Mapping kernels defined over countably infinite mapping systems and their application

Kilho Shin[*] YSHIN@AI.U-HYOGO.AC.JP
Graduate School of Applied Informatics, University of Hyogo, Kobe, Japan

Editor: Chun-Nan Hsu and Wee Sun Lee

Abstract

The mapping kernel is a generalization of Haussler's convolution kernel, and has a wide range of application including kernels for higher degree structures such as trees. Like Haussler's convolution kernel, a mapping kernel is a *finite* sum of values of a primitive kernel. One of the major reasons to use the mapping kernel template in engineering novel kernels is because a strong theorem is known for positive definiteness of the resulting mapping kernels. If the mapping kernel meets the *transitivity* condition and if the primitive kernel is positive definite, the mapping kernel is also positive definite. In this paper, we generalize this theorem by showing, even when we extend the definition of mapping kernels so that a mapping kernel can be a converging sum of countably infinite primitive kernel values, the transitivity condition is still a criteria to determine positive definiteness of mapping kernels according to the extended definition. Interestingly, this result is also useful to investigate positive definiteness of mapping kernels determined as finite sums, when they do not meet the transitivity condition. For this purpose, we introduce a general method that we call *covering technique*.

Keywords: kernel, discrete structure, alignment

1. Introduction

The mapping kernel (Shin and Kuboyama, 2008, 2010) is a generalization of Haussler's convolution kernel (Haussler, 1999), and is known to have a wide application range for analysis of discrete structures such as strings and trees. A mapping kernel is defined by

$$K(x, y) = \sum_{(x',y') \in M_{x,y}} \kappa(\gamma_x(x'), \gamma_y(y')),$$

where $M_{x,y}$ and κ are a finite set and a kernel. Shin and Kuboyama (2008) showed that K is positive definite for any positive definite κ, if, and only if, the family of finite sets $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ meets the certain condition referred to as *transitivity* (Definition 2 and Theorem 3). Although this result is significantly useful to design positive definite mapping kernels for various structures, it still has a couple of constraints. First, the constraint that $M_{x,y}$ should be finite will be an obstacle to extend the result to the continuous applications. Secondly, it is also a fact that a mapping kernel can become positive definite for a non-transitive $\{M_{x,y}\}$ and/or a non-positive-definite κ. As we will see in Section 4.2, we have some important practical examples of this case. In this paper, we first show that Theorem 3

[*] A note

by Shin and Kuboyama (2008) can extend to the case where $M_{x,y}$ is countably infinite (Theorem 6). This not only relaxes the first constraint, but also is useful to address the second problem. In fact, we present a method based on our *covering theorem* (Theorem 12), a corollary to Theorem 6, with which we can investigate positive definiteness of mapping kernels with non-transitive $\{M_{x,y}\}$ and/or non-positive definite κ.

We start with a brief review of Haussler's convolution kernel. Let $\mathcal{X}$ be a space of objects and $\mathcal{X}'$ be a domain over which a kernel κ is defined. To define an $\boldsymbol{R}$-convolution kernel, Haussler (1999) assumed a *finite* relation $\boldsymbol{R} \subsetneqq \mathcal{X}'^n \times \mathcal{X}$, and let

$$K(x, y) = \sum_{(\xi_1,\ldots,\xi_n,x)\in\boldsymbol{R}} \left(\sum_{(\eta_1,\ldots,\eta_n,y)\in\boldsymbol{R}} \prod_{i=1}^{n} \kappa'(\xi_i, \eta_i) \right).$$

With respect to positive definiteness of $\boldsymbol{R}$-convolution kernels, Haussler (1999) showed the following theorem

Theorem 1 (Haussler (1999)) *If κ' is positive definite, K is also positive definite.*

Shin and Kuboyama (2008) simplified this definition, and showed that any $\boldsymbol{R}$-convolution kernel can be derived from kernels determined by the following simpler formula

$$K(x, y) = \sum_{(x',y')\in\mathcal{X}'_x\times\mathcal{X}'_y} \kappa(x', y').$$

Both $\mathcal{X}'_x$ and $\mathcal{X}'_y$ are finite subsets of $\mathcal{X}'$, and are determined according to x and y. To see how an $\boldsymbol{R}$-convolution kernel can be derived from this formulation, we have only to let $\mathcal{X}'_x = \{(\xi_1, \ldots, \xi_n) \mid (\xi_1, \ldots, \xi_n, x) \in \boldsymbol{R}\}$, $\mathcal{X}'_y = \{(\eta_1, \ldots, \eta_n) \mid (\eta_1, \ldots, \eta_n, x) \in \boldsymbol{R}\}$ and $\kappa((\xi_1, \ldots, \xi_n), (\eta_1, \ldots, \eta_n)) = \prod_{i=1}^{n} \kappa'(\xi_i, \eta_i)$. Furthermore, since the formula determines nothing other than $\boldsymbol{R}$-convolution kernels with $n = 1$, Haussler's theorem holds true for the kernels determined by this formula as well.

Starting from the aforementioned simple formalization of the convolution kernels, Shin and Kuboyama (2008) generalized the concept of the convolution kernel, and introduced the *mapping kernel*. A mapping kernel is defined by the formula

$$K(x, y) = \sum_{(x',y')\in M_{x,y}} \kappa(\gamma_x(x'), \gamma_y(y')).$$

In the above, $M_{x,y}$ is a *finite* subset of $\mathcal{X}'_x \times \mathcal{X}'_y$ and γ_x is an arbitrary mapping from $\mathcal{X}'_x$ to $\mathcal{X}'$. The mapping kernel generalizes Haussler's convolution kernel in two ways.

1. The range of the pair (x', y') can be a subset $M_{x,y}$ instead of the entire $\mathcal{X}'_x \times \mathcal{X}'_y$.
2. The set $\mathcal{X}'_x$ can be an arbitrary set not limited to a subset of $\mathcal{X}'$.

In other words, the convolution kernel is the special case of the mapping kernel when $M_{x,y}$ is $\mathcal{X}'_x \times \mathcal{X}'_y$ and γ_x is an inclusion mapping (this implies that $\mathcal{X}'_x$ is a subset of $\mathcal{X}'$).

In addition, the family of sets $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ are called a *mapping system* and the triplet $(\mathcal{X}', \kappa, \{\gamma_x \mid x \in \mathcal{X}\})$ an *evaluation system*. An evaluation system is said to be positive definite, if, and only if, $\kappa : \mathcal{X}' \times \mathcal{X}' \to \mathbb{R}$ is positive definite. In relation to positive definiteness of mapping kernels, Shin and Kuboyama (2008) introduced the following important notion and theorem.

Definition 2 *A mapping system* $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ *is* transitive*, if, and only if, the following conditions are met*

- *If* $(x', y') \in M_{x,y}$*, then* $(y', x') \in M_{y,x}$ *holds.*
- *If* $(x', y') \in M_{x,y}$ *and* $(y', z') \in M_{y,z}$*, then* $(x', z') \in M_{x,z}$ *holds.*

Theorem 3 (Shin and Kuboyama (2008)) *The following conditions are equivalent.*

1. *The mapping system* $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ *is* transitive*.*
2. *For an arbitrary positive definite evaluation system* $(\mathcal{X}', \kappa, \{\gamma_x \mid x \in \mathcal{X}\})$*, the mapping kernel derived from it is positive definite.*

One of the major advantages of using the mapping kernel template to design positive definite kernels is due to this beautiful theorem regarding positive definiteness. In fact, the mapping kernel has a wide range of application. Many string kernels in the literature fall within this range. The spectrum kernel (Leslie et al., 2002) is a typical example. On tree kernels, Shin and Kuboyama (2010) performed an exhaustive survey, and reported that, 18 of 19 tree kernels of different types from the literature can be defined using the mapping kernel template in a straightforward manner. The popular examples of tree kernels of the parsing tree kernel (Collins and Duffy, 2001) and the elastic tree kernel (Kashima and Koyanagi, 2002) belong to this category. Also, Zhang and Chan (2009) defined a new tree kernel based on local alignments of subtrees, and proved its positive definiteness taking advantage of the fact that it is a mapping kernel.

A remark that we should make here is that, although $\mathcal{X}'_x$ and $\mathcal{X}'_y$ do not have to be finite, Shin and Kuboyama (2008) assumed $M_{x,y}$ being finite. This was to make mapping kernels have definite values on one hand and to prove Theorem 3 on the other hand.

In this regard, this paper will generalize the mapping kernel template and Theorem 3 by eliminating this constraint of finiteness. In fact, we show that the condition that $M_{x,y}$ must be finite is not necessary to have Theorem 3 hold. To be precise, if a mapping kernel function converges to a definite value with countably infinite $M_{x,y}$, the condition that the mapping system is transitive is equivalent to the condition that the resulting mapping kernel is positive definite for any positive definite evaluation system. This is the assertion of our main theorem (Theorem 6).

This generalization can have important practical application in particular to deal with mapping kernels with finite but non-transitive mapping system. In Section 4, we propose a new technique, namely *covering technique*, to examine positive definiteness of such mapping kernels taking advantage of Theorem 6. In Section 4.2, we see two examples to show how this technique can work. In the first example, we convert the soft minimum version of the well known Levenshtein string edit distance into a kernel. Although the resulting kernel is a mapping kernel, the associated mapping system is not transitive, and hence, Theorem 3 is not helpful to prove that the kernel is positive definite. However, thanks to the covering technique, its positive definiteness turns out to be reduced to Theorem 6. To be specific, we first expand the original mapping system $\{M_{x,y}\}$ in a certain way, and obtain a new mapping system $\{\overline{M}_{x,y}\}$. The new mapping system becomes transitive but countably infinite. Moreover, it is a *covering* of the original one in the sense that a surjective mapping from $\overline{M}_{x,y}$ to $M_{x,y}$ is given. We view the kernel in question as a mapping kernel defined over $\overline{M}_{x,y}$, and prove that it is positive definite by Theorem 6. In the second example, focusing on a different kernel, the time series kernel introduced by Cuturi et al. (2007), we prove that the kernel is positive definite in a similar way to the first example. Since Cuturi et al. (2007) proved the same result in a specific way, our proof is an alternative proof. The value of our proof probably exists in that we could show that the problem can be dealt with in a more general framework.

2. Extension to countably infinite mapping systems

The objective of this section is to introduce our main theorem (Theorem 6). We first see some conditions for mapping kernels to converge with countably infinite $M_{x,y}$, and then show the assertion of Theorem 6. A proof to Theorem 6 will be given in Section 3.

2.1. Conditions for convergence

With a countably infinite $M_{x,y}$, the kernel

$$K(x,y) = \sum_{(x',y') \in M_{x,y}} \kappa(\gamma_x(x'), \gamma_y(y'))$$

always converges to the same value regardless of the order of calculating the sum, if, and only if, we have the following hold.

$$\sum_{(x',y') \in M_{x,y}} \left| \kappa(\gamma_x(x'), \gamma_y(y')) \right| < \infty$$

This property is a direct consequence from Proposition 4.

Proposition 4 *When $\{a_i\}_{i \in \mathbb{N}}$ is an infinite sequence of real numbers, the following conditions are equivalent.*

1. $\sum_i |a_i| < \infty$ *holds.*
2. $M = \sum_{a_i > 0} a_i < \infty$ *and* $m = \sum_{a_i < 0} a_i > -\infty$ *hold.*
3. $|\sum_{j=1}^{\infty} a_{i(j)}| < \infty$ *holds for an arbitrary bijection* $i : \mathbb{N} \to \mathbb{N}$.

Furthermore, we have $\sum_{j=1}^{\infty} a_{i(j)} = M + m$, *when these conditions hold.*

Proof Since the equivalence between 1 and 2 is evident, we see that 2 and 3 are equivalent to each other. First, we show that 2 implies 3. By definition, for an arbitrary $\varepsilon > 0$, there exists an integer n_ε such that, for any $n \geq n_\varepsilon$,

$$M - \varepsilon \leq \sum_{j \leq n, a_j > 0} a_j \leq M \qquad \text{and} \qquad m \leq \sum_{j \leq n, a_j < 0} a_j \leq m + \varepsilon$$

hold. When we let $N_\varepsilon = \max\{i^{-1}(j) \mid j \leq n_\varepsilon\}$, for any $n \geq N_\varepsilon$, we have

$$M - \varepsilon \leq \sum_{j \leq n, a_{i(j)} > 0} a_{i(j)} \leq M \qquad \text{and} \qquad m \leq \sum_{j \leq n, a_{i(j)} < 0} a_{i(j)} \leq m + \varepsilon,$$

and hence,

$$M + m - \varepsilon \leq \sum_{j=1}^{n} a_{i(j)} \leq M + m + \varepsilon.$$

To prove that 3 implies 2, we prove its contra-position. Without loss of generality, we can assume $M = \infty$, and then, we can define $i : \mathbb{N} \to \mathbb{N}$ such that

$$\sum_{j \leq n, a_{i(j)} > 0} a_{i(j)} \geq 2 \cdot \left| \sum_{j \leq n, a_{i(j)} < 0} a_{i(j)} \right|$$

for any $n > 0$. Evidently, $\sum_{j=1}^{\infty} a_{i(j)}$ diverges to ∞. ■

Even if $\sum_{j=1}^{\infty} a_{i(j)}$ converges with some bijective $i : \mathbb{N} \to \mathbb{N}$, the condition 1 or 2 of Proposition 4 does not necessarily hold. In fact, when we let $a_{2n-1} = \frac{1}{n}$ and $a_{2n} = -\frac{1}{n}$, $\sum_{n=1}^{\infty} a_n = 0$ holds. Nevertheless, we have

$$\sum_{a_i > 0} a_i = \sum_{i=1}^{\infty} \frac{1}{i} > \int_1^{\infty} \frac{1}{x} dx = \lim_{x \to \infty} \ln x = \infty.$$

In general, if the condition 1 or 2 of Proposition 4 does not hold, the value of $\sum_{j=1}^{n} a_{i(j)}$ can vary according to the order of taking sum, that is, $i : \mathbb{N} \to \mathbb{N}$.

2.2. Main theorem

We first introduce the notion of *converging* evaluation systems.

Definition 5 *We let $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ be a mapping system such that $M_{x,y}$ is countable. An evaluation system $(\mathcal{X}', \kappa, \{\gamma_x \mid x \in \mathcal{X}\})$ is said to be* converging *with respect to $\{M_{x,y}\}$, if, and only if, $\sum_{(x',y') \in M_{x,y}} |\kappa(\gamma_x(x'), \gamma_y(y'))| < \infty$ holds.*

By Proposition 4, the condition that an evaluating system is converging is necessary and sufficient that the associated mapping kernel $K(x, y) = \sum_{(x',y') \in M_{x,y}} \kappa(\gamma_x(x'), \gamma_y(y'))$ always converges to the same value and is well defined. Hence, our main theorem can be described as follows.

Theorem 6 *For a mapping system $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ such that $M_{x,y}$ is countable, the following conditions are equivalent.*

1. *The mapping system is* transitive. *That is, $(x', y') \in M_{x,y}$ implies $(y', x') \in M_{y,x}$, and $(x', y') \in M_{x,y} \wedge (y', z') \in M_{y,z}$ implies $(x', z') \in M_{x,z}$.*

2. *For an arbitrary positive definite converging evaluation system $(\mathcal{X}', \kappa, \{\gamma_x \mid x \in \mathcal{X}\})$, the mapping kernel derived from it is positive definite.*

3. A proof to Theorem 6

In this section, we assume $|\mathcal{X}| < \infty$ without any loss of generality. This is because positive definiteness of a mapping kernel K is defined so that, for an arbitrary finite subset $\mathcal{X}_0 \subseteqq \mathcal{X}$, the Gram matrix of K associated with $\mathcal{X}_0$ is positive definite.

3.1. 1 implies 2

Since $M_{x,y}$ is assumed to be countable, we have an injective (one-to-one) numbering scheme $n_{x,y} : M_{x,y} \to \mathbb{N}$. Furthermore, we let

$$M_{x,y}^{(n)} = \{(x',y') \in M_{x,y} \mid n_{x,y}((x',y')) \leq n\}.$$

To start with, we see the following preliminary proposition.

Proposition 7 *If $\{M'(x,y) \subseteqq \mathcal{X}'_x \times \mathcal{X}'_y \mid x,y \in \mathcal{X}\}$ and $\{M''(x,y) \subseteqq \mathcal{X}'_x \times \mathcal{X}'_y \mid x,y \in \mathcal{X}\}$ are both transitive, then $\{M'(x,y) \cap M''(x,y) \mid x,y \in \mathcal{X}\}$ is also transitive.*

Proof If $(x',y') \in M'(x,y) \cap M''(x,y)$, both $(y',x') \in M'(y,x)$ and $(y',x') \in M''(y,x)$ hold, since $\{M'(x,y)\}$ and $\{M''(x,y)\}$ are transitive. $(y',x') \in M'(y,x) \cap M''(y,x)$ follows.

In the same way, if $(x',y') \in M'(x,y) \cap M''(x,y)$ and $(y',z') \in M'(y,z) \cap M''(y,z)$, $(x',z') \in M'(x,z)$ and $(x',z') \in M''(x,z)$ hold, and hence, we have $(x',z') \in M'(x,z) \cap M''(x,z)$.

The assertion of the proposition immediately follows. ■

Lemma 8 will play a crucial role to prove that 1 implies 2.

Lemma 8 *Assume $\mathcal{X}$ is finite and $\{M_{x,y}\}$ is transitive. For any $\left\{M_{x,y}^{(\#)} \subseteqq M_{x,y} \mid x,y \in \mathcal{X}\right\}$ such that $|M_{x,y}^{(\#)}| < \infty$, a transitive mapping system $\{\overline{M}_{x,y} \mid x,y \in \mathcal{X}\}$ exists, and satisfies $M_{x,y}^{(\#)} \subseteqq \overline{M}_{x,y} \subseteqq M_{x,y}$ and $|\overline{M}_{x,y}| < \infty$ for $\forall x,y \in \mathcal{X}$.*

Proof We define $\overline{\mathcal{X}}'_x \subseteqq \mathcal{X}'_x$ by $\overline{\mathcal{X}}'_x = \{x' \mid \exists(y)\exists(y')[(x',y') \in M_{x,y}^{(\#)}]\}$. Since $\mathcal{X}$ and $M_{x,y}^{(\#)}$ are finite, $|\overline{\mathcal{X}}'_x| < \infty$ holds for $\forall x \in \mathcal{X}$. Also, it is evident that $\{\overline{\mathcal{X}}'_x \times \overline{\mathcal{X}}'_y \mid x,y \in \mathcal{X}\}$ is transitive. Therefore, to show the assertion, it suffices to define $\overline{M}_{x,y} = M_{x,y} \cap \left(\overline{\mathcal{X}}'_x \times \overline{\mathcal{X}}'_y\right)$ for $\forall x,y \in \mathcal{X}$. Its transitivity follows from Proposition 7. ■

Now, we are ready to complete the proof of $1 \Rightarrow 2$.

Since $\left|M_{x,y}^{(n)}\right| \leq n$, we can apply Lemma 8 to $\{M_{x,y}^{(n)} \mid x,y \in \mathcal{X}\}$ to obtain $\{\overline{M}_{x,y}^{(n)} \mid x,y \in \mathcal{X}\}$. $\{\overline{M}_{x,y}^{(n)} \mid x,y \in \mathcal{X}\}$ is transitive, and $|\overline{M}_{x,y}^{(n)}| < \infty$ holds. By Theorem 3,

$$\overline{K}^{(n)}(x,y) = \sum_{(x',y') \in \overline{M}_{x,y}^{(n)}} \kappa(\gamma_x(x'), \gamma_y(y'))$$

turns out positive definite.

On one hand, $M_{x,y}^{(m)} \subseteqq M_{x,y}^{(n)}$ for $m \leq n$ and $\bigcup_{n=1}^{\infty} M_{x,y}^{(n)} = M_{x,y}$ hold, and $K(x,y)$ is a point-wise limitation of $K^{(n)}(x,y) = \sum_{(x',y') \in M_{x,y}^{(n)}} \kappa(\gamma_x(x'), \gamma_y(y'))$. That is to say,

$$K(x,y) = \lim_{n \to \infty} K^{(n)}(x,y)$$

holds. On the other hand, $M_{x,y}^{(n)} \subseteqq \overline{M}_{x,y}^{(n)}$ implies

$$\lim_{n \to \infty} K^{(n)}(x,y) = \lim_{n \to \infty} \overline{K}^{(n)}(x,y).$$

Thus, $K(x,y)$ turns out to be a point-wise limitation of positive definite kernels, and therefore, we can conclude that $K(x,y)$ is positive definite as well.

3.2. 2 implies 1

Since $M_{x,y}$ is countable and $\mathcal{X}$ is finite, without loss of generality, we can assume that $\mathcal{X}'_x$ is also countable. In fact, we can redefine $\mathcal{X}'_x$ by $\mathcal{X}'_x = \{x' \mid \exists(y)\exists(y')[(x',y') \in M_{x,y}\}$. Hence, we have an infinite increasing sequence of finite subsets of $\mathcal{X}'_x$ such that

$$\mathcal{X}'^{(1)}_x \subseteqq \mathcal{X}'^{(2)}_x \subseteqq \cdots \subseteqq \mathcal{X}'_x \text{ and } \bigcup_{n=1}^{\infty} \mathcal{X}'^{(n)}_x = \mathcal{X}'_x.$$

In addition, we let

$$M_{x,y}^{(n)} = M_{x,y} \cap \left(\mathcal{X}'^{(n)}_x \times \mathcal{X}'^{(n)}_y \right).$$

With this setting, we have the following lemma.

