Journal of Machine Learning Research 7 (2006) 2603-2620 Submitted 9/05; Revised 10/06; Published 12/06

On Representing and Generating Kernels by Fuzzy Equivalence Relations

Bernhard Moser BERNHARD.MOSER@SCCH.AT
Knowledge-Based Technology
Software Competence Center Hagenberg
A-4232 Hagenberg, Austria

Editor: Ralf Herbrich

Abstract

Kernels are two-placed functions that can be interpreted as inner products in some Hilbert space. It is this property which makes kernels predestinated to carry linear models of learning, optimization or classification strategies over to non-linear variants. Following this idea, various kernel-based methods like support vector machines or kernel principal component analysis have been conceived which prove to be successful for machine learning, data mining and computer vision applications. When applying a kernel-based method a central question is the choice and the design of the kernel function. This paper provides a novel view on kernels based on fuzzy-logical concepts which allows to incorporate prior knowledge in the design process. It is demonstrated that kernels mapping to the unit interval with constant one in its diagonal can be represented by a commonly used fuzzy-logical formula for representing fuzzy rule bases. This means that a great class of kernels can be represented by fuzzy-logical concepts. Apart from this result, which only guarantees the existence of such a representation, constructive examples are presented and the relation to unlabeled learning is pointed out.

Keywords: kernel, triangular norm, T-transitivity, fuzzy relation, residuum

1. Motivation

Positive-definiteness plays a prominent role especially in optimization and machine learning due to the fact that two-place functions with this property, so-called kernels, can be represented as inner products in some Hilbert space. Thereby, optimization techniques conceived on the basis of linear models can be extended to non-linear algorithms. For a survey of applications see, for example, Jolliffe (1986), Schölkopf and Smola (2002) and Schölkopf et al. (1998).

Recently in Moser (2006) it was shown that kernels with values from the unit interval can be interpreted as fuzzy equivalence relations motivated by the idea that kernels express a kind of similarity. This means that the concept of fuzzy equivalence relations, or synonymously fuzzy similarity relations, is more general than that of kernels, provided only values in the unit interval are considered. Fuzzy equivalence relations distinguish from Boolean equivalence relations by a many-valued extension of transitivity which can be interpreted as many-valued logical model of the statement "*IF x is similar to y AND y is similar to z THEN x is similar to z*". In contrast to the Boolean case, in many-valued logics the set of truth values is extended such that also assertions, for example, whether two elements x and y are similar, can be treated as a matter of degree. The standard model for the set of (quasi) truth values of fuzzy logic and other many-valued logical systems is the unit interval. If $E(x,y)$ represents the (quasi) truth value of the statement that x is

©2006 Bernhard Moser.

similar to y, then the many-valued version of transitivity is modeled by

$$T(E(x,y),E(y,z)) \leq E(x,z)$$

where T is a so-called triangular norm which is an extension of the Boolean conjunction. This many-valued concept for transitivity is called T-transitivity. For a survey on triangular norms see, for example, Dubois and Prade (1985), Gottwald (1986), Gottwald (1993) and Klement et al. (2000), and for fuzzy equivalence relations and T-transitivity see, for example, Bodenhofer (2003), Höhle (1993), Höhle (1999), Klement et al. (2000), and Zadeh (1971).

Based on the semantics of fuzzy logic, this approach allows to incorporate knowledge-based models for the design of kernels. From this perspective, the most interesting mathematical question is how positive-semidefinite fuzzy equivalence relations can be characterized or at least constructed under some circumstances. At least for some special cases, proofs are provided in Section 4, which motivate further research aiming at establishing a more general theory on the positive-definiteness of fuzzy equivalence relations. These cases are based on the most prominent representatives of triangular norms, that is the Minimum, the Product and the Łukasiewicz t-norm.

The paper is structured as follows. First of all, in Section 2, some basic prerequisites concerning kernels and fuzzy relations are outlined. In Section 3, a result about the T-transitivity of kernels from Moser (2006) is cited and interpreted as existence statement that guarantees a representation of kernels mapping to the unit interval with constant 1 in its diagonal by a certain, commonly used, fuzzy-logical construction of a fuzzy equivalence relation. Finally, in contrast to the pure existence theorem of Section 3, in Section 4 constructive examples of fuzzy equivalence relations are provided which are proven to be kernels. In a concluding remark, the relationship to the problem of labeled and unlabeled learning is pointed out.

2. Prerequisites

This section summarizes definitions and facts from the theory of kernels as well as from fuzzy set theory which are needed later on.

2.1 Kernels and Positive-Semidefiniteness Preserving Functions

There is an extensive literature concerning kernels and kernel-based methods like support vector machines or kernel principal component analysis especially in the machine learning, data mining and computer vision communities. For an overview and introduction, see, for example, Schölkopf and Smola (2002). Here we present only what is needed later on. For completeness let us recall the basic definition for kernels and positive-semidefiniteness.

Definition 1 *Let $\mathcal{X}$ be a non-empty set. A real-valued function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be a kernel iff it is symmetric, that is, $k(x,y) = k(y,x)$ for all $x,y \in \mathcal{X}$, and positive-semidefinite, that is, $\sum_{i,j=1}^{n} c_i c_j k(x_i,x_j) \geq 0$ for any $n \in \mathbb{N}$, any choice of $x_1,\dots,x_n \in \mathcal{X}$ and any choice of $c_1,\dots,c_n \in \mathbb{R}$.*

One way to generate new kernels from known kernels is to apply operations which preserve the positive-semidefiniteness property. A characterization of such operations is provided by C. H. FitzGerald (1995).

Theorem 2 ***(Closeness Properties of Kernels)*** *Let $f : \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, then $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ given by*

$$k(x,y) := f(k_1(x,y),\dots,k_n(x,y))$$

is a kernel for any choice of kernels $k_1,\ldots,k_n$ on $\mathcal{X}\times\mathcal{X}$ iff f is the real restriction of an entire function on $\mathbb{C}^n$ of the form

$$f(x_1,\ldots,x_n) = \sum_{r_1\geq 0,\ldots,r_n\geq 0} c_{r_1,\ldots,r_n} x_1^{r_1}\cdots x_n^{r_n} \tag{1}$$

where $c_{r_1,\ldots,r_n}\geq 0$ for all nonnegative indices $r_1,\ldots,r_n$.

2.2 Triangular Norms

Triangular norms have been originally studied within the framework of probabilistic metric spaces, see Schweizer and Sklar (1961) and Schweizer and Sklar (1983). In this context, t-norms proved to be an appropriate concept when dealing with triangle inequalities. Later on, t-norms and their dual version, t-conorms, have been used to model conjunction and disjunction for many-valued logic, see Dubois and Prade (1985), Gottwald (1986), Gottwald (1993) and Klement et al. (2000).

