COMPLEX KERNELS FOR PROPER COMPLEX-VALUED SIGNALS: A REVIEW

Rafael Boloix-Tortosa,F. Javier Payán-Somet, Eva Arias-de-Reyna and Juan José Murillo-Fuentes

Department of Signal Theory and Communications, University of Seville, Spain. e-mail: rboloix@us.es

ABSTRACT

In this paper we investigate the conditions that complex kernels must satisfy for proper complex-valued signals. We study the structure that complex kernels for proper complex-valued signals must have. Also, we demonstrate that complex kernels that have been previously proposed and used in adaptive filtering of complex-valued signals assume that those signals are proper, i.e, they are not correlated with their complex conjugate. We provide an example of how a complex-valued kernel suitable for a particular model is designed, with a procedure that could help in other designs. The experiments included show the good behavior of the proposed kernel in the task of nonlinear channel equalization.

***Index Terms*—** Gaussian processes, regression, proper complex processes, kernel methods.

1. INTRODUCTION

Complex-valued signals model a vast range of nowadays systems in science and engineering. The nonlinear processing of complex-valued signals has been recently addressed using reproducing kernel Hilbert spaces (RKHS) [5]. Some complex kernels have been lately proposed for classification [6], kernel principal component analysis and regression [1–4]. Regarding regression, in [2] the authors propose a complex-valued kernel based on the results in [6]. The same kernel is adopted in [1], and its convergence behavior is studied in [7]. As discussed later in this paper, the resulting approach involves properness of the complex-valued signals, i.e., they are uncorrelated with their complex conjugate. Besides, the kernel used is neither stationary nor isotropic, and it may suffer from numerical problems. In [4] the authors review the kernel design to improve the previous solution with a kernel they denote as independent. The resulting kernel yields also proper complex-valued outputs. The kernel is stationary, but again it is not isotropic in the complex-valued input space, as the real and imaginary parts of the input are split and fed to different real valued kernels. Hence, these previous works do not report results for isotropic and stationary kernels that may better model the underlying physics of some systems. Also, the structure of the kernel remains more rigid than needed. These drawbacks make these solutions not powerful enough to learn a wide range of systems.

We study in this paper the conditions that complex positive definite kernels must satisfy for proper complex-valued signals, to improve previous solutions. The starting point is the complex nonlinear regression problem $y = f(\mathbf{x}) + \epsilon$, where the output, $y \in \mathbb{C}$, the input vector, $\mathbf{x} \in \mathbb{C}^d$, and the unknown nonlinear latent function, $f \in \mathbb{C}$, are complex valued, and the error, ϵ is modeled as additive zero-mean complex Gaussian noise. We analyze the structure of the covariance matrix of the complex-valued vector $\mathbf{y} = [y(\mathbf{x}(1)), ..., y(\mathbf{x}(n))]^\top$ when it is proper. The covariance function or kernel must produce the entries of that covariance matrix. We will show that the real part of the kernel is given by the covariance of the real part plus the covariance of the imaginary part of the outputs, while the imaginary part of the kernel describes the cross-covariance between real and imaginary parts of the outputs. We prove that the real and imaginary parts of the kernel can be designed with different features. But we conclude that the imaginary part, in addition to be skew-symmetric, must be constructed to ensure the whole covariance to be semi-definite positive, i.e. a reproducing kernel or covariance matrix [8]. We also pay attention to the modeling of physical systems. As an example, we propose the construction of a complex kernel that explains a positive and negative correlation of the real part of the output with the imaginary part for a positive and negative delay, canceling at the origin. Also, in order to measure similarity between inputs our example makes use of the absolute value of the complex difference between inputs. Therefore, the kernel is isotropic and stationary. We resort to the convolution approach [9, 10] to ensure that the produced kernel is a valid covariance function. The procedure followed could help in other complex-valued kernel designs for proper complex-valued outputs.

2. COMPLEX COVARIANCE FUNCTIONS

Consider a zero-mean complex vector $\mathbf{y} = \mathbf{y}_\mathrm{r} + \mathrm{j}\mathbf{y}_\mathrm{j} \in \mathbb{C}^n$, with $\mathbf{y}_\mathrm{r}$ its real part and $\mathbf{y}_\mathrm{j}$ its imaginary part. The covariance matrix $\mathbf{K} = \mathbb{E}\left[\mathbf{y}\mathbf{y}^\mathrm{H}\right]$ is [11]:

$$\mathbf{K} = \mathbf{K}_\mathrm{rr} + \mathbf{K}_\mathrm{jj} + \mathrm{j}\left(\mathbf{K}_\mathrm{jr} - \mathbf{K}_\mathrm{rj}\right), \quad (1)$$