Lemma 9 *$\{M_{x,y} \mid x,y \in \mathcal{X}\}$ is transitive, if $\{M_{x,y}^{(n)} \mid x,y \in \mathcal{X}\}$ is transitive for all $n > 0$.*

Proof Assume $(x',y') \in M_{x,y}$ and $(y',z') \in M_{y,z}$. By definition, there exists an integer n such that $x' \in \mathcal{X}'^{(n)}_x$, $y' \in \mathcal{X}'^{(n)}_y$ and $z' \in \mathcal{X}'^{(n)}_z$, and hence, we have $(x',y') \in M_{x,y}^{(n)}$ and $(y',z') \in M_{y,z}^{(n)}$. $(x',z') \in M_{x,z}^{(n)}$ follows, since $\{M_{x,y}^{(n)}\}$ is transitive. ■

To prove that 2 implies 1, we claim that, if $M_{x,y}$ is not transitive, then there exists an converging evaluation system $(\mathcal{X}', \kappa, \{\gamma_x \mid x \in \mathcal{X}\})$ such that κ is positive definite but the mapping kernel K derived from it is not.

By Lemma 9, $\{M_{x,y}^{(n)} \mid x,y \in \mathcal{X}\}$ is not transitive for some $n > 0$. Since $\left|\mathcal{X}'^{(n)}_x\right| < \infty$, by Theorem 3, we have a *partial* evaluation system $(\mathcal{X}', \kappa, \{\gamma_x \mid x \in \mathcal{X}\})$ such that:

- The domain of definition of γ_x is $\mathcal{X}'^{(n)}_x$;
- κ is positive definite;
- $K(x,y) = \sum_{(x',y') \in M_{x,y}^{(n)}} \kappa(\gamma_x(x'), \gamma_y(y'))$ is not positive definite.

To complete the proof, we first add the *null* element $\bullet$ to $\mathcal{X}'$, and let $\kappa(\bullet, \xi) = \kappa(\xi, \bullet) = 0$ for $\forall \xi \in \mathcal{X}'$. This modification to κ does not harm its positive definiteness. Next, we expand the domain of definition of γ_x so that $\gamma_x(x') = \bullet$ for $x' \in \mathcal{X}'_x \setminus \mathcal{X}'^{(n)}_x$. This modification does not change the definition of the resulting mapping kernel. In fact, we have

$$\sum_{(x',y') \in M_{x,y}} \kappa(\gamma_x(x'), \gamma_y(y')) = \sum_{(x',y') \in M^{(n)}_{x,y}} \kappa(\gamma_x(x'), \gamma_y(y')).$$

Evidently, these modifications determine a positive definite converging evaluation system that results in a non-positive-definite mapping kernel.

4. Covering technique

Theorem 6 enables us to investigate positive definiteness of mapping kernels, even if they are defined over non-transitive mapping systems or with evaluation systems that are not positive definite. In this section, we introduce a new technique, namely *covering technique*, that fill the gap between Theorem 6 and such mapping kernels.

4.1. Covering theorem

In this section, we will see a technique to examine positive definiteness of a mapping kernel when the mapping system $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ is not transitive and/or the kernel κ of the evaluation system is not positive definite. Although Theorems 3 and 6 assert that, if a mapping system is transitive, the resulting mapping kernel derived from a positive definite evaluation system is always positive definite, they do not deny the possibility that mapping kernels with non-transitive mapping systems and/or non-positive-definite evaluation systems can be positive definite.

In the following, we let $\{M_{x,y} \subseteqq \mathcal{X}'_x \times \mathcal{X}'_y \mid x, y \in \mathcal{X}\}$ be a mapping system that is not necessarily transitive and $\{\overline{M}_{x,y} \subseteqq \overline{\mathcal{X}}'_x \times \overline{\mathcal{X}}'_y \mid x, y \in \mathcal{X}\}$ be a transitive mapping system such that each $\overline{M}_{x,y}$ is countable but not necessarily finite.

Definition 10 *A mapping system $\{\overline{M}_{x,y} \subseteqq \overline{\mathcal{X}}'_x \times \overline{\mathcal{X}}'_y \mid x, y \in \mathcal{X}\}$ is a* covering *of $\{M_{x,y} \subseteqq \mathcal{X}'_x \times \mathcal{X}'_y \mid x, y \in \mathcal{X}\}$, if, and only if, there exists a family of mappings $\boldsymbol{\varphi} = \{\varphi_{x,y} \mid x, y \in \mathcal{X}\}$ such that each $\varphi_{x,y}$ is a surjective mapping from $\overline{M}_{x,y}$ onto $M_{x,y}$.*

In Definition 10, although $M_{x,y}$ is not necessarily finite, the existence of $\overline{M}_{x,y}$ implies that $M_{x,y}$ is countable for $\forall x, y \in \mathcal{X}$.

Definition 11 *The pair of a mapping system $\mathcal{M} = \{M_{x,y} \mid x, y \in \mathcal{X}\}$ and an evaluation system $\mathcal{E} = (\{\gamma_x \mid x \in \mathcal{X}\}, \mathcal{X}', \kappa)$ is* resolvable*, if, and only if, there exist a transitive mapping system $\overline{\mathcal{M}} = \{\overline{M}_{x,y} \subseteqq \overline{\mathcal{X}}'_x \times \overline{\mathcal{X}}'_y \mid x, y \in \mathcal{X}\}$ and a evaluation system $\overline{\mathcal{E}} = (\{\overline{\gamma}_x \mid x \in \mathcal{X}\}, \overline{\mathcal{X}}', \overline{\kappa})$ such that:*

1. *$\overline{\mathcal{M}}$ is a covering of $\mathcal{M}$;*
2. *$\overline{\kappa}$ is positive definite and $\kappa(x', y') = \sum_{(\bar{x}', \bar{y}') \in \varphi^{-1}_{x,y}((x',y'))} \overline{\kappa}(\bar{x}', \bar{y}')$.*

Theorem 12 provides us with a method to cope with the problem to examine whether a mapping kernel is positive definite when the associated mapping system is not transitive and/or sub-kernels κ is not positive definite.

Theorem 12 *If the pair of a mapping system $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ and a converging evaluation system $\mathcal{E} = (\{\gamma_x \mid x \in \mathcal{X}\}, \mathcal{X}', \kappa)$ is resolvable, the resulting mapping kernel is positive definite.*

Proof Let $\overline{\mathcal{M}}$ and $\overline{\mathcal{E}}$ be as defined in Definition 11. By definition,

$$\sum_{(x',y') \in M_{x,y}} \kappa(x', y') = \sum_{(x',y') \in M_{x,y}} \left(\sum_{(\bar{x}',\bar{y}') \in \varphi^{-1}_{x,y}((x',y'))} \overline{\kappa}(\bar{x}', \bar{y}') \right) = \sum_{(\bar{x}',\bar{y}') \in \overline{M}_{x,y}} \overline{\kappa}(\bar{x}', \bar{y}')$$

holds, and the assertion of the theorem follows from Theorem 6. ∎

4.2. Examples

In this section, we see two examples where a mapping kernel defined over a non-transitive mapping system is proved to be positive definite using the covering technique.

4.2.1. Levenshtein edit distance kernel

We start with a brief review of the well known Levenshtein edit distance for strings, and then define a kernel which is tightly relating to it. The positive definiteness of the kernel is proved taking advantage of the covering technique (Section 4.1).

We let Σ be an alphabet, and let x and y be strings over Σ. Hence, x and y are elements of Σ^*, and an *alignment* of x and y is a two-row table to represent an *edit script* to convert x into y. The following is an example of alignments when we let $x = \xi_1 \dots \xi_5$ and $y = \eta_1 \dots \eta_7$.

$$\begin{array}{cccccccccc} \xi_1 & \xi_2 & - & \xi_3 & - & - & - & \xi_4 & - & \xi_5 \\ - & \eta_1 & \eta_2 & - & \eta_3 & \eta_4 & \eta_5 & - & \eta_6 & \eta_7 \end{array}$$

The columns of an alignment determine a set of *edit operations* that comprise the edit script that the alignment determines. Each edit operation is either deletion, insertion or replacement of characters. An alignment determines an edit script as follows: A column $\begin{pmatrix} \xi_i \\ - \end{pmatrix}$ indicates deletion of ξ_i; A column $\begin{pmatrix} - \\ \eta_j \end{pmatrix}$ indicates insertion of η_j; A column $\begin{pmatrix} \xi_i \\ \eta_j \end{pmatrix}$ indicates replacement of ξ_i with η_j; The column $\begin{pmatrix} - \\ - \end{pmatrix}$ must not appear. For example, the edit script depicted by the diagram above deletes ξ_1, ξ_3 and ξ_4, inserts $\eta_2, \eta_3, \eta_4, \eta_5$ and η_6, and replaces ξ_2 and ξ_5 with η_1 and η_7.

Also, when we let $\tilde{\Sigma} = \Sigma \cup \{-\}$, each row of an alignment determines a string over $\tilde{\Sigma}$. Thus, an alignment is viewed as a pair of strings $(x', y') \in \tilde{\Sigma}^* \times \tilde{\Sigma}^*$, and we let $M_{x,y}$ denote the subset of $\tilde{\Sigma}^* \times \tilde{\Sigma}^*$ that consists of all of the possible alignments from x to y.

Next, we see the definition of the *cost* of edit scripts and then that of the Levenshtein distance. For an alignment $\alpha = (\tilde{\xi}_1 \dots \tilde{\xi}_\nu, \tilde{\eta}_1 \dots \tilde{\eta}_\nu) \in \tilde{\Sigma}^* \times \tilde{\Sigma}^*$, the cost of the corresponding edit script σ_α is calculated by $\mathsf{cost}(\sigma_\alpha) = \sum_{i=1}^{\nu} \mathsf{cost}(\tilde{\xi}_i \to \tilde{\eta}_i)$, where $\mathsf{cost}(\tilde{\xi} \to \tilde{\eta})$ represents the cost assigned to the edit operation $\tilde{\xi} \to \tilde{\eta}$. Although the assignment of costs to edit operations is not unique, a typical way to define $\mathsf{cost}(\tilde{\xi} \to \tilde{\eta})$ is to let $\mathsf{cost}(\tilde{\xi} \to \tilde{\eta}) = 1 - \delta_{\tilde{\xi},\tilde{\eta}}$. For simplicity, we adopt this typical definition in this section. Finally, the Levenshtein edit distance $d(x, y)$ between x and y is defined by

$$d(x, y) = \min\{\mathsf{cost}(\sigma_\alpha) \mid \alpha \in M_{x,y}\}.$$

Levenshtein edit distance has proved very useful to measure the similarity between strings.

Now, we will try to convert Levenshtein edit distance into a positive definite kernel. By this, we would be benefited by taking advantage of the kernel-based classifiers such as SVM. A common method to convert a distance $d(x, y)$ into a kernel is to define the kernel as $e^{-\gamma d(x,y)}$ for $\gamma > 0$. If Levenshtein edit distance were negative definite, the kernel derived in this way would be positive definite. However, in the reality, this is not the case.

In order to obtain a positive definite kernel, we consider the soft minimum version of Levenshtein edit distance. That is to say, we define a new distance by evaluating the *soft minimum* of edit costs instead of evaluating their hard minimum. The soft minimum of $a_1, \ldots, a_\nu$ is defined by $-\frac{1}{\gamma} \log\left(\sum_{i=1}^{\nu} e^{-\gamma a_i}\right)$, and approximates the hard minimum $\min\{a_1, \ldots, a_\nu\}$. In fact,

$$\min\{a_1, \ldots, a_\nu\} = \lim_{\gamma \to \infty} -\frac{1}{\gamma} \log\left(\sum_{i=1}^{\nu} e^{-\gamma a_i}\right)$$

holds. The soft minimum has several advantages that would benefit us, and one of them is that it is an analytic function in $a_1, \ldots, a_{\nu-1}$ and a_ν. This is a clear contrast with the hard minimum.

Thus, the soft minimum version of Levenshtein edit distance should be defined by

$$\tilde{d}(x, y) = -\frac{1}{\gamma} \log\left(\sum_{\alpha \in M_{x,y}} e^{-\gamma \cdot \mathsf{cost}(\sigma_\alpha)}\right).$$

Hence, we apply the aforementioned method to convert this new distance into a kernel, and define the kernel by

$$K(x, y) = e^{-\gamma\left[-\frac{1}{\gamma}\log\left(\sum_{\alpha \in M_{x,y}} e^{-\gamma \cdot \mathsf{cost}(\sigma_\alpha)}\right)\right]} = \sum_{\alpha \in M_{x,y}} e^{-\gamma \cdot \mathsf{cost}(\sigma_\alpha)}. \tag{1}$$

The kernel K defined by Equation (1) is an instance of the mapping kernel. In the following, we will formally determine a mapping system and an evaluation system for K.

- Since $M_{x,y} \subset \tilde{\Sigma}^* \times \tilde{\Sigma}^*$ holds as described before, we let $\mathcal{X}'_x = \tilde{\Sigma}^*$ for all $x \in \Sigma^*$ to view $\{M_{x,y} \mid x, y \in \Sigma^*\}$ as a mapping system.
- To determine the evaluation system $(\mathcal{X}', \kappa, \{\gamma_x\})$, we let $\mathcal{X}'$ be $\tilde{\Sigma}^*$ and $\gamma_x : \mathcal{X}'_x \to \mathcal{X}'$ be the identity. Although the domain of definition of the function $e^{-\gamma \cdot \mathsf{cost}(\sigma_\alpha)}$ is $M_{x,y}$, it is easily expanded to $\mathcal{X}' \times \mathcal{X}'$ by taking advantage of the property of

$$e^{-\gamma \cdot \mathsf{cost}(\sigma_\alpha)} = \prod_{i=1}^{\nu} e^{-\gamma \cdot \mathsf{cost}(\tilde{\xi}_i \to \tilde{\eta}_i)} = \prod_{i=1}^{\nu} e^{-\gamma\left(1 - \delta_{\tilde{\xi}_i, \tilde{\eta}_i}\right)},$$

when $\alpha = (\tilde{\xi}_1 \dots \tilde{\xi}_\nu, \tilde{\eta}_1 \dots \tilde{\eta}_\nu)$. We can indeed define $\kappa : \tilde{\Sigma}^* \times \tilde{\Sigma}^* \to \mathbb{R}$ as follows.

$$\kappa(\tilde{\xi}_1 \dots \tilde{\xi}_\mu, \tilde{\eta}_1 \dots \tilde{\eta}_\nu) = \begin{cases} \prod_{i=1}^{\nu} e^{-\gamma\left(1-\delta_{\tilde{\xi}_i,\tilde{\eta}_i}\right)}, & \text{if } \mu = \nu; \\ 0, & \text{if } \mu \neq \nu. \end{cases}$$

Now, we see that K is a mapping kernel.