Definition 3 *A function $T:[0,1]^2\rightarrow[0,1]$ is called* t-norm *(triangular norm), if it satisfies the following conditions:*

$$\begin{array}{llll}
(i) & \forall x,y\in[0,1]: & T(x,y)=T(y,x) & (\textit{commutativity})\\
(ii) & \forall x,y,z\in[0,1]: & T(x,T(y,z))=T(T(x,y),z) & (\textit{associativity})\\
(iii) & \forall x,y,z\in[0,1]: & y\leq z\Longrightarrow T(x,y)\leq T(x,z) & (\textit{monotonicity})\\
(iv) & \forall x,y\in[0,1]: & T(x,1)=x\wedge T(1,y)=y & (\textit{boundary condition})
\end{array}$$

Further, a t-norm is called Archimedean *if it is continuous and satisfies*

$$x\in(0,1)\Rightarrow T(x,x)<x.$$

Due to its associativity, many-placed extensions $T_n:[0,1]^n\rightarrow[0,1]$, $n\in\mathbb{N}$, of a t-norm T are uniquely determined by

$$T_n(x_1,\ldots,x_n)=T(x_1,T_{n-1}(x_2,\ldots,x_n)).$$

Archimedean t-norms are characterized by the following representation theorem due to Ling (1965):

Theorem 4 *Let $T:[0,1]^2\rightarrow[0,1]$ be a t-norm. Then T is Archimedean if, and only if, there is a continuous, strictly decreasing function $f:[0,1]\rightarrow[0,\infty]$ with $f(1)=0$ such that for $x,y\in[0,1]$,*

$$T(x,y)=f^{-1}(\min(f(x)+f(y),f(0))).$$

By setting $g(x)=\exp(-f(x))$, Ling's characterization yields an alternative representation with a multiplicative generator function

$$T(x,y)=g^{-1}(\max(g(x)g(y),g(0))).$$

For $g(x)=x$ we get the product $T_P(x,y)=xy$. The setting $f(x)=1-x$ yields the so-called Łukasiewcz t-norm $T_L(x,y)=\max(x+y-1,0)$. Due to Ling's theorem 4 an Archimedean t-norm T is isomorphic either to T_L or T_P, depending on whether the additive generator takes a finite value at 0 or not. In the former case, the Archimedean t-norm is called *non-strict*, in the latter it is called *strict*.

A many-valued model of an implication is provided by the so-called residuum given by

$$\vec{T}(a,b) = \sup\{c \in [0,1] | T(a,c) \leq b\} \tag{2}$$

where T is a left-continuous t-norm. Equation (2) is uniquely determined by the so-called adjunction property

$$\forall a,b,c \in [0,1] : T(a,b) \leq c \Leftrightarrow a \leq \vec{T}(b,c). \tag{3}$$

Consequently, the operator

$$\overleftrightarrow{T}(a,b) = \min\left\{\vec{T}(a,b), \vec{T}(b,a)\right\} \tag{4}$$

models a biimplication. For details, for example, see Gottwald (1986) and Klement et al. (2000). Tables 1 and 2 list examples of t-norms with their induced residuum $\vec{T}$. For further examples see, for example, Klement et al. (2000).

$T_{\cos}(a,b)$	=	$\max(ab - \sqrt{1-a^2}\sqrt{1-b^2}, 0)$
$T_L(a,b)$	=	$\max(a+b-1, 0)$
$T_P(a,b)$	=	$a\,b$
$T_M(a,b)$	=	$\min(a,b)$

Table 1: Examples of t-norms

$\vec{T}_{\cos}(a,b)$	=	$\begin{cases} \cos(\arccos(b) - \arccos(a)) & \text{if } a > b, \\ 1 & \text{else} \end{cases}$
$\vec{T}_L(a,b)$	=	$\min(b-a+1, 1)$
$\vec{T}_P(a,b)$	=	$\begin{cases} \frac{b}{a} & \text{if } a > b, \\ 1 & \text{else} \end{cases}$
$\vec{T}_M(a,b)$	=	$\begin{cases} b & \text{if } a > b, \\ 1 & \text{else} \end{cases}$

Table 2: Examples of residuums

2.3 T-Equivalences

If we want to classify based on a notion of similarity or indistinguishability, we face the problem of transitivity. For instance, let us consider two real numbers to be indistinguishable if and only if they differ by at most a certain bound $\varepsilon > 0$, this is modeled by the relation $\sim_\varepsilon$ given by $x \sim_\varepsilon y :\Leftrightarrow |x-y| < \varepsilon$, $\varepsilon > 0$, $x,y \in \mathbb{R}$. Note that the relation $\sim_\varepsilon$ is not transitive and, therefore, not an equivalence relation. The transitivity requirement turns out to be too strong for this example. The problem of identification and transitivity in the context of similarity of physical objects was early pointed out and discussed philosophically by Poincaré (1902) and Poincaré (1904). In the framework of fuzzy logic, the way to overcome this problem is to model similarity by fuzzy relations based on a many-valued concept of transitivity, see Bodenhofer (2003), Höhle (1993), Höhle (1999), Klement et al. (2000) and Zadeh (1971).

Definition 5 *A function $E : X^2 \longrightarrow [0,1]$ is called a* fuzzy equivalence relation*, or synonymously,* T-equivalence *with respect to the t-norm T if it satisfies the following conditions:*

$$\begin{array}{llll} (i) & \forall x \in X: & E(x,x) = 1 & (\textit{reflexivity}) \\ (ii) & \forall x,y \in X: & E(x,y) = E(y,x) & (\textit{symmetry}) \\ (iii) & \forall x,y,z \in X: & T(E(x,y),E(y,z)) \le E(x,z) & (\textit{T-transitivity}). \end{array}$$

The value $E(x,y)$ can be also looked at as the (quasi) truth value of the statement "*x is equal to y*". Following this semantics, *T-transitivity* can be seen as a many-valued model of the proposition, "*If x is equal to y and y is equal to z, then x is equal to z*". T-equivalences for Archimedean t-norms are closely related to metrics and pseudo-metrics as shown by Klement et al. (2000) and Moser (1995).

Theorem 6 *Let T be an Archimedean t-norm given by*

$$\forall a,b \in [0,1] : T(a,b) = f^{-1}(\min(f(a) + f(b), f(0))),$$

where $f : [0,1] \to [0,\infty]$ is a strictly decreasing, continuous function with $f(1) = 0$.
(i) If $d : X^2 \to [0,\infty[$ is a pseudo-metric, then the function $E_d : X^2 \to [0,1]$ defined by

$$E_d(x,y) = f^{-1}(\min(d(x,y), f(0)))$$

is a T-equivalence with respect to the t-norm T.
(ii) If $E : X^2 \to [0,1]$ is a T-equivalence relation, then the function $d_E : X^2 \to [0,\infty]$ defined by

$$d_E(x,y) = f(E(x,y))$$

is a pseudo-metric.

Another way to construct T-equivalences is to employ $\overrightarrow{T}$-operators. The proof of the following assertion can be found in Trillas and Valverde (1984), Kruse et al. (1993) and Kruse et al. (1994).

Theorem 7 *Let T be a left-continuous t-norm, $\overleftrightarrow{T}$ its induced biimplication, $\mu_i : \mathcal{X} \to [0,1]$, $i \in I$, I non-empty; then $E : \mathcal{X} \times \mathcal{X} \to [0,1]$ given by*

$$E(x,y) = \inf_{i \in I} \overleftrightarrow{T}(\mu_i(x), \mu_i(y)) \tag{5}$$

is a T-equivalence relation.

