Beyond Standard Classes of Generalized Joint Signal Representations of Arbitrary Variables: Mercer Kernel-Based Representations

Julien Gosme and Cédric Richard, *Member, IEEE*

Abstract—In this letter, we present a new approach for extending the scope of standard covariant signal representations by means of implicit nonlinear mappings applied to signals via Mercer kernels. One of the advantages of using such kernels is that we do not need to exhibit the underlying nonlinear maps to be able to compute signal representations. This gives increased computational efficiency. Finally, conditions on kernels to preserve covariance properties are finally discussed.

Index Terms—Nonlinear equations, signal representations, time-frequency analysis.

I. INTRODUCTION

RECENT research in nonstationary signal processing has introduced the concept of displacement operators and characterized the classes of all linear and bilinear representations covariant to a given displacement operator [1], [2]. Cohen [3], [4], Baraniuk [5], [6], and Sayeed and Jones [7] have proposed alternative approaches to the fore-mentioned covariance theory proposed by Hlawatsch *et al.*, see also [5], [8]–[10]. All these advances facilitate nonstationary signal processing which, compared to standard linear and bilinear packages of time-frequency and time-scale representations, can be well adapted to larger classes of phenomena. In this letter, we present a new approach to extend the scope of covariant signal representations of arbitrary variables by means of *implicit* nonlinear mappings applied to signals. One of the benefits of this generalization is its formalism based on Mercer kernels [11]. In the spirit of the *kernel trick* that has fully contributed to the success of kernel-based leaning methods such as SVM, see e.g., [12], this strategy gives a straightforward calculation of representations.

In the first part, we present some key points of the approach proposed by Hlawatsch *et al.* to design linear and bilinear covariant representations. In the second part, we propose to transform signals by means of nonlinear mappings in order to get new generalized joint signal representations of arbitrary variables. After a brief reminder of the interest of using some of the properties of reproducing kernel Hilbert spaces, we show how to implicitly apply nonlinear mappings to signals in an elegant

and computationally efficient way. Next, we express generalized joint signal representations in a form where reproducing kernels appear, yielding a wide choice of nonlinearities. A noteworthy property of this formulation is that the computational cost is independent of the underlying nonlinear mapping. Finally, conditions on kernels to preserve the covariance properties of generalized joint signal representations are discussed.

II. LINEAR AND BILINEAR COVARIANT REPRESENTATIONS

In recent years, general approaches have been proposed for designing joint representations in signal analysis that are covariant with respect to the action of specific operators, e.g., time shift $\mathbf{T}_\tau$, frequency shift $\mathbf{F}_\nu$ and scaling $\mathbf{C}_\sigma$ defined as $(\mathbf{T}_\tau x)(t) = x(t - \tau)$, $(\mathbf{F}_\nu x)(t) = e^{i2\pi\nu t}x(t)$ and $(\mathbf{C}_\sigma x)(t) = (1/\sqrt{\sigma})x(t/\sigma)$, respectively. Joint signal representations are based on pairs of operators, e.g., time and frequency shifts $\mathbf{F}_\nu \mathbf{T}_\tau$ and time shift and dilation $\mathbf{T}_\tau \mathbf{C}_\sigma$. A unified covariance theory of time-frequency analysis has been proposed in [1], [2]. It is based on operator pairs of the form $\mathbf{D}_\theta = \mathbf{A}_\alpha \mathbf{B}_\beta$ (possibly up to a phase factor), called *displacement operators*. The authors call *affine* displacement operator (respectively, *dual* displacement operator) the displacement operator generalizing the time shift/scaling pair of operators $\mathbf{T}_\tau \mathbf{C}_\sigma$ (respectively, the time-frequency shift pair $\mathbf{F}_\nu \mathbf{T}_\tau$) by the commutation relation they satisfy. With $\mathbf{T}_\tau$, $\mathbf{F}_\nu$, and $\mathbf{C}_\sigma$, this approach is equivalent to the method proposed in [5], [8] and discussed in [10], which uses an unitary warping applied to time-frequency and time-scale operator pairs. These operators differ from the hermitian operators used in [3], [4]. It should be noted that there are connections with [1], [2], as explained in [9].

In [1] and [2], covariance with respect to a displacement operator in joint signal representations is defined as follows.

Definition 1: Let $\{\mathbf{D}_\theta\}_{\theta \in (\mathcal{D}, \star)}$ be a family of displacement operators for a group $(\mathcal{D}, \star)$ on a Hilbert space $\mathcal{X}$. A linear θ-representation $L_x(\theta)$ is called covariant to $\mathbf{D}_\theta$ if it satisfies the following relation for all $x \in \mathcal{X}$:

$$L_{\mathbf{D}_{\theta_1} x}(\theta) = c\left(\theta \star \theta_1^{-1}, \theta_1\right) L_x\left(\theta \star \theta_1^{-1}\right), \quad \forall \theta, \theta_1 \in \mathcal{D} \quad (1)$$

where $c : \mathcal{D} \times \mathcal{D} \to \mathbb{C}$ is a continuous function called cocycle. A bilinear θ-representation $B_{x,y}(\theta)$ is called covariant to $\mathbf{D}_\theta$ if the following relation is verified for all $x, y \in \mathcal{X}$:

$$B_{\mathbf{D}_{\theta_1} x, \mathbf{D}_{\theta_1} y}(\theta) = B_{x,y}\left(\theta \star \theta_1^{-1}\right), \quad \forall \theta, \theta_1 \in \mathcal{D}. \quad (2)$$

The above definition means that a covariant representation reflects the action of the displacement operator on the analyzed

signal. The cocycle c in (1) results from the composition law $\mathbf{D}_{\theta_2}\mathbf{D}_{\theta_1} = c(\theta_1, \theta_2)\mathbf{D}_{\theta_1 \star \theta_2}$. It corresponds to a phase factor for dual operators. It is equal to 1 for the affine operators. See, e.g., [1], [2] and references therein for more details. Let us now characterize joint covariant linear and bilinear representations in signal analysis with the following theorems [1], [2].

Theorem 1: All linear representations covariant to a displacement operator $\mathbf{D}_\theta$ on a Hilbert space $\mathcal{X}$ are given by

$$L_x(\theta) = \langle x, \mathbf{D}_\theta \rangle_{\mathcal{X}} = \int x(t)(\mathbf{D}_\theta h)^*(t)dt \qquad (3)$$

for all $\theta \in \mathcal{D}$, where h is an arbitrary function in $\mathcal{X}$.

Theorem 2: All bilinear representations covariant to a displacement operator $\mathbf{D}_\theta$ on a Hilbert space $\mathcal{X}$ are given by the following expression:

$$\begin{aligned} B_{x,y}(\theta) &= \left\langle x, \mathbf{D}_\theta \mathbf{K} \mathbf{D}_\theta^{-1} y \right\rangle_{\mathcal{X}} \\ &= \int\int x(t_1)y^*(t_2) \\ &\quad \times \left[\mathbf{D}_\theta \mathbf{K} \mathbf{D}_\theta^{-1} \right]^* (t_1, t_2)dt_1 dt_2 \end{aligned} \qquad (4)$$

for all $\theta \in \mathcal{D}$, where $\mathbf{D}_\theta^{-1}$ is the adjoint operator of $\mathbf{D}_\theta$ and $\left[\mathbf{D}_\theta \mathbf{K} \mathbf{D}_\theta^{-1} \right]^* (t_1, t_2)$ denotes the kernel of $\mathbf{D}_\theta \mathbf{K} \mathbf{D}_\theta^{-1}$, with $\mathbf{K}$ an arbitrary linear operator on $\mathcal{X}$.

III. REPRODUCING KERNEL HILBERT SPACES AND MERCER'S THEOREM

A convenient way to model similarity between the elements of a space $\mathcal{X}$ is through an inner product function $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ which has the useful property that its value is maximal whenever its arguments are equal. Definitions (3) and (4) can then be viewed as similarity measures between x and $\mathbf{D}_\theta h$ and between x and $\mathbf{D}_\theta \mathbf{K} \mathbf{D}_\theta^{-1} y$, respectively. A possible way to extend the scope of linear and bilinear covariant distributions is to apply a nonlinear mapping function $\Psi : \mathcal{X} \longrightarrow \mathcal{H}$ to the arguments of the inner product function $\langle \cdot, \cdot \rangle_{\mathcal{X}}$. We shall now explore this basic idea, taking the theory of reproducing kernels as a starting point [11], [13] to propose a new formalism for linear and bilinear distributions and their nonlinear extensions. One of the very interesting properties about using kernels is that we do not need to exhibit the underlying nonlinear mapping Ψ in order to be able to design and compute distributions. The rest of this section reviews the main definitions and the properties of reproducing kernel Hilbert spaces.