To prove that K is positive definiteness, it might be helpful to take advantage of Theorem 3. That is to say, if κ is positive definite and the mapping system is transitive, we can conclude that K is positive definite.

On one hand, it is easy to see $e^{-\gamma \cdot \mathsf{cost}(\tilde{\xi} \to \tilde{\eta})} = e^{-\gamma\left(1-\delta_{\tilde{\xi},\tilde{\eta}}\right)}$ is a positive definite kernel defined over $\tilde{\Sigma}$, and positive definiteness of κ follows.

On the other hand, the mapping system turns out not to be transitive. This is because, even if $(x', y') \in M_{x,y}$ and $(y', z') \in M_{y,z}$ hold, (x', z') may include $\begin{pmatrix} - \\ - \end{pmatrix}$ as columns, and if this is the case, (x', z') cannot be an alignment by definition. For example, in the following example, (x', y') or (y', z') includes no $\begin{pmatrix} - \\ - \end{pmatrix}$, whereas (x', z') includes five $\begin{pmatrix} - \\ - \end{pmatrix}$.

x'	ξ_1	ξ_2	$-$	ξ_3	$-$	$-$	$-$	ξ_4	$-$	ξ_5
y'	$-$	η_1	η_2	$-$	η_3	η_4	η_5	$-$	η_6	η_7
z'	ζ_1	ζ_2	$-$	ζ_3	$-$	$-$	$-$	ζ_4	$-$	ζ_5

Thus, we cannot rely on Theorem 3 in a straightforward manner to prove that K is positive definite. Speaking the consequence first, the covering technique described in Section 4 is effective, and Theorem 12 enables us to prove that K is positive definite. In the remainder of this section, we will determine a covering $\{\overline{M}_{x,y}\}$ of $\{M_{x,y}\}$ and a positive definite evaluation system $(\overline{\mathcal{X}}', \overline{\kappa}, \{\overline{\gamma}_x\})$ to show that $\{M_{x,y}\}$ and $(\mathcal{X}', \kappa, \{\gamma_x\})$ is resolvable.

To determine $\overline{M}_{x,y}$, letting $\overline{\mathcal{X}}'_x = \mathcal{X}'_x = \tilde{\Sigma}^*$, we expand $M_{x,y}$ so that each $(\bar{x}', \bar{y}') \in \overline{M}_{x,y}$ is obtained by adding zero or more of the column $\begin{pmatrix} - \\ - \end{pmatrix}$ to some alignment $(x', y') \in M_{x,y}$. For example, the following $(\bar{x}', \bar{y}') \in \overline{M}_{x,y}$ is obtained by adding three $\begin{pmatrix} - \\ - \end{pmatrix}$ to (x', y') in the previous example.

$\bar{x}'$	$\mathbf{-}$	ξ_1	ξ_2	$-$	ξ_3	$-$	$\mathbf{-}$	$-$	$-$	ξ_4	$-$	$\mathbf{-}$	ξ_5
$\bar{y}'$	$\mathbf{-}$	$-$	η_1	η_2	$-$	η_3	$\mathbf{-}$	η_4	η_5	$-$	η_6	$\mathbf{-}$	η_7

Contrarily, to determine the associated $\varphi_{x,y}$, we define $\varphi_{x,y}(\bar{x}', \bar{y}')$ by eliminating all $\begin{pmatrix} - \\ - \end{pmatrix}$ from $(\bar{x}', \bar{y}')$. $\varphi_{x,y}(\bar{x}', \bar{y}')$ includes no $\begin{pmatrix} - \\ - \end{pmatrix}$, and hence fall within $M_{x,y}$. For example, the aforementioned pair $(\bar{x}', \bar{y}') \in \overline{M}_{x,y}$ is evidently mapped to $(x', y') \in M_{x,y}$ by $\varphi_{x,y}$.

To determine the positive definite evaluation system $(\overline{\mathcal{X}}', \overline{\kappa}, \{\overline{\gamma}_x\})$, we let $\overline{\mathcal{X}}'$ be $\tilde{\Sigma}^*$ and $\overline{\gamma}_x : \mathcal{X}'_x \to \overline{\mathcal{X}}'$ be the identity in the same way as for $(\mathcal{X}', \kappa, \{\gamma_x\})$. What is left to define is the positive definite $\overline{\kappa}$. We first introduce a new positive definite kernel $\lambda : \tilde{\Sigma} \times \tilde{\Sigma} \to \mathbb{R}$ as follows.

$$\lambda(\tilde{\xi}, \tilde{\eta}) = \begin{cases} \epsilon, & \text{if } (\tilde{\xi}, \tilde{\eta}) = (-,-); \\ (1-\epsilon)e^{-\gamma\left(1-\delta_{\tilde{\xi},\tilde{\eta}}\right)} & \text{if } (\tilde{\xi}, \tilde{\eta}) \neq (-,-). \end{cases}$$

A constant ϵ is determined so that $1 > \epsilon \geq (1-\epsilon)e^{-\gamma}$ holds, and hence, it is selected from the interval $[e^{-\gamma}/(1+e^{-\gamma}), 1)$. By this choice, λ becomes positive definite. In fact, the Gram matrix for $\lambda(\tilde{\xi}, \tilde{\eta}) - (1-\epsilon)e^{-\gamma}$ is such that all of the diagonal elements are non-negative and all of the other non-diagonal elements are zero. Finally, when we define $\overline{\kappa}$ over $\tilde{\Sigma}^* \times \tilde{\Sigma}^*$ by

$$\overline{\kappa}(\tilde{\xi}_1 \ldots \tilde{\xi}_m, \tilde{\eta}_1 \ldots \tilde{\eta}_\nu) = \begin{cases} (1-\epsilon)\prod_{i=1}^{\nu} \lambda(\tilde{\xi}_i, \tilde{\eta}_i), & \text{if } m = n; \\ \\ 0, & \text{if } m \neq n, \end{cases}$$

it is also positive definite.