For further details on T-equivalences see also Boixader and Jacas (1999), Höppner et al. (2002), Jacas (1988), Trillas et al. (1999) and Valverde (1985).

3. Representing Kernels by T-Equivalences

It is interesting that the concept of kernels, which is motivated by geometric reasoning in terms of inner products and mappings to Hilbert spaces and which is inherently formulated by algebraic terms, is closely related to the concept of fuzzy equivalence relations as demonstrated and discussed in more detail in Moser (2006). In this section, we start with the result that any kernel $k : \mathcal{X} \times \mathcal{X} \to [0,1]$ with $k(x,x) = 1$ for all $x \in \mathcal{X}$ is T-transitive and, therefore, a fuzzy equivalence relation. The proof can be found in Moser (2006), see also Appendix A.1.

Theorem 8 *Any kernel $k : \mathcal{X} \times \mathcal{X} \to [0,1]$ with $k(x,x) = 1$ is (at least) $T_{\cos}$-transitive, where*

$$T_{\cos}(a,b) = \max\{a\,b - \sqrt{1-a^2}\sqrt{1-b^2}, 0\}. \tag{6}$$

The nomenclature is motivated by the fact that the triangular norm defined by Equation (6) is an Archimedean t-norm which is generated by the arcosine function as its additive generator. From this result, the following existence theorem can be derived, which guarantees that any kernel under consideration can be represented by the fuzzy-logical formula given by (5). In fuzzy systems, this formula is commonly used for modeling rule bases (see, for example, Kruse et al., 1993, 1994).

Theorem 9 *Let $\mathcal{X}$ be a non-empty universe of discourse, $k : \mathcal{X} \times \mathcal{X} \to [0,1]$ a kernel in the sense of Definition 1 and $k(x,x) = 1$ for all $x \in \mathcal{X}$; then there is a family of membership functions $\mu_i : \mathcal{X} \to [0,1]$, $i \in I$, I non-empty and a t-norm T, such that*

$$\forall x,y \in \mathcal{X} : k(x,y) = \inf_{i \in I} \overset{\leftrightarrow}{T}(\mu_i(x), \mu_i(y)). \tag{7}$$

Proof. Let us set $I := \mathcal{X}$, $\mu_{x_0}(x) = k(x,x_0)$ and let us choose $T_{\cos}$ as t-norm. For convenience let us denote

$$h(x,y) = \inf_{x_0 \in \mathcal{X}} \overset{\leftrightarrow}{T}_{\cos}(\mu_{x_0}(x), \mu_{x_0}(y)),$$

which is equivalent to

$$h(x,y) = \inf_{x_0 \in \mathcal{X}} \overset{\leftrightarrow}{T}_{\cos}(k(x_0,x), k(x_0,y)).$$

According to Theorem 8, k is $T_{\cos}$-transitive, that is,

$$\forall x_0,x,y \in \mathcal{X} : \overset{\leftrightarrow}{T}_{\cos}(k(x_0,x), k(x_0,y)) \le k(x,y).$$

This implies that $h(x,y) \le k(x,y)$ for all $x,y \in \mathcal{X}$. Now let us consider the other inequality. Due to the adjunction property (3), we obtain

$$T_{\cos}(k(x,y), k(x_0,y)) \le k(x,x_0) \Leftrightarrow k(x,y) \le \vec{T}_{\cos}(k(x_0,y), k(x,x_0))$$

and

$$T_{\cos}(k(x,y), k(x_0,x)) \le k(y,x_0) \Leftrightarrow k(x,y) \le \vec{T}_{\cos}(k(x_0,x), k(y,x_0)),$$

from which it follows that

$$\forall x,y,x_0 \in \mathcal{X} : k(x,y) \le \min\{\vec{T}_{\cos}(k(x_0,y), k(x,x_0)), \vec{T}_{\cos}(k(x_0,x), k(y,x_0))\}.$$

Hence by Definition 4,

$$\forall x,y \in \mathcal{X} : k(x,y) \le h(x,y)$$

which ends the proof. ■

For an arbitrary choice of fuzzy membership functions, there is no necessity that the resulting relation (7) implies positive-semidefiniteness and, therefore, a kernel. For an example of a $T_{\cos}$-equivalence which is not a kernel see Appendix A.4. Theorem 9 guarantees only the existence of a representation of the form (5) but it does not tell us how to construct the membership functions μ_i. In the following section, we provide examples of fuzzy equivalence relations which yield kernels for any choice of membership functions.

4. Constructing Kernels by Fuzzy Equivalence Relations

In the Boolean case, positive-definiteness and equivalence are synonymous, that is, a Boolean relation $R : \mathcal{X} \times \mathcal{X} \to \{0,1\}$ is positive-definite if and only if R is the indicator function of an equivalence relation $\cong$, that is, $R(x,y) = 1$ if $x \cong y$ and $R(x,y) = 0$ if $x \ncong y$. For a proof, see Appendix A.2. This relationship can be used to obtain an extension to fuzzy relations as given by the next theorem whose proof can be found in the Appendix A.3.

Theorem 10 *Let $\mathcal{X}$ be a non-empty universe of discourse, $\mu_i : \mathcal{X} \to [0,1]$, $i \in I$, I non-empty; then the fuzzy equivalence relation $E_M : \mathcal{X} \times \mathcal{X} \to [0,1]$ given by*

$$E_M(x,y) = \inf_{i \in I} \overset{\leftrightarrow}{T}_M(\mu_i(x), \mu_i(y))$$

is positive-semidefinite.

In the following, the most prominent representatives of Archimedean t-norms, the Product T_P and the Łukasiewicz t-norm T_L, are used to construct positive-semidefinite fuzzy similarity relations. Though the first part can also be derived from a result due to Yaglom (1957) that characterizes isotropic stationary kernels by its spectral representation, here we prefer to present a direct, elementary proof. Compare also Bochner (1955) and Genton (2001).

Theorem 11 *Let $\mathcal{X}$ be a non-empty universe of discourse, $\nu : \mathcal{X} \to [0,1]$ and let $h : [0,1] \to [0,1]$ be an isomorphism of the unit interval that can be expanded in the manner of Equation (1), that is $h(x) = \sum_k c_k x^k$ with $c_k \geq 0$; then the fuzzy equivalence relations $E_{L,h}, E_{P,h} : \mathcal{X} \times \mathcal{X} \to [0,1]$ given by*

$$E_{L,h}(x,y) = h\left(\overset{\leftrightarrow}{T}_L\left(h^{-1}(\nu(x)), h^{-1}(\nu(y))\right)\right) \tag{8}$$

and

$$E_{P,h}(x,y) = h\left(\overset{\leftrightarrow}{T}_P\left(h^{-1}(\nu(x)), h^{-1}(\nu(y))\right)\right) \tag{9}$$

are positive-semidefinite.