Thanks to Spanish government (Ministerio de Educación y Ciencia TEC2012-38800-C03-02) and European Union (FEDER) for funding.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

where $\mathbf{K}_{\mathrm{rr}}$ and $\mathbf{K}_{\mathrm{jj}} \in \mathbb{R}_{+}^{n\times n}$ are the covariance matrices of real and imaginary parts of $\mathbf{y}$, respectively, and $\mathbf{K}_{\mathrm{rj}} = \mathbb{E}\left[\mathbf{y}_{\mathrm{r}}\mathbf{y}_{\mathrm{j}}^{\top}\right] = \mathbf{K}_{\mathrm{jr}}^{\top} \in \mathbb{R}^{n\times n}$ is the cross-covariance matrix of real and imaginary parts. From (1) it is clear that a kernel or covariance function, $k(\mathbf{x},\mathbf{x}')$, which gives rise to $\mathbf{K}$ can be composed by three real-valued covariance functions, $k_{\mathrm{rr}}(\mathbf{x},\mathbf{x}')$, $k_{\mathrm{jj}}(\mathbf{x},\mathbf{x}')$ and $k_{\mathrm{rj}}(\mathbf{x},\mathbf{x}')$, that give rise to the three real covariance matrices $\mathbf{K}_{\mathrm{rr}}$, $\mathbf{K}_{\mathrm{jj}}$ and $\mathbf{K}_{\mathrm{rj}}$, respectively:

$$k(\mathbf{x},\mathbf{x}') = k_{\mathrm{rr}}(\mathbf{x},\mathbf{x}') + k_{\mathrm{jj}}(\mathbf{x},\mathbf{x}') + \mathrm{j}\left(k_{\mathrm{rj}}(\mathbf{x}',\mathbf{x}) - k_{\mathrm{rj}}(\mathbf{x},\mathbf{x}')\right). \quad (2)$$

For the particular case of a proper complex Gaussian vector, the pseudo-covariance matrix is zero, $\mathbb{E}\left[\mathbf{y}\mathbf{y}^{\top}\right] = \mathbf{0}$, and the following relations hold: $\mathbf{K}_{\mathrm{rr}} = \mathbf{K}_{\mathrm{jj}}$ and $\mathbf{K}_{\mathrm{jr}} = \mathbf{K}_{\mathrm{rj}}^{\top} = -\mathbf{K}_{\mathrm{rj}}$ [11]. Therefore, in such a case functions $k_{\mathrm{rr}}(\mathbf{x},\mathbf{x}') = k_{\mathrm{jj}}(\mathbf{x},\mathbf{x}')$, and $k_{\mathrm{rj}}(\mathbf{x},\mathbf{x}')$ must yield either a null or a skew-symmetric cross-covariance matrix $\mathbf{K}_{\mathrm{rj}}$. The covariance matrix (1) for a proper vector yields

$$\mathbf{K} = \mathbf{K}_{\mathrm{r}} + \mathrm{j}\mathbf{K}_{\mathrm{j}}, \quad (3)$$

where $\mathbf{K}_{\mathrm{r}} = 2\mathbf{K}_{\mathrm{rr}}$ and $\mathbf{K}_{\mathrm{j}} = -2\mathbf{K}_{\mathrm{rj}}$. Following the guidelines in [12] we may conclude that a proposed kernel is a valid reproducing kernel if (3) is a covariance matrix. Hence, it must be a Hermitian positive semi-definite matrix [8]. Also, $\mathbf{K}_{\mathrm{rr}}$ must be a symmetric and positive semi-definite matrix, since the marginals of the joint probability functions of the real and imaginary parts must be also covariance matrices. It follows that the condition $\mathbf{v}^{\mathrm{H}}\mathbf{K}\mathbf{v} \geq 0$ for any $\mathbf{v} \in \mathbb{C}^n$ yields,

$$\mathbf{v}^{\mathrm{H}}\mathbf{K}\mathbf{v} = \mathbf{v}_{\mathrm{r}}^{\top}\mathbf{K}_{\mathrm{r}}\mathbf{v}_{\mathrm{r}} + \mathbf{v}_{\mathrm{j}}^{\top}\mathbf{K}_{\mathrm{r}}\mathbf{v}_{\mathrm{j}} - 2\mathbf{v}_{\mathrm{r}}^{\top}\mathbf{K}_{\mathrm{j}}\mathbf{v}_{\mathrm{j}} \geq 0 \quad (4)$$

where the first terms to the right of the equality are greater or equal to zero, since $\mathbf{K}_{\mathrm{r}}$ is positive semi-definite and we used $\mathbf{v}_{\mathrm{r}}^{\top}\mathbf{K}_{\mathrm{j}}\mathbf{v}_{\mathrm{r}} = \mathbf{v}_{\mathrm{j}}^{\top}\mathbf{K}_{\mathrm{j}}\mathbf{v}_{\mathrm{j}} = 0$ since $\mathbf{K}_{\mathrm{j}}$ is skew-symmetric.

3. PREVIOUSLY PROPOSED COMPLEX KERNELS

In [2, 6] a complex-valued Gaussian kernel is proposed as an extension of the real Gaussian kernel:

$$k_{\mathbb{C}}(\mathbf{x},\mathbf{x}') = \exp\left(-(\mathbf{x}-\mathbf{x}'^{*})^{\top}(\mathbf{x}-\mathbf{x}'^{*})/\gamma\right). \quad (5)$$

Further operations allow us to show that this kernel gives rise to a covariance matrix for the proper case as in (3):

$$\begin{aligned} k_{\mathbb{C}}(\mathbf{x},\mathbf{x}') &= \exp\left(-(|\mathbf{x}_{\mathrm{r}}-\mathbf{x}'_{\mathrm{r}}|^2 - |\mathbf{x}_{\mathrm{j}}+\mathbf{x}'_{\mathrm{j}}|^2)/\gamma\right) \\ &\cdot \exp\left(-2\mathrm{j}(\mathbf{x}_{\mathrm{r}}-\mathbf{x}'_{\mathrm{r}})^{\top}(\mathbf{x}_{\mathrm{j}}+\mathbf{x}'_{\mathrm{j}})/\gamma\right) \\ &= \exp\left(-(|\mathbf{x}_{\mathrm{r}}-\mathbf{x}'_{\mathrm{r}}|^2/\gamma\right)\exp\left(|\mathbf{x}_{\mathrm{j}}+\mathbf{x}'_{\mathrm{j}}|^2/\gamma\right) \\ &\cdot \left(\cos(2(\mathbf{x}_{\mathrm{r}}-\mathbf{x}'_{\mathrm{r}})^{\top}(\mathbf{x}_{\mathrm{j}}+\mathbf{x}'_{\mathrm{j}})/\gamma)\right. \\ &\left.-\mathrm{j}\sin(2(\mathbf{x}_{\mathrm{r}}-\mathbf{x}'_{\mathrm{r}})^{\top}(\mathbf{x}_{\mathrm{j}}+\mathbf{x}'_{\mathrm{j}})/\gamma)\right), \end{aligned} \quad (6)$$

where $\mathbf{x} = \mathbf{x}_{\mathrm{r}} + \mathrm{j}\mathbf{x}_{\mathrm{j}}$, $\mathbf{x}' = \mathbf{x}'_{\mathrm{r}} + \mathrm{j}\mathbf{x}'_{\mathrm{j}}$, and $|\cdot|$ is the ℓ^2-norm. Notice that the kernel in (6) corresponds to assuming the output is proper with non-null skew-symmetric cross-covariance matrix. This kernel measures similarities between real parts while measures dissimilarity between imaginary ones and it is not stationary. It also has an oscillatory behavior. In addition, the exponent in the kernel may easily grow large and positive. These characteristics may not be very useful when modeling the underlying physics of many systems.

In [4], the authors propose the so-called independent kernel to improve the previous one:

$$k_{ind}(\mathbf{x},\mathbf{x}') = \kappa_{\mathbb{R}}\left(\mathbf{x}_{\mathrm{r}},\mathbf{x}'_{\mathrm{r}}\right) + \kappa_{\mathbb{R}}\left(\mathbf{x}_{\mathrm{j}},\mathbf{x}'_{\mathrm{j}}\right) + \mathrm{j}\left(\kappa_{\mathbb{R}}\left(\mathbf{x}_{\mathrm{r}},\mathbf{x}'_{\mathrm{j}}\right) - \kappa_{\mathbb{R}}\left(\mathbf{x}_{\mathrm{j}},\mathbf{x}'_{\mathrm{r}}\right)\right), \quad (7)$$

where $\kappa_{\mathbb{R}}$ is a real kernel of real inputs. Again, this kernel follows the structure for the proper case. The kernel in (7) solves the measure of similarity between inputs by checking for the real and imaginary parts independently. At this point, we lose the intuition about the complex nature of inputs. The kernel assumes again that the output process to model is proper complex valued, where the imaginary part is non-null and skew-symmetric. One of the main drawbacks of this kernel is that it is not isotropic, due to the way real and imaginary parts of the inputs have been split in the kernel. For example, if a real Gaussian kernel is used in (7) as proposed in [4], $\kappa_{\mathbb{R}} = \alpha\exp(-|\mathbf{x}-\mathbf{x}'|^2/\beta)$ for some hyperparameters α and β, whenever two inputs are distant enough tthe kernel vanishes except for several cases, for example similar imaginary parts. For any $\mathbf{x}_{\mathrm{j}} = \mathbf{x}'_{\mathrm{j}}$, the covariance yields the maximum value for $\kappa_{\mathbb{R}}$, $k_{ind}(\mathbf{x},\mathbf{x}') = \alpha$. In the imaginary part of the kernel we have a similar behavior.

4. A COMPLEX-VALUED KERNEL FOR PROPER COMPLEX-VALUED SIGNALS

If a dependence exists between the real and the imaginary parts of the output, it should be included in the kernel design. However, in many problems it is not easy to know the characteristics of such dependence. For that reason the complex Gaussian kernel in (5) or the independent kernel in (7) might not be the best choice for some models. The proposed kernel structure in (2) much better adapts to the problem at hand. The real part of the kernel can be designed with a structure different from the structure of the imaginary part, adapting the design to the problem to model. Also, if the cross-covariance is known to be null or negligible the imaginary part of the kernel can be set to zero.

In the following we provide an example of how a complex-valued kernel suitable for a particular model is produced. Here we propose just one kernel, but the procedure followed could help in other designs. The design should meet condition (4), that encodes intuitive facts such as the maximum absolute value of the cross-covariance being lower or equal to the maximum absolute value of the covariance. On the other hand, the kernel must be able to explain the dependencies between real and imaginary parts of the output, if known, while $\mathbf{K}_{\mathrm{rj}}$ is restricted to be skew-symmetric.

We propose to model a system where the real and imaginary parts of the outputs are correlated for delayed points with delay $\mu \in \mathbb{C}^d$. Since the covariance matrix must be skew symmetric, the correlation will be positive (or negative) for a delay of μ and negative (or positive) for a delay of $-\mu$. We must ensure that the kernel corresponds to a covariance matrix. To this end we bring here the convolution approach [9, 10]. We model the output of the process as the filtering of two independent white Gaussian noises and compute the kernels from the filter responses. This way we meet the condition of being a covariance matrix. The filters are designed to model the proposed system, fulfilling the condition of a skew symmetric imaginary part of the kernel. Also, we propose to use the inner product of the inputs, $\mathbf{x}^{\mathrm{H}}\mathbf{x}$, a simple metric in complex numbers, to cope with isotropy.

The convolution process is sketched in Fig. 1. Consider two independent, real, stationary, Gaussian white noise processes $S_{\mathrm{r}}(\mathbf{x})$ and $S_{\mathrm{j}}(\mathbf{x})$, where $\mathbf{x} \in \mathbb{C}^d$, producing an output $Y(\mathbf{x}) = U(\mathbf{x}) + W(\mathbf{x})$, where $W(\mathbf{x})$ is a stationary Gaussian white noise, and $U(\mathbf{x})$ is defined by the sum of convolutions

$$\begin{aligned} U(\mathbf{x}) &= (h_{\mathrm{r}}(\mathbf{x}) + \mathrm{j}h_{\mathrm{rj}}(\mathbf{x})) \star S_{\mathrm{r}}(\mathbf{x}) \\ &\quad + (h_{\mathrm{r}}(\mathbf{x}) + \mathrm{j}h_{\mathrm{jr}}(\mathbf{x})) \star S_{\mathrm{j}}(\mathbf{x}) = \sum_{m=1}^{4} h_m(\mathbf{x}) \star S_m(\mathbf{x}), \end{aligned} \tag{8}$$

where we have used the following notation: $h_1(\mathbf{x}) = h_{\mathrm{r}}(\mathbf{x})$, $h_2(\mathbf{x}) = \mathrm{j}h_{\mathrm{rj}}(\mathbf{x})$, $h_3(\mathbf{x}) = h_{\mathrm{r}}(\mathbf{x})$, $h_4(\mathbf{x}) = \mathrm{j}h_{\mathrm{jr}}(\mathbf{x})$, $S_1(\mathbf{x}) = S_2(\mathbf{x}) = S_{\mathrm{r}}(\mathbf{x})$, and $S_3(\mathbf{x}) = S_4(\mathbf{x}) = S_{\mathrm{j}}(\mathbf{x})$. The covariance of $Y(\mathbf{x})$ is derived as $\mathbf{C}(\mathbf{x}_a, \mathbf{x}_b) = \mathbf{C}_U(\mathbf{x}_a, \mathbf{x}_b) + \sigma_W^2 \delta_{ab}$, where σ_W^2 is the variance of $W(\mathbf{x})$, and

$$\begin{aligned} \mathbf{C}_U(\mathbf{x}_a, \mathbf{x}_b) &= \mathbb{E}\left[U(\mathbf{x}_a)U^*(\mathbf{x}_b)\right] \\ &= \mathbb{E}\Bigg[\sum_{m=1}^{4} \int_{\mathbb{C}^d} h_m(\boldsymbol{\alpha}) S_m(\mathbf{x}_a - \boldsymbol{\alpha}) \mathrm{d}^d\boldsymbol{\alpha} \\ &\quad \cdot \sum_{n=1}^{4} \int_{\mathbb{C}^d} h_n^*(\boldsymbol{\beta}) S_n(\mathbf{x}_b - \boldsymbol{\beta}) \mathrm{d}^d\boldsymbol{\beta}\Bigg] \\ &= \sum_{m=1}^{4}\sum_{n=1}^{4} \Bigg\{ \int_{\mathbb{C}^d}\int_{\mathbb{C}^d} h_m(\boldsymbol{\alpha}) h_n^*(\boldsymbol{\beta}) \\ &\quad \cdot \mathbb{E}\left[S_m(\mathbf{x}_a - \boldsymbol{\alpha}) S_n(\mathbf{x}_b - \boldsymbol{\beta})\right] \mathrm{d}^d\boldsymbol{\alpha}\mathrm{d}^d\boldsymbol{\beta} \Bigg\}. \end{aligned} \tag{9}$$

Since $S_1(\mathbf{x}) = S_2(\mathbf{x}) = S_{\mathrm{r}}(\mathbf{x})$, and $S_3(\mathbf{x}) = S_4(\mathbf{x}) = S_{\mathrm{j}}(\mathbf{x})$, processes $S_m(\mathbf{x}_a - \boldsymbol{\alpha})$ and $S_n(\mathbf{x}_b - \boldsymbol{\beta})$ covary only if $m, n \in \{1, 2\}$ or $m, n \in \{3, 4\}$, and $(\mathbf{x}_a - \boldsymbol{\alpha}) = (\mathbf{x}_b - \boldsymbol{\beta})$. In such cases, $\mathbb{E}\left[S_m(\mathbf{x}_a - \boldsymbol{\alpha}) S_n(\mathbf{x}_b - \boldsymbol{\beta})\right] = \delta(\boldsymbol{\alpha} - (\mathbf{x}_a - \mathbf{x}_b + \boldsymbol{\beta}))$,

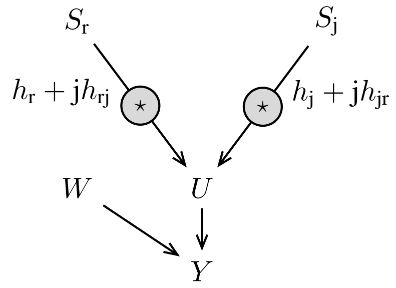

Fig. 1. Convolution system model to design a kernel for proper complex-valued signals.

and the integrals in (9) yield

$$\begin{aligned} &\int_{\mathbb{C}^d}\int_{\mathbb{C}^d} h_m(\boldsymbol{\alpha}) h_n^*(\boldsymbol{\beta}) \delta(\boldsymbol{\alpha} - (\mathbf{x}_a - \mathbf{x}_b + \boldsymbol{\beta})) \mathrm{d}^d\boldsymbol{\alpha}\mathrm{d}^d\boldsymbol{\beta} \\ &= \int_{\mathbb{C}^d} h_m(\boldsymbol{\beta} + \boldsymbol{d}_{\mathbf{x}}) h_n^*(\boldsymbol{\beta}) \mathrm{d}^d\boldsymbol{\beta}, \end{aligned} \tag{10}$$

where $\mathbf{d}_{\mathbf{x}} = \mathbf{x}_a - \mathbf{x}_b$. Hence,

$$\begin{aligned} \mathbf{C}_U(\mathbf{x}_a, \mathbf{x}_b) &= \sum_{m=1}^{2}\sum_{n=1}^{2} \int_{\mathbb{C}^d} h_m(\boldsymbol{\beta} + \mathbf{d}_{\mathbf{x}}) h_n^*(\boldsymbol{\beta}) \mathrm{d}^d\boldsymbol{\beta} \\ &\quad + \sum_{m=3}^{4}\sum_{n=3}^{4} \int_{\mathbb{C}^d} h_m(\boldsymbol{\beta} + \mathbf{d}_{\mathbf{x}}) h_n^*(\boldsymbol{\beta}) \mathrm{d}^d\boldsymbol{\beta}. \end{aligned} \tag{11}$$

If we set the filter responses to parameterized exponentials as $h_i(\mathbf{x}) = (\mathrm{j})^{c_i} v_i \exp(-(\mathbf{x} - \mu_i)^{\mathrm{H}}(\mathbf{x} - \mu_i)/\gamma)$, the integrals in (11) are as follows

$$\begin{aligned} &\int_{\mathbb{C}^d} h_m(\boldsymbol{\beta} + \mathbf{d}_{\mathbf{x}}) h_n^*(\boldsymbol{\beta}) \mathrm{d}\boldsymbol{\beta}^d \\ =&(\mathrm{j})^{c_m}(-\mathrm{j})^{c_n} v_m v_n \int_{\mathbb{C}^d} \exp\left(-\frac{(\boldsymbol{\beta} - \mu_n)^{\mathrm{H}}(\boldsymbol{\beta} - \mu_n)}{\gamma}\right) \\ &\cdot \exp\left(-\frac{(\boldsymbol{\beta} + \mathbf{d}_{\mathbf{x}} - \mu_m)^{\mathrm{H}}(\boldsymbol{\beta} + \mathbf{d}_{\mathbf{x}} - \mu_m)}{\gamma}\right) \mathrm{d}\boldsymbol{\beta}^d \\ =&(\mathrm{j})^{c_m}(-\mathrm{j})^{c_n} v_m v_n \left(\int_{\mathbb{C}^d} \exp\left(-\frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{H}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{0.5\gamma}\right) \mathrm{d}\boldsymbol{\beta}^d\right) \\ &\cdot \exp\left(-\frac{(\mathbf{d}_{\mathbf{x}} - \mu_m + \mu_n)^{\mathrm{H}}(\mathbf{d}_{\mathbf{x}} - \mu_m + \mu_n)}{2\gamma}\right) \\ =&(\mathrm{j})^{c_m}(-\mathrm{j})^{c_n} \left(\frac{\pi\gamma}{2}\right)^d v_m v_n \\ &\cdot \exp\left(-\frac{(\mathbf{d}_{\mathbf{x}} - \mu_m + \mu_n)^{\mathrm{H}}(\mathbf{d}_{\mathbf{x}} - \mu_m + \mu_n)}{2\gamma}\right), \end{aligned} \tag{12}$$

where $\hat{\boldsymbol{\beta}} = 0.5(\mu_n - (\mathbf{d}_{\mathbf{x}} - \mu_m))$.

We propose the following parameter values. For $h_1(\mathbf{x}) = h_3(\mathbf{x}) = h_{\mathrm{r}}(\mathbf{x})$, we set $v_1 = v_3 = v_{\mathrm{r}}$, $\mu_1 = \mu_3 = 0$ and $c_1 = c_3 = 0$. For $h_2(\mathbf{x}) = \mathrm{j}h_{\mathrm{rj}}(\mathbf{x})$, we set $\mu_2 = \mu$, $v_2 = v_{\mathrm{rj}}$ and $c_2 = 1$. And for $h_4(\mathbf{x}) = \mathrm{j}h_{\mathrm{jr}}(\mathbf{x})$, we set $\mu_4 = -\mu$, $v_4 = -v_{\mathrm{rj}}$ and $c_4 = 1$. By making use of these values and

(12), (11) yields

$$
\begin{aligned}
&\mathbf{C}_U(\mathbf{x}_a, \mathbf{x}_b) = \mathbf{C}_U(\mathbf{d_x}) \\
&= \left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{r}}^2 \exp\left(-\frac{\mathbf{d_x}^{\mathrm{H}}\mathbf{d_x}}{2\gamma}\right) \\
&-\mathrm{j}\left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{r}}v_{\mathrm{rj}} \exp\left(-\frac{(\mathbf{d_x}+\mu)^{\mathrm{H}}(\mathbf{d_x}+\mu)}{2\gamma}\right) \\
&+\mathrm{j}\left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{r}}v_{\mathrm{rj}} \exp\left(-\frac{(\mathbf{d_x}-\mu)^{\mathrm{H}}(\mathbf{d_x}-\mu)}{2\gamma}\right) \\
&+\left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{rj}}^2 \exp\left(-\frac{\mathbf{d_x}^{\mathrm{H}}\mathbf{d_x}}{2\gamma}\right) + \left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{r}}^2 \exp\left(-\frac{\mathbf{d_x}^{\mathrm{H}}\mathbf{d_x}}{2\gamma}\right) \\
&+\mathrm{j}\left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{r}}v_{\mathrm{rj}} \exp\left(-\frac{(\mathbf{d_x}-\mu)^{\mathrm{H}}(\mathbf{d_x}-\mu)}{2\gamma}\right) \\
&-\mathrm{j}\left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{r}}v_{\mathrm{rj}} \exp\left(-\frac{(\mathbf{d_x}+\mu)^{\mathrm{H}}(\mathbf{d_x}+\mu)}{2\gamma}\right) \\
&+\left(\frac{\pi\gamma}{2}\right)^d v_{\mathrm{rj}}^2 \exp\left(-\frac{\mathbf{d_x}^{\mathrm{H}}\mathbf{d_x}}{2\gamma}\right) \\
&= \left(\frac{\pi\gamma}{2}\right)^d (2v_{\mathrm{r}}^2 + 2v_{\mathrm{rj}}^2) \exp\left(-\frac{\mathbf{d_x}^{\mathrm{H}}\mathbf{d_x}}{2\gamma}\right) \\
&+\mathrm{j}\left(\frac{\pi\gamma}{2}\right)^d 2v_{\mathrm{r}}v_{\mathrm{rj}} \left(\exp\left(-\frac{(\mathbf{d_x}-\mu)^{\mathrm{H}}(\mathbf{d_x}-\mu)}{2\gamma}\right)\right. \\
&\left.-\exp\left(-\frac{(\mathbf{d_x}+\mu)^{\mathrm{H}}(\mathbf{d_x}+\mu)}{2\gamma}\right)\right). \qquad (13)
\end{aligned}
$$

Hence, from this covariance, up to a multiplying constant the kernel in (2) can be set as,

$$
\begin{aligned}
k(\mathbf{x}, \mathbf{x}') = (v_{\mathrm{r}}^2 + v_{\mathrm{rj}}^2) \exp\left(-\frac{\mathbf{d_x}^{\mathrm{H}}\mathbf{d_x}}{2\gamma}\right) \\
+\mathrm{j}v_{\mathrm{r}}v_{\mathrm{rj}}\left(\exp\left(-\frac{(\mathbf{d_x}-\mu)^{\mathrm{H}}(\mathbf{d_x}-\mu)}{2\gamma}\right)\right. \\
\left.-\exp\left(-\frac{(\mathbf{d_x}+\mu)^{\mathrm{H}}(\mathbf{d_x}+\mu)}{2\gamma}\right)\right). \qquad (14)
\end{aligned}
$$

where $\mathbf{d_x} = \mathbf{x}' - \mathbf{x}$. Note that the constants $v_{\mathrm{rj}} \in \mathbb{R}$, $v_{\mathrm{r}} \in \mathbb{R}$ and $\mu \in \mathbb{C}^d$ could be set according to the problem at hand or learned as hyperparameters.

If the cross-covariance cancels or is negligible, then in (14) we must set $v_{\mathrm{rj}} = 0$.

5. EXPERIMENTS

We bring here the nonlinear channel equalization task in [2]. The channels consisted of a linear filter $t(n) = (-0.9 + 0.8\mathrm{j}) \cdot s(n) + (0.6 - 0.7\mathrm{j}) \cdot s(n-1)$ and a strong memoryless nonlinearity, $q(n) = t(n) + (0.2 + 0.25\mathrm{j}) \cdot t^2(n) + (0.12 + 0.09\mathrm{j}) \cdot t^3(n)$. The input signals had the form $s(n) = 0.70(\sqrt{1-\rho^2}X(n) + \mathrm{j}\rho Y(n))$, as in [2], and $X(n)$ and $Y(n)$ were Gaussian random variables. Note that the real and the imaginary parts of the input signals were generated independently and, therefore, had null cross-covariances. Also note that the input signals are circular for $\rho = 1/\sqrt{2}$ and highly noncircular if ρ approaches 0 or 1. As in [2] at the receiver end of the channel, the signal $q(n)$ was corrupted by additive white circular Gaussian noise with the SNR set to 16 dB, the filter length was $L = 5$ and the equalization time delay $D = 2$. In all cases the results were averaged over 100 trials where the input signals $s(n)$ and noise output were generated randomly. Figs. 2 and 3 show the learning curves of the NCKLMS2 algorithm in [2]. This algorithm uses the complex Gaussian kernel in (5) and the code to run the experiment is available in [13]. The tunable parameters were set to the values in [2]: $\gamma = 5^2$, the step update parameter to $1/4$, and the novelty criterion was used for the sparsification with $\delta_1 = 0.15$ and $\delta_2 = 0.2$. We observed stability problems in the learning process of the NCKLMS2, due to the kernel used. This problem was alleviated by using sparsification. We also used the independent kernel (7) in the NCKLMS2 approach, with $\kappa_{\mathbb{R}}$ being the real Gaussian kernel, as proposed in [4]. We labeled this algorithm as NCKLMS2-i. The tunable parameter were set to the same values as for the

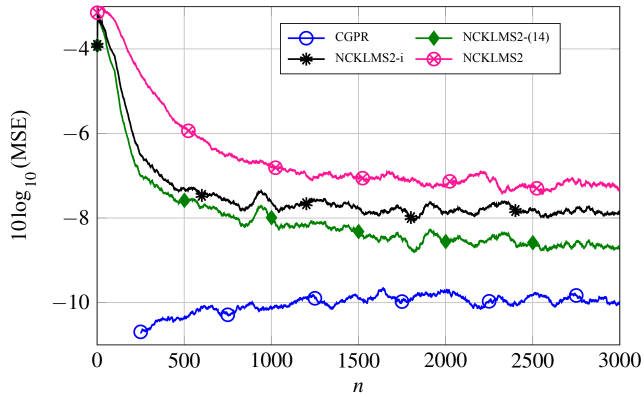

Fig. 2. Learning curves for NCKLMS2, NCKLMS2-i, NCKLMS2-(14), and the proper complex GPR (CGPR) for the circular input case.

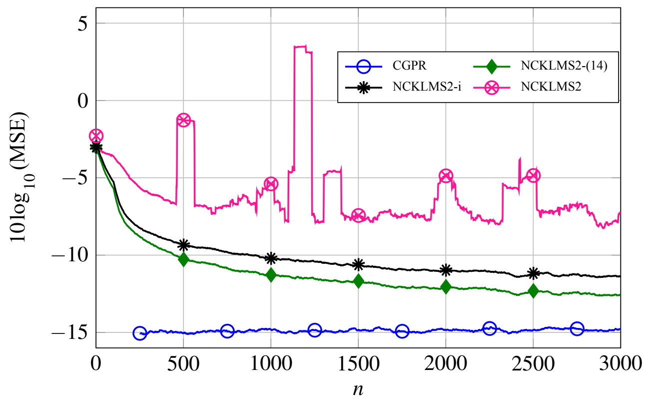

Fig. 3. Learning curves for NCKLMS2, NCKLMS2-i, NCKLMS2-(14), and the proper complex GPR (CGPR) for the noncircular input case ($\rho = 0.1$).