Let $\mathcal{H}$ be a reproducing kernel Hilbert space consisting of mappings ψ from a signal space $\mathcal{X}$ to $\mathbb{R}$ and $< \cdot, \cdot >_{\mathcal{H}}$ the inner product defined on $\mathcal{H}$. As the Riesz representation theorem states, there is a unique function $\kappa(\cdot, y)$ of $\mathcal{H}$ for which

$$\psi(y) = \langle \psi, \kappa(\cdot, y) \rangle_{\mathcal{H}}, \quad \forall \psi \in \mathcal{H} \qquad (5)$$

for every fixed $y \in \mathcal{X}$. A proof of this may be found in [13]. Here $\kappa(\cdot, y)$ is the representer of evaluation at y and $\kappa(\cdot, \cdot)$ is the reproducing kernel associated with $\mathcal{H}$. In particular, $\{\kappa(\cdot, x) : x \in \mathcal{X}\}$ spans $\mathcal{H}$ and the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ has just to be defined on it. Denoting the function $\kappa(\cdot, x)$ by $\Psi(x)$, (5) implies

$$\kappa(x, y) = \langle \Psi(x), \Psi(y) \rangle_{\mathcal{H}} \qquad (6)$$

for all $x, y \in \mathcal{X}$. The kernel κ then evaluates the inner product of all the pairs of elements of $\mathcal{X}$ mapped to $\mathcal{H}$, without any explicit knowledge of either Ψ or $\mathcal{H}$. The key idea of the kernel technique used in this letter is to choose a kernel κ rather than a mapping Ψ for designing joint representations in signal analysis. Of course, not every function κ can serve as a kernel. According to the Hilbert-Schmidt Theory [14], any continuous hermitian-symmetric function κ can be expanded as follows:

$$\kappa(x, y) = \sum_{i=1}^{\infty} \lambda_i \psi_i(x) \psi_i^*(y) \qquad (7)$$

where λ_i and ψ_i are eigenvalues and eigenfunctions given by

$$\int \kappa(x, y)\psi_i(x)dx = \lambda_i \psi_i(y). \qquad (8)$$

A sufficient condition to ensure that κ is an inner product in a Hilbert space $\mathcal{H}$ is that all the λ_i's in (7) are positive. According to Mercer's theorem [11], this condition is achieved if and only if

$$\int\int \phi(x)\kappa(x, y)\phi^*(y)dxdy > 0 \qquad (9)$$

for all ϕ fulfilling $\int \phi(x)^2 dx < \infty$. From (7), it is straightforward to construct a map Ψ into a potentially infinite-dimensional space which satisfies (6). For instance, we may use $\Psi(x) = (\sqrt{\lambda_1}\psi_1(x), \sqrt{\lambda_2}\psi_2(x), \ldots)$. Examples of kernel satisfying Mercer's theorem are the complete polynomial kernel $\kappa(x, y) = [1 + \langle x, y \rangle_{\mathcal{X}}]^q$ and the qth degree polynomial kernel $\kappa(x, y) = [\langle x, y \rangle_{\mathcal{X}}]^q$, with $q \in \mathbb{N}^*$. Radial basis functions are also Mercer kernels that have received significant attention in the statistical and learning machine communities [12]. The two main ones are the radially gaussian kernel given by $\kappa(x, y) = \exp\left(-\|x - y\|_{\mathcal{X}}^2 / \beta_0^2\right)$ and the Laplace kernel defined as $\kappa(x, y) = \exp\left(-\|x - y\|_{\mathcal{X}} / \beta_0\right)$, with β_0 the kernel bandwidth. The above-mentioned examples are instances of the two most widely used families of kernels: projective kernels involving the inner product $\langle x, y \rangle_{\mathcal{X}}$, and radial kernels depending on the norm $\|x - y\|_{\mathcal{X}}$. Other examples may be found in Table I, where η_0 is a normalization constant and $\alpha_0 \in \mathbb{R}$, $\beta_0 \in \mathbb{R}_+^*$, $q \in \mathbb{N}^*$ are parameters to be set. See also, e.g., [15], [16]. Moreover, there exist simple rules for designing valid kernels on the basis of given Mercer kernels κ_1 and κ_2, e.g., $\kappa_1(x, y) + \kappa_2(x, y)$ and $\kappa_1(x, y) \cdot \kappa_2(x, y)$ are also Mercer kernels [16].

IV. KERNELIZED REPRESENTATIONS

We have suggested two equivalent ways of extending the scope of linear and bilinear signal representations.

1) *Explicitly* choose a mapping Ψ and apply it to the arguments of the inner product function $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ in (3) and (4).
1) Select a Mercer kernel κ which *implicitly* corresponds to a fixed mapping Ψ.

Though mathematically equivalent, kernels are often much easier to define and avoid computational problems through the implicit mapping. Therefore, we propose the following definition.

TABLE I
SOME EXAMPLES OF TYPICAL MERCER KERNELS

Projective kernels		Radial kernels	
Sigmoid	$\frac{1}{\eta_0}\tanh(\beta_0 <x,y>_{\mathcal{X}} - \alpha_0)$	Uniform	$\frac{1}{\eta_0}\mathbb{1}_{\|x-y\|_{\mathcal{X}}\leq\beta_0}$
qth degree polynomial	$[<x,y>_{\mathcal{X}}]^q$	Epanechnikov	$\frac{1}{\eta_0}(\beta_0^2 - \|x-y\|_{\mathcal{X}}^2)\,\mathbb{1}_{\|x-y\|_{\mathcal{X}}\leq\beta_0}$
Complete polynomial	$[1+<x,y>_{\mathcal{X}}]^q$	Cauchy	$\frac{1}{\eta_0}\frac{1}{1+\|x-y\|_{\mathcal{X}}^2/\beta_0^2}$

Definition 2: Let $\{\mathbf{D}_\theta\}_{\theta\in(\mathcal{D},\star)}$ be a family of displacement operators on a Hilbert space $\mathcal{X}$, where $(\mathcal{D},\star)$ is a group. Let κ be any Mercer kernel. Distributions given by

$$KL_x(\theta) = \kappa(x, \mathbf{D}_\theta h) \tag{10}$$

with h an arbitrary function of $\mathcal{X}$, are called *kernelized linear distributions*.

In the following, we shall consider only projective kernels and radial kernels. In both cases, it is worth noting that the adjoint operator $\mathbf{D}_\theta^\dagger$ with respect to the kernel κ, i.e., the operator such that $\kappa(\mathbf{D}_\theta x, y) = \kappa\left(x, \mathbf{D}_\theta^\dagger y\right)$, corresponds to the adjoint $\mathbf{D}_\theta^{-1}$ related to the inner product $\langle\cdot,\cdot\rangle_{\mathcal{X}}$. In fact, it can be written

$$\kappa(\mathbf{D}_\theta x, y) = g(\langle \mathbf{D}_\theta x, y\rangle_{\mathcal{X}})$$
$$= g\left(\langle x, \mathbf{D}_\theta^{-1} y\rangle_{\mathcal{X}}\right) = \kappa\left(x, \mathbf{D}_\theta^\dagger y\right) \tag{11}$$

in the case of projective kernels. For radial kernels, it suffices to note that $\|x - y\|_{\mathcal{X}}^2 = \|x\|_{\mathcal{X}}^2 + \|y\|_{\mathcal{X}}^2 - 2\text{Re}(\langle x, y\rangle_{\mathcal{X}})$ and apply (11). At this point, one can ask whether kernelized linear distributions are covariant to displacement operators.

We know that $KL_{\mathbf{D}_{\theta_1}x}(\theta) = \kappa\left(x, \mathbf{D}_{\theta_1}^\dagger \mathbf{D}_\theta h\right)$. Accordingly, every kernelized linear distribution derived from a family of displacement operators is covariant if it satisfies the following sufficient condition:

$$\mathbf{D}_{\theta_1}^\dagger \mathbf{D}_\theta = \mathbf{D}_{\theta\star\theta_1^{-1}}, \quad \forall \theta, \theta_1 \in \mathcal{D}. \tag{12}$$

This condition ensures that $KL_{\mathbf{D}_{\theta_1}x}(\theta) = KL_x\left(\theta\star\theta_1^{-1}\right)$, which means that the representation of the displaced signal $\mathbf{D}_{\theta_1}x$ equals the displaced representation of the original signal x. If we take a closer look at (12), substituting $\theta = \theta_0$ and $\theta_1 = \theta$, where θ_0 denotes the identity element of the group $(\mathcal{D},\star)$, we obtain: $\mathbf{D}_\theta^\dagger = \mathbf{D}_{\theta^{-1}}$. This result implies that the cocycle c of the displacement operator $\mathbf{D}_\theta$ equals 1. Therefore displacement operators are constrained by the above condition to be *affine*. We can formulate the following proposition.

Proposition 1: Every kernelized linear distribution derived from a family of affine operators is covariant.

As an example, consider the following displacement operator $\mathbf{R}_\theta = \mathbf{T}_\tau \mathbf{C}_\sigma$ with $\theta = (\sigma, \tau)$ and the kernelized linear distributions

$$KL_x^{(1)}(\theta) = \exp\left(-\frac{\|x - \mathbf{R}_\theta h\|_{\mathcal{X}}^2}{\rho^2}\right) \tag{13}$$

$$KL_x^{(2)}(\theta) = [1 + \langle x, \mathbf{R}_\theta h\rangle_{\mathcal{X}}]^q. \tag{14}$$

These distributions involve a radial kernel and a projective kernel, respectively, which implies that $\mathbf{R}_\theta^\dagger = \mathbf{R}_\theta^{-1}$ in both cases. In addition, $\mathbf{R}_\theta$ is an affine displacement operator [1],

[2]. Therefore, we can conclude from Proposition 1 that the above distributions are covariant, meaning that

$$KL_{\mathbf{R}_{\sigma',\tau'}x}(\sigma, \tau) = KL_x\left(\frac{\sigma}{\sigma'}, \frac{\tau - \tau'}{\sigma'}\right). \tag{15}$$

Let us now discuss the case of dual displacement operators, which differ from affine displacement operators by the cocycle $c \neq 1$. A natural approach to obtain a similar result as in Proposition 1 might be to modify (12) as follows:

$$\mathbf{D}_{\theta_1}^\dagger \mathbf{D}_\theta = c\left(\theta\star\theta_1^{-1}, \theta_1\right)\mathbf{D}_{\theta\star\theta_1^{-1}}, \quad \forall \theta, \theta_1 \in \mathcal{D}. \tag{16}$$

The above relation is compatible with the composition law of displacement operators. However the condition (16) leads to $KL_{\mathbf{D}_{\theta_1}x}(\theta) = \kappa\left(x, c\left(\theta\star\theta_1^{-1}, \theta_1\right)\mathbf{D}_{\theta\star\theta_1^{-1}}h\right)$, showing that it cannot ensure covariance of $KL_x(\theta)$ without an additional property that would make it possible to remove the cocycle parameter $c\left(\theta\star\theta_1^{-1}, \theta_1\right)$ out of κ. For example, the covariance property is valid up to the phase factor $c\left(\theta\star\theta_1^{-1}, \theta_1\right)^q$ for the following kernelized linear distributions:

$$KL_x^{(3)}(\theta) = [\langle x, \mathbf{S}_\theta h\rangle_{\mathcal{X}}]^q, \quad q \in \mathbb{N}^* \tag{17}$$

where $\mathbf{S}_\theta = \mathbf{F}_\nu \mathbf{T}_\tau$ denotes the time-frequency shift operator. Its cocycle is $c\left((\tau_1, \nu_1), (\tau_2, \nu_2)\right) = \exp\left(-j2\pi\nu_1\tau_2\right)$. For $q = 1$, it is worth noting that $KL_x^{(3)}(\theta)$ corresponds to linear distributions defined in (3). Remember that they satisfy the covariance property (1).

Next, we consider kernelized bilinear distributions in a similar way as we did for linear ones.

Definition 3: Let $\{\mathbf{D}_\theta\}_{\theta\in(\mathcal{D},\star)}$ be a family of displacement operators on a Hilbert space $\mathcal{X}$, where $(\mathcal{D},\star)$ is a group. Let κ be any Mercer kernel. Distributions given by

$$KB_{x,y}(\theta) = \kappa\left(x, \mathbf{D}_\theta \mathbf{K}\mathbf{D}_\theta^\dagger y\right) \tag{18}$$

where $\mathbf{K}$ is an arbitrary linear operator on $\mathcal{X}$ are called *kernelized bilinear distributions*.

We then study the covariance of kernelized bilinear distributions derived from projective and radial kernels. Using the unitarity of $\mathbf{D}_\theta$, as well as the composition law $\mathbf{D}_{\theta_2}\mathbf{D}_{\theta_1} = c(\theta_1, \theta_2)\mathbf{D}_{\theta_1\star\theta_2}$ and $\mathbf{D}_\theta^\dagger = c^*(\theta, \theta^{-1})\mathbf{D}_{\theta^{-1}}$, one can verify that

$$KB_{\mathbf{D}_{\theta_1}x, \mathbf{D}_{\theta_1}y}(\theta) = \kappa\left(\mathbf{D}_{\theta_1}x, \mathbf{D}_\theta \mathbf{K}\mathbf{D}_\theta^\dagger \mathbf{D}_{\theta_1}y\right)$$
$$= \kappa\left(x, \mathbf{D}_{\theta_1}^\dagger \mathbf{D}_\theta \mathbf{K}\mathbf{D}_\theta^\dagger \mathbf{D}_{\theta_1}y\right)$$
$$= \kappa\left(x, \mathbf{D}_{\theta_1}^\dagger \mathbf{D}_\theta \mathbf{K}\left[\mathbf{D}_{\theta_1}^\dagger \mathbf{D}_\theta\right]^\dagger y\right)$$
$$= \kappa\left(x, \mathbf{D}_{\theta\star\theta_1^{-1}}\mathbf{K}[\mathbf{D}_{\theta\star\theta_1^{-1}}]^\dagger y\right).$$
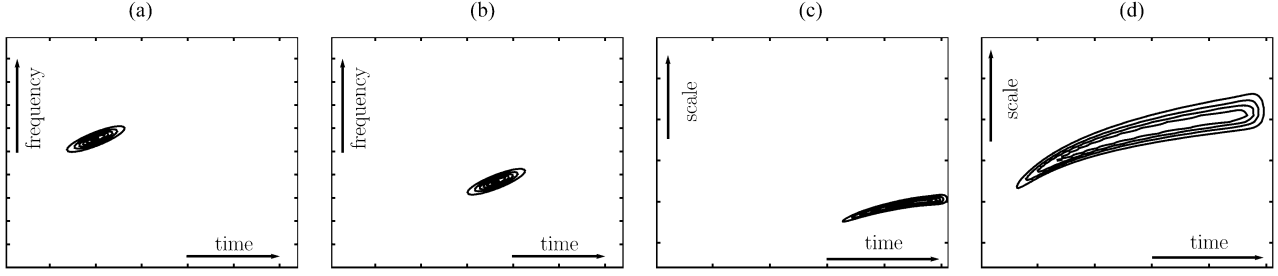
Consequently, we can put forward the following proposition.

Fig. 1. Kernelized smoothed pseudo-Wigner distributions derived from (a)–(b) a sigmoid Mercer kernel ($\alpha_0 = 15$, $\beta_0 = 0.15$, $\eta_0 = 1$) and (c)–(d) a complete polynomial Mercer kernel ($q = 5$). These figures illustrate covariance of kernelized bilinear distributions with respect to the action of the pairs of operators (a)–(b) $\mathbf{F}_\nu \mathbf{T}_\tau$ and (c)–(d) $\mathbf{T}_\tau \mathbf{C}_\sigma$ on analyzed signals.

Proposition 2: Let $\{\mathbf{D}_\theta\}_{\theta \in (\mathcal{D}, \star)}$ be a family of displacement operators on a Hilbert space $\mathcal{X}$. Kernelized bilinear distributions based on radial or projective kernels are covariant to $\mathbf{D}_\theta$ for all $x, y \in \mathcal{X}$

$$KB_{\mathbf{D}_{\theta_1} x, \mathbf{D}_{\theta_1} y}(\theta) = KB_{x,y}\left(\theta \star \theta_1^{-1}\right), \quad \forall\, \theta, \theta_1 \in \mathcal{D}. \quad (19)$$

Compared to the covariance property of kernelized linear distributions, we note that the above proposition involves both dual and affine displacement operators. As an example, for $\mathbf{S}_{\tau,\nu} = \mathbf{F}_\nu \mathbf{T}_\tau$ and $\mathbf{R}_{\sigma,\tau} = \mathbf{T}_\tau \mathbf{C}_\sigma$, it means that

$$KB_{\mathbf{S}_{\tau',\nu'} x, \mathbf{S}_{\tau',\nu'} y}(\tau, \nu) = KB_{x,y}(\tau - \tau', \nu - \nu') \quad (20)$$

$$KB_{\mathbf{R}_{\sigma',\tau'} x, \mathbf{R}_{\sigma',\tau'} y}(\sigma, \tau) = KB_{x,y}\left(\frac{\sigma}{\sigma'}, \frac{\tau - \tau'}{\sigma'}\right). \quad (21)$$

Fig. 1 illustrates the covariance property of kernelized bilinear distributions.