Proof. To prove the positive-definiteness of the two-placed functions $E_{L,h}$ and $E_{P,h}$ given by equations (8) and (9) respectively, we have to show that

$$\sum_{i,j=1}^{n} E_{L,h}(x_i, x_i)\, c_i c_j \geq 0, \sum_{i,j=1}^{n} E_{P,h}(x_i, x_j)\, c_i c_j \geq 0$$

for any $n \in \mathbb{N}$ and any choice of $x_1, \ldots, x_n \in \mathcal{X}$, respectively. According to an elementary result from Linear Algebra this is equivalent to the assertion that the determinants ($1 \leq m \leq n$)

$$D_m = det\left[(E(x_i, x_j))_{i,j \in \{1,\ldots,m\}}\right]$$

of the minors of the matrix $(E(x_i, x_j))_{i,j}$ satisfy

$$\forall m \in \{1, \ldots, n\} : D_m \geq 0,$$

where E denotes either $E_{L,h}$ or $E_{P,h}$. Recall that the determinant of a matrix is invariant with respect to renaming the indices, that is, if $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a permutation then

$$det\left[(a_{ij})_{i,j}\right] = det\left[(a_{\sigma(i)\sigma(j)})_{i,j}\right].$$

For convenience, let denote $\mu_i = h^{-1}(\nu(x_i))$. Then, without loss of generality, we may assume that the values μ_i are ordered monotonically decreasing, that is,

$$\mu_i \geq \mu_j \text{ for } i < j. \tag{10}$$

Case T_L: Note that $\overleftrightarrow{T}_L(a,b) = \min\{\overrightarrow{T}_L(a,b), \overrightarrow{T}_L(b,a)\} = 1 - |a-b|$. Then we have to show that for all dimensions $n \in \mathbb{N}$, the determinant of

$$E^{(n)} = (1 - |\mu_i - \mu_j|)_{i,j \in \{1,\ldots,n\}}$$

is non-negative, that is

$$det[E^{(n)}] \geq 0.$$

Due to the assumption (10), we have

$$1 - |\mu_i - \mu_j| = \begin{cases} 1 - (\mu_i - \mu_j) & \text{if } i \leq j, \\ 1 - (\mu_j - \mu_i) & \text{else} \end{cases}$$

which yields

$$E^{(n)} = \begin{pmatrix} 1 & 1-(\mu_1-\mu_2) & \ldots & 1-(\mu_1-\mu_{n-1}) & 1-(\mu_1-\mu_n) \\ 1-(\mu_1-\mu_2) & 1 & \ldots & 1-(\mu_2-\mu_{n-1}) & 1-(\mu_2-\mu_n) \\ 1-(\mu_1-\mu_3) & 1-(\mu_2-\mu_3) & \ldots & 1-(\mu_3-\mu_{n-1}) & 1-(\mu_3-\mu_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-(\mu_1-\mu_{n-1}) & 1-(\mu_2-\mu_{n-1}) & \ldots & 1 & 1-(\mu_{n-1}-\mu_n) \\ 1-(\mu_1-\mu_n) & 1-(\mu_2-\mu_n) & \ldots & 1-(\mu_{n-1}-\mu_n) & 1 \end{pmatrix}.$$

Now let us apply determinant-invariant elementary column operations to simplify this matrix by subtracting the column with index $i-1$ from the column with index i, $i \geq 2$. This yields

$$\tilde{E}^{(n)} = \begin{pmatrix} 1 & \mu_2-\mu_1 & \ldots & \mu_{n-1}-\mu_{n-2} & \mu_n-\mu_{n-1} \\ 1-(\mu_1-\mu_2) & -(\mu_2-\mu_1) & \ldots & \mu_{n-1}-\mu_{n-2} & \mu_n-\mu_{n-1} \\ 1-(\mu_1-\mu_3) & -(\mu_2-\mu_1) & \ldots & \mu_{n-1}-\mu_{n-2} & \mu_n-\mu_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-(\mu_1-\mu_{n-1}) & -(\mu_2-\mu_1) & \ldots & -(\mu_{n-2}-\mu_{n-1}) & \mu_n-\mu_{n-1} \\ 1-(\mu_1-\mu_n) & -(\mu_2-\mu_1) & \ldots & -(\mu_{n-2}-\mu_{n-1}) & -(\mu_{n-1}-\mu_n) \end{pmatrix}.$$

Therefore,

$$\begin{aligned} \alpha &= \prod_{i=2}^{n}(\mu_{i-1}-\mu_i) \geq 0 \qquad (11) \\ det[E^{(n)}] &= det[\tilde{E}^{(n)}] = \alpha\, det[\hat{E}_n], \end{aligned}$$

where

$$\hat{E}^{(n)} = \begin{pmatrix} 1 & -1 & \ldots & -1 & -1 \\ 1-(\mu_1-\mu_2) & +1 & \ldots & -1 & -1 \\ 1-(\mu_1-\mu_3) & +1 & \ldots & -1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1-(\mu_1-\mu_{n-1}) & +1 & \ldots & +1 & -1 \\ 1-(\mu_1-\mu_n) & +1 & \ldots & +1 & +1 \end{pmatrix}. \tag{12}$$

Let us apply Laplacian determinant expansion by minors to the first column of matrix (12), that is

$$det[A] = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} det[A_{ij}]$$

where $A = (a_{ij})$ is an $n \times n$-matrix, j arbitrarily chosen from $\{1, \ldots, n\}$ and A_{ij} is the matrix corresponding to the cofactor a_{ij} obtained by canceling out the i-th row and the j-th column from A (see, for example, Muir, 1960). For $n = 1$, we get the trivial case $det[\hat{E}^{(1)}] = 1$. Note that the first and the last rows of the matrices $\hat{E}^{(n)}_{i,1}$ for $1 < i < n$ only differ by their signum, consequently the minors $det[\hat{E}^{(n)}_{i,1}]$ for $1 < i < n$, $n \geq 2$, are vanishing, that is,

$$det[A_{i,1}] = 0, \text{ for } 1 < i < n.$$

Therefore, according to the Laplacian expansion, we get

$$det[\hat{E}^{(n)}] = 1 \cdot det[\hat{E}^{(n)}_{1,1}] + (-1)^n (1 - (\mu_1 - \mu_n)) \cdot det[\hat{E}^{(n)}_{1,n}]. \quad (13)$$

Observe that

$$\begin{aligned} det[\hat{E}^{(n)}_{1,1}] &= 2^{n-2} \\ det[\hat{E}^{(n)}_{1,n}] &= (-1)^{n-1} 2^{n-2}. \end{aligned}$$

Consequently, Equation (13) simplifies to

$$\begin{aligned} det[\hat{E}^{(n)}] &= 2^{n-2} \left(1 + (-1)^n (-1)^{n-1} 2^{n-2} (1 - (\mu_1 - \mu_n))\right) \\ &= 2^{n-2} (1 - (1 - (\mu_1 - \mu_n))) \\ &= 2^{n-2} (\mu_1 - \mu_n) \\ &\geq 0 \end{aligned}$$

which together with (11) proves the first case.