Finally, to prove that the pair $(\{M_{x,y}\}, (\mathcal{X}', \kappa, \{\gamma_x\}))$ is resolvable, we claim

$$\sum_{(\bar{x}', \bar{y}') \in \varphi^{-1}_{x',y'}(x',y')} \overline{\kappa}(\bar{x}', \bar{y}') = \kappa(x', y').$$

We will prove this claim. Let $(x', y') \in M_{x,y} = (\tilde{\xi}_1 \ldots \tilde{\xi}_\nu, \tilde{\eta}_1 \ldots \tilde{\eta}_\nu)$ and $(\bar{x}', \bar{y}')$ satisfy $\varphi_{x,y}(\bar{x}', \bar{y}') = (x', y')$. As two-row tables, $(\bar{x}', \bar{y}')$ can be obtained by inserting zero or more of the column $\begin{pmatrix} - \\ - \end{pmatrix}$ to (x', y'). Since (x', y') has ν columns, there are $\nu + 1$ positions in (x', y') for the insertion. Hence, we have

$$\begin{aligned} \sum_{(\bar{x}', \bar{y}') \in \varphi^{-1}_{x',y'}(x',y')} \overline{\kappa}(\bar{x}', \bar{y}') &= (1-\epsilon)\left(\sum_{i=0}^{\infty} \lambda(-,-)^i\right)^{\nu+1} \prod_{i=1}^{\nu} \lambda(\tilde{\xi}_i, \tilde{\eta}_i) \\ &= (1-\epsilon)\left(\frac{1}{1-\epsilon}\right)^{\nu+1} \prod_{i=1}^{\nu} \lambda(\tilde{\xi}_i, \tilde{\eta}_i) = \prod_{i=1}^{\nu} e^{-\gamma\left(1-\delta_{\tilde{\xi}_i,\tilde{\eta}_i}\right)} = \kappa(x', y'). \end{aligned}$$

4.2.2. Global alignment kernel for time series

Cuturi et al. (2007) introduced an alignment kernel that is similar to the kernel introduced in Section 4.2.1 but is different in the definition of alignments. This kernel is introduced in relation to the well known Dynamic Time Warping (DTW) family of distance, and is to handle time series such as speech data.

In the following, we see the definition of alignments for this kernel. We let $\mathcal{X} = \Sigma^*$. For $x = \xi_1 \ldots \xi_\mu$ and $y = \eta_1 \ldots \eta_\nu$ in $\mathcal{X}$, we first consider a two-dimensional $\mu \times \nu$ lattice. The left-bottom point of the lattice corresponds to (ξ_1, η_1), and the right-top point does to (ξ_μ, η_ν). An alignment between x and y is a path from (ξ_1, η_1) to (ξ_μ, η_ν) such that the next point of a point on the path is a point adjacent to the point in the lattice that is located in the direction of either north, east or north-east of the point. The following diagram displays an example of alignments in the case of $\mu = 5$ and $\nu = 7$.

$$\begin{array}{cccccccc}
\xi_5 & & & & & & & \bullet \\
\xi_4 & & & & & & & \uparrow \\
\xi_3 & & & & \rightarrow & \rightarrow & \nearrow & \\
\xi_2 & & & \rightarrow & \uparrow & & & \\
\xi_1 & \rightarrow & \nearrow & & & & & \\
 & \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 & \eta_6 & \eta_7
\end{array}$$

Also, we can uniquely represent any alignment as a pair of strings of the same length. In fact, letting $x = \xi_1 \dots \xi_\mu$ and $y = \eta_1 \dots \eta_\nu$, we determine $(\xi_{i(1)} \dots \xi_{i(p)}, \eta_{j(1)} \dots \eta_{j(p)})$ so that $(\xi_{i(k)}, \eta_{j(k)})$ is the k-th point of the path. For example, the alignment depicted by the previous chart corresponds to $(\xi_1\xi_1\xi_2\xi_2\xi_3\xi_3\xi_3\xi_4\xi_5, \eta_1\eta_2\eta_3\eta_4\eta_4\eta_5\eta_6\eta_7\eta_7)$. Thus, we define the entire set of alignments between x and y by

$$M_{x,y} = \Bigg\{ (\xi_{i(1)} \dots \xi_{i(p)}, \eta_{j(1)} \dots \eta_{j(p)}) \mid 1 = i(1) \leq i(2) \dots \leq i(p) = m, 1 = j(1) \leq j(2) \dots \leq j(p) = n, \begin{pmatrix} i(a+1) - i(a) \\ j(a+1) - j(a) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \Bigg\}.$$

Then, the kernel in question is defined as a mapping kernel on this $M_{x,y}$ as follows, when a local kernel $\lambda : \Sigma \times \Sigma \to \mathbb{R}$ is given.

$$K(x, y) = \sum_{(\xi_{i(1)} \dots \xi_{i(p)}, \eta_{j(1)} \dots \eta_{j(p)}) \in M_{x,y}} \prod_{k=1}^{p} \lambda(\xi_{i(k)}, \eta_{j(k)})$$

Note that we let $\mathcal{X} = \mathcal{X}' = \mathcal{X}'_x = \Sigma^*$ and γ_x be the identity.

For a reason similar to as seen in Section 4.2.1, this mapping system $\{M_{x,y} \mid x, y \in \mathcal{X}\}$ is not transitive. Hence, we cannot rely on Theorem 3, to conclude that K is positive definite even if λ is positive definite.

To investigate when K is positive definite, we take advantage of the covering technique again. Speaking the consequence first, we let $\overline{\mathcal{X}} = \overline{\mathcal{X}}' = \overline{\mathcal{X}}'_x = \Sigma^*$ and $\overline{\gamma}_x$ be the identity.

Furthermore, we define $\overline{M}_{x,y}$ by

$$\overline{M}_{\xi_1 \dots \xi_\mu, \eta_1 \dots \eta_\nu} = \Bigg\{ (\xi_{i(1)} \dots \xi_{i(p)}, \eta_{j(1)} \dots \eta_{j(\overline{p})}) \mid 1 = i(1) \leq i(2) \dots \leq i(\overline{p}) = m, 1 = j(1) \leq j(2) \dots \leq j(\overline{p}) = n \; \begin{pmatrix} i(a+1) - i(a) \\ j(a+1) - j(a) \end{pmatrix} \in \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \Bigg\}.$$

By adding $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ to the last condition, $\{\overline{M}_{x,y} \mid x, y \in \mathcal{X}\}$ becomes transitive.

The associated mapping $\varphi_{x,y} : \overline{M}_{x,y} \to M_{x,y}$ can be defined as follows. If $(\xi_{i(1)} \dots \xi_{i(p)}, \eta_{j(1)} \dots \eta_{j(p)})$ includes more than one (ξ_a, η_b) in it, that is, if $i(k) = \dots = i(\ell) = a$ and $j(k) = \dots = j(\ell) =$

b holds for $\ell > k$, we eliminate $\ell - k$ of them, and leave one. By applying this operation to all of the multiple occurrences of (ξ_a, η_b), we finally obtain an alignment in $M_{x,y}$. To determine $\overline{\kappa}$, we assume that it is of the form

$$\overline{\kappa}(\xi_1 \dots \xi_\mu, \xi_1 \dots \xi_\nu) = \begin{cases} \prod_{i=1}^m \overline{\lambda}(\xi_i, \eta_i), & \text{if } m = n; \\ 0, & \text{if } m \neq n, \end{cases}$$

for some kernel $\overline{\lambda}$. The condition to apply Theorem 12 is that $\sum_{(\bar{x}',\bar{y}') \in \varphi^{-1}_{x',y'}(x',y')} \overline{\kappa}(\bar{x}', \bar{y}') = \kappa(x', y')$ holds for any $(x', y') \in M_{x,y}$. On one hand, when we let $x = \xi_1 \dots \xi_\mu$, $y = \eta_1 \dots \eta_\nu$, $x' = \xi_{i(1)} \dots \xi_{i(p)}$ and $y' = \eta_{j(1)} \dots \eta_{j(p)}$,

$$\sum_{(\bar{x}',\bar{y}') \in \varphi^{-1}_{x,y}(x',y')} \overline{\kappa}(\bar{x}', \bar{y}') = \prod_{k=1}^{p} \left(\sum_{l=1}^{\infty} \overline{\lambda}(\xi_{i(k)}, \eta_{j(k)})^{\ell} \right) = \prod_{k=1}^{p} \frac{\overline{\lambda}(\xi_{i(k)}, \eta_{j(k)})}{1 - \overline{\lambda}(\xi_{i(k)}, \eta_{j(k)})}$$