V. CONCLUSION AND FUTURE DIRECTIONS

We have proposed a new approach to broaden the range of tools of joint signal representations of arbitrary variables. This generalization is significant because it not only extends the scope of traditional time-frequency and time-scale representations, but also presents a new formalism based on Mercer kernels which allows a straightforward calculation of representations to be made. We have also focused on the property of covariance with respect to the practically important transformations of time shifts, frequency shifts and scale changes. In particular, we have shown that kernelized bilinear representations derived from two essential classes of kernels satisfy the covariance property with respect to the action of displacement operators.

The present letter is a first study of kernelized joint signal representations, and further extensions of this concept are possible. Future work should be focused on developing efficient numerical schemes, studying the properties of representations inherited from kernels or bringing together our approach and existing concepts such as polynomial and higher order representations [17]–[19]. In addition, possible applications include extensions of statistical methods developed for arbitrary joint signal representations to handle a wider variety of problems in nonstationary statistical signal processing [20].

REFERENCES

[1] F. Hlawatsch and G. Taubӧck, "The covariance theory of time-frequency analysis," in *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. Oxford, U.K.: Elsevier, 2003, pp. 102–113.

[2] F. Hlawatsch, G. Taubӧck, and T. Twaroch, *Wavelets and Signal Processing*. Boston, MA: Birkhauser, 2002, ch. Covariant Time-Frequency Analysis.

[3] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[4] ——, "A general approach for obtaining joint representations in signal analysis—Part 1: Characteristic function operator method," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1080–1090, May 1996.

[5] R. G. Baraniuk, "Covariant time-frequency representations through unitary equivalence," *IEEE Signal Processing Lett.*, vol. 3, no. 3, pp. 79–81, Mar. 1996.

[6] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. Signal Processing*, vol. 43, no. 10, pp. 2269–2282, Oct. 1995.

[7] A. M. Sayeed and D. L. Jones, "Integral transforms covariant to unitary operators and their applications for joint signal representations," *IEEE Trans. Signal Processing*, vol. 44, no. 6, pp. 1365–1377, Jun. 1996.

[8] R. G. Baraniuk, "Marginals vs. covariance in joint distribution theory," in *Proc. IEEE ICASSP'95*, 1995, pp. 1021–1024.

[9] A. M. Sayeed, "On the equivalence of the operator and kernel method for joint distributions of arbitrary variables," *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 1067–1070, Apr. 1997.

[10] A. M. Sayeed and D. L. Jones, "Equivalence of generalized joint signal representations of arbitrary variables," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 2959–2970, Dec. 1996.

[11] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," in *Philos. Trans. Roy. Soc. London A*, vol. 209, 1909, pp. 415–446.

[12] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer, 1995.

[13] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.

[14] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. New York: Interscience, 1953.

[15] M. Genton, "Classes of kernels for machine learning: A statistics perspective," *J. Mach. Learn. Res.*, pp. 299–312, 2001.

[16] R. Herbrich, *Learning Kernel Classifiers. Theory and Algorithms*. Cambridge, MA: The MIT Press, 2001.

[17] o B. Boashash and P. O'Shea, "Polynomial wigner-ville distributions and their relationship to time-varying higher order spectra," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 216–220, Jan. 1994.

[18] B. Boashash and B. Ristić, *Higher Order Statistical Signal Processing*. Cheshire, : Longman Cheshire, 1993, ch. Polynomial WVD's and time-varying polyspectra.

[19] L. Stanković, "L-class of time-frequency distributions," *IEEE Signal Processing Lett.*, vol. 3, no. 1, pp. 22–25, Jan. 1996.

[20] A. M. Sayeed and D. L. Jones, "Optimal quadratic detection and estimation using generalized joint signal representations," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3031–3043, Dec. 1996.