Case T_P: First of all, let us compute $\overset{\leftrightarrow}{T}_P(a,b) = \min\{\vec{T}_P(a,b), \vec{T}_L(b,a)\}$. Hence,

$$\overset{\leftrightarrow}{T}_P(a,b) = \begin{cases} \min\{\frac{b}{a}, \frac{a}{b}\} & \text{if } a, b > 0, \\ 0 & \text{if } a = 0 \text{ and } b > 0 \text{ ,} \\ 0 & \text{if } b = 0 \text{ and } a > 0 \text{ ,} \\ 1 & \text{if } a = 0 \text{ and } b = 0 \text{ .} \end{cases}$$

Again, without loss of generality, let us suppose that the values μ_i, $i \in \{1, \ldots, n\}$ are ordered monotonically decreasing, that is $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$. Before checking the general case, let us consider the special case of vanishing μ-values. For this, let us assume for the moment that

$$\mu_i = \begin{cases} > 0 & \text{if } i < i_0 \text{ ,} \\ 0 & \text{else} \end{cases}$$

which implies that $\overset{\leftrightarrow}{T}_P(\mu_i, \mu_j) = 0$ for $i < i_0$ and $j \geq i_0$ and $\overset{\leftrightarrow}{T}_P(\mu_i, \mu_j) = 1$ for $i \geq i_0$ and $j \geq i_0$. This leads to a decomposition of the matrix

$$E^{(n)} = \left(\overset{\leftrightarrow}{T}_P(\mu_i, \mu_j)\right)_{ij}$$

such that

$$det[E^{(n)}] = det[E^{(i_0-1)}] \cdot det[I_{n-i_0-1}]$$

where I_k denotes the $k \times k$-matrix with constant entries 1, hence $det[I_{n-i_0-1}] \in \{0,1\}$. Therefore, we may assume that

$$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n > 0.$$

Then we have to show that for all dimensions $n \in \mathbb{N}$, the determinant of

$$E^{(n)} = \left(\min\left\{\frac{\mu_i}{\mu_j}, \frac{\mu_j}{\mu_i}\right\}\right)_{i,j\in\{1,\ldots,n\}}$$

is non-negative, that is

$$det[E^{(n)}] \geq 0.$$

Consider

$$E^{(n)} = \begin{pmatrix} 1 & \frac{\mu_2}{\mu_1} & \ldots & \frac{\mu_{n-1}}{\mu_1} & \frac{\mu_n}{\mu_1} \\ \frac{\mu_2}{\mu_1} & 1 & \ldots & \frac{\mu_{n-1}}{\mu_2} & \frac{\mu_n}{\mu_2} \\ \frac{\mu_3}{\mu_1} & \frac{\mu_3}{\mu_2} & \ldots & \frac{\mu_{n-1}}{\mu_3} & \frac{\mu_n}{\mu_3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\mu_{n-1}}{\mu_1} & \frac{\mu_{n-1}}{\mu_2} & \ldots & 1 & \frac{\mu_n}{\mu_{n-1}} \\ \frac{\mu_n}{\mu_1} & \frac{\mu_n}{\mu_2} & \ldots & \frac{\mu_n}{\mu_{n-1}} & 1 \end{pmatrix}. \tag{14}$$

Now, multiply the i-th column by $-\mu_{i+1}/\mu_i$ and add it to the $(i+1)$-th column of matrix (14), $1 \leq i < n$, then we get

$$\tilde{E}^{(n)} = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ * & 1-\left(\frac{\mu_2}{\mu_1}\right)^2 & \ldots & 0 & 0 \\ * & * & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \ldots & 1-\left(\frac{\mu_{n-1}}{\mu_{n-2}}\right)^2 & 0 \\ * & * & \ldots & * & 1-\left(\frac{\mu_n}{\mu_{n-1}}\right)^2 \end{pmatrix} \tag{15}$$

where $*$ is a placeholder for any real value. By this, the determinant of the matrix in Equation (15) readily turns out to be

$$det[E^{(n)}] = det[\tilde{E}^{(n)}] = \prod_{i=1}^{n-1}\left(1-\left(\frac{\mu_{i+1}}{\mu_i}\right)^2\right) \geq 0$$

which together with Theorem (2) ends the proof. ■

Note that relations (8) and (9) are T-transitive with respect to the corresponding isomorphic Archimedean t-norms,

$$T_{L,h}(x,y) = h(T_L(h^{-1}(x),h^{-1}(x))) \text{ and } T_{P,h}(x,y) = h(T_P(h^{-1}(x),h^{-1}(x))),$$

respectively.

Corollary 12 *Let $\mathcal{X}$ be a non-empty universe of discourse, $\mu_i : \mathcal{X} \to [0,1]$, $\lambda_i \in]0,1]$ with $\sum_i \lambda_i = 1$ where $i \in \{1,\dots,n\}$, $n \in \mathbb{N}$, then the fuzzy equivalence relations $\tilde{E}_L, \tilde{E}_P : \mathcal{X} \times \mathcal{X} \to [0,1]$ given by*

$$\tilde{E}_L(x,y) = \sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(x), \mu_i(y)) \tag{16}$$

and

$$\tilde{E}_P(x,y) = \prod_{i=1}^{n} \left(\overleftrightarrow{T}_P(\mu_i(x), \mu_i(y))\right)^{\lambda_i} \tag{17}$$

are T_L- and T_P-equivalences, respectively, and kernels.

Proof. First of all, let us check the T_L-transitivity of formula (16). This can readily be shown by means of the definition of T_L and the T_L-transitivity of $\overleftrightarrow{T}_L$ due to the following inequalities:

$$\begin{aligned}
T_L\left(\sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(x),\mu_i(y)), \sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(y),\mu_i(yz)\right) &= \\
\max\left\{\sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(x),\mu_i(y)) + \sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(y),\mu_i(z)) - 1, 0\right\} &= \\
\max\left\{\sum_{i=1}^{n} \lambda_i \left(\overleftrightarrow{T}_L(\mu_i(x),\mu_i(y)) + \sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(y),\mu_i(z)) - 1\right), 0\right\} &\leq \\
\max\left\{\sum_{i=1}^{n} \lambda_i T_L\left(\overleftrightarrow{T}_L(\mu_i(x),\mu_i(y)), \sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(y),\mu_i(z))\right), 0\right\} &\leq \\
\max\left\{\sum_{i=1}^{n} \lambda_i \overleftrightarrow{T}_L(\mu_i(x),\mu_i(z)), 0\right\} &= \\
\lambda_i \overleftrightarrow{T}_L(\mu_i(x),\mu_i(z)).&
\end{aligned}$$

This, together with the T_P-transitivity of $\overleftrightarrow{T}_P$, proves that the formulas given by (16) and (17) are T_L- and T_P-equivalences, respectively.

Expanding the factors of formula (17) yields

$$\left(\overleftrightarrow{T}_P(\mu_i(x),\mu_i(y))\right)^{\lambda_i} = \begin{cases} 1 & \text{if } \mu_i(x) = \mu_i(y) = 0, \\ \frac{\min(\mu_i^{\lambda_i}(x), \mu_i^{\lambda_i}(y))}{\max(\mu_i^{\lambda_i}(x), \mu_i^{\lambda_i}(y))} & \text{else} \end{cases} \tag{18}$$

which by comparing case T_P of the proof of Theorem 11 shows that the left-hand side of Equation (18) is positive-semidefinite.

As the convex combination and the product are special cases of positive-semidefiniteness preserving functions according to Theorem 1, the functions defined by equations (16) and (17) prove to be again positive-semidefinite and, therefore, kernels. ■

It is interesting to observe that both formulas (16) and (17) can be expressed in the form, $f(\|\tau(x) - \tau(y)\|_1)$, where $f : I \to [0,1]$, I some interval, is a strictly decreasing function, $\tau : \mathcal{X} \to I^n$, I some interval, $\tau(x) = (\tau_1(x), \dots, \tau_n(x))$ and $\|\tau(x)\|_1 = \sum_{i=1}^{n} |\tau_i(x)|$. Indeed, for Equation (16) let us define

$$\begin{aligned}
f_L : [0,1] \to [0,1], \quad f_L(a) &= 1 - a \\
\tau_L : \mathcal{X} \to [0,1]^n, \quad \tau_L(x) &= (\lambda_1 \mu_1(x), \dots, \lambda_n \mu_n(x))
\end{aligned}$$

and for Equation (17) and positive membership functions μ_i, $\mu_i(x) > 0$ for all $x \in \mathcal{X}$, let us define

$$\begin{aligned} f_P : [0,\infty[\rightarrow [0,1], \quad f_P(a) &= e^{-a} \\ \tau_P : \mathcal{X} \rightarrow]-\infty,1]^n, \quad \tau_P(x) &= (\lambda_1 \ln(\mu_1(x)),\ldots,\lambda_n \ln(\mu_n(x))) \end{aligned}$$

Therefore, we get

$$\tilde{E}_L(x,y) = 1 - \|\tau_L(x) - \tau_L(y)\|_1 \tag{19}$$

$$\tilde{E}_P(x,y) = e^{-\|\tau_P(x)-\tau_P(y)\|_1}. \tag{20}$$

While formulas (19) and (20) provide a geometrical interpretation by means of the norm $\|.\|_1$, the corresponding formulas (16) and (17) yield a semantical model of the assertion

"*IF* x is equal to y with respect to feature μ_1 *AND* ... *AND* x is equal to y with respect to feature μ_n *THEN* x is equal to y"

as aggregation of biimplications in terms of fuzzy logic. While in the former case, the aggregation has some compensatory effect, the latter is just a conjunction in terms of the Product triangular norm. For details on aggregation operators see, for example, Saminger et al. (2002) and Calvo et al. (2002).

The formulas (16) and (17) coincide for the following special case. If the membership functions μ_i are indicator functions of sets $A_i \subseteq \mathcal{X}$ which form a partition of $\mathcal{X}$, then the kernels (16) and (17) reduce to the indicator function characterizing the Boolean equivalence relation induced by this partition $\{A_1,\ldots,A_n\}$.

The formulas (16) and (17) for general membership functions therefore provide kernels which can be interpreted to be induced by a family of fuzzy sets and, in particular, by fuzzy partitions, that is, families of fuzzy sets fulfilling some criteria which extend the axioms for a Boolean partition in a many-valued logical sense. For definitions and further details on fuzzy partitions see, for example, De Baets and Mesiar (1998), Demirci (2003) and Höppner and Klawonn (2003).

It is a frequently used paradigm that the decision boundaries for a classification problem lie between clusters rather than intersecting them. Due to this cluster hypothesis, the problem of designing kernels based on fuzzy partitions is closely related to the problem of learning kernels from unlabeled data. For further details on semi-supervised learning see, for example, Seeger (2002), Chapelle et al. (2003) and T. M. Huang (2006). It is left to future research to explore this relationship to the problem of learning from labeled and unlabeled data and related concepts like covariance kernels.

5. Conclusion

In this paper, we have presented a novel view on kernels from a fuzzy logical point of view. Particularly, the similarity-measure aspect of a kernel is addressed and investigated by means of the so-called T-transitivity which is characteristic for fuzzy equivalence relations. As a consequence, we derived that a large class of kernels can be represented in a way that is commonly used for representing fuzzy rule bases. In addition to this proof for the existence of such a representation, constructive examples are presented. It is the idea of this research to look for a combination of knowledge-based strategies with kernel-based methods in order to facilitate a more flexible designing process of kernels which also allows to incorporate prior knowledge. Further research aims at analyzing the behavior of kernels constructed in this way when applied in the various kernel methods like support vector machines, kernel principal components analysis and others. In particular, it is intended to focus on the problem of learning kernels from unlabeled data where the fuzzy partitions are induced by appropriate clustering principles.

Acknowledgments

Bernhard Moser gratefully acknowledges partial support by the Austrian Government, the State of Upper Austria, and the Johannes Kepler University Linz in the framework of the K*plus* Competence Center Program. Furthermore special thanks go to the anonymous reviewers who gave helpful suggestions and to Felix Kossak for careful proof-reading.

Appendix A.

For sake of completeness the following sections provide proofs regarding Theorem 8, the characterization of kernels in the Boolean case and the construction of kernels by means of the minimum t-norm T_M. Furthermore, in Section A.4 an example of a non-positive-semidefinite $T_{\cos}$-equivalence is given.

A.1 Proof of Theorem 8

Let us start with the analysis of 3-dimensional matrices.

Lemma 13 *Let $M = (m_{ij})_{ij} \in [0,1]^{3\times 3}$ be a 3×3 symmetric matrix with $m_{ii} = 1$, $i = 1,2,3$; then M is positive-semidefinite iff for all $i, j, k \in \{1,2,3\}$ there holds*

$$m_{ij}m_{jk} - \sqrt{1-m_{ij}^2}\sqrt{1-m_{jk}^2} \leq m_{ik}$$

Proof. For simplicity, let $a = m_{1,2}$, $b = m_{1,3}$ and $c = m_{2,3}$. Then the determinant of M, $Det(M)$, is a function of the variables a,b,c given by

$$D(a,b,c) = 1 + 2abc - a^2 - b^2 - c^2.$$

For any choice of a,b, the quadratic equation $D(a,b,c) = 0$ can be solved for c, yielding two solutions $c_1 = c_1(a,b)$ and $c_2 = c_2(a,b)$ as functions of a and b,

$$\begin{aligned} c_1(a,b) &= ab - \sqrt{1-a^2}\sqrt{1-b^2} \\ c_2(a,b) &= ab + \sqrt{1-a^2}\sqrt{1-b^2}. \end{aligned}$$

Obviously, for all $|a| \leq 1$ and $|b| \leq 1$, the values $c_1(a,b)$ and $c_2(a,b)$ are real. By substituting $a = \cos\alpha$ and $b = \cos(\beta)$ with $\alpha, \beta \in [0, \frac{\pi}{2}]$, it becomes readily clear that

$$\begin{aligned} c_1(a,b) &= c_1(\cos(\alpha), \cos(\beta)) \\ &= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \\ &= \cos(\alpha+\beta) \in [-1,1] \end{aligned}$$

and, analogously,

$$\begin{aligned} c_2(a,b) &= c_2(\cos(\alpha),\cos(\beta)) \\ &= \cos(\alpha)\cos(\beta)+\sin(\alpha)\sin(\beta) \\ &= \cos(\alpha-\beta)\in[-1,1]. \end{aligned}$$

As for all $a,b\in[-1,1]$ the determinant function $D_{a,b}(c):=D(a,b,c)$ is quadratic in c with negative coefficient for c^2, there is a uniquely determined maximum at $c_0(a,b)=ab$. Note that for all $a,b\in[-1,1]$, we have

$$c_1(a,b)\leq c_0(a,b)\leq c_2(a,b)$$

and

$$D(a,b,c_0(a,b))=1+2ab(ab)-a^2-b^2-(ab)^2=(1-a^2)(1-b^2)\geq 0.$$

Therefore, $D(a,b,c)\geq 0$ if and only if $c\in[c_1(a,b),c_2(a,b)]$.