On the other hand, we have $\kappa(x', y') = \prod_{k=1}^{p} \lambda(\xi_{i(k)}, \eta_{j(k)})$. Hence, to have $\sum_{(\bar{x}',\bar{y}') \in \varphi^{-1}_{x',y'}(x',y')} \overline{\kappa}(\bar{x}', \bar{y}') = \kappa(x', y')$ hold, it suffices that we require $\lambda = \overline{\lambda}/(1 - \overline{\lambda})$, equivalently, $\overline{\lambda} = \lambda/(1 + \lambda)$.

Unfortunately, even if λ is positive definite, $\overline{\lambda}$ is not necessarily positive definite. Cuturi et al. (2007) defined that λ is *geometrically divisible*, if, and only if, $\lambda/(1+\lambda)$ is positive definite. Also, Cuturi (2011) proved the following important lemma with respect to geometrical divisibility.

Lemma 13 (Cuturi (2011), Lemma 3) *For any infinitely divisible kernel* $\overline{\lambda}$ *such that* $0 < \overline{\lambda} < 1$*, then* $\overline{\lambda}/(1 - \overline{\lambda})$ *is geometrically and infinitely divisible.*

By taking advantage of Lemma 13, starting with an infinitely divisible kernel $\overline{\lambda}$, we can obtain an infinitely divisible kernel λ such that $\overline{\lambda} = \lambda/(1 + \lambda)$. The property of infinite divisibility provides us with a method to avoid the situation where the diagonal elements of a Gram matrix is dominating.

Although a proof of Lemma 13 is given in a specific manner in (Cuturi, 2011), it can be proved based on a more general property of infinitely divisible kernels as shown in Appendix A.

5. Conclusion

We saw that the theorem for positive definiteness of mapping kernels can be extended to the case where mapping systems consist of countable sets that are not necessarily finite. This extension provides us with a method to examine positive definiteness of mapping kernels with non-transitive mapping systems and/or with non-positive-definite evaluation systems. In fact, in this paper, we saw two important practical examples of positive definite mapping kernels defined over non-transitive mapping systems. For future study, we will investigate such mapping kernels more closely, and will also explore novel mapping kernels that are positive definite but defined over non-transitive mapping systems and/or with non-positive-definite evaluation systems.

Appendix A. An alternative proof of Lemma 13

It is easy to see the assertion of Lemma 13 can be reduced to the following proposition.

Proposition 14 *If a negative definite kernel $\psi : X \times X \to \mathbb{C}$ satisfies $\Re\psi > 0$, then $1/\psi$ is indefinitely divisible and $\log\psi$ is negative definite.*

Remark 15 *For a negative definite ψ, $\Re\psi \geq 0$ (resp. $\Re\psi > 0$) is equivalent to $\psi\ |_\Delta \geq 0$ (resp. $\psi\ |_\Delta > 0$). This follows from $0 \geq \psi(x,x) + \psi(y,y) - 2\Re\psi(x,y)$.*

To prove the proposition, we extend the assertion of Corollary 2.10 of (Berg et al., 1984) as follows.

Lemma 16 *If a negative definite kernel $\psi : X \times X \to \mathbb{C}$ satisfies $\Re\psi \geq 0$, then $\log(t + \psi)$ is negative definite for arbitrary $t > 0$.*

For an arbitrary $t > 0$, we have

$$\log(t + \psi) = \log t + \log(1 + \psi/t).$$

By Corollary 2.10, $\log(1 + \psi/t)$ is negative definite, and hence, so is $\log(t + \psi)$.

Now, we prove the proposition. For $t > 0$, we define X_t by

$$X_t = \{x \in X \mid \Re\psi(x,x) \geq t\}.$$

Apparently, $X_t \subsetneqq X_s$ holds for $t > s$, and X is identical to $\bigcup_{t>0} X_t$. Moreover, when we let $\psi_t = \psi|_{X_t \times X_t}$, we have $\Re\psi_t \geq t$. This follows from

$$0 \geq \psi(x,x) + \psi(y,y) - \psi(x,y) - \overline{\psi(x,y)} \geq 2t - 2\Re\psi(x,y).$$

Thus, Theorem 2.3 of Berg et al. (1984) implies that

$$\frac{1}{\psi_t} = \frac{1}{t + (\psi_t - t)}$$

is positive definite, since $\Re(\psi_t - t) \geq 0$. The conclusion that $1/\psi$ is positive definite follows from $\lim_{t\to 0} 1/\psi_t = 1/\psi$.

Following the same discussion, we can reach the conclusion that $\log\psi$ is negative definite. Here, we use the lemma stated in the above instead of Theorem 2.3.

The infinite divisibility of $1/\psi$ follows from

$$-\log\frac{1}{\psi} = \log\psi.$$

References

C. Berg, J. P. R. Christensen, and R. Ressel. *Harmonic Analysis on semigroups. Theory of positive definite and related functions.* Springer, 1984.

M. Collins and N. Duffy. Convolution kernels for natural language. In *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001]*, pages 625–632. MIT Press, 2001.

M. Cuturi. Fast global alignment kernels. In *ICML 2011*, 2011.

M. Cuturi, J. P. Vert, O. Birkenes, and T. Matsui. A kernel for time series based on global alignments. In *ICASSP 2007*, pages II–413 – 416, 2007.

D. Haussler. Convolution kernels on discrete structures. UCSC-CRL 99-10, Dept. of Computer Science, University of California at Santa Cruz, 1999.

H. Kashima and T. Koyanagi. Kernels for semi-structured data. In *the 9th International Conference on Machine Learning (ICML 2002)*, pages 291–298, 2002.

C. S. Leslie, E. Eskin, and W. Stafford Noble. The spectrum kernel: A string kernel for SVM protein classification. In *Pacific Symposium on Biocomputing*, pages 566–575, 2002.

K. Shin and T. Kuboyama. A generalization of Haussler's convolution kernel - mapping kernel. In *ICML 2008*, 2008.

K. Shin and T. Kuboyama. Generalization of haussler's convolution kernel - mapping kernel and its application to tree kernels. *J. Comput. Sci. Technol*, 25(5)::1040–1054, 2010.

L. Zhang and K. P. Chan. A study of convolution tree kernel with local alignment. In *EACL 2009*, pages 25 – 32, 2009.