

# Gaussian Processes Regressors for Complex Proper Signals in Digital Communications

Rafael Boloix-Tortosa, F. Javier Payán-Somet and Juan José Murillo-Fuentes

Dep. Teoría de la Señal y Comunicaciones

Escuela Superior de Ingeniera. Universidad de Sevilla

Camino de los Descubrimientos s/n, 41092, Sevilla, Spain

Email: rboloix@us.es

**Abstract**—In this paper we develop the complex-valued version of the Gaussian processes for regression (GPR) for proper complex signals. This tool has proved to be useful in the nonlinear detection in digital communications in real valued models. GPRs can be cast as nonlinear MMSE where hyperparameters can be tuned optimizing a marginal likelihood (ML). This feature allows for a flexible kernel that can easily adapt either to a linear or nonlinear solution. We introduce the complex-valued form of the GPR, and develop it for the proper complex case. We also deal with the optimization of the ML. Some experiments included illustrate the good performance of the proposal.

## I. INTRODUCTION

In digital communications the optimal detection is usually nonlinear and unfeasible due to its computational complexity and the need of non-observable data. Hence approximate solutions are considered. The MMSE criteria is widely used, where the linear form is preferred if computational complexity is to be kept to a minimum. Gaussian processes for regression (GPR) have been proposed as nonlinear MMSE solutions that provide a good compromise between performance and the pair complexity plus training data needed [1], [2]. Their major advantage being the optimization of its hyperparameters through the marginal likelihood. These solutions have been developed for real valued models, assuming that an augmented input space with real and complex parts could be used to cope with the complex-valued model [3]. In [4] we may find a good description of the digital communication model deeply revised from the complex-valued point of view. Besides, results on the kernel design for the complex case have been reported recently [5], [6]. Based on these works, we propose a full description of the GPR for proper complex-valued systems. We develop the marginal likelihood needed in the optimization of the hyperparameters, and its derivation to optimize it. Some experimental results have been included to illustrate the good performance of this solution.

## II. SYSTEM MODEL AND DETECTION

We propose a general model where we transmit  $K$  symbols every sample time,  $\mathbf{b}_t = [b_t(1), b_t(2), \dots, b_t(K)]^\top$ , and receive  $N$  signals arranged in the column-vector  $\mathbf{r}_t$ :

$$\mathbf{r}_t = \mathbf{P} \cdot \mathbf{v}_t + \mathbf{n}_t, \quad (1)$$

where  $\mathbf{v}_t = [\mathbf{b}_t^\top, \mathbf{b}_{t-1}^\top \dots \mathbf{b}_{t-M_s+1}^\top]^\top$  and  $\mathbf{n}_t$  is an  $N$ -dimensional column-vector with additive white Gaussian

noise, where  $M_s$  takes into account the memory or length of the channel response. Unless needed, subindex  $t$  will be omitted hereafter.

As an example, in synchronous DS-CDMA the matrix  $\mathbf{P}$  in (1) summarizes the effect of the channel, the length- $N$  spreading codes and the different amplitudes for each of the  $K$  users, and  $\mathbf{v}_t$  is a  $KM_s$ -dimensional vector including the transmitted bits ( $M_s = 2$  if  $2 \leq M_c \leq K$ , where  $M_c$  is the length of the channel in chip periods) [7], [1]. Many other communication systems can be modeled as (1) aswell.

At the receiver, we learn the system model using training data. In wireless communications short training sequences are mandatory to increase the number of information bits. After the training stage, for every new received vector, denoted as  $\mathbf{r}_*$ , a detector estimates the  $j$ -th transmitted symbol as  $\hat{b}_*(j) = f(\mathbf{r}_*)$  for some  $j$ , where this index will be omitted hereafter. Linear detectors work well when the inter-symbolic interference (ISI) and the users interference are negligible. In the general case, the solution is highly nonlinear and nonlinear detectors are necessary to get a good estimation.