Recall from linear algebra that by renaming the indices, the determinant does not change. Therefore, without loss of generality, we may assume that

$$a\geq b\geq c.$$

For convenience, let $Q=\{(x,y,z)\in[0,1]^3 | x\geq y\geq z\}$. Then, obviously, for any choice of $a,b\in[0,1]$ there holds $(a,b,c_1(a,b))\in Q$. Elementary algebra shows that $(a,b,c_2(a,b))\in Q$ is only the case for $a=b=1$. As for $a=b=1$ the two solutions c_1, c_2 coincide, that is, $c_1(1,1)=c_2(1,1)=1$, it follows that for any choice of $(a,b,c)\in Q$, there holds $D(a,b,c)\geq 0$ if and only if

$$c_1(a,b)=ab-\sqrt{1-a^2}\sqrt{1-b^2}\leq c. \tag{21}$$

If $(a,b,c)\notin Q$, then the inequality (21) is trivially satisfied which together with (21) proves the lemma ■

Now Theorem 8 immediately follows from Definition (1), Lemma (13) and the characterizing inequality (21).

A.2 Characterization of Kernels in the Boolean Case

The following lemma and proposition can also be found as an exercise in Schölkopf and Smola (2002).

Lemma 14 *Let $\sim$ be an equivalence relation on $\mathcal{X}$ and let $k:\mathcal{X}\times\mathcal{X}\to\{0,1\}$ be induced by $\sim$ via $k(x,y)=1$ if and only if $x\sim y$; then k is a kernel.*

Proof. By definition of positive-definiteness, let us consider an arbitrary sequence of elements $x_1,\ldots,x_n$. Then there are at most n equivalence classes $Q_1,\ldots,Q_m$ on the set of indices $\{1,\ldots,n\}$, $m\leq n$, where $\bigcup_{i=1,\ldots,m}Q_i=\{1,\ldots,n\}$ and $Q_i\cap Q_j=\emptyset$ for $i\neq j$. Note that $k(x_i,x_j)=0$ if the indices

i, j belong to different equivalence classes. Then, for any choice of reals $c_1, \ldots, c_n$, we obtain

$$\begin{aligned}\sum_{i,j} c_i c_j k(x_i, x_j) &= \sum_{p=1}^{m} \sum_{i,j \in Q_p} c_i c_j k(x_i, x_j) \\ &= \sum_{p=1}^{m} \sum_{i,j \in Q_p} c_i c_j \cdot 1 \\ &= \sum_{p=1}^{m} \left(\sum_{i \in Q_p} c_i \right)^2 \\ &\geq 0\end{aligned}$$

■

Proposition 15 $k : \mathcal{X} \times \mathcal{X} \to \{0,1\}$ *with* $k(x,x) = 1$ *for all* $x \in \mathcal{X}$ *is a kernel if and only if it is induced by an equivalence relation.*

Proof. It only remains to be shown that if k is a kernel, then it is the indicator function of an equivalence relation, that is, it is induced by an equivalence relation. If k is a kernel, according to Lemma 13, for all $x, y, z \in \mathcal{X}$, it has to satisfy $T_{\cos}(k(x,y), k(y,z)) \leq k(x,z)$, which implies,

$$k(x,y) = 1, \quad k(y,z) = 1 \Longrightarrow k(x,z) = 1.$$

Obviously, we have $k(x,x) = 1$ and $k(x,y) = k(y,x)$ due to the reflexivity and symmetry assumption of k, respectively. ■

A.3 Constructing Kernels by T_M

For convenience let us recall the basic notion of an α-cut from fuzzy set theory:

Definition 16 *Let* $\mathcal{X}$ *be a non-empty set and* $\mu : \mathcal{X} \to [0,1]$*; then*

$$[\mu]_\alpha = \{x \in \mathcal{X} \,|\, \mu(x) \geq \alpha\}$$

is called the α*-cut of the membership function* μ*.*

Lemma 17 $k : \mathcal{X} \times \mathcal{X} \to [0,1]$ *is a* T_M*-equivalence if and only if all* α*-cuts of* k *are Boolean equivalence relations.*

Proof.

(i) Let us assume that k is a T_M-equivalence. Let $\alpha \in [0,1]$, then by definition,

$$[k]_\alpha = \{(x,y) \in \mathcal{X} \times \mathcal{X} \,|\, k(x,y) \geq \alpha\}.$$

In order to show that $[k]_\alpha$ is a Boolean equivalence, the axioms for reflexivity, symmetry and transitivity have to be shown. Reflexivity and symmetry are trivially satisfied as for all $x, y \in \mathcal{X}$, there holds by assumption that $k(x,x) = 1$ and $k(x,y) = k(y,x)$. In order to show transitivity, let us consider $(x,y), (y,z) \in [k]_\alpha$, that means $k(x,y) \geq \alpha$ and $k(y,z) \geq \alpha$; then by the T_M-transitivity assumption it follows that

$$\alpha \leq \min(k(x,y), k(y,z)) \leq k(x,z),$$

hence $(x,z) \in [k]_\alpha$.

(ii) Suppose now that all α-cuts of k are Boolean equivalence relations. Then, in particular, $[k]_\alpha$ with $\alpha = 1$ is reflexive, hence $k(x,x) = 1$ for all $x \in \mathcal{X}$. The symmetry of k follows from the fact that for all $\alpha \in [0,1]$ and pairs $(x,y) \in [k]_\alpha$, by assumption, we have $(y,x) \in [k]_\alpha$. In order to show the T_M-transitivity property, let us consider arbitrarily chosen elements $x,y,z \in \mathcal{X}$. Let $\alpha = \min(k(x,y),k(y,z))$; then by the transitivity assumption of $[k]_\alpha$, it follows that $(x,z) \in [k]_\alpha$, consequently

$$k(x,z) \geq \alpha = \min(k(x,y),k(y,z)).$$

■

Proposition 18 *If $k : \mathcal{X} \times \mathcal{X} \to [0,1]$ is a T_M-equivalence then it is positive-semidefinite.*

Proof. Choose arbitrary elements $x_1,\ldots,x_n \in \mathcal{X}$ and consider the set of values which are taken by all combinations $k(x_i,x_j)$, $i,j \in \{1,\ldots,n\}$ and order them increasingly, that is

$$\left\{k(x_i,x_j) | \, i,j \in \{1,\ldots,n\}\right\} = \left\{\alpha_1,\ldots,\alpha_m\right\},$$

where $0 \leq \alpha_1 \leq \cdots \alpha_m \leq 1$. Observe that for all pairs (x_i,x_j), $i,j \in \{1,\ldots,n\}$ there holds

$$k(x_i,x_j) = \sum_{\nu=2}^{m} (\alpha_\nu - \alpha_{\nu-1}) \mathbf{1}_{[k]_{\alpha_\nu}}(x_i,x_j) + \alpha_1 \mathbf{1}_{[k]_{\alpha_1}}(x_i,x_j)$$

showing that on the set $\{x_1,\ldots,x_n\} \times \{x_1,\ldots,x_n\}$, the function k is a linear combination of indicator functions of Boolean equivalences (which are positive-semidefinite by Proposition 15) with non-negative coefficients and, consequently, it has to be positive semidefinite. ■

A.4 Example of a Non-Positive-Semidefinite $T_{\cos}$-Equivalence

For dimensions $n > 3$, the $T_{\cos}$-transitivity is no longer sufficient to guarantee positive semi-definiteness. Consider, for example $A_n = (a_{ij}^{(n)})_{ij}$ where

$$a_{ij}^{(n)} = \begin{cases} \lambda & \text{if } \min(i,j) = 1, \max(i,j) > 1\ , \\ 1 & \text{if } i = j, \\ 0 & \text{else}\ . \end{cases} \tag{22}$$

Choose $\lambda = 1/\sqrt{2}$, then $T_{\cos}(\lambda,\lambda) = 0$, hence we have $T_{\cos}(a_{ij}^{(n)}, a_{jk}^{(n)}) \leq a_{ik}^{(n)}$ for all indices $i,j,k \in 1,\ldots,n$. As $det(A_n) < 0$ for $n > 3$, the matrix A_n cannot be positive-semidefinite though the $T_{\cos}$-transitivity conditions are satisfied.

References

S. Bochner. *Harmonic Analysis and the Theory of Probability*. University of California Press, Los Angeles, California, 1955.

U. Bodenhofer. A note on approximate equality versus the Poincaré paradox. *Fuzzy Sets and Systems*, 133(2):155–160, 2003.

D. Boixader and J. Jacas. T-indistinguishability operators and approximate reasoning via CRI. In D. Dubois, E. P. Klement, and H. Prade, editors, *Fuzzy Sets, Logics and Reasoning about Knowledge*, volume 15 of *Applied Logic Series*, pages 255–268. Kluwer Academic Publishers, Dordrecht, 1999.

A. Pinkus C. H. FitzGerald, C.A. Micchelli. Functions that preserve families of positive semidefinite matrices. *Linear Alg. and Appl.*, 221:83–102, 1995.

T. Calvo, G. Mayor, and R. Mesiar, editors. *Aggregation Operators*, volume 97 of *Studies in Fuzziness and Soft Computing*. Physica-Verlag, Heidelberg, 2002.

O. Chapelle, J. Weston, and B. Schölkopf. Cluster kernels for semi-supervised learning. volume 15 of *NIPS*. 2003.

B. De Baets and R. Mesiar. T-partitions. *Fuzzy Sets and Systems*, 97:211–223, 1998.

M. Demirci. On many-valued partitions and many-valued equivalence relations. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 11(2):235–253, 2003.

D. Dubois and H. Prade. A review of fuzzy set aggregation connectives. *Inform. Sci.*, 36:85–121, 1985.

M. G. Genton. Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2:299–312, 2001.

S. Gottwald. Fuzzy set theory with t-norms and Φ-operators. In A. Di Nola and A. G. S. Ventre, editors, *The Mathematics of Fuzzy Systems*, volume 88 of *Interdisciplinary Systems Research*, pages 143–195. Verlag TÜV Rheinland, Köln, 1986.

S. Gottwald. *Fuzzy Sets and Fuzzy Logic*. Vieweg, Braunschweig, 1993.

U. Höhle. Fuzzy equalities and indistinguishability. In *Proc. 1st European Congress on Fuzzy and Intelligent Technologies*, volume 1, pages 358–363, Aachen, 1993.

U. Höhle. The Poincaré paradox and non-classical logics. In D. Dubois, E. P. Klement, and H. Prade, editors, *Fuzzy Sets, Logics and Reasoning about Knowledge*, volume 15 of *Applied Logic Series*, pages 7–16. Kluwer Academic Publishers, Dordrecht, 1999.

F. Höppner and F. Klawonn. Improved fuzzy partitions for fuzzy regression models. *Internat. J. Approx. Reason.*, 32:85–102, 2003.

F. Höppner, F. Klawonn, and P. Eklund. Learning indistinguishability from data. *Soft Computing*, 6 (1):6–13, 2002.

J. Jacas. On the generators of T-indistinguishability operators. *Stochastica*, 12:49–63, 1988.

I. T. Jolliffe. *Principal Component Analysis*. Springer Verlag, New York, 1986.

E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*, volume 8 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, 2000.

R. Kruse, J. Gebhardt, and F. Klawonn. *Fuzzy-Systeme*. B. G. Teubner, Stuttgart, 1993.

R. Kruse, J. Gebhardt, and F. Klawonn. *Foundations of Fuzzy Systems*. John Wiley & Sons, New York, 1994.

C. H. Ling. Representation of associative functions. *Publ. Math. Debrecen*, 12:189–212, 1965.

B. Moser. On the t-transitivity of kernels. *Fuzzy Sets and Systems*, 157:1787–1796, 2006.

B. Moser. *A New Approach for Representing Control Surfaces by Fuzzy Rule Bases*. PhD thesis, Johannes Kepler Universität Linz, October 1995.

T. Muir. *A Treatise on the Theory of Determinants*. Dover, New York, 1960.

H. Poincaré. *La Science et l'Hypothése*. Flammarion, Paris, 1902.

H. Poincaré. *La Valeur de la Science*. Flammarion, Paris, 1904.

S. Saminger, R. Mesiar, and U. Bodenhofer. Domination of aggregation operators and preservation of transitivity. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 10(Suppl.):11–35, 2002.

B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, Cambridge, 2002.

B. Schölkopf, A. J. Smola, and K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 1998.

B. Schweizer and A. Sklar. Associative functions and statistical triangle inequalities. *Publ. Math. Debrecen*, 8:169–186, 1961.

B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North-Holland, Amsterdam, 1983.

M. Seeger. Covariance kernels from bayesian generative models. *Neural Information Processing Systems*, 14:905–912, 2002.

I. Kopriva T. M. Huang, V. Kecman. *Kernel Based Algorithms for Mining Huge Data Sets, Supervised, Semi-supervised, and Unsupervised Learning*. Springer-Verlag, Berlin, 2006.

E. Trillas and L. Valverde. An inquiry into indistinguishability operators. In H. J. Skala, S. Termini, and E. Trillas, editors, *Aspects of Vagueness*, pages 231–256. Reidel, Dordrecht, 1984.

E. Trillas, S. Cubillo, and E. Castiñeira. Menger and Ovchinnikov on indistinguishabilities revisited. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 7(3):213–218, 1999.

L. Valverde. On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3): 313–328, 1985.

A. M. Yaglom. Some classes of random fields in n-dimensional space, related to stationary random processes. *Theory of Probability and its Applications*, 2:273–320, 1957.

L. A. Zadeh. Similarity relations and fuzzy orderings. *Inform. Sci.*, 3:177–200, 1971.