

# Channel Identification of Non-linear Systems with Binary-Valued Output Observations Based on Positive Definite Kernels

Rachid Fateh<sup>1,\*</sup>, Anouar Darif<sup>1</sup>, and Said Safi<sup>1</sup>

<sup>1</sup>Laboratory of Innovation in Mathematics, Applications and Information Technologies (LIMATI), Polydisciplinary Faculty, Sultan Moulay Slimane University, Po. Box 592, 23000 Beni Mellal, Morocco

## Abstract.

Nowadays, the kernel methods are increasingly developed, they are a significant source of advances, not only in terms of computational cost but also in terms of the obtained efficiencies in solving complex tasks, they are founded on the theory of reproducing kernel Hilbert spaces (RKHS). In this paper, we propose an algorithm for recursive identification of finite impulse response (FIR) nonlinear systems, whose outputs are detected by binary value sensors. This algorithm is based on a nonlinear transformation of the data using a kernel function. This transformation performs a basic change that allows the data to be projected into a new space where the relationships between the variables are linear. To test the accuracy of the proposed algorithm, we have compared it with another algorithm proposed in the literature, for that, we employ the practical frequency selective fading channel, called Broadband Radio Access Network (BRAN). Monte Carlo simulation results, in noisy environment and for various data length, demonstrate that the proposed algorithm can give better precision.

## 1 Introduction

The field of systems identification has recently become an active area of research that has attracted the attention of a considerable number of researchers [1–5]. The identification consists in searching the parameters of mathematical models of a system, based on experimental data and a priori available knowledge. The identification of linear systems has been carried out for decades using recursive type method. Nevertheless, today's systems are becoming increasingly non-linear as a result of their complexity. According to the progressive usage of non-linear models in real systems, numerous resolution methods for the identification of nonlinear systems, such as Volterra filters, neural networks and kernel methods have been developed.

Nowadays, kernel methods are increasingly developed, they are a significant source of advances, not only in terms of computational cost but also in terms of the obtained efficiencies in solving complex tasks. As these methods largely determine the efficiency of the treatments, by their ability to reveal existing similarities in the processed data. They are based on a central principle called “kernel trick”, exploited for the first time with the Support Vector Machine (SVM)[6, 7], then also used to transform numerous linear dimensionality reduction algorithms into non-linear algorithms [8].

The kernel's trick allows to give a non-linear character to several originally linear methods and, without restriction, they can be expressed only in terms of inner products of the observations. The several kernel adap-

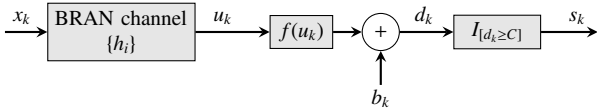
tive filtering algorithms, including the kernel least mean square (KLMS)[9] and kernel affine projection algorithms (KAPA) [10], have recently been noted for nonlinear signal processing [11]. Furthermore, to optimize the quality of the basic kernel adaptive filtering algorithms, subtypes of these algorithms have also been mentioned [12–14] for channel identification [15] and equalization for nonlinear systems. In this paper, we propose an algorithm based, on positive definite kernel function. In order to test its efficiency, we have compared with the Pouliquen et al. algorithm, where the goal is to identify a practical, i.e. measured, frequency-selective fading channel, called Broadband Radio Access Network (BRAN C), representing the outdoor propagation. This channel of the model is normalized by the European Telecommunications Standards Institute (ETSI)[16, 17].

The present paper is arranged as follows: In the next section, we introduced the structure of the nonlinear model using quantized observations. In Section 3, we presented the derivation of the Pouliquen et al. algorithm. In Section 4 we review some basic concepts of kernel methods, and the proