^{\frac{1}{2}}, \mathrm{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mathrm{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{2.1}$$

for any n- dimensional vectors in a metric on $\mathbb{R}^n$ called the **Euclidean metric** on $\mathbb{R}^n$.

(ii) Let (X, g) be an n- dimensional Riemannian manifold. Then the function $d_g : X \times X \to \mathbb{R}$ given by

$$d_g(x,y) := \inf_{\gamma \in \Gamma_{xy}} \int_a^b \Big[g(\gamma(t), \dot{\gamma}(t))\Big]^{\frac{1}{2}} dt, \tag{2.2}$$

is a metric on X, where $\Gamma_{xy} := \{\gamma : [a,b] \to X \mid \gamma \text{ (piecewise) } C^\infty\text{-curve}, \gamma(a) = x, \gamma(b) = y\}$ is the set of (piecewise) C^∞-curves joining points x and y, $\dot{\gamma}(t) := \frac{d\gamma(t)}{dt}$ the tangent vector to γ at $\gamma(t)$.

Definition 2.3. Let X be an arbitrary set. A function $\rho : X \times X \to \mathbb{R}$ is called a **quasi-metric** on X if the following conditions are satisfied for all $x, y, z \in X$:

(i) Positiveness : $\rho(x,y) > 0$ if $x \neq y, \rho(x,x) = 0$.

(ii) Triangle inequality : $\rho(x,z) \le \rho(x,y) + \rho(y,z)$.

(iii) Separation axiom : $\rho(x,y) = \rho(y,x) = 0 \Rightarrow x = y$.

Then (X, ρ) is called a **quasi-metric space**.

Example 2.4. (i) The u^L - distance in $\mathbb{R}$ is defined as

$$u^L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, u^L := \max\{x - y, 0\}.$$

Clearly this is a quasi - metric because
$u^L(1,2) := \max\{-1, 0\} = 0$, while $u^L(2,1) := \max\{1, 0\} = 1$.

(ii) Let (X, F) be on n-dimensional Finsler metric. Then the function $d_F : X \times X \to \mathbb{R}$, given by

$$d_F(x,y) := \inf_{\gamma \in \Gamma_{xy}} \int_a^b F(\gamma(t), \dot{\gamma}(t))dt, \tag{2.3}$$

is a quasi-metric on X, where $\Gamma_{xy}, \gamma, \dot{\gamma}$ have the same meanings as in the Riemannian case.

Moreover, an important class of quasi-metric space are the so-called **weighted quasi-metric space**

Definition 2.5. A weighted quasi-metric space is a triple (X, ρ, w), where X is a non-empty set, $\rho : X \times X \to [0, \infty)$ and $w : X \to [0, \infty)$ satisfying the following conditions

(i) $\rho(x,x) = 0$ for every $x \in X$

(ii) $\rho(x,z) \le \rho(x,y) + \rho(y,z)$ for every $x, y, z \in X$

(iii) if $\rho(x,y) = \rho(y,x) = 0$, then $x = y$

(iv) $\rho(x,y) + w(x) = \rho(y,x) + w(y)$ for every $x, y \in X$.

The function ρ is called **quasi-metric**, and w is the **weight function**.

Example 2.6. (i) The u^L metric restricted to $[0, \infty)$, i.e.

$$u^L : [0,\infty) \times [0,\infty) \to \mathbb{R}, \qquad u^L(x,y) = \begin{cases} y - x & ,\text{if } x \le y \\ 0 & ,\text{if } y < x \end{cases}$$

We can see that this metric u^L is a quasi-metric, with the weight function $w(x) := x$. Indeed, one can easily see that if $x \le y$, then $u^L(x,y) \neq u^L(y,x)$ and

$$u^L(x,y) + w(x) = y - x + x = y$$
$$u^L(y,x) + w(y) = 0 + y = y.$$

(ii) The metric induced by a Finsler metric of Randers type $F = \alpha + \beta$, where $\alpha(x,y) := \sqrt{a_{ij}(x)y^iy^j}$, $\alpha(x,y)$ Riemannian metric on X, $\beta(x,y) := b_i(x)\cdot y^i$ an exact linear 1-form on X (see [2] for details).

(iii) Distances in the geometry of sequence comparison in bioinformatics.

Definition 2.7. If (X,ρ) is a quasi-metric space then the function $d : X \times X \to [0,\infty)$ given by

$$d(x,y) := \frac{1}{2}[\rho(x,y) + \rho(y,x)], \qquad \forall x,y \in X \tag{2.4}$$

is called **the symmetrization** of ρ.

Lemma 2.8. If (X,ρ) is a quasi-metric space and d is symmetrization of ρ, then (X,d) is a metric space.

Proof. Let $x,y,z \in X$. We verify the conditions in definition 2.1.

(i) Positiveness:
$d(x,y) = \frac{1}{2}[\rho(x,y) + \rho(y,x)]$
Since, $\rho(x,y) \geq 0$ and $\rho(y,x) \geq 0$, it is clear that $d(x,y) \geq 0$.
Hence, $\rho(x,x) = 0$, so

$$d(x,x) = \frac{1}{2}[\rho(x,x) + \rho(x,x)] = \rho(x,x) = 0$$

(ii) Symmetry:

$$\begin{aligned} d(x,y) &= \frac{1}{2}[\rho(x,y) + \rho(y,x)] \\ &= \frac{1}{2}[\rho(y,x) + \rho(x,y)] = d(y,x) \end{aligned}$$

(iii) The triangle inequality:

$$\begin{aligned} d(x,y) &= \frac{1}{2}[\rho(x,y) + \rho(y,x)] \\ &\leq \frac{1}{2}[\rho(x,z) + \rho(z,y) + \rho(y,z) + \rho(z,x)] \\ &= \frac{1}{2}[\rho(x,z) + \rho(z,x)] + \frac{1}{2}[\rho(z,y) + \rho(y,z)] \\ &= d(x,z) + d(z,y). \end{aligned}$$

Therefore, (X,d) is metric space. □

Proposition 2.9. Using the symmetrization d of weight of quasi-metric ρ with the weight function $w : X \to [0, \infty)$ we have

$$\rho(x, y) = d(x, y) + \frac{1}{2}[w(y) - w(x)], \forall x, y \in X.$$

Moreover, we have

$$\frac{1}{2}|w(x) - w(y)| \leq d(x, y), \forall x, y \in X.$$

Indeed, we have

$$\begin{aligned}
\rho(x, y) + w(x) &= \rho(y, x) + w(y) \\
\rho(x, y) - \rho(y, x) &= w(y) - w(x) \\
2\rho(x, y) - (\rho(x, y) + \rho(y, x)) &= w(y) - w(x) \\
\rho(x, y) &= \frac{1}{2}[\rho(x, y) + \rho(y, x)] + \frac{1}{2}[w(y) - w(x)] \\
d(x, y) &= \rho(x, y) + \frac{1}{2}[w(y) - w(x)].
\end{aligned}$$

Proposition 2.10. If (X, ρ, w) is a weighted quasi-metric space, then the perimeter length of any geodesic triangle on X does not depend on the orientation, that is

$$\rho(x, y) + \rho(y, z) + \rho(z, x) = \rho(x, z) + \rho(z, y) + \rho(y, x), \quad \forall x, y, z \in X. \tag{2.5}$$

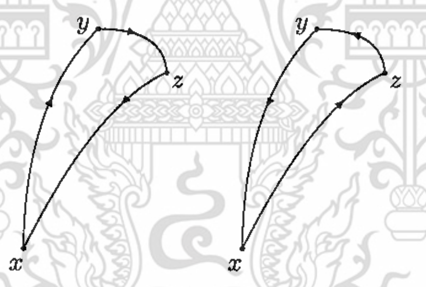

Figure 2.1: The perimeter of the triangle Δxyz is independent of the orientation.

Proof. Let $x, y, z \in X$ and w be weigth function.
From a weightability, we have

$$\rho(x, y) + w(x) = \rho(y, x) + w(y) \tag{2.6}$$
$$\rho(y, z) + w(y) = \rho(z, y) + w(z) \tag{2.7}$$
$$\rho(z, x) + w(z) = \rho(x, z) + w(x). \tag{2.8}$$

Then (2.6)+(2.7)+(2.8), we have

$$\rho(x, y) + \rho(y, z) + \rho(z, x) = \rho(x, z) + \rho(z, y) + \rho(y, x).$$

Therefore, the proof is complete. □

Remark 2.11. It should be clear that not any quasi-metric space is weightable. In fact, it can be shown that the class of weightable quasi-metric spaces are exactly those quasi-metric spaces that satisfy relation (2.5) (see [7]).

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

$$\begin{aligned} p(\varphi(x),\varphi(y)) &\leq \rho(x,y), \qquad \forall x,y \in X && (2.9)\\ u(\varphi(x)) &\leq w(x), \qquad \forall x \in X && (2.10)\end{aligned}$$

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

In the case we have equality in relation (2.9), 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, i.e.

$$p(\varphi(x),\varphi(y)) = \rho(x,y), \qquad \forall x,y \in X \tag{2.11}$$

$$u(\varphi(x)) = w(x), \qquad \forall x \in X. \tag{2.12}$$

Finally, an **embedding** of (X,ρ,w) into (G,Q,W) is an isomorphism of (X,ρ,w) onto a subspace of (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 2.12 (The product of a metric space with a half ray). Consider a metric space (X,d) and the half ray $I := [0,\infty)$. Then the product space $G := X \times I$ inherits a natural structure of (generalized) weighted quasi-metric space (G,Q,W), where

$$\begin{aligned} &Q : G \times G \to [0,\infty), \qquad Q(u,v) := d(x,y) + \eta - \xi,\\ &W : G \to [0,\infty), \qquad W(u) := 2\xi, \qquad \forall u = (x,\xi), v = (y,\eta) \in X \times I.\end{aligned} \tag{2.13}$$

That is, let $(x,\xi),(y,\eta),(z,\zeta) \in X \times [0,\infty)$, then

(i) It is clear that

$$Q((x,\xi),(x,\xi)) = d(x,x) + \xi - \xi = 0$$

(ii) From (2.13), we have

$$\begin{aligned} Q((x,\xi),(z,\zeta)) &= d(x,z) + \zeta - \xi\\ &\leq d(x,y) + d(y,z) + \zeta - \xi\\ &= Q((x,\xi),(y,\eta)) + \xi - \eta + Q((y,\eta),(z,\zeta)) + \eta - \zeta + \zeta - \xi\\ &= Q((x,\xi),(y,\eta)) + Q((y,\eta),(z,\zeta)),\end{aligned}$$

so, $Q((x,\xi),(z,\zeta)) \leq Q((x,\xi),(y,\eta)) + Q((y,\eta),(z,\zeta))$

(iii) Suppose that $Q((x,\xi),(y,\eta)) = 0$ and $Q((y,\eta),(x,\xi)) = 0$, that is

$$Q((x,\xi),(y,\eta)) = d(x,y) + \eta - \xi = 0 \tag{2.14}$$

$$Q((y,\eta),(x,\xi)) = d(y,x) + \xi - \eta = 0 \tag{2.15}$$

by adding these two equality we get, $2d(x,y) = 0$, whence $x = y$, so that $\xi = \eta$.

(iv) Let $W((x,\xi)) = 2 \cdot \pi(x,\xi) = 2\xi$, we have

$$\begin{aligned} Q((x,\xi),(y,\eta)) + W((x,\xi)) &= d(x,y) + \eta - \xi + 2\xi \\ &= d(x,y) + \eta + \xi \\ &= d(y,x) + \xi - \eta + 2\eta \\ &= Q((y,\eta),(x,\xi)) + W((y,\eta)). \end{aligned}$$

Therefore, (G, Q, W) is a generalized weighted quasi-metric space.

Remark 2.13. The generalized weighted quasi-metric space (G, Q, W) constructed in Example 2.12 is sometimes called **the bundle over** (X, d) (see [7]).

Example 2.14 (The Graph of a function). We consider the case of the graph of a non-negative valued function $f : X \to [0,\infty)$ defined on a metric space (X, d).

Indeed, if we denote the graph of f by

$$G_f := \{(x, f(x)) : x \in X\}, \tag{2.16}$$

then (G_f, Q, W) is a naturally induced weighted quasi-metric space structure defined by

$$\begin{aligned} &Q : G_f \times G_f \to [0,\infty), \qquad Q(u,v) := d(x,y) + f(y) - f(x), \\ &W : G_f \to [0,\infty), \qquad W(u) := 2f(x), \qquad \forall u = (x, f(x)), v = (y, f(y)) \in G_f. \end{aligned} \tag{2.17}$$

Based on these, one has

Theorem 2.15 ([7]). Every weighted quasi-metric space (Y, p, u) is embeddable in a bundle over a suitable metric space (X, d).

Proof. Let a metric space $(X, d) := (Y, \widehat{p})$ the symmetrization of (Y, p, u).
Then, a weighted quasi-metric space (T, P, U) the corresponding bundle, where $T := Y \times [0,\infty)$. Define a mapping $\varphi : Y \to T$ as follows:

$$\varphi(x) := (x, \frac{1}{2}u(x)). \tag{2.18}$$

We want to show that this is an embedding.
Then, for any $x, y \in Y$, we have

$$\begin{aligned}
\widehat{p}(x,y) &= \widehat{p}(y,x) \\
\widehat{p}(x,y) + \frac{1}{2}u(y) + \frac{1}{2}u(x) &= \widehat{p}(y,x) + \frac{1}{2}u(x) + \frac{1}{2}u(y) \\
\widehat{p}(x,y) + \frac{1}{2}u(y) - \frac{1}{2}u(x) + u(x) &= \widehat{p}(y,x) + \frac{1}{2}u(y) - \frac{1}{2}u(x) + u(y) \\
P(\varphi(x),\varphi(y)) + U(\varphi(x)) &= P(\varphi(y),\varphi(x)) + U(\varphi(y)),
\end{aligned}$$

where $P(\varphi(x),\varphi(y)) = \widehat{p}(x,y) + \frac{1}{2}u(y) - \frac{1}{2}u(x) = p(x,y)$.
Clearly we also have $U(\varphi(x)) = u(x) = 2(\frac{1}{2}w(x))$ for each $x \in Y$.
Thus, φ is an embedding. □

Theorem 2.16 ([7]). (i) Let (X,d) be a metric space and $f : X \to [0,\infty)$ a 1-Lipschitz function. Then the graph of f is a weighted quasi-metric space (G_f, Q, W).