In digital communications, linear MMSE estimators are typically used, due to their simplicity and nearly optimal results in many standard scenarios. The linear,  $f(\mathbf{r}) = \mathbf{r}^\top \mathbf{w}$ , MMSE detector can be expressed as a Wiener filter [8]:

$$\mathbf{w}_{\text{mmse}} = \underset{\mathbf{w}}{\text{argmin}} \mathbb{E} \left[ (b - \mathbf{r}^\top \mathbf{w})^2 \right] = (\mathbf{C}_{\mathbf{r}\mathbf{r}})^{-1} \mathbf{C}_{\mathbf{r}b}, \quad (2)$$

where  $\mathbf{C}_{\mathbf{r}\mathbf{r}}$  and  $\mathbf{C}_{\mathbf{r}b}$  are, respectively, the autocorrelation of the inputs and the cross-correlation between the inputs and outputs. This solution can be estimated from training data by replacing  $\mathbf{C}_{\mathbf{r}\mathbf{r}}$  and  $\mathbf{C}_{\mathbf{r}b}$  by their sampled versions:

$$\mathbf{w}_{\text{smmse}} = (\mathbf{R} \cdot \mathbf{R}^\top)^{-1} \mathbf{R} \cdot \mathbf{b}, \quad (3)$$

where the columns of  $\mathbf{R}$  are the received signals at every symbol in the training set. We denote this solution as linear sampled MMSE (SMMSE).

However, in the general case, the MMSE is not linear. It is widely known that the MMSE solution for AWGN channels is given by the conditional mean of  $b$  given  $\mathbf{r}$ :

$$\hat{b}_* = f_{\text{mmse}}(\mathbf{r}_*) = \mathbb{E} [b_* | \mathbf{r}_*], \quad (4)$$

which is a nonlinear function of the inputs, unless the inputs are Gaussian distributed, that it is not the case.

The computation of the optimal solution is hard to accomplish, given the great number of possible combinations of transmitted symbols due to the number of points in the transmitted constellations, the number of users and the channel memory. In [1], [2] the real valued GPR is proposed as a nonlinear tool to estimate the MMSE response. However, to our knowledge, there is no formulation available for complex-valued systems. In the following we develop a complete formulation of the GPR and hyperparameters adaptation for the complex proper case, which is the main contribution of this work.

### III. COMPLEX GAUSSIAN PROCESSES

A complex Gaussian process  $f(\mathbf{x})$  is a collection of complex random variables, any finite number of which have a joint complex Gaussian distribution.

The general pdf of a complex Gaussian random vector  $\mathbf{z} : \Omega \rightarrow \mathbb{C}^D$  is [4]

$$p(\mathbf{z}) = \frac{1}{\pi^n \det^{1/2} \underline{\mathbf{R}}_{zz}} \exp \left\{ -\frac{1}{2} (\underline{\mathbf{z}} - \underline{\mu}_z)^H \underline{\mathbf{R}}_{zz}^{-1} (\underline{\mathbf{z}} - \underline{\mu}_z) \right\}, \quad (5)$$

where the augmented vector  $\underline{\mathbf{z}}$  is obtained by stacking  $\mathbf{z}$  on top of its complex conjugate  $\mathbf{z}^*$ . Similarly, the augmented mean vector  $\underline{\mu}_z$  is obtained by stacking vector  $\mu_z$  on top of its complex conjugate  $\mu_z^*$ .  $\underline{\mathbf{R}}_{zz}$  is the augmented covariance matrix of  $\mathbf{z}$

$$\underline{\mathbf{R}}_{zz} = \begin{bmatrix} \mathbf{R}_{zz} & \tilde{\mathbf{R}}_{zz} \\ \tilde{\mathbf{R}}_{zz}^* & \mathbf{R}_{zz}^* \end{bmatrix}, \quad (6)$$

where  $\mathbf{R}_{zz} = \mathbb{E}[(\mathbf{z} - \mu_z)(\mathbf{z} - \mu_z)^H]$  is the usual covariance matrix and  $\tilde{\mathbf{R}}_{zz} = \mathbb{E}[(\mathbf{z} - \mu_z)(\mathbf{z} - \mu_z)^T]$  is the complementary covariance or pseudo-covariance matrix. It is important to note that both  $\mathbf{R}_{zz}$  and  $\tilde{\mathbf{R}}_{zz}$  are required for a complete second-order characterization of the complex Gaussian random vector  $\mathbf{z} \sim \mathcal{N}(\mu_z, \mathbf{R}_{zz}, \tilde{\mathbf{R}}_{zz})$ .