NCKLMS2, except the step update parameter that was set to $1/8$. We proposed to improve the NCKLMS2 by using the kernel derived in this paper in (14) with $v_{\mathrm{rj}} = 0$. This algorithm is labeled as NCKLMS2-(14) in Figs. 2 and 3. We set all the parameters for this algorithm to the same values as for the NCKLMS2. Notice that the parameters were not optimized so results for the proposed NCKLMS2-(14) may be suboptimal. However, while the NCKLMS2-i algorithm exhibits improved performance compared to the NCKLMS2 algorithm, the NCKLMS2-(14) algorithm outperforms both. In this equalization problem the cross-covariance between real and imaginary parts of the signals to-be-learned, $s(n)$, is null. Therefore, we have set the imaginary part of the kernel to zero with much better results. Also, the measure of similarity of inputs through the simple norm of the complex difference between inputs endows the kernel with useful properties such as isotropy and stationarity, better fitting the underlying model.

Finally, in order to illustrate the good performance of the proposed kernel when used in other kernel-based complex-valued regression algorithm, why tried it with the complex Gaussian processes for regression for proper complex signals (CGPR) algorithm proposed in [14]. The first 250 samples were used as training set to estimate the kernel hyperparameters (γ and noise variance) by means of the maximization of the log marginal likelihood. Then, for each new input the CGPR predicted the corresponding $s(n)$ using all the previous input-output pairs as training set. The results for this algorithm considerably outperform the NCKLMS2 algorithm because of the good GPR capabilities together with the suitable kernel selection and hyperparameters estimation.

6. CONCLUSION

In the proper case, the imaginary part of the complex-valued kernel must satisfy skewness. This particular condition makes the kernel design a difficult task. We found that kernels proposed in the literature were designed simply as a complex-valued kernel satisfying skewness. But without further insight on the meaning of the imaginary part. In this paper we review this topic with a twofold conclusion. On the one hand, the imaginary part is to be used if we have a control of its role, as a tool to model a known particular feature of the system. Otherwise, it is better to resort to a real-value kernel. We propose a simple but powerful Gaussian kernel with distance between complex inputs as argument. On the other, we show how to include known features in the imaginary part of the kernel, by means of the convolution approach.

REFERENCES

[1] T. Ogunfunmi and T. K. Paul, "On the complex kernel-based adaptive filter.," in *IEEE Int. Symp. on Circuits and Systems, ISCAS*, 2011, pp. 1263–1266.

[2] P. Bouboulis and S. Theodoridis, "Extension of wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," *IEEE Trans. Signal Processing*, vol. 59, no. 3, pp. 964–978, 2011.

[3] P. Bouboulis, S. Theodoridis, and M. Mavroforakis, "The augmented complex kernel LMS," *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4962–4967, 2012.

[4] F.A. Tobar, A. Kuh, and D.P. Mandic, "A novel augmented complex valued kernel LMS," in *IEEE 7th Sensor Array and Multichannel Signal Processing Workshop, SAM*, 2012, pp. 473–476.

[5] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Adaptive computation and machine learning. MIT Press, Cambridge, Massachusetts, 2002.

[6] I. Steinwart, D. Hush, and C. Scovel, "An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels," *IEEE Trans. Information Theory*, vol. 52, no. 10, pp. 4635–4643, Oct 2006.

[7] T.K. Paul and T. Ogunfunmi, "Study of the convergence behavior of the complex kernel least mean square algorithm," *IEEE Trans. Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1349–1363, Sept 2013.

[8] G. Tauboöck, "Complex-valued random vectors and channels: Entropy, divergence, and capacity," *IEEE Trans. Information Theory*, vol. 58, no. 5, pp. 2729–2744, May 2012.

[9] P. Boyle and M. Frean, "Dependent gaussian processes," in *In Advances in Neural Information Processing Systems 17*. 2005, pp. 217–224, MIT Press.

[10] C.A. Calder and N. Cressie, "Some topics in convolution-based spatial modeling," in *Proceedings of the 56th Session of the International Statistics Institute*, Lisbon, Portugal, August 2007.

[11] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data. The Theory of Improper and Noncircular Signals*, Cambridge University Press, Cambridge, UK, 2010.

[12] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," *Found. Trends Mach. Learn.*, vol. 4, no. 3, pp. 195–266, Mar. 2012.

[13] P. Bouboulis, Code for [2], Available: http://bouboulis.mysch.gr/code/CKAPSM.rar

[14] R. Boloix-Tortosa, F.J. Payan-Somet, and J.J. Murillo-Fuentes, "Gaussian processes regressors for complex proper signals in digital communications," in *IEEE 8th Sensor Array and Multichannel Signal Processing Workshop, SAM*, June 2014, pp. 137–140.