(ii) Conversely, every weighted quasi-metric space (Y,p,u) can be constructed in this way.

Indeed, given a weighted quasi-metric space (Y,p,u) one can construct

- a metric space $(X,d) := (Y,\widehat{p})$, where $\widehat{p}$ is the symmetrization of p,
- a Lipschitz function $f : X \to [0,\infty)$, $f(x) := \frac{1}{2}u(x)$.

Next, we recall the differential manifold structure of the graph of a smooth function.

Let us consider a C^∞ function $f : X \to [0,\infty)$, $x \mapsto f(x)$ and the graph of f denoted by $G_f = \{(x,f(x)) : x \in X\} \subset X \times \mathbb{R}$. Then it is known that G_f is a C^∞ submanifold of the product manifold $X \times \mathbb{R}$ that is actually diffeomorphic to X. Indeed, the mapping

$$\varphi : X \to G_f, \qquad x \mapsto \varphi(x) = (x, f(x)) \tag{2.19}$$

with the inverse

$$\psi : G_f \to X, \qquad u = (x,f(x)) \mapsto \psi(x,f(x)) = x \tag{2.20}$$

is a diffeomophism. Remark that ψ is nothing else than the projection onto the first factor.

Any given weighted quasi-metric space (X,ρ,w) that satisfies some supplementary metrizability condition induces a Finsler structure $(X, F = F_0 + df)$ on X.

2.2 Hilbert spaces

In this section, we give definition and some properties about Hilbert spaces (see [1]).

Let $\mathcal{H}$ be a vector space over $\mathbb{C}$(or $\mathbb{R}$) with additivity "$+$" and scalar multiplication "$\cdot$". Such a vector space $\mathcal{H}$ is called a **Hilbert space** if

(i) It is endowed with **an inner product** $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that is **a hermitian form** , i.e. It has the properties

(a) $\langle \alpha x + \beta y, z \rangle = \alpha \cdot \langle x, z \rangle + \beta \cdot \langle y, z \rangle, \qquad \alpha, \beta \in \mathbb{C}, x, y, z \in \mathcal{H}$

(b) $\langle x, y \rangle = \overline{\langle y, x \rangle}$, where $\bar{\cdot}$ means complex conjugate,

(c) $\langle x, x \rangle \in (0, \infty)$ for $x \neq 0$.

(ii) $(\mathcal{H}, \|\cdot\|)$ is a complete norm vector space, where the norm on $\mathcal{H}$ is induced by the inner product $\langle \cdot, \cdot \rangle$, i.e. $\|x\| = \sqrt{\langle x, x \rangle}, \forall x \in \mathcal{H}$.

Proof. If $\|x\|$ is a norm on $\mathcal{H}$, then it is satisfies three properties,

(a) By $\langle x, x \rangle \geq 0$, hence $\|x\| = \sqrt{\langle x, x \rangle} \geq 0$. Furthermore, $\|x\| = 0$ if and only if $\langle x, x \rangle = 0$, and this holds if and only if $x = 0$.

(b) Let $\lambda \in \mathbb{R}$, then we have

$$\begin{aligned} \|\lambda \cdot x\|^2 &= (\sqrt{\langle \lambda \cdot x, \lambda \cdot x \rangle})^2 = \langle \lambda x, \lambda x \rangle \\ &= \lambda \bar{\lambda} \langle x, x \rangle = |\lambda|^2 \|x\|^2. \end{aligned}$$

Taking the square root of both sides gives $|\lambda| \|x\|$.

(c) Suppose that $x, y \in \mathcal{H}$. Using the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} \|x + y, x + y\|^2 &= \langle x + y, x + y \rangle \\ &= \langle x, x \rangle + \langle x, y \rangle + \overline{\langle x, y \rangle} + \langle y, y \rangle \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 \\ &= \left(\|x\| + \|y\| \right)^2 \end{aligned}$$

Taking square root of both sides gives, $\|x + y\| \leq \|x\| + \|y\|$.

$\square$

In other words, a Hilbert space is a complex (or real) inner product space this is also a complete metric space with respect to the distance function

$$d(x, y) := \|x - y\| = \sqrt{\langle x - y, x - y \rangle}, \forall x, y \in \mathcal{H},$$

induced by the inner product.

Remark 2.17. The complex inner product $\langle\cdot,\cdot\rangle$ is linear in the first argument, but anti-linear on the second argument, i.e.

$$\langle x, \alpha y + \beta z\rangle = \overline{\alpha}\cdot\langle x, y\rangle + \overline{\beta}\cdot\langle x, z\rangle, \qquad \alpha,\beta \in \mathbb{C}. \tag{2.21}$$

Proof. Let $x, y, z \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{C}$. Then

$$\begin{aligned}\langle x, \alpha y + \beta z\rangle &= \overline{\langle \alpha y + \beta z, x\rangle}\\ &= \overline{\alpha\cdot\langle y, x\rangle + \beta\cdot\langle z, x\rangle}\\ &= \overline{\alpha\cdot\langle y, x\rangle} + \overline{\beta\cdot\langle z, x\rangle}\\ &= \overline{\alpha}\cdot\overline{\langle y, x\rangle} + \overline{\beta}\cdot\overline{\langle z, x\rangle}\\ &= \overline{\alpha}\cdot\langle x, y\rangle + \overline{\beta}\cdot\langle x, z\rangle.\end{aligned}$$

□

Proposition 2.18. If $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is a Hilbert space, then

$$|Re\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle}\cdot\sqrt{\langle y, y\rangle} \tag{2.22}$$

(Cauchy-Schwartz inequality)

Remark 2.19. Any complex (or real) vector space with inner product is called a **pre-Hilbert space**. If a pre-Hilbert space is complete, then it is called a Hilbert space.

Completeness means that any Cauchy sequence converges with respect to the induced norm to an element in the space.

We recall that a vector space V over $\mathbb{R}$ is called a **normed space** if it has a norm $\|\cdot\| : V \to [0,\infty)$, $x \longmapsto \|x\|$ that satisfies

(i) $\|x\| > 0$ if $x \neq 0$, and $\|0\| = 0$, where 0 is zero element of V.

(ii) $\|\lambda\cdot x\| = |\lambda|\cdot\|x\|$, $\lambda \in \mathbb{R}, x \in V$.

(iii) $\|x + y\| \leq \|x\| + \|y\|$.

A Hilbert space is normed space $(\mathcal{H}, \|\cdot\|)$ that satisfies the following condition of **completeness**. If for any sequence $\{x_n\}$ in V, s.t. $\lim_{m,n\to\infty}\|x_m - x_n\| = 0$, there exists $x \in V$ s.t. $\lim_{n\to\infty}\|x - x_n\| = 0$, then the normed space is called a **Banach space**.

Example 2.20. (i) Any finite dimensional vector space with a (real or complex) inner product is a Hilbert space. Hence all finite dimensional vector are Banach spaces. Recall that a **real inner product space** is obtained in the same way on a complex are, except that it is a real vector space and that the inner product takes real values, i.e. $\langle\cdot,\cdot\rangle : \mathcal{H}\times\mathcal{H} \to \mathbb{R}$.

(ii) $\mathcal{H} = \mathbb{C}^n$ with inner product

$$\langle x, y\rangle := x \cdot \overline{y} = \sum_{i=1}^{n} x_i \overline{y}_i, \tag{2.23}$$

where $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{C}$.

(iii) Consider l^2 to be the space of all infinite sequences $\mathbf{z} = (z_1, z_2, \ldots)$ of complex numbers s.t. the series $\sum_{n=1}^{\infty} |\mathbf{z}_n|$ is convergent.

Then $\left(l^2, \langle\cdot,\cdot\rangle\right)$ is a Hilbert space with the inner product $\langle \mathbf{z}, w\rangle = \sum_{n=1}^{\infty} \mathbf{z}_n \overline{w}_n$.

2.3 Symmetric kernels

In this section, 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 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 (see [5] and [8]).

A **topological** (X, τ) is a set X with topology τ, i.e. a collection of subsets of X with the following properties: (i) $X, \phi \in \tau$

(ii) if $A, B \in \tau$, then $A \cap B \in \tau$

(iii) for any collection $\{A_\alpha\}_\alpha$ if all $A_\alpha \in \tau$, then $\bigcup_\alpha A_\alpha \in \tau$ (see [4]).

Definition 2.21. ([3]) A function $K : X \times X \to \mathbb{C}$ or $\mathbb{R}$ and a map $\phi : X \to \mathbb{R}^n$ satisfying, for all $x, y \in X$,

$$K(x, y) = \langle \phi(x), \phi(y)\rangle$$

is a **kernel function.**

In other word, kernel function is a function that take as its inputs vectors in the original space and return the inner product between the images of two data points in the feature space.

Definition 2.22. ([8]) A continuous function $\Phi : X \times X \to \mathbb{C}$ or $\mathbb{R}$ such that $\Phi(x, x) = 0$ and $\Phi(x, y) = \Phi(y, x)$ is called a **symmetric kernel** on X.

We recall the following definition.

Definition 2.23. ([8]) The kernel K on the topological space X is called **positive definite (PD)** if for any $n \in \mathbb{N}$, and elements $x_1, \ldots, x_n \in X$ and any scalars $c_1, \ldots, c_n \in \mathbb{R}$ we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K(x_i, x_j) \geq 0. \tag{2.24}$$

Example 2.24. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $f : X \to \mathcal{H}$ a continuous function. Then the kernel $\Phi(x,y) = \langle f(x), f(y)\rangle$, for any $x, y \in X$ is positive definite.

Indeed, we have

$$\begin{aligned}
\mathbf{c}^T\Phi\mathbf{c} &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\bar{c}_j\Phi(x_i,x_j), \qquad \forall i,j \in \{1,...,n\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\bar{c}_j\langle f(x_i), f(x_j)\rangle \\
&= \left\langle \sum_{i=1}^{n} c_i f(x_i), \sum_{i=1}^{n} c_i f(x_i) \right\rangle \\
&= \left\| \sum_{i=1}^{n} c_i f(x_i) \right\|^2 \geq 0.
\end{aligned} \tag{2.25}$$

We will show that formula (2.25) is true for all $n \in \mathbb{N}$.
For $n = 2$, we have

$$\begin{aligned}
\mathbf{c}^T\Phi\mathbf{c} &= \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} \Phi(x_1,x_1) & \Phi(x_1,x_2) \\ \Phi(x_2,x_1) & \Phi(x_2,x_2) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\
&= \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} \langle f(x_1), f(x_1)\rangle & \langle f(x_1), f(x_2)\rangle \\ \langle f(x_2), f(x_1)\rangle & \langle f(x_2), f(x_2)\rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\
&= (c_1)^2\langle f(x_1), f(x_1)\rangle + 2c_1\bar{c}_2\langle f(x_1), f(x_2)\rangle + (c_2)^2\langle f(x_2), f(x_2)\rangle \\
&= \Big\langle c_1 f(x_1) + c_2 f(x_2), c_1 f(x_1) + c_2 f(x_2) \Big\rangle \\
&= \Big\| c_1 f(x_1) + c_2 f(x_2) \Big\|^2 \geq 0
\end{aligned}$$

and by induction step we can easily to prove formula (2.25).
Induction step: Let k be positive integer and suppose (2.25) is true for $n = k$, we will show that (2.25) is true for $n = k + 1$.
Thus

$$\begin{aligned}
\mathbf{c}^T\Phi\mathbf{c} = {}& \begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix} \begin{pmatrix} \Phi(x_1,x_1) & \cdots & \Phi(x_1,x_k) \\ \Phi(x_2,x_1) & \cdots & \Phi(x_2,x_k) \\ \vdots & \ddots & \vdots \\ \Phi(x_k,x_1) & \cdots & \Phi(x_k,x_k) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} + 2c_1\bar{c}_{k+1}\Phi(x_1,x_{k+1}) \\
& + 2c_2\bar{c}_{k+1}\Phi(x_2,x_{k+1}) + \cdots + 2c_k\bar{c}_{k+1}\Phi(x_k,x_{k+1}) + c_{k+1}^2\Phi(x_{k+1},x_{k+1}) \\
= {}& \begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix} \begin{pmatrix} \langle f(x_1), f(x_1)\rangle & \cdots & \langle f(x_1), f(x_k)\rangle \\ \langle f(x_2), f(x_1)\rangle & \cdots & \langle f(x_2), f(x_k)\rangle \\ \vdots & \ddots & \vdots \\ \langle f(x_k), f(x_1)\rangle & \cdots & \langle f(x_k), f(x_k)\rangle \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} + 2c_1\bar{c}_{k+1}\langle f(x_1), f(x_{k+1})\rangle \\
& + 2c_2\bar{c}_{k+1}\langle f(x_2), f(x_{k+1})\rangle + \cdots + 2c_k\bar{c}_{k+1}\langle f(x_k), f(x_{k+1})\rangle + c_{k+1}^2\langle f(x_{k+1}), f(x_{k+1})\rangle
\end{aligned}$$

$$= (c_1)^2\langle f(x_1), f(x_1)\rangle + 2c_1\bar{c}_2\langle f(x_1), f(x_2)\rangle + 2c_1\bar{c}_3\langle f(x_1), f(x_3)\rangle$$
$$+ 2c_1\bar{c}_4\langle f(x_1), f(x_4)\rangle + \cdots + 2c_k\bar{c}_k\langle f(x_k), f(x_k)\rangle + (c_k)^2\langle f(x_k), f(x_k)\rangle$$
$$+ 2c_1\bar{c}_{k+1}\langle f(x_1), f(x_{k+1})\rangle + \cdots + 2c_k\bar{c}_{k+1}\langle f(x_k), f(x_{k+1})\rangle + (c_{k+1})^2\langle f(x_{k+1}), f(x_{k+1})\rangle$$
$$= \Big\langle c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{k+1} f(x_{k+1}), c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{k+1} f(x_{k+1})\Big\rangle$$
$$= \Big\| c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{k+1} f(x_{k+1})\Big\|^2 \geq 0.$$

Thus, formula (2.25) holds for $n = k+1$.

Remark 2.25. Obviously kernel $\Phi(x,y) = \langle x, y\rangle$ is positive definite kernel.

Definition 2.26. ([8]) The kernel K on the topological space X is called **conditionally negative definite (CND)** if it satisfies

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

(ii) $K(x,y) = K(y,x)$ 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 $c_1 + c_2 + \cdots + c_n = 0$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) \leq 0. \tag{2.26}$$

Example 2.27. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a real Hilbert space, and let $d : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$, $d(x,y) = \|x - y\|$ be the induced Hilbert distance. The kernel $\Psi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$,

$$\Psi(x,y) = d^2(x,y) = \|x-y\|^2, \qquad \forall x, y \in \mathcal{H} \tag{2.27}$$

is CND.

Indeed, for any $x_1, \ldots, x_n \in \mathcal{H}$ and $c_1, \ldots, c_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} c_i = 0$, we have

$$\mathbf{c}^T\Psi\mathbf{c} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \|x_i - x_j\|^2 = -2\Big\|\sum_{i=1}^{n} c_i x_i\Big\|^2 \leq 0. \tag{2.28}$$

We will show that formula (2.28) is true for all $n \in \mathbb{N}$.
For $n = 2$, we have

$$\mathbf{c}^T\Psi\mathbf{c} = \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} 0 & \Psi(x_1,x_2) \\ \Psi(x_2,x_1) & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$
$$= c_1c_2\Big[\Psi(x_1,x_2) + \Psi(x_2,x_1)\Big] = c_1c_2\Big[\|x_1 - x_2\|^2 + \|x_2 - x_1\|^2\Big]$$
$$= 2c_1c_2\|x_1 - x_2\|^2$$

Since $c_1 + c_2 = 0, c_1c_2 = -c_1^2$. Then

$$\mathbf{c}^T\Psi\mathbf{c} = -2(c_1)^2\|x_1 - x_2\|^2 \leq 0.$$

For $n = 3$, we have

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c} &= \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} 0 & \Psi(x_1,x_2) & \Psi(x_1,x_3) \\ \Psi(x_2,x_1) & 0 & \Psi(x_2,x_3) \\ \Psi(x_3,x_1) & \Psi(x_3,x_2) & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \\
&= c_1c_2\Big[\Psi(x_1,x_2) + \Psi(x_2,x_1)\Big] + c_1c_3\Big[\Psi(x_1,x_3) + \Psi(x_3,x_1)\Big] \\
&\quad + c_2c_3\Big[\Psi(x_2,x_3) + \Psi(x_3,x_2)\Big] \\
&= 2c_1c_2\|x_1 - x_2\|^2 + 2c_1c_3\|x_1 - x_3\|^2 + 2c_2c_3\|x_2 - x_3\|^2 \\
&= 2\Big[c_1c_2\|x_1\|^2 - 2c_1c_2\langle x_1, x_2\rangle + c_1c_2\|x_2\|^2 \\
&\quad + c_1c_3\|x_1\|^2 - 2c_1c_3\langle x_1, x_3\rangle + c_1c_3\|x_3\|^2 \\
&\quad + c_2c_3\|x_2\|^2 - 2c_2c_3\langle x_2, x_3\rangle + c_2c_3\|x_3\|^2\Big] \\
&= 2\Big[(c_1c_2 + c_1c_3)\|x_1\|^2 + (c_1c_2 + c_2c_3)\|x_2\|^2 + (c_1c_3 + c_2c_3)\|x_3\|^2 \\
&\quad - 2c_1c_2\langle x_1, x_2\rangle - 2c_1c_3\langle x_1, x_3\rangle - 2c_2c_3\langle x_2, x_3\rangle\Big].
\end{aligned}$$

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

$$\begin{aligned}
c_1c_2 &= -c_1^2 - c_1c_3, \qquad & c_1c_2 + c_1c_3 &= -c_1^2 \\
c_2c_3 &= -c_2^2 - c_1c_2, \qquad & c_1c_2 + c_2c_3 &= -c_2^2 \\
c_1c_3 &= -c_3^2 - c_2c_3, \qquad & c_1c_3 + c_2c_3 &= -c_3^2.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c} &= -2\Big[c_1^2\|x_1\|^2 + c_1^2\|x_2\|^2 + c_3^2\|x_3\|^2 + 2c_1c_2\langle x_1, x_2\rangle + 2c_1c_3\langle x_1, x_3\rangle + 2c_2c_3\langle x_2, x_3\rangle\Big] \\
&= -2\|c_1x_1 + c_2x_2 + c_3x_3\|^2 \leq 0.
\end{aligned}$$

Induction step: Let k be positive integer and suppose (2.28) is true for $n = k$, we will show that (2.28) is true for $n = k + 1$.
Thus

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c} &= \begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix} \begin{pmatrix} 0 & \Psi(x_1,x_2) & \cdots & \Psi(x_1,x_k) \\ \Psi(x_2,x_1) & 0 & \cdots & \Psi(x_2,x_k) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi(x_k,x_1) & \Psi(x_k,x_2) & \cdots & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \\
&\quad + c_1c_{k+1}\Big[\Psi(x_1,x_{k+1}) + \Psi(x_{k+1},x_k)\Big] + \cdots + c_kc_{k+1}\Big[\Psi(x_k,x_{k+1}) + \Psi(x_{k+1},x_k)\Big] \\
&= c_1c_2\Big[\Psi(x_1,x_2) + \Psi(x_2,x_1)\Big] + \cdots + c_kc_{k+1}\Big[\Psi(x_k,x_{k+1}) + \Psi(x_{k+1},x_k)\Big] \\
&= 2c_1c_2\|x_1 - x_2\|^2 + 2c_1c_3\|x_1 - x_3\|^2 + \cdots + 2c_kc_{k+1}\|x_k - x_{k+1}\|^2
\end{aligned}$$

$$\begin{aligned}
&= 2\Big[c_1c_2\|x_1\|^2 - 2c_1c_2\langle x_1, x_2\rangle + c_1c_2\|x_2\|^2 + \cdots \\
&\quad + c_kc_{k+1}\|x_k\|^2 - 2c_kc_{k+1}\langle x_k, x_{k+1}\rangle + c_kc_{k+1}\|x_{k+1}\|^2\Big] \\
&= 2\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1})\|x_1\|^2 + (c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1})\|x_2\|^2 + \cdots \\
&\quad + (c_1c_k + c_2c_k + \cdots + c_kc_{k+1})\|x_k\|^2 + (c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1})\|x_{k+1}\|^2 \\
&\quad - 2c_1c_2\langle x_1, x_2\rangle - \cdots - 2c_kc_{k+1}\langle x_{k-1}, x_k\rangle\Big].
\end{aligned}$$

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

$$\begin{aligned}
c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1} &= -c_1^2 \\
c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1} &= -c_2^2 \\
&\vdots \\
c_1c_k + c_2c_k + \cdots + c_kc_k &= -c_k^2 \\
c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1} &= -c_{k+1}^2.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c} &= -2\Big[c_1^2\|x_1\|^2 + c_2^2\|x_2\|^2 + \cdots + c_k^2\|x_k\|^2 + c_{k+1}^2\|x_{k+1}\|^2 \\
&\quad + 2c_1c_2\langle x_1, x_2\rangle + \cdots + 2c_kc_{k+1}\langle x_k, x_{k+1}\rangle\Big] \\
&= -2\|c_1x_1 + c_2x_2 + \cdots + c_kx_k + c_{k+1}x_{k+1}\|^2 \leq 0.
\end{aligned}$$

Therefore, Ψ is CND for all $n \in \mathbb{N}$.

Example 2.28. Let X be a topological space, $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ a real Hilbert space and $f : X \to \mathcal{H}$ a continuous mapping. Then the kernel $\Psi : X \times X \to \mathbb{R}$,

$$\Psi(x, y) = d^2(f(x), f(y)) = \|f(x) - f(y)\|^2, \qquad \forall x, y \in X, \tag{2.29}$$

is CND.

Indeed, for any $x_1, ..., x_n \in X$ and $c_1, ..., c_n \in \mathbb{R}$ such that $\sum_{i=1}^n c_i = 0$, we have

$$\mathbf{c}^T\Psi\mathbf{c} = \sum_{i=1}^n\sum_{j=1}^n c_ic_j\|f(x_i) - f(x_j)\|^2 = -2\Big\|\sum_{i=1}^n c_if(x_i)\Big\|^2 \leq 0. \tag{2.30}$$

We will show that formula (2.30) is true for all $n \in \mathbb{N}$.
For $n = 2$, we have

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c} &= \begin{pmatrix} c_1 & c_2 \end{pmatrix}\begin{pmatrix} 0 & \Psi(x_1, x_2) \\ \Psi(x_2, x_1) & 0 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\
&= c_1c_2\Big[\Psi(x_1, x_2) + \Psi(x_2, x_1)\Big] = c_1c_2\Big[\|f(x_1) - f(x_2)\|^2 + \|f(x_2) - f(x_1)\|^2\Big] \\
&= 2c_1c_2\|f(x_1) - f(x_2)\|^2
\end{aligned}$$

Since $c_1+c_2=0, c_1c_2=-c_1^2$. Then

$$\mathbf{c}^T\Psi\mathbf{c}=-2(c_1)^2\|f(x_1)-f(x_2)\|^2\le 0.$$

For $n=3$, we have

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c}&=\begin{pmatrix}c_1 & c_2 & c_3\end{pmatrix}\begin{pmatrix}0 & \Psi(x_1,x_2) & \Psi(x_1,x_3)\\ \Psi(x_2,x_1) & 0 & \Psi(x_2,x_3)\\ \Psi(x_3,x_1) & \Psi(x_3,x_2) & 0\end{pmatrix}\begin{pmatrix}c_1\\ c_2\\ c_3\end{pmatrix}\\
&=c_1c_2\Big[\Psi(x_1,x_2)+\Psi(x_2,x_1)\Big]+c_1c_3\Big[\Psi(x_1,x_3)+\Psi(x_3,x_1)\Big]\\
&\quad+c_2c_3\Big[\Psi(x_2,x_3)+\Psi(x_3,x_2)\Big]\\
&=2c_1c_2\|f(x_1)-f(x_2)\|^2+2c_1c_3\|f(x_1)-f(x_3)\|^2+2c_2c_3\|f(x_2)-f(x_3)\|^2\\
&=2\Big[c_1c_2\|f(x_1)\|^2-2c_1c_2\langle f(x_1),f(x_2)\rangle+c_1c_2\|f(x_2)\|^2\\
&\quad+c_1c_3\|f(x_1)\|^2-2c_1c_3\langle f(x_1),f(x_3)\rangle+c_1c_3\|f(x_3)\|^2\\
&\quad+c_2c_3\|f(x_2)\|^2-2c_2c_3\langle f(x_2),f(x_3)\rangle+c_2c_3\|f(x_3)\|^2\Big]\\
&=2\Big[(c_1c_2+c_1c_3)\|f(x_1)\|^2+(c_1c_2+c_2c_3)\|f(x_2)\|^2+(c_1c_3+c_2c_3)\|f(x_3)\|^2\\
&\quad-2c_1c_2\langle f(x_1),f(x_2)\rangle-2c_1c_3\langle f(x_1),f(x_3)\rangle-2c_2c_3\langle f(x_2),f(x_3)\rangle\Big].
\end{aligned}$$

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

$$\begin{aligned}
c_1c_2&=-c_1^2-c_1c_3, & c_1c_2+c_1c_3&=-c_1^2\\
c_2c_3&=-c_2^2-c_1c_2, & c_1c_2+c_2c_3&=-c_2^2\\
c_1c_3&=-c_3^2-c_2c_3, & c_1c_3+c_2c_3&=-c_3^2.
\end{aligned}$$

Then

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c}&=-2\Big[c_1^2\|f(x_1)\|^2+c_1^2\|f(x_2)\|^2+c_3^2\|f(x_3)\|^2+2c_1c_2\langle f(x_1),f(x_2)\rangle\\
&\quad+2c_1c_3\langle f(x_1),f(x_3)\rangle+2c_2c_3\langle f(x_2),f(x_3)\rangle\Big]\\
&=-2\|c_1f(x_1)+c_2f(x_2)+c_3f(x_3)\|^2\le 0.
\end{aligned}$$

Induction step: Let k be positive integer and suppose (2.30) is true for $n=k$, we will show that (2.30) is true for $n=k+1$.
Thus

$$\begin{aligned}
\mathbf{c}^T\Psi\mathbf{c}&=\begin{pmatrix}c_1 & \cdots & c_k\end{pmatrix}\begin{pmatrix}0 & \Psi(x_1,x_2) & \cdots & \Psi(x_1,x_k)\\ \Psi(x_2,x_1) & 0 & \cdots & \Psi(x_2,x_k)\\ \vdots & \vdots & \ddots & \vdots\\ \Psi(x_k,x_1) & \Psi(x_k,x_2) & \cdots & 0\end{pmatrix}\begin{pmatrix}c_1\\ \vdots\\ c_k\end{pmatrix}\\
&\quad+c_1c_{k+1}\Big[\Psi(x_1,x_{k+1})+\Psi(x_{k+1},x_k)\Big]+\cdots+c_kc_{k+1}\Big[\Psi(x_k,x_{k+1})+\Psi(x_{k+1},x_k)\Big]\\
&=c_1c_2\Big[\Psi(x_1,x_2)+\Psi(x_2,x_1)\Big]+\cdots+c_kc_{k+1}\Big[\Psi(x_k,x_{k+1})+\Psi(x_{k+1},x_k)\Big]\\
&=2c_1c_2\|f(x_1)-f(x_2)\|^2+2c_1c_3\|f(x_1)-f(x_3)\|^2+\cdots+2c_kc_{k+1}\|f(x_k)-f(x_{k+1})\|^2
\end{aligned}$$

$$= 2\Big[c_1c_2\|f(x_1)\|^2 - 2c_1c_2\langle f(x_1), f(x_2)\rangle + c_1c_2\|f(x_2)\|^2 + \cdots$$
$$+ c_kc_{k+1}\|f(x_k)\|^2 - 2c_kc_{k+1}\langle f(x_k), f(x_{k+1})\rangle + c_kc_{k+1}\|f(x_{k+1})\|^2\Big]$$
$$= 2\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1})\|f(x_1)\|^2 + (c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1})\|f(x_2)\|^2 + \cdots$$
$$+ (c_1c_k + c_2c_k + \cdots + c_kc_{k+1})\|f(x_k)\|^2 + (c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1})\|f(x_{k+1})\|^2$$
$$- 2c_1c_2\langle f(x_1), f(x_2)\rangle - \cdots - 2c_kc_{k+1}\langle f(x_{k-1}), f(x_k)\rangle\Big].$$

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

$$c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1} = -c_1^2$$
$$c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1} = -c_2^2$$
$$\vdots$$
$$c_1c_k + c_2c_k + \cdots + c_kc_k = -c_k^2$$
$$c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1} = -c_{k+1}^2.$$

Then

$$\mathbf{c}^T\Psi\mathbf{c} = -2\Big[c_1^2\|f(x_1)\|^2 + c_2^2\|f(x_2)\|^2 + \cdots + c_k^2\|f(x_k)\|^2 + c_{k+1}^2\|f(x_{k+1})\|^2$$
$$+ 2c_1c_2\langle x_1, f(x_2)\rangle + \cdots + 2c_kc_{k+1}\langle f(x_k), f(x_{k+1})\rangle\Big]$$
$$= -2\|c_1f(x_1) + c_2f(x_2) + \cdots + c_kf(x_k) + c_{k+1}f(x_{k+1})\|^2 \le 0.$$

Therefore, Ψ is CND for all $n \in \mathbb{N}$.
For symmetric kernels the following results are fundamental.

Theorem 2.29 (The GNS construction for PD kernels, [8]). 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 $\Phi(x, y) = \langle f(x), f(y)\rangle$.

Proof. For every $x \in X$, we denote by Φ_x the continuous function

$$\Phi_x : X \to \mathbb{R} \text{ or } \mathbb{C}, \qquad y \mapsto \Phi_x(y) := \Phi(x, y).$$

Let

$$V := \text{span}\Big\{\Phi_x : x \in X\Big\} = \Big\{\sum_{i=1}^{n} a_i\Phi_{x_i} : a_1, ..., a_n \in \mathbb{R} \text{ or } \mathbb{C}, x_1, ..., x_n \in X\Big\}$$

be the subset of $C(X)$, the space of continuous functions on X. For any two elements φ, ψ of V, i.e.,

$$\varphi = \sum_{i=1}^{m} a_i\Phi_{x_i} \qquad \psi = \sum_{j=1}^{n} b_j\Phi_{x_j}$$

define

$$\langle \varphi, \psi \rangle_\Phi := \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \Phi(x_i, x_j). \tag{2.31}$$

Firstly observe that

$$\begin{aligned} \langle \varphi, \psi \rangle_\Phi &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \Phi(x_i, x_j) = \sum_{j=1}^{n} b_j \sum_{i=1}^{m} a_i \Phi(x_i, x_j) \\ &= \sum_{j=1}^{n} b_j \sum_{i=1}^{m} a_i \Phi_{x_i}(x_j) = \sum_{j=1}^{n} b_j \varphi(x_j), \end{aligned}$$

and observe that the analogous relation

$$\begin{aligned} \langle \varphi, \psi \rangle_\Phi &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \Phi(x_i, x_j) = \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j \Phi(x_i, x_j) \\ &= \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j \Phi(x_j, x_i) = \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j \Phi_{x_j}(x_i) \\ &= \sum_{i=1}^{m} a_i \psi(x_i). \end{aligned}$$

That is,

$$\langle \varphi, \psi \rangle_\Phi = \sum_{j=1}^{n} b_j \varphi(x_j) = \sum_{i=1}^{m} a_i \psi(x_i). \tag{2.32}$$

The pairing (2.31) is independent of the choice of the representation of φ and ψ on V, i.e. independent of the choice of $(a_1, ..., a_n), (x_1, ..., x_n)$ for φ and $(b_1, ..., b_n), (y_1, ..., y_n)$ for ψ, respectively.

Observe that

$$\langle \varphi, \Phi_x \rangle_\Phi = \left\langle \sum_{i=1}^{m} a_i \Phi_{x_i}, \Phi_x \right\rangle = \sum_{i=1}^{m} a_i \Phi(x_i, x) = \varphi(x). \tag{2.33}$$

We also observe that in the case Φ is symmetric, hence the Cauchy-Schwartz inequality

$$|\varphi(x)|^2 \leq \Phi(x, x) \cdot \langle \varphi, \varphi \rangle_\Phi \tag{2.34}$$

holds good for all $x \in X$.

Taking into account that Φ is PD, it follows that the pair $(V, \langle \cdot, \cdot \rangle_\Phi)$ is a pre-Hilbert space.

We define now the Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_\Phi)$ to be the **completion** of $(V, \langle \cdot, \cdot \rangle_\Phi)$. In other words, the Hilbert space $\mathcal{H}$ is a subspace of the space of all functions φ on X which are pointwise limits of Cauchy sequences in $(V, || \cdot ||_\Phi)$.

Let $f : X \to V, f(x) := \Phi_x$. Then

$$\langle f(x), f(y) \rangle_\Phi = \Phi(x, y) \tag{2.35}$$

and

$$\begin{aligned} \|f(x) - f(y)\|_\Phi^2 &= \|f(x)\|_\Phi^2 - 2\langle f(x), f(y) \rangle_\Phi + \|f(y)\|_\Phi^2 \\ &= \langle f(x), f(x) \rangle_\Phi - 2\langle f(x), f(y) \rangle_\Phi + \langle f(y), f(y) \rangle_\Phi \\ &= \Phi(x, x) - 2\Phi(x, y) + \Phi(y, y). \end{aligned}$$

Since Φ is continuous, then f is continuous. $\square$

We are not going to use this result. The following result is important for applications and will be extensively used in the present research.

Theorem 2.30 (The GNS construction for CND kernels, [8]). 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 $\Psi(x, y) = \|f(x) - f(y)\|^2$.

Proof. For every $x \in X$, let $\delta_x : X \to \mathbb{R}$ be the Dirac function at x.
Let us consider the real vector space V given by

$$V := \Big\{ \varphi = \sum_{i=1}^{m} c_i \delta_{x_i} \ : \ \sum_{i=1}^{m} c_i = 0, \ x_i \in X, \ c_i \in \mathbb{R} \Big\}.$$

For any two elements φ, ψ of V, that is

$$\varphi = \sum_{i=1}^{m} a_i \delta_{x_i} \qquad \psi = \sum_{j=1}^{n} b_j \delta_{x_j}$$

we define

$$\langle \varphi, \psi \rangle_\Psi := -\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \Psi(x_i, x_j), \tag{2.36}$$

and observe that $\langle \cdot, \cdot \rangle_\Psi$ is a positive symmetric form on V.

Moreover, we consider the null space

$$N_\Psi := \{ \varphi \in V \ : \ \langle \varphi, \psi \rangle_\Psi = 0 \},$$

which obviously is a linear subspace of V.

On the quotient space V/N_Ψ, we define the pairing

$$\langle [\varphi], [\psi] \rangle_\Psi := \langle \varphi, \psi \rangle_\Psi,$$

where $[\varphi]$, and $[\psi]$ are the equivalence classes of φ and ψ in V. This is a well defined scalar product in V/N_Ψ.

We consider again the Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_\Psi)$ to be the completion of V/N_Ψ. Let x_0 be any fixed point in X and let

$$f : X \to \mathcal{H}, \qquad f(x) := [\delta_x - \delta_{x_0}].$$

It follows

$$||f(x) - f(y)||^2_\Psi = \Psi(x, y),$$

for any $x, y \in X$, since $\Psi(x, x) = \Psi(y, y) = 0$ and $\Psi(x, y) = \Psi(y, x)$. Moreover, f is continuous, since Ψ is continuous. $\square$

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

Theorem 2.31 (Schoenberg, [8]). If X is a topological space and $\Psi : X \times X \to \mathbb{R}$ continuous kernel 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 following two properties are equivalent

(A) Ψ is CND kernel

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

Next we will consider about the symmetric kernels induced by geodesic distance. In particular, if (X,d) is a metric space, then it is customary to consider kernels induced by d given in the form:

$$K(x,y) = \exp\Big(- \lambda \cdot d^q(x,y)\Big), \quad \text{for any } \lambda, q > 0.$$

Special cases are:

- if $q = 1$, then $K(x,y) = \exp\Big(- \lambda \cdot d(x,y)\Big)$ is called the **Laplacian kernel** of (X,d)
- if $q = 2$, then $K(x,y) = \exp\Big(- \lambda \cdot d^2(x,y)\Big)$ is called the **Gaussian kernel** of (X,d).

The Schoenberg theorem and GNS constructon for CND kernels imply

Theorem 2.32. If the Gaussian kernel $K(x,y) = \exp\Big(- \lambda \cdot d^2(x,y)\Big)$ of the metric space (X,d) is PD, then there exist

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

such that $d(x,y) = \|f(x) - f(y)\|_{\mathcal{H}} = d_{\mathcal{H}}(f(x), f(y))$, where $d_{\mathcal{H}}$ is the induced distance of $(\mathcal{H}, \langle \cdot, \cdot \rangle)$.

Indeed, if

$$K_d(x,y) := \exp\Big(- \lambda \cdot d^2(x,y)\Big) = \exp\Big(- \lambda \cdot \Psi(x,y)\Big)$$

is PD, with $\Psi(x,y) = d^2(x,y)$, then Schoenberg theorem implies $\Psi(x,y) = d^2(x,y)$ must be CND.

On the other hand, if Ψ is CND, then the GNS construction implies the existence of $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ and f as in hypothesis such that

$$\Psi(x,y) = \|f(x) - f(y)\|^2_{\mathcal{H}} = d^2(x,y)$$

and hence

$$d(x,y) = \|f(x) - f(y)\|_{\mathcal{H}}.$$

Remark 2.33. In other words, if the geodesic Gaussian kernel is PD, then the metric space (X, d) 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], [5].

Theorem 2.34. Let (X, d) be a geodesic metric space and assume that the Gaussian kernel $K(x, y) = \exp(-\lambda d^2(x, y))$ is PD 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 2.35. Let X be a complete, smooth Riemannian manifold with Riemannian distance function d on X. Let us assume that the Riemannian distance induced Gaussian kernel $K(x, y) = \exp(-\lambda d^2(x, y))$ is PD for all $\lambda > 0$. Then the Riemannian manifold X is isometric to an Euclidean space.

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

The geometrical reason behind this unexpected result actually comes from the injectivity of the isometric embedding. Indeed, we recall from [2] 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

$$d_Y(\varphi(x), \varphi(y)) = d_X(x, y), \qquad \forall x, y \in X. \tag{2.37}$$

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 (2.37) 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 between two length spaces is called an **arcwise isometry** if

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

An injective arcwise isometry is called an **isometric embedding.**

Remark 2.36. (i) 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.

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

Example 2.37. Let us consider the unit sphere

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

Then

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

(ii) On the other hand, there exists an embedding $\phi : \mathbb{S}^2 \to \mathbb{R}^3$ (or $\mathbb{R}^n$) that would be an isometry onto its image.

Indeed, consider the coordinate system (θ, φ) on $\mathbb{S}^2$ 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

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

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

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

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 2.38. It is interesting to see that actually we can embed $\mathbb{S}^2$ as metric space into an infinite dimensional Hilbert space [2].

Chapter 3

Non-Symmetric Kernels

In this chapter, 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 research.

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

We will consider in this chapter another type of kernels, namely non-symmetric kernels. In special, we are interested in non-symmetric kernels induced by weighted quasi-metrics.

If (X, ρ, w) is a weighted quasi-metric, we will 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 3.1. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $f : X \to \mathcal{H}$ a continuous function. Then the function

$$K(x,y) := \langle f(x), f(y)\rangle + \frac{1}{2}[w(y) - w(x)], \qquad \forall x, y \in X \tag{3.1}$$

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

Indeed, we have

$$\begin{aligned}
\mathbf{c}^T K \mathbf{c} &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j} K(x_i, x_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}\left\{\langle f(x_i), f(x_j)\rangle + \frac{1}{2}\Big[w(x_j) - w(x_i)\Big]\right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}\langle f(x_i), f(x_j)\rangle + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big[w(x_j) - w(x_i)\Big] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}\langle f(x_i), f(x_j)\rangle + \frac{1}{2}\left[\sum_{j=1}^{n} w(x_j) - \sum_{i=1}^{n} w(x_i)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}\langle f(x_i), f(x_j)\rangle \\
&= \left\|\sum_{i=1}^{n} c_i f(x_i)\right\|^2 \geq 0.
\end{aligned} \tag{3.2}$$

We will show that formula (3.2) is true for all $n \in \mathbb{N}$.
For $n = 2$, we have

$$
\begin{aligned}
\mathbf{c}^T K \mathbf{c} &= \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} K(x_1,x_1) & K(x_1,x_2) \\ K(x_2,x_1) & K(x_2,x_2) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\
&= (c_1)^2 K(x_1,x_1) + c_1\bar{c}_2 \Big[K(x_1,x_2) + K(x_2,x_1) \Big] + (c_2)^2 K(x_2,x_2) \\
&= (c_1)^2 \langle f(x_1), f(x_1) \rangle + c_1\bar{c}_2 \Big[\langle f(x_1), f(x_2) \rangle + \frac{1}{2}[w(x_2) - w(x_1)] \\
&\quad + \langle f(x_2), f(x_1) \rangle + \frac{1}{2}[w(x_1) - w(x_2)] \Big] + (c_2)^2 \langle f(x_2), f(x_2) \rangle \\
&= \Big\langle c_1 f(x_1) + c_2 f(x_2), c_1 f(x_1) + c_2 f(x_2) \Big\rangle \\
&= \Big\| c_1 f(x_1) + c_2 f(x_2) \Big\|^2 \geq 0.
\end{aligned}
$$

Induction step: Let k be positive integer and suppose (3.2) is true for $n = k$, we will show that (3.2) is true for $n = k+1$.
Thus

$$
\begin{aligned}
\mathbf{c}^T K \mathbf{c} = &\begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix} \begin{pmatrix} K(x_1,x_1) & \cdots & K(x_1,x_k) \\ K(x_2,x_1) & \cdots & K(x_2,x_k) \\ \vdots & \ddots & \vdots \\ K(x_k,x_1) & \cdots & K(x_k,x_k) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \\
&+ 2c_1\bar{c}_{k+1} \Big[K(x_1,x_{k+1}) + K(x_{k+1},x_1) \Big] + \cdots + 2c_k\bar{c}_{k+1} \Big[K(x_k,x_{k+1}) + K(x_{k+1},x_k) \Big] \\
&+ (c_{k+1})^2 K(x_{k+1},x_{k+1}) \\
= &(c_1)^2 K(x_1,x_1) + c_1\bar{c}_2 \Big[K(x_1,x_2) + K(x_2,x_1) \Big] + c_1\bar{c}_3 \Big[K(x_1,x_3) + K(x_3,x_1) \Big] \\
&+ c_1\bar{c}_4 \Big[K(x_1,x_4) + K(x_4,x_1) \Big] + \cdots + c_{k-1}\bar{c}_k \Big[K(x_{k-1},x_k) + K(x_k,x_{k-1}) \Big] \\
&+ (c_2)^2 K(x_2,x_2) + (c_3)^2 K(x_3,x_3) + \cdots + (c_k)^2 K(x_k,x_k) \\
&+ 2c_1\bar{c}_{k+1} \Big[K(x_1,x_{k+1}) + K(x_{k+1},x_1) \Big] + \cdots + 2c_k\bar{c}_{k+1} \Big[K(x_k,x_{k+1}) + K(x_{k+1},x_k) \Big] \\
&+ (c_{k+1})^2 K(x_{k+1},x_{k+1}) \\
= &(c_1)^2 \langle f(x_1), f(x_1) \rangle + c_1\bar{c}_2 \Big[\langle f(x_1), f(x_2) \rangle + \frac{1}{2}[w(x_2) - w(x_1)] + \langle f(x_2), f(x_1) \rangle \\
&+ \frac{1}{2}[w(x_1) - w(x_2)] \Big] + \cdots + c_k\bar{c}_{k+1} \Big[\langle f(x_k), f(x_{k+1}) \rangle + \frac{1}{2}[w(x_{k+1}) - w(x_k)] \\
&+ \langle f(x_{k+1}), f(x_k) \rangle + \frac{1}{2}[w(x_k) - w(x_{k+1})] \Big] + (c_2)^2 \langle f(x_2), f(x_2) \rangle \\
&+ (c_3)^2 \langle f(x_3), f(x_3) \rangle + \cdots + (c_{k+1})^2 \langle f(x_{k+1}), f(x_{k+1}) \rangle \\
= &\Big\langle c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{k+1} f(x_{k+1}), c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{k+1} f(x_{k+1}) \Big\rangle \\
= &\Big\| c_1 f(x_1) + c_2 f(x_2) + \cdots + c_{k+1} f(x_{k+1}) \Big\|^2 \geq 0.
\end{aligned}
$$

Thus, formula (3.2) holds for $n = k+1$.

Example 3.2. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $f : X \to \mathcal{H}$ a continuous function. Then the function

$$K(x,y) := \Big[d_{\mathcal{H}}(x,y) + \frac{1}{2}\big(w(y) - w(x)\big)\Big]^2 \tag{3.3}$$

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

Indeed, for any $x_1, ..., x_n \in X$ and $c_1, ..., c_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} c_i = 0$, we compute

$$\mathbf{c}^T K \mathbf{c} = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) = -2\Big\|\sum_{i=1}^{n} c_i x_i\Big\|^2 - \frac{1}{2}\Big[\sum_{i=1}^{n} c_i w(x_i)\Big]^2 \leq 0. \tag{3.4}$$

We will show that formula (3.4) is true for all $n \in \mathbb{N}$.
For $n = 2$, we have

$$\begin{aligned}
\mathbf{c}^T K \mathbf{c} &= \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} 0 & K(x_1,x_2) \\ K(x_2,x_1) & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \\
&= c_1 c_2 \Big[K(x_1,x_2) + K(x_2,x_1)\Big] \\
&= c_1 c_2 \bigg\{ \Big[\|x_1 - x_2\| + \frac{1}{2}\big[w(x_2) - w(x_1)\big]\Big]^2 + \Big[\|x_2 - x_1\| + \frac{1}{2}\big[w(x_1) - w(x_2)\big]\Big]^2 \bigg\} \\
&= c_1 c_2 \Big\{ 2\|x_1 - x_2\|^2 + \frac{1}{2}\big[w(x_1) - w(x_2)\big]^2 \Big\} \\
&= 2c_1 c_2 \|x_1 - x_2\|^2 + \frac{1}{2} c_1 c_2 \big[w(x_1) - w(x_2)\big]^2
\end{aligned}$$

Since $c_1 + c_2 = 0, c_1 c_2 = -(c_1)^2$ it follows

$$\mathbf{c}^T K \mathbf{c} = -2(c_1)^2 \|x_1 - x_2\|^2 - \frac{1}{2}(c_1)^2 \big[w(x_1) - w(x_2)\big]^2 \leq 0.$$

Induction step: Let k be positive integer and suppose (3.4) is true for $n = k$, we will show that (3.4) is true for $n = k+1$.
Thus

$$\begin{aligned}
\mathbf{c}^T K \mathbf{c} &= \begin{pmatrix} c_1 & \cdots & c_k \end{pmatrix} \begin{pmatrix} 0 & K(x_1,x_2) & \cdots & K(x_1,x_k) \\ K(x_2,x_1) & 0 & \cdots & K(x_2,x_k) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_k,x_1) & K(x_k,x_2) & \cdots & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} \\
&\quad + c_1 c_{k+1}\Big[K(x_1,x_{k+1}) + K(x_{k+1},x_1)\Big] + \cdots + c_k c_{k+1}\Big[K(x_k,x_{k+1}) + K(x_{k+1},x_k)\Big] \\
&= c_1 c_2 \Big[K(x_1,x_2) + K(x_2,x_1)\Big] + \cdots + c_{k-1} c_k \Big[K(x_{k-1},x_k) + K(x_k,x_{k-1})\Big] \\
&\quad + c_1 c_{k+1}\Big[K(x_1,x_{k+1}) + K(x_{k+1},x_1)\Big] + \cdots + c_k c_{k+1}\Big[K(x_k,x_{k+1}) + K(x_{k+1},x_k)\Big] \\
&= c_1 c_2 \Big\{ 2\|x_1 - x_2\|^2 + \frac{1}{2}\big[w(x_1) - w(x_2)\big]^2 \Big\} + \cdots \\
&\quad + c_k c_{k+1} \Big\{ 2\|x_k - x_{k+1}\|^2 + \frac{1}{2}\big[w(x_k) - w(x_{k+1})\big]^2 \Big\}
\end{aligned}$$

$$\begin{aligned}
&= 2\Big[c_1c_2\|x_1 - x_2\|^2 + \cdots + c_kc_{k+1}\|x_k - x_{k+1}\|^2\Big] \\
&\quad + \frac{1}{2}\Big[c_1c_2\Big[w(x_1) - w(x_2)\Big]^2 + \cdots + c_kc_{k+1}\Big[w(x_k) - w(x_{k+1})\Big]^2\Big] \\
&= 2\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1})\|x_1\|^2 \\
&\quad + (c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1})\|x_2\|^2 + \cdots \\
&\quad + (c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1})\|x_k\|^2 \\
&\quad - 2c_1c_2\langle x_1, x_2\rangle - \cdots - 2c_kc_{k+1}\langle x_k, x_{k+1}\rangle\Big] \\
&\quad + \frac{1}{2}\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1})w(x_1)^2 \\
&\quad + (c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1})w(x_2)^2 + \cdots \\
&\quad + (c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1})w(x_{k+1})^2 \\
&\quad - 2c_1c_2w(x_1)w(x_2) - \cdots - 2c_kc_{k+1}w(x_k)w(x_{k+1})\Big]
\end{aligned}$$

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

$$\begin{aligned}
c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1} &= -c_1^2 \\
c_1c_2 + c_2c_3 + \cdots + c_2c_{k+1} &= -c_2^2 \\
&\vdots \\
c_1c_k + c_2c_k + \cdots + c_kc_{k+1} &= -c_k^2 \\
c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1} &= -c_{k+1}^2,
\end{aligned}$$

it follows

$$\begin{aligned}
c^TKc = -2\|c_1x_1 + c_2x_2 + \cdots + c_{k+1}x_{k+1}\|^2 \\
-\frac{1}{2}\Big[c_1w(x_1) + c_2w(x_2) + \cdots + c_{k+1}w(x_{k+1})\Big]^2 \leq 0.
\end{aligned}$$

Thus, formula (3.4) holds for $n = k+1$.

In order to prove this example, we get the following lemma

Lemma 3.3. If $w : X \to \mathbb{R}$ is a continuous function and $c_1 + \cdots + c_n = 0$, then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_ic_j\big[w(x_i) - w(x_j)\big]^2 = -2\Big[\sum_{i=1}^{n} c_iw(x_i)\Big]^2 \leq 0. \tag{3.5}$$

for all $x_1, ..., x_n \in X$.

Proof. We will show that formula (3.5) is true for all $n \in \mathbb{N}$.

For $n = 2$, we have

$$\begin{aligned}
\sum_{i=1}^{2}\sum_{j=1}^{2} c_ic_j\big[w(x_i) - w(x_j)\big]^2 &= c_1^2\big[w(x_1) - w(x_1)\big]^2 + c_1c_2\big[w(x_1) - w(x_2)\big]^2 \\
&\quad + c_1c_2\big[w(x_2) - w(x_1)\big]^2 + c_2^2\big[w(x_2) - w(x_2)\big]^2 \\
&= 2c_1c_2\big[w(x_1) - w(x_2)\big]^2
\end{aligned}$$

Since $c_1+c_2=0, c_1c_2=-c_1^2$ or $-c_2^2$. Then

$$\sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_j\left[w(x_i)-w(x_j)\right]^2=-2c_1^2\left[w(x_1)-w(x_2)\right]^2\leq 0.$$

For $n=3$, we have

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=1}^{3}c_ic_j\left[w(x_i)-w(x_j)\right]^2 &= \sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_j\left[w(x_i)-w(x_j)\right]^2+2c_1c_3\left[w(x_1)-w(x_3)\right]^2\\
&\quad+2c_2c_3\left[w(x_2)-w(x_3)\right]^2\\
&=2c_1c_2\left[w(x_1)-w(x_2)\right]^2+2c_1c_3\left[w(x_1)-w(x_3)\right]^2\\
&\quad+2c_2c_3\left[w(x_2)-w(x_3)\right]^2\\
&=2\Big[c_1c_2\left[w(x_1)-w(x_2)\right]^2+c_1c_3\left[w(x_1)-w(x_3)\right]^2\\
&\quad+c_2c_3\left[w(x_2)-w(x_3)\right]^2\Big]\\
&=2\Big[c_1c_2w(x_1)^2+c_1c_2w(x_2)^2+c_1c_3w(x_1)^2+c_1c_3w(x_3)^2\\
&\quad+c_2c_3w(x_2)^2+c_2c_3w(x_3)^2-2c_1c_2w(x_1)w(x_2)\\
&\quad-2c_1c_3w(x_1)w(x_3)-2c_2c_3w(x_2)w(x_3)\Big]\\
&=2\Big[(c_1c_2+c_1c_3)w(x_1)^2+(c_1c_2+c_2c_3)w(x_2)^2+(c_1c_3+c_2c_3)w(x_3)^2\\
&\quad-2c_1c_2w(x_1)w(x_2)-2c_1c_3w(x_1)w(x_3)-2c_2c_3w(x_2)w(x_3)\Big]\end{aligned}$$

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

$$\begin{aligned}c_1c_2+c_1c_3&=-c_1^2\\ c_1c_2+c_2c_3&=-c_2^2\\ c_1c_3+c_2c_3&=-c_3^2.\end{aligned}$$

Then

$$\begin{aligned}\sum_{i=1}^{3}\sum_{j=1}^{3}c_ic_j\left[w(x_i)-w(x_j)\right]^2 &= -2\Big[c_1^2w(x_1)^2+c_2^2w(x_2)^2+c_3^2w(x_3)^2\\
&\quad+2c_1c_2w(x_1)w(x_2)+2c_1c_3w(x_1)w(x_3)+2c_2c_3w(x_2)w(x_3)\Big]\\
&=-2\Big[c_1w(x_1)+c_2w(x_2)+c_3w(x_3)\Big]^2\leq 0.\end{aligned}$$

Induction step: Let k be positive integers and suppose (3.5) is CND for $n=k$, we will show that (3.5) is CND for $n=k+1$.
Thus

$$\begin{aligned}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\left[w(x_i)-w(x_j)\right]^2 &= \sum_{i=1}^{k}\sum_{j=1}^{k}c_ic_j\left[w(x_i)-w(x_j)\right]^2+2c_1c_{k+1}\left[w(x_1)-w(x_{k+1})\right]^2\\
&\quad+2c_2c_{k+1}\left[w(x_2)-w(x_{k+1})\right]^2+\cdots\\
&\quad+2c_kc_{k+1}\left[w(x_k)-w(x_{k+1})\right]^2\end{aligned}$$

$$
\begin{aligned}
&= 2\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_k)w(x_1)^2 + \cdots \\
&\quad + (c_1c_k + c_2c_k + \cdots + c_{k-1}c_k)w(x_k)^2 \\
&\quad - 2c_1c_2w(x_1)w(x_2) - \cdots - 2c_{k-1}c_kw(x_{k-1})w(x_k)\Big] \\
&\quad + 2c_1c_{k+1}\big[w(x_1) - w(x_{k+1})\big]^2 + \cdots \\
&\quad + 2c_kc_{k+1}\big[w(x_k) - w(x_{k+1})\big]^2 \\
&= 2\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_k)w(x_1)^2 + c_1c_{k+1}w(x_1)^2 + \cdots \\
&\quad + (c_1c_k + c_2c_k + \cdots + c_{k-1}c_k)w(x_k)^2 + c_kc_{k+1}w(x_k)^2 \\
&\quad + (c_1c_{k+1} + c_2c_{k+1} \cdots + c_kc_{k+1})w(x_{k+1})^2 \\
&\quad - 2c_1c_2w(x_1)w(x_2) - \cdots - 2c_kc_{k+1}w(x_{k-1})w(x_k)\Big] \\
&= 2\Big[(c_1c_2 + c_1c_3 + \cdots + c_1c_{k+1})w(x_1)^2 + \cdots \\
&\quad + (c_1c_k + c_2c_k + \cdots + c_kc_{k+1})w(x_k)^2 \\
&\quad + (c_1c_{k+1} + c_2c_{k+1} \cdots + c_kc_{k+1})w(x_{k+1})^2 \\
&\quad - 2c_1c_2w(x_1)w(x_2) - \cdots - 2c_kc_{k+1}w(x_{k-1})w(x_k)\Big]
\end{aligned}
$$

Since $c_1 + c_2 + \cdots + c_k = 0$, we have

$$
\begin{aligned}
c_1c_2 + c_1c_3 + \cdots + c_1c_k &= -c_1^2 \\
c_1c_2 + c_2c_3 + \cdots + c_2c_k &= -c_2^2 \\
&\vdots \\
c_1c_k + c_2c_k + \cdots + c_kc_{k+1} &= -c_k^2 \\
c_1c_{k+1} + c_2c_{k+1} + \cdots + c_kc_{k+1} &= -c_{k+1}^2
\end{aligned}
$$

Then

$$
\begin{aligned}
\sum_{i=1}^{k+1}\sum_{j=1}^{k+1} c_ic_j\big[w(x_i) - w(x_j)\big]^2 &= -2\Big[c_1^2w(x_1)^2 + c_2^2w(x_2)^2 + \cdots c_{k+1}^2w(x_{k+1})^2 \\
&\qquad + 2c_1c_2w(x_1)w(x_2) + \cdots + 2c_kc_{k+1}w(x_k)w(x_{k+1})\Big] \\
&= -2\Big[c_1w(x_1) + c_2w(x_2) + \cdots + c_{k+1}w(x_{k+1})\Big]^2 \leq 0.
\end{aligned}
$$

The proof is complete. □

Proposition 3.4. (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 X is a convex cone.

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

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

Proof. (i) Let K_1 and K_2 are CND, then for any $x_1, ..., x_n \in X$ and $c_1, ..., c_n \in \mathbb{R}$ such that $c_1 + \cdots + c_n = 0$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_1(x_i, x_j) \leq 0,$$
$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_2(x_i, x_j) \leq 0.$$

Then,

$$\begin{aligned} 0 &\geq s \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_1(x_i, x_j) + t \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_2(x_i, x_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \cdot (s \cdot K_1 + t \cdot K_2)(x_i, x_j). \end{aligned}$$

Therefore, the set of CND kernels on X is a convex cone.

(ii) Let $\{K_t\}_t$ is a family of CND kernels converging point-wise on $X \times X$ to a continuous kernel $K : X \times X \to \mathbb{R}$, then for any $(x, y) \in X \times X$, we have for any $\varepsilon > 0$, there exists $T > 0$ such that $|K_t(x, y) - K(x, y)| < \varepsilon$, for $t > T$.
For any $x_1, ..., x_n \in X$ we have,

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

and hence, for any $c_1, ..., c_n \in \mathbb{R}$, we obtain

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \cdot \varepsilon < \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_t(x_i, x_j) - \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) < \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \cdot \varepsilon$$
$$-\Big(\sum_{i=1}^{n} c_i\Big) \cdot \sum_{j=1}^{n} c_j \cdot \varepsilon < \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_t(x_i, x_j) - \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) < \Big(\sum_{i=1}^{n} c_i\Big) \cdot \sum_{j=1}^{n} c_j \cdot \varepsilon$$

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

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

Therefore, K is also CND.

(iii) Let K is PD kernel on X. For any $c_1, ..., c_n \in \mathbb{R}$, such that $\sum_{i=1}^{n} c_i = 0$ and $x_1, ..., x_n \in X$, we compute

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j H(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) \\ &= \Big(\sum_{j=1}^{n} c_j\Big) \cdot \Big[\sum_{i=1}^{n} c_i K(x_i, x_i)\Big] - \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) \leq 0 \end{aligned}$$

Therefore, H is CND.

$\square$

Remark 3.5. Let $K : X \times X \to \mathbb{R}$ be an arbitrary non-symmetric kernel and let us denote

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

the **average symmetrized kernel**.

Then, an elementary computation shows that for any $x_1, ..., x_n \in X$ and any $c_1, ..., c_n \in \mathbb{R}$, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j H(x_i, x_j).$$

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 3.6. (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 onto account that $d(x,y) = \frac{1}{2}\Big[\rho(x,y) + \rho(y,x)\Big]$.
That is,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \rho(x_i, x_j), \tag{3.6}$$

for any $x_1, ..., x_n \in X$ and any $c_1, ..., c_n \in \mathbb{R}$.

Proof. We will show that formula (3.6) is true for all $n \in \mathbb{N}$.
For $n = 2$, we have

$$\begin{aligned}\sum_{i=1}^{2}\sum_{j=1}^{2} c_i c_j d(x_i, x_j) &= \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} c_i c_j \Big[\rho(x_i, x_j) + \rho(x_j, x_i)\Big] \\ &= \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} c_i c_j \rho(x_i, x_j) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} c_i c_j \rho(x_j, x_i)\end{aligned}$$

Since,

$$\begin{aligned}\sum_{i=1}^{2}\sum_{j=1}^{2} c_i c_j \rho(x_i, x_j) &= c_1^2\rho(x_1,x_1) + c_1c_2\rho(x_1,x_2) + c_2c_1\rho(x_2,x_1) + c_2^2\rho(x_2,x_2) \\ &= c_1^2\rho(x_1,x_1) + c_1c_2\rho(x_2,x_1) + c_2c_1\rho(x_1,x_2) + c_2^2\rho(x_2,x_2) \\ &= \sum_{i=1}^{2}\sum_{j=1}^{2} c_i c_j \rho(x_j, x_i).\end{aligned}$$

Then,

$$\sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_jd(x_i,x_j) = \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_j\rho(x_i,x_j) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_j\rho(x_j,x_i)$$
$$= \sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_j\rho(x_i,x_j).$$

Induction step: Let k be positive integers and suppose (3.6) is CND for $n = k$, we will show that (3.6) is CND for $n = k+1$.

Thus

$$\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_jd(x_i,x_j) = \frac{1}{2}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\Big[\rho(x_i,x_j)+\rho(x_j,x_i)\Big]$$
$$= \frac{1}{2}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_i,x_j) + \frac{1}{2}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_j,x_i)$$

Since,

$$\begin{aligned}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_i,x_j) &= c_1^2\rho(x_1,x_1) + c_1c_2\rho(x_1,x_2) + \cdots + c_1c_{k+1}\rho(x_1,x_{k+1}) + \cdots\\
&\quad + c_{k+1}c_1\rho(x_{k+1},x_1) + c_{k+1}c_2\rho(x_{k+1},x_2) + \cdots + c_{k+1}c_k\rho(x_{k+1},x_k)\\
&\quad + c_2^2\rho(x_2,x_2) + c_3^2\rho(x_3,x_3) + \cdots + c_{k+1}^2\rho(x_{k+1},x_{k+1})\\
&= c_1^2\rho(x_1,x_1) + c_1c_2\rho(x_2,x_1) + \cdots + c_1c_{k+1}\rho(x_{k+1},x_1) + \cdots\\
&\quad + c_{k+1}c_1\rho(x_1,x_{k+1}) + c_{k+1}c_2\rho(x_2,x_{k+1}) + \cdots + c_{k+1}c_k\rho(x_k,x_{k+1})\\
&\quad + c_2^2\rho(x_2,x_2) + c_3^2\rho(x_3,x_3) + \cdots + c_{k+1}^2\rho(x_{k+1},x_{k+1})\\
&= \sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_j,x_i).\end{aligned}$$

Then,

$$\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_jd(x_i,x_j) = \frac{1}{2}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_i,x_j) + \frac{1}{2}\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_j,x_i)$$
$$= \sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\rho(x_i,x_j).$$

From remark 3.5, we can conclude that $\rho(x,y)$ is PD if and only if $d(x,y)$ is PD and $\rho(x,y)$ is CND if and only if $d(x,y)$ is CND. □

Example 3.7. 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 $u = (x,\xi)$ and $v = (y,\eta)$ are points in $\mathcal{H}\times[0,\infty)$. Obviously $\bar{Q}$ is a quasi-metric on $\mathcal{H}\times[0,\infty)$.

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

One can see easily now 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 3.8. 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 (X, ρ, w) 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 $\psi : X \to \mathcal{H}$ such that

$$d(x,y) = \langle \psi(x), \psi(y) \rangle, \qquad \forall x,y \in X.$$

We will extend this mapping to $G_f := \{(x, f(x) : x \in X\}$: graph of f, that is

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

by defining

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

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

$$\bar{Q}\big((x,a),(y,b)\big) = \langle x, y \rangle + b - a, \text{ and } \quad \bar{W}(x,a) = 2 \cdot a.$$

We compute,

$$\begin{aligned} Q(u,v) &= Q\big((x,f(x)),(y,f(y))\big) = d(x,y) + f(y) - f(x) \\ &= \langle \psi(x), \psi(y) \rangle + f(y) - f(x) = \bar{Q}(\varphi(u), \varphi(v)) \end{aligned}$$

and similarly $W(u) = W((x, f(x))) = 2 \cdot f(x) = W((\psi(x), f(x)) = \bar{W}(\varphi(u))$. ☐

We have the following general result.

Proposition 3.9. Let $\psi : X \times X \to \mathbb{R}$ be a (non-symmetric) kernel on X. If $\exp(-\lambda \cdot \psi)$ is PD, then ψ is CND.

Proof. We assume $\mathcal{L}(x,y) = \exp(-\lambda \cdot \psi(x,y))$ is PD. Then we observe that

$$\Phi = 1 - \exp(-\lambda \cdot \psi) = 1 - \mathcal{L}$$

is CND.

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

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j [1 - \exp(-\lambda \cdot \psi(x_i, x_j))] = -\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \exp(-\lambda \cdot \psi(x_i, x_j)) \leq 0$$

by hypothesis.

Using now the fact that the set of CND kernels on X is a closed set we have that

$$\lim_{t \to 0} \frac{1 - \mathcal{L}}{t} = \lim_{t \to 0} \frac{1 - \exp(-t \cdot \psi(x,y))}{t}$$

is also CND kernel.

Recalling from L'Hospital's Theorem

$$\lim_{t \to 0} \frac{1 - \exp(-t \cdot \psi(x,y))}{t} = \psi(x,y)$$

it follows that $\psi(x,y)$ is CND kernel. □

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

Theorem 3.10. (i) The weighted quasi-metric space (X, ρ, w) is CND if and only if Laplacian kernel $\exp(-\lambda \cdot \rho(x,y))$ 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 $\mathcal{L}(x,y) = \exp(-\lambda \cdot \rho(x,y))$ is PD. Then we observe that

$$\Phi = 1 - \exp(-\lambda \cdot \rho) = 1 - \mathcal{L}$$

is CND.

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

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

by hypothesis.

Using now the fact that the set of CND kernels on X is a closed set we have that

$$\lim_{t \to 0} \frac{1 - \mathcal{L}}{t} = \lim_{t \to 0} \frac{1 - \exp(-t \cdot \rho(x,y))}{t}$$

is also CND kernel.

Recalling from L'Hospital's Theorem

$$\lim_{t\to 0}\frac{1-\exp(-t\cdot\psi(x,y))}{t}=\rho(x,y)$$

it follows that $\rho(x,y)$ is CND kernel.

Conversely, we assume now that (X,ρ) is CND. Then it follows (X,d) is CND. Form

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

A simple computation shows that

$$\exp(-\lambda\cdot\rho(x,y))=\exp(-\lambda\cdot d(x,y))\cdot\exp(-\frac{\lambda}{2}[w(y)-w(x)]).$$

For any $x_1,...,x_n$ and any $c_1,...,c_n$ we compute

$$\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\exp(-\lambda\cdot\rho(x_i,x_j))=\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\exp(-\lambda\cdot d(x_i,x_j))\cdot c_ic_j\exp(-\frac{\lambda}{2}[w(x_j)-w(x_i)])$$

For $n=2$ we have,

$$\begin{aligned}
&\sum_{i=1}^{2}\sum_{j=1}^{2}c_ic_j\exp(-\lambda\cdot\rho(x_i,x_j))\\
&=c_1^2\exp(-\lambda\cdot\rho(x_1,x_1))+c_1c_2\exp(-\lambda\cdot\rho(x_1,x_2))\\
&\quad+c_2c_1\exp(-\lambda\cdot\rho(x_2,x_1))+c_2^2\exp(-\lambda\cdot\rho(x_2,x_2))\\
&=c_1^2\exp(-\lambda\cdot d(x_1,x_1))+c_1c_2\exp(-\lambda\cdot d(x_1,x_2))\cdot c_ic_j\exp(-\frac{\lambda}{2}[w(x_2)-w(x_1)])\\
&\quad+c_2c_1\exp(-\lambda\cdot d(x_2,x_1))\cdot c_ic_j\exp(-\frac{\lambda}{2}[w(x_1)-w(x_2)])+c_2^2\exp(-\lambda\cdot d(x_2,x_2))\\
&=c_1^2\exp(-\lambda\cdot d(x_1,x_1))+c_1c_2\cdot\exp(-\lambda\cdot d(x_1,x_2))\Big\{\exp(-\frac{\lambda}{2}[w(x_2)-w(x_1)])\\
&\quad+\exp(-\frac{\lambda}{2}[w(x_1)-w(x_2)])\Big\}+c_2^2\exp(-\lambda\cdot d(x_2,x_2))\\
&=c_1^2\exp(-\lambda\cdot d(x_1,x_1))+c_1c_2\exp(-\lambda\cdot d(x_1,x_2))+c_2^2\exp(-\lambda\cdot d(x_2,x_2))\\
&\quad+c_1c_2\exp(-\lambda\cdot d(x_1,x_2))\Big\{\exp(-\frac{\lambda}{2}[w(x_2)-w(x_1)])+\exp(-\frac{\lambda}{2}[w(x_1)-w(x_2)])-1\Big\}.
\end{aligned}$$

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

$$c_1^2\exp(-\lambda\cdot d(x_1,x_1))+c_1c_2\exp(-\lambda\cdot d(x_1,x_2))+c_2^2\exp(-\lambda\cdot d(x_2,x_2))\geq 0. \quad (3.7)$$

On the other hand, observe that

$$\exp(-\frac{\lambda}{2}[w(x_2)-w(x_1)])+\exp(-\frac{\lambda}{2}[w(x_1)-w(x_2)])-1>0. \quad (3.8)$$

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

$$\begin{aligned}
\exp(-\frac{\lambda}{2}[w(x_2)-w(x_1)])+\exp(-\frac{\lambda}{2}[w(x_1)-w(x_2)])-1&=Y+\frac{1}{Y}-1\\
&=\frac{Y^2-Y+1}{Y}>0
\end{aligned}$$

because $Y^2-Y+1>0, Y>0$ for any Y and from (3.7), (3.8) it results that $\exp(-\lambda\cdot\rho(x,y))$ is PD.

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

$$d(x,y) = d^2_{\mathcal{H}}(\psi(x),\psi(y)) = \langle \psi(x)-\psi(y), \psi(x)-\psi(y)\rangle \tag{3.9}$$

We will extend this mapping to $G_f := \{(x, f(x) : x \in X\}$: graph of f, that is

$$Q: G_f \times G_f \to [0,\infty), \qquad Q(u,v) := \sqrt{d(x,y)} + f(y) - f(x),$$
$$W: G_f \to [0,\infty), \qquad W(u) := 2\cdot f(x), \qquad \forall u = (x,f(x)), v = (y, f(y)) \in G_f,$$

by defining

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

We will show now that this is an isomorphism of (G_f, Q, W) with $(\mathcal{H}\times[0,\infty), \bar{Q}, \bar{W})$, where

$$\bar{Q}\big((x,a),(y,b)\big) = d_{\mathcal{H}}(x,y) + b - a, \text{ and } \qquad \bar{W}(x,a) = 2\cdot a.$$

We compute,

$$\begin{aligned}\bar{Q}(\varphi(u),\varphi(v)) &= \bar{Q}\big((\psi(x), f(x)), (\psi(x), f(y))\big)\\ &= d_{\mathcal{H}}(\psi(x),\psi(y)) + f(y) - f(x)\\ &= \sqrt{d(x,y)} + f(y) - f(x) = Q(u,v)\end{aligned}$$

and similarly $W(u) = W((x, f(x))) = 2\cdot f(x) = W((\psi(x), f(x)) = \bar{W}(\varphi(u))$.

□

We turn now to Gaussian kernels $\mathcal{G}(x,y) = \exp(-\lambda\cdot\rho^2(x,y)), \lambda > 0$.

Theorem 3.11. If (X, d^2) is CND then,

(i) (X,ρ^2) is CND, for $\rho(x,y) = d(x,y) + f(y) - f(x)$

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

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

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

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

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j [f(x_i) - f(x_j)]^2$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) - 2\Big[\sum_{i=1}^{n} c_i f(x_i)\Big]^2,$$

due to lemma 3.3. Since d^2 is CND it result $\frac{1}{2}[\rho^2(x,y) + \rho^2(y,x)]$ is CND and due to lemma 3.6, ρ^2 is CND.

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

$$d^2(x,y) = \|\psi(x) - \psi(y)\|^2_{\mathcal{H}} = d^2_{\mathcal{H}}(\psi(x), \psi(y)), \tag{3.10}$$

that is $d(x,y) = d_{\mathcal{H}}(\psi(x), \psi(y))$.
Then, we will extend this mapping to $G_f := \{(x, f(x) : x \in X\}$: graph of f, that is

$$Q : G_f \times G_f \to [0,\infty), \qquad Q(u,v) := d(x,y) + f(y) - f(x),$$
$$W : G_f \to [0,\infty), \qquad W(u) := 2 \cdot f(x), \qquad \forall u = (x, f(x)), v = (y, f(y)) \in G_f,$$

and defining

$$\varphi : (G_f, Q, W) \to (\mathcal{H} \times [0,\infty), \bar{Q}, \bar{W}), \qquad \varphi(u) = \varphi(x, f(x)) = (\psi(x), f(x))$$

We will show now that this is an isomorphism of (G_f, Q, W) with $(\mathcal{H} \times [0,\infty), \bar{Q}, \bar{W})$, where

$$\bar{Q}\big((x,a),(y,b)\big) = d_{\mathcal{H}}(x,y) + b - a, \text{ and } \qquad \bar{W}(x,a) = 2 \cdot a.$$

By computing, we get

$$Q(u,v) = d(x,y) + f(y) - f(x) = d_{\mathcal{H}}(\varphi(x), \varphi(y)) + w(y) - w(x)$$
$$= \bar{Q}(\varphi(u), \varphi(v)),$$

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

□

Remark 3.12. 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 3.11 gives

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \rho^2(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) + 2\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d(x_i, x_j)[f(x_j) - f(x_i)]$$
$$+ \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j [f(x_j) - f(x_i)]^2.$$

It can be proved by induction that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d(x_i, x_j)[f(x_j) - f(x_i)] = 0.$$

Then, we get

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \rho^2(x_i, x_j) &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j [f(x_i) - f(x_j)]^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) - 2\Big[\sum_{i=1}^{n} c_i f(x_i)\Big]^2.\end{aligned}$$

and hence

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j d^2(x_i, x_j) = 2\Big[\sum_{i=1}^{n} c_i f(x_i)\Big]^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \rho^2(x_i, x_j). \tag{3.11}$$

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 is CND do not implies d^2 is CND in general.

Therefore, we conclude that ρ^2 is CND do not implies 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, ρ, w) and (Y, p, u), if $p(\varphi(x), \varphi(y)) = \rho(x, y)$ and $u(\varphi(x)) = w(x)$, for any $x, y \in X$ (see chapter 2 and [7]).

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

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

leads to $\rho(x, y) = \rho(y, x)$ if and only if $w(y) - w(x) = 0$,
i.e. $\rho(x, y) = \rho(y, x) = d(x, y)$. 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

$$p(\varphi(x_1), \varphi(x_2)) = \rho(x_1, x_2),$$
$$u(\varphi(x_1)) = w(x_1),$$
$$u(\varphi(x_2)) = w(x_2),$$

and hence

$$p(\varphi(x_1), \varphi(x_2)) = 0 = \rho(x_1, x_2),$$
$$p(\varphi(x_2), \varphi(x_1)) = 0 = \rho(x_2, x_1)$$

Then $\rho(x_1, x_2) = \rho(x_2, x_1) = 0$ 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 conclusion is 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 3.13. Consider $X = \mathbb{S}^2$ the unit 2-dimensional sphere in $\mathbb{R}^2$ or $\mathbb{R}^n$ with distance function ρ, and let $w : \mathbb{S}^2 \to \mathbb{R}$ be the function $w(\theta, \varphi) = \tan\theta$ for the parametrization $(\theta, \varphi) \in [0, \pi] \times [0, 2\pi]$. Clearly $(\mathbb{S}, \rho, w)$ obtained from d and w is a (generalized) weighted quasi-metric built on the total space $\mathbb{S}^2 \times \mathbb{R}$.

Chapter 4

Conclusions and Suggestions

4.1 Conclusions

If ψ is a non-symmetric kernel, then the underlying space do not need to be flat. In other words, the non-symmetry of the kernel is able to solve the contradiction between the curvature and the positive definiteness of the kernel.

The main conclusion that follows from here is that the non-symmetric kernels allow us a good statistical analysis of sets of data obtained from other curved spaces. The statistical analysis of this type of data is very important for mathematical modelling and especially for time periodic data sets, data in computer graphics, biology, etc.

Another important conclusion is that the real world is not a flat, Euclidean one, but rather a quite curved one and that new mathematical tools are necessary for exploring this world. Since locally any topological space on manifold is flat, the analysis of curved spaces brings us from the world of local data sets to the longer world of global data sets.

Nevertheless, our findings show that the world not need to be thought as being Euclidean or Riemannian. The real world is Finslerian, non-symmetric and curved.

The research of non-symmetric kernels do not ends here, yet this is only the beginning. Many other statistical and geometrical properties remain to be studied. Applications of non-symmetric kernels to SVM and other classical statistical analysis techniques are still open and it would be very interesting to compare the results of symmetric kernels analysis with non-symmetric kernels analysis. My conclusion is that further research will show in what cases symmetric kernels should be used and when we must consider non-symmetric ones. Such methodology selection skills will be necessary in the further for any researcher working in data analysis.

4.2 Suggestions

Our research suggests that more study concerning weighted quasi metrics and non-symmetric kernels is necessary.

For instance, it would be very interesting and useful to have an implementation of SVM and other machine learning algorithms using non-symmetric kernels in the open-source statistical programming language R.

Another idea suggested by our research would be an implementation of the well-known k-means clustering method replacing the usual symmetric metric (Euclidean, correlation, metric, etc) with a weighted quasi-metric. A comparison of clustering results will point out advantages and disadvantages of both methods. We hope to study some of these in the future.

References

[1] Bierens, H.J., 2007. **Introduction to Hilbert Spaces.** Pennsylvania State University.

[2] Burago, D., Burago, Y. and Ivanov, S., 2001. **A course on Metric Geometry**. American Mathematical Society 33.

[3] Cristisnini, N., Shawe-Taylor, J., Saunders, C., 2007. **kernel Methods: A Paradigm for Pattern Analysis**. Kernel methods in Bioengineering, Signal and Image Processing, 1-40.

[4] Deza, M., Deza, E., 2006. **Dictionary of Distances**, Elsevier Science.

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

[6] 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.

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

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

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

Appendices

Appendix A

Research Paper

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 10520, Thailand
[2]School of Biological Science, Department of Biology, Tokai University,
Sapporo Campus 5-1-1-1 Minamisawa, Minamiku, Sapporo, 005-8601 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

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

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 $x_1,\dots,x_n \in X$ and any scalars $c_1,\dots,c_n \in \mathbb{R}$ we have

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

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 $\Phi\left(x,y\right) = \langle f\left(x\right), f\left(y\right)\rangle,\ \forall x,y \in X$ is positive definite.

Remark 2.4 Obviously $\Phi\left(x,y\right) = \langle x,y\rangle$ 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. $K\left(x,x\right) = 0$ for all $x \in X$.
2. $K\left(x,y\right) = K\left(y,x\right)$ for all $x,y \in X$.
3. Forany $n \in \mathbb{N}$, and elements $x_1,\dots,x_n \in X$ and any real numbers $c_1,\dots,c_n$ with $c_1 + c_2 + \dots + c_n = 0$ we have

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

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

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

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: $K(x, y) = \exp(-\lambda \cdot \rho^q(x, y))$, $\lambda, q > 0$ .The Schoenberg theorem and GNS construction for CND kernels imply.

Theorem 2.9 *If the Gaussiankernel* $K(x, y) = \exp(-\lambda \cdot \rho^2(x, y))$ *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 $\rho(x, y) = \| f(x) - f(y) \|_{\mathcal{H}} = d_{\mathcal{H}}(f(x), f(y))$,

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* $K(x, y) = \exp(-\lambda \cdot d^2(x, y))$ *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* $K(x, y) = \exp(-\lambda \cdot d^2(x, y))$ *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.

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

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

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

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 $\phi:\ S^2 \to \mathbb{R}^3\ (or\ \mathbb{R}^n)$ 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

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

$d: M \times M \to \mathbb{R}$ 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

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

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* $\rho(x, y) = d(x, y) + \frac{1}{2}[w(y) - w(x)]$ *for any* $x, y \in M$.

Moreover, we have $\frac{1}{2}|w(x) - w(y)| \leq d(x, y)$ *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

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

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

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

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

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

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

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

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 $(\mathcal{H}, \langle\cdot,\ \cdot\rangle)$.

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

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*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\left[w(x_i)-w(x_j)\right]^2=-2\left[\sum_{i=1}^{n}c_iw(x_i)\right]^2\leq 0 \text{ for all } x_1,\dots,x_n\in\mathcal{H}. \tag{4.3}$$

Proof. For $n=2$:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\left[w(x_i)-w(x_j)\right]^2=c_1^2\left[w(x_1)-w(x_1)\right]^2+c_1c_2\left[w(x_1)-w(x_2)\right]^2+$$
$$c_1c_2\left[w(x_2)-w(x_1)\right]^2+c_2^2\left[w(x_2)-w(x_2)\right]^2$$
$$=2c_1c_2\left[w(x_1)-w(x_2)\right]^2.$$

Since $c_1+c_2=0,\ c_1c_2=-c_1^2$. Then

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

For $n=3$:

$$\sum_{i=1}^{3}\sum_{j=1}^{3}c_ic_j\left[w(x_i)-w(x_j)\right]^2=2\Big[(c_1c_2+c_1c_3)w(x_1)^2+(c_1c_2+c_2c_3)w(x_2)^2+(c_1c_3+c_2c_3)w(x_1)^2$$
$$-2c_1c_2w(x_1)w(x_2)-2c_1c_3w(x_1)w(x_3)-2c_2c_3w(x_2)w(x_3)\Big].$$

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

$$c_1c_2+c_1c_3=-c_1^2,\ c_1c_2+c_2c_3=-c_2^2,\ c_1c_3+c_2c_3=-c_3^2.$$

Then

$$\sum_{i=1}^{3}\sum_{j=1}^{3}c_ic_j\left[w(x_i)-w(x_j)\right]^2=-2\Big[c_1^2w(x_1)^2+c_2^2w(x_2)^2+c_3^2w(x_3)^2$$
$$+2c_1c_2w(x_1)w(x_2)+2c_1c_3w(x_1)w(x_3)+2c_2c_3w(x_2)w(x_3)\Big]$$
$$=-2\left[c_1w(x_1)+c_2w(x_2)+c_3w(x_3)\right]^2\leq 0.$$

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

$$\sum_{i=1}^{k+1}\sum_{j=1}^{k+1}c_ic_j\left[w(x_i)-w(x_j)\right]^2=2\Big[(c_1c_2+c_1c_3+\cdots+c_1c_{k+1})w(x_1)^2+(c_1c_2+c_2c_3+\cdots+c_2c_{k+1})w(x_2)^2$$
$$+\cdots+(c_1c_{k+1}+c_2c_{k+1}+\cdots+c_kc_{k+1})w(x_{k+1})^2-2c_1c_2w(x_1)w(x_2)$$
$$-2c_1c_3w(x_1)w(x_3)-\cdots-2c_kc_{k+1}w(x_k)w(x_{k+1})\Big].$$

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

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

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

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* $\{K_i\}_i$ *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.*

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

(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 $x_1,\ldots, x_n$ and $c_1,\ldots, c_n$ 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,

$$0\ge s\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jK_1\left(x_i,x_j\right)+t\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jK_2\left(x_i,x_j\right)$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_j\cdot\left(s\cdot K_1+t\cdot K_2\right)\left(x_i,x_j\right).$$

(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 $x_1,\ldots, x_n\in M$, 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 $c_1,\ldots,c_n\in\mathbb{R}$ 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 $c_1,\ldots, c_n\in\mathbb{R}$, $\sum_{i=1}^{n}c_i=0$, and $x_1,\ldots, x_n\in M$, 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 $x_1,\ldots,x_n\in M$ and any $c_1,\ldots,c_n\in\mathbb{R}$, 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

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

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

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

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

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 $c_1,\dots,c_n$ such that $c_1+\cdots+c_n=0$ we have

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

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 $x_1,\ldots,x_n$ and any $c_1,\ldots,c_n$ we compute

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

For $n=2$ we have,

$$c_1^2e^{-\lambda\cdot\rho(x_1,x_1)}+c_1c_2e^{-\lambda\cdot\rho(x_1,x_2)}+c_2c_1e^{-\lambda\cdot\rho(x_2,x_1)}+c_2^2e^{-\lambda\cdot\rho(x_2,x_2)}$$
$$=c_1^2e^{-\lambda\cdot d(x_1,x_1)}+c_1c_2e^{-\lambda\cdot d(x_1,x_2)}+c_2^2e^{-\lambda\cdot d(x_2,x_2)}+c_1c_2e^{-\lambda\cdot d(x_1,x_2)}\left\{e^{-\frac{\lambda}{2}[w(x_2)-w(x_1)]}+e^{-\frac{\lambda}{2}[w(x_1)-w(x_2)]}-1\right\}.$$

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

$$c_1^2e^{-\lambda\cdot d(x_1,x_1)}+c_1c_2e^{-\lambda\cdot d(x_1,x_2)}+c_2c_1e^{-\lambda\cdot d(x_2,x_1)}+c_2^2e^{-\lambda\cdot d(x_2,x_2)}\geq 0. \quad (4.4)$$

On the other hand, observe that

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

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

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

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

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

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

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 (M,d^2) is CND then,*

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

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

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

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

And for any $x_1,\ldots,x_n$ and any $c_1,\ldots,c_n$ 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 (M,d^2) is CND, by the GNS construction for symmetric kernels it follows that there exists $\varphi:M\to(\mathcal{H},\langle\cdot,\cdot\rangle)$ continuous function such that

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

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

Then, we can extend this function to $\varphi:(G_f,Q,W)\to(\mathcal{H}\times[0,\infty),\bar{Q},\bar{W})$ by $\varphi(u)=\varphi(x,w(x))=(\varphi(x),2f(x))$. By computing, we get

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

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

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 $\sum_{i=1}^{n}\sum_{j=1}^{n}c_ic_jd(x_i,x_j)\cdot\left[f(x_j)-f(x_i)\right]=0$.

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

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

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

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

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

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

and hence $p(\varphi(x_1), \varphi(x_2)) = 0 = q(x_1, x_2)$, $p(\varphi(x_2), \varphi(x_1)) = 0 = q(x_2, x_1)$ $\Rightarrow q(x_1, x_2) = q(x_2, x_1) = 0$ 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.

Appendix B

Positive Definite Matrices

Positive Definite Matrices

1. Symmetric and Hermitian matrices

In linear algebra a real symmetric $n \times n$ matrix $M = (M_{ij})$ is said to be *positive definite* if the scalars

$$c^T \cdot M \cdot c = \sum_{i=1}^{m} \sum_{j=1}^{n} c_i c_j M_{ij} > 0$$

for any non-zero column vector $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$, where c^T denotes the transpose of c.

The *negative definite, positive semi-definite* and *negative semi-definite* matrices are definite in the same way accept that $c^T Mc$ is $< 0, \geq 0$ and ≤ 0, respectively.

We recall that the extension to complex matrices of a symmetric matrix is *the Hermitian matrix* (we always think complex Hermitian matrices are equivalent to real symmetric matrices).

An $n \times n$ matrix $A = (a_{ij})$ with complex entries such that $a_{ij} = \overline{a_{ji}}$ for $\forall i, j \in \{1, ..., n\}$ is called a *Hermitian matrix*, where $\overline{a}$ is the conjugate of the complex numbers a.

Obviously if A is real matrix, then Hermitian means symmetric. If the conjugate transpose matrix of A is denoted $A^\dagger$, then the Hermitian property can be written as

$$A = A^\dagger$$

Properties of Hermitian matrices

(i) The entries on the main diagonal of a Hermitian matrix A must be real because $a = \overline{a}$ if and only if a is real.

(ii) Any Hermitian matrix A can be diagonalized by a unitary matrix, i.e. for any Hermitian matrix A there exist a unitary matrix U and a diagonal real matrix P s.t.

$$A = U^{-1} \cdot P \cdot U,$$

here unitary matrix means $U^\dagger U = UU^\dagger = I$. Since P is real matrix it follows that the eigenvalues of a Hermitian matrix A are real and since U is unitary it follows A has n linearly independent eigenvectors (the eigenvectors of A are columns in U).

For a Hermitian matrix M (i.e. equivalent of a real symmetric matrix) we can define *positive definiteness* by asking

$$x^* Mx > 0, \qquad \forall x \in \mathbb{C}^n,$$

where x^* is conjugate transpose of x.

Properties of positive definite Hermitian (or symmetric real) matrices

(i) **All eigenvalues are positive.** Indeed, an explained above, if M is Hermitian then there exists an eigen decomposition of M, i.e. $M = U^{-1} \cdot P \cdot U$, where U is unitary complex and P is diagonal with real entries. The columns of U give eigenvectors of M and diagonal elements of P give the eigenvalues of M. Moreover, if M is positive definite, then P must be positive definite (simple calculate). But a real diagonal matrix P is positive definite if and only if the element on the diagonal are positive.

(ii) **The associated linear form is an inner product.** Indeed, we associated to any complex matrix M the linear form (also called sesquilinear form)

$$\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}, \qquad \langle x, y \rangle := y^* M x,$$

for $\forall x, y \in \mathbb{C}^n$, where y^* is the complex conjugate of y.

Obviously $\langle \cdot, \cdot \rangle$ is linear in both arguments. However $\langle \cdot, \cdot \rangle$ is an inner product in $\mathbb{C}^n$ if and only if $\|x\|^2 = \langle x, x \rangle$ is real and positive for $\forall x \in \mathbb{C}^n$, and this is equivalent to M positive definite.

Moreover, any inner product in $\mathbb{C}^n$ arises in this way from a Hermitian positive definite matrix.

A Hermitian matrix is called negative-definite, negative semi-definite or positive semi-definite if and only if its eigenvalues are negative, non-positive or non-negative, respectively.

Indeed, we define an $n \times n$ Hermitian matrix M to be *negative definite* if $x^* M x < 0$, for all non-zero $x \in \mathbb{C}^n$ (or all non-zero $x \in \mathbb{R}^n$ for symmetric real matrices), where x^* is the conjugate transpose of x. The definition of positive semi-definite and negative semi-definite Hermitian matrices is the same accept that $x^* M x$ is ≥ 0 and ≤ 0, respectively.

We also remark that a Hermitian matrix that is neither positive definite, negative definite, positive semi-definite, negative semi-definite is called indefinite. Indefinite matrices always have both positive and negative eigenvalues.

2. Non-symmetric matrices.

The notion of positive definiteness can be easily extended to non-symmetric matrices.

A complex matrix M is called *positive definite* if $Re(x^* M x) > 0$, for all non-zero complex $x \in \mathbb{C}^n$, where $Re(\cdot)$ denotes the real part of a complex number.

This is a weaker definition and in can be used when the matrix M is non-Hermitian, like $M = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ for example.

For real matrices M, we call M positive definite if $c^T Mc > 0$, for all non-zero vector $c \in \mathbb{R}^n$, even M is not symmetric matrix.

Remark 4.1. (i) Observe that for an arbitrary complex matrix M, we have $Re(x^* Mx) > 0$ if and only if the symmetrized matrix $\overline{M} := \dfrac{M + M^*}{2}$ that is a Hermitian matrix, is positive definite as a Hermitian matrix, for any non-zero $x \in \mathbb{C}^n$.

(ii) If M is an arbitrary real matrix then $c^T Mc > 0$ if and only if the symmetrized matrix $\overline{M} := \dfrac{1}{2}(M + M^*)$, that is a symmetric matrix, is positive definite in the usual sense.

Author Biography

Name	Miss Uraiwan Somboon
Date of Birth	27 April 1992
Address	75/114 Village No.10 Suwinthawong Road, Lampakchee, Nongchok, Bangkok, 10530
Education	(2015) Bachelor of Science in Applied Mathematics GPA 3.56 King Mongkut's Institute of Technology Ladkrabang
Scholarship	Scholarships for Graduate Student, the Faculty of Science, King Mongkut's Institute of Technology Ladkrabang.

Academic Publication

Somboon, U., Pantaragphong, P., Sabau, S.V., 2017. "Geodesic Distance Kernels." KMITL Science and Technology Journal Volume 17 No. 1 January-June 2017 page 29-41