A complex Gaussian process  $f(\mathbf{x})$  is completely specified by its augmented mean function,  $\underline{\mu}(\mathbf{x})$ , its covariance function,  $k(\mathbf{x}_i, \mathbf{x}_j)$ , and its complementary covariance function,  $\tilde{k}(\mathbf{x}_i, \mathbf{x}_j)$ , i.e.  $f(\mathbf{x}) \sim \mathcal{GP}(\underline{\mu}(\mathbf{x}), k(\mathbf{x}_i, \mathbf{x}_j), \tilde{k}(\mathbf{x}_i, \mathbf{x}_j))$ .

There is an important special case in which the complementary covariance for a complex Gaussian vector vanishes.

*Definition 3.1:* [4] If the complementary covariance matrix vanishes,  $\tilde{\mathbf{R}}_{zz} = 0$ ,  $\mathbf{z}$  is called proper, otherwise it is called improper.

The pdf (5) of a complex proper Gaussian random vector  $\mathbf{z} : \Omega \rightarrow \mathbb{C}^D$  simplifies to

$$p(\mathbf{z}) = \frac{1}{\pi^n \det \mathbf{R}_{zz}} \exp \left\{ -(\mathbf{z} - \mu_z)^H \mathbf{R}_{zz}^{-1} (\mathbf{z} - \mu_z) \right\}. \quad (7)$$

We can propose now the following definition:

*Definition 3.2:* If the complementary covariance function of a complex Gaussian process vanishes, the process is called proper, otherwise it is called improper.

### IV. COMPLEX PROPER GAUSSIAN PROCESSES FOR REGRESSION

We consider the model

$$f(\mathbf{x}) = \phi(\mathbf{x})^H \mathbf{w}, \quad (8)$$

where  $\mathbf{x}$  is the complex input vector of dimension  $D$  and we specify a prior over the parameters  $\mathbf{w}$ : a zero mean proper complex Gaussian with covariance matrix  $\Sigma_p$ , i.e.,  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_p, \mathbf{0})$ . This linear regression model gives an example of a zero mean complex Gaussian process,  $\mu(\mathbf{x}) = \phi(\mathbf{x})^H \mathbb{E}[\mathbf{w}] = 0$ , with covariance function  $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^H \mathbb{E}[\mathbf{w}\mathbf{w}^H] \phi(\mathbf{x}_j) = \phi(\mathbf{x}_i)^H \Sigma_p \phi(\mathbf{x}_j) = \varphi(\mathbf{x}_i)^H \varphi(\mathbf{x}_j)$ . The complementary covariance function vanishes,  $\tilde{k}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^H \mathbb{E}[\mathbf{w}\mathbf{w}^T] \phi^*(\mathbf{x}_j) = 0$ , i.e.,  $f(\mathbf{x})$  is a complex proper Gaussian process:  $f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}_i, \mathbf{x}_j), 0)$ .

We can specify a covariance function, take  $n$  input vectors  $\mathbf{x}_i, i = 1, \dots, n$  aggregated column-wise in the  $D \times n$  matrix  $\mathbf{X}$ , and write out the corresponding  $n \times n$  covariance matrix  $\mathbf{K}(\mathbf{X}, \mathbf{X})$  elementwise from  $k(\mathbf{x}_i, \mathbf{x}_j)$ . For those inputs we generate a complex proper Gaussian random vector  $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}(\mathbf{X}, \mathbf{X}), \mathbf{0})$ .

In the more realistic situation

$$y = f(\mathbf{x}) + \epsilon = \phi(\mathbf{x})^H \mathbf{w} + \epsilon, \quad (9)$$

the observed values  $y$  differ from  $f(\mathbf{x})$  by additive complex proper Gaussian noise with zero mean and variance  $\sigma_n^2$ , i.e.,  $\epsilon \sim \mathcal{N}(0, \sigma_n^2, 0)$ . Hence, the complex random vector  $\mathbf{y}$  that results from the inputs in  $\mathbf{X}$  is also proper Gaussian:  $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}, \mathbf{0})$ , with  $\mathbf{C} = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}$ . Furthermore,  $\mathbf{y}$  and  $\mathbf{f}$  are cross-proper, as  $\tilde{\mathbf{R}}_{yf} = \mathbb{E}[\mathbf{y}\mathbf{f}^T] = \mathbb{E}[\mathbf{f}\mathbf{f}^T] + \mathbb{E}[\epsilon\mathbf{f}^T] = 0$ . Hence,  $\mathbf{y}$  and  $\mathbf{f}$  are jointly proper [4], i.e., the composite complex random vector  $[\mathbf{y}^T, \mathbf{f}^T]^T$  is Gaussian proper. If we take now a number of test input vectors aggregated column-wise in matrix  $\mathbf{X}_*$ , the joint distribution of the training outputs  $\mathbf{y}$  and the test outputs  $\mathbf{f}_*$  is

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} \mathbf{C} & \mathbf{K}(\mathbf{X}, \mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*, \mathbf{X}) & \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}, \mathbf{0} \right), \quad (10)$$

and the conditional distribution of  $\mathbf{f}_*$  given  $\mathbf{y}$  is

$$f_* | \mathbf{x}_*, \mathbf{X}, \mathbf{y} \sim \mathcal{N}(\bar{f}_*, \text{cov}(f_*), 0), \quad (11)$$

where

$$\begin{aligned} \bar{f}_* &= \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \mathbf{C}^{-1} \mathbf{y}, \\ \text{cov}(f_*) &= \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X}) \mathbf{C}^{-1} \mathbf{K}(\mathbf{X}, \mathbf{X}_*). \end{aligned} \quad (12)$$

In the case that there is only one test input vector  $\mathbf{x}_*$ ,

$$\bar{f}_* = \mathbf{k}_*^T \mathbf{C}^{-1} \mathbf{y}, \quad (14)$$

$$\text{cov}(f_*) = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T \mathbf{C}^{-1} \mathbf{k}_*, \quad (15)$$

where  $\mathbf{k}(\mathbf{x}_*) = \mathbf{k}_*$  is the vector of covariances between the test point and the  $n$  training points.

### A. The marginal likelihood

The marginal likelihood (or evidence) is the integral of the likelihood times the prior

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{f}, \mathbf{X}) p(\mathbf{f}|\mathbf{X}) d\mathbf{f}, \quad (16)$$

Here, the prior is complex proper Gaussian,  $\mathbf{f}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}(\mathbf{X}, \mathbf{X}), \mathbf{0})$ . Since  $\mathbf{y}$  and  $\mathbf{f}$  are jointly proper, the likelihood is the factorized proper Gaussian,  $\mathbf{y}|\mathbf{f} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2 \mathbf{I}, \mathbf{0})$ . Now, using (7) in (16) the log marginal likelihood yields

$$L = \log p(\mathbf{y}|\mathbf{X}) = -\mathbf{y}^H \mathbf{C}^{-1} \mathbf{y} - \log \det \mathbf{C} - n \log \pi, \quad (17)$$

i.e.,  $\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}, \mathbf{0})$ .

### B. Model selection for complex proper GP regression

In order to turn complex Gaussian processes into powerful practical tools it is essential to develop methods that address the model selection problem. It is common to use a hierarchical specification of models, with the parameters,  $\mathbf{w}$  at the first level and some hyperparameters  $\theta_i$  at the second level. In our model, the covariance function  $k(\mathbf{x}_i, \mathbf{x}_j)$  can be parameterized in terms of the hyperparameters  $\theta_i$ .

It is a property of the marginal likelihood that it automatically incorporates a trade-off between model fit and model complexity [9]. This is the reason why the marginal likelihood is valuable in solving the model selection problem.

The log marginal likelihood in (17) can be explicitly written conditioned on the hyperparameters,  $L(\theta)$ , with  $\mathbf{C} = \mathbf{C}(\theta) = \mathbf{K}(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}$ , where  $\theta$  is the vector of hyperparameters. We can maximize (17) in order to set those hyperparameters. Notice that the variance of the additive noise,  $\sigma_n^2$ , is also treated as another hyperparameter and can be estimated from this optimization.

$L(\theta)$  in (17) is a real function of a complex Hermitian matrix  $\mathbf{C}(\theta)$  that depends on the hyperparameters. The chain rule can be used to seek the derivative of the log marginal likelihood with respect to  $\mathbf{C}(\theta)$  and then the derivative of  $\mathbf{C}(\theta)$  with respect to the hyperparameters. However, the Cauchy-Riemann conditions dictate that nonconstant real valued functions that are defined on complex domains are not holomorphic, so the complex derivative cannot be used. Instead, we must seek formal or Wirtinger derivatives [5], [10].

The chain rule for the Wirtinger derivative of the function  $L(\theta)$  with respect to the hyperparameter  $\theta_i$  yields [10]

$$\frac{\partial L}{\partial \theta_i} = \sum_{k=1}^n \sum_{l=1}^n \left( \frac{\partial L}{\partial C_{lk}} \frac{\partial C_{lk}}{\partial \theta_i} + \frac{\partial L}{\partial C_{lk}^*} \frac{\partial C_{lk}^*}{\partial \theta_i} \right), \quad (18)$$

where  $C_{lk}$  is the element  $(l, k)$  of matrix  $\mathbf{C}$ . When finding the Wirtinger derivatives  $\frac{\partial L}{\partial C_{lk}}$  and  $\frac{\partial L}{\partial C_{lk}^*}$  the elements  $C_{lk}$  and  $C_{lk}^*$  are treated as independent variables. Therefore,

$$\frac{\partial L}{\partial \theta_i} = \sum_{k=1}^n \sum_{l=1}^n \frac{\partial L}{\partial C_{lk}} \frac{\partial C_{lk}}{\partial \theta_i} = \text{Tr} \left[ \left( \frac{\partial L}{\partial \mathbf{C}} \right)^\top \frac{\partial \mathbf{C}}{\partial \theta_i} \right]. \quad (19)$$

The derivative of the first term of the function  $L$  in (17) with respect to  $\mathbf{C}$  yields

$$\begin{aligned} \frac{\partial}{\partial \mathbf{C}} (-\mathbf{y}^H \mathbf{C}^{-1} \mathbf{y}) &= -\frac{\partial}{\partial \mathbf{C}} \text{Tr} (\mathbf{y}^H \mathbf{C}^{-1} \mathbf{y}) \\ &= \mathbf{C}^{-\top} (\mathbf{y} \mathbf{y}^H)^\top \mathbf{C}^{-\top}. \end{aligned} \quad (20)$$

The derivative of the second term of the function  $L$  in (17) with respect to  $\mathbf{C}$  yields

$$\frac{\partial}{\partial \mathbf{C}} (-\log \det \mathbf{C}) = -\mathbf{C}^{-\top}. \quad (21)$$

Substitution of (20) and (21) in (19) yield

$$\frac{\partial L}{\partial \theta_i} = \text{Tr} \left[ (\mathbf{C}^{-1} (\mathbf{y} \mathbf{y}^H) \mathbf{C}^{-1} - \mathbf{C}^{-1}) \frac{\partial \mathbf{C}}{\partial \theta_i} \right]. \quad (22)$$

### C. Complex covariance functions

We recall here the complex kernel function used in the Experiments section of this study, the *independent Gaussian kernel* [6]. This kernel is defined as

$$\begin{aligned} \kappa(\mathbf{x}_i, \mathbf{x}_j) &= \kappa_{\mathbb{R}}(\alpha_i, \alpha_j) + \kappa_{\mathbb{R}}(\beta_i, \beta_j) \\ &\quad + j(\kappa_{\mathbb{R}}(\alpha_i, \beta_j) - \kappa_{\mathbb{R}}(\beta_i, \alpha_j)), \end{aligned} \quad (23)$$

where  $\mathbf{x}_l = \alpha_l + j\beta_l$  and  $\kappa_{\mathbb{R}}$  is the well-known real Gaussian kernel. This kernel is a generic extension of  $\kappa_{\mathbb{R}}$  and inherits the intuitive physical meaning of a measure of similarity of the samples that  $\kappa_{\mathbb{R}}$  has.

## V. EXPERIMENTS

By replacing  $\mathbf{r}$  and  $b$  in (1) by  $\mathbf{x}$  and  $y$ , respectively, in (9) we yield the complex GPR model to estimate  $b$ . That estimate can be found as  $\hat{f}_*$  in (14) for each new received vector  $\mathbf{x}_* = \mathbf{r}_*$ . In these experiments we focus on different scenarios to show the good behavior of this detector. We consider here QPSK modulations. Note that they are proper complex-valued and so are the received signals in (1).

In Fig. 1 we have included the BER for the complex GPR with the kernel in (23) for the experiment in Section V.A in [1]. In this experiment we have a DS-CDMA system with  $K = 2$  equal power users and signatures  $[+1+1-1-1]$  and  $[+1-1-1+1]$ . The channel is given by  $c(z) = 0.3 + 0.7z^{-1} + 0.3z^{-2}$ . The four received chips after the two matched filters are the inputs,  $\mathbf{x}$ , i.e.,  $N = 2$ . The number of training samples used was  $n = 64$ . In Fig. 2 the decision boundary is included for the complex GPR with 64 training samples and  $SNR = 16$  dB, along with the points corresponding to the bits of the real part of the inputs for both users. This experiment illustrates a scenario where the GPR is of major interest, since the linear response fails to detect the transmitted symbols. The complex GPR successfully computes the nonlinear solution, even for the short number of training samples used.

In Fig. 3 we include a DS-CDMA scenario with  $K = 10$  equal power users and signatures with  $N = 31$  chips and a channel of 10 zero mean complex Gaussian iid taps. We depict the BER along the number of training samples,  $n$ , for a  $SNR = 12$  dB. In this experiment, the inputs to the detector are  $N = 31$  received chips for one symbol period. Since the

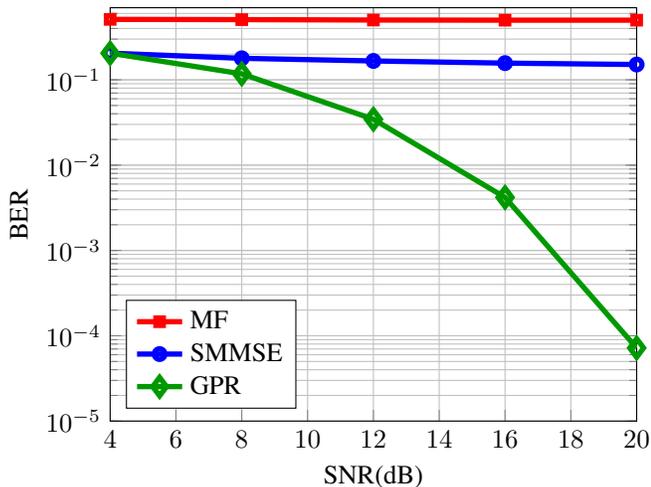


Fig. 1. BER along the SNR in a DS-CDMA scenario with  $K = 2$  users and signatures  $[+1 + 1 - 1 - 1]$  and  $[+1 - 1 - 1 + 1]$  for SMMSE ( $\circ$ ), the matched filter ( $\square$ ) and GPR ( $\diamond$ ) detectors for  $n = 64$  training samples.

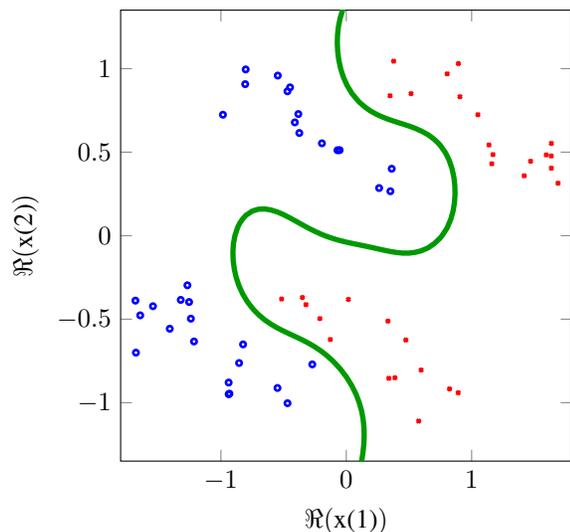


Fig. 2. Decision boundary for the GPR in a DS-CDMA scenario with  $K = 2$  users and signatures  $[+1 + 1 - 1 - 1]$  and  $[+1 - 1 - 1 + 1]$  for  $n = 64$  training samples and  $SNR = 16$  dB.

large dimension, the solution is nearly linear and the linear SMMSE exhibits a good performance. The GPR with the nonlinear complex kernel adapts well to this linear solution, even for a short number of samples. In the short length scenario the regularization term learned is high to compensate for the lack of information, and the GPR performs better than the SMMSE. As the number of training samples increases, the nonlinear part of the kernel learns the linear solution and the GPR is able to follow the SMMSE. The slight mismatch between both curves is again due to the regularization term.

## VI. CONCLUSION

In this paper we develop a complete formulation of the GPR and hyperparameters adaptation for the complex proper case.

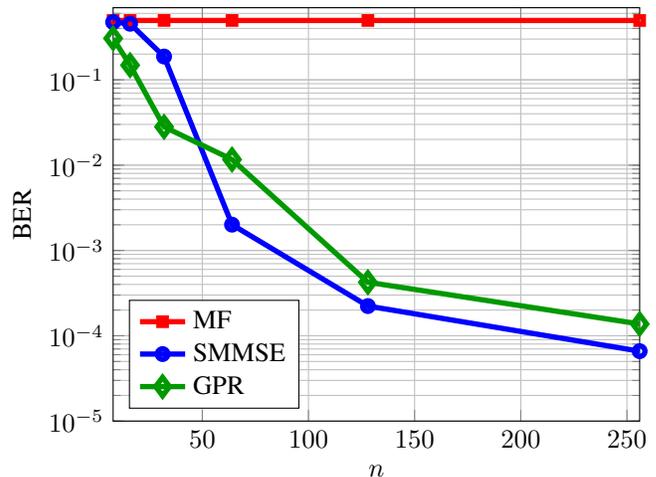


Fig. 3. BER along  $n$ , in a DS-CDMA scenario with  $K = 10$  users and signatures with  $N = 31$  for SMMSE ( $\circ$ ), the matched filter ( $\square$ ) and GPR ( $\diamond$ ) detectors for  $SNR = 12$  dB.

Complex proper GPR can adapt either to linear or nonlinear solutions with short training sequences, therefore they are of interest as detectors in communication applications, as shown in the included experiments.

## ACKNOWLEDGMENT

This work was partially funded by Spanish government (Ministerio de Educación y Ciencia, Consolider-Ingenio 2010 CSD2008-00010, TEC2012-38800-C03-{02}) and European Union (FEDER).

## REFERENCES

- [1] J. Murillo-Fuentes and F. Pérez-Cruz, "Gaussian process regressors for multiuser detection in DS-CDMA systems," *IEEE Transactions on Communications*, vol. 57, no. 8, pp. 2339–2347, 2009.
- [2] F. Pérez-Cruz, S. Van Vaerenbergh, J. Murillo-Fuentes, M. Lazarogredilla, and I. Santamaria, "Gaussian processes for nonlinear signal processing: An overview of recent advances," *Signal Processing Magazine, IEEE*, vol. 30, no. 4, pp. 40–50, July 2013.
- [3] F. Pérez-Cruz, J. Murillo-Fuentes, and S. Caro, "Nonlinear channel equalization with Gaussian Processes for Regression," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5283–5286, Oct. 2008.
- [4] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data. The Theory of Improper and Noncircular Signals*. Cambridge, UK: Cambridge University Press, 2010.
- [5] P. Bouboulis and S. Theodoridis, "Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS," *Signal Processing, IEEE Transactions on*, vol. 59, no. 3, pp. 964–978, 2011.
- [6] F. Tobar, A. Kuh, and D. Mandic, "A novel augmented complex valued kernel LMS," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2012 IEEE 7th*, 2012, pp. 473–476.
- [7] S. Chen, A. K. Samingan, and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels," *IEEE Transactions on Neural Network*, vol. 12, no. 3, pp. 604–611, December 2001.
- [8] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.
- [9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology, Cambridge, Massachusetts: MIT Press, 2006.
- [10] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 2740–2746, June 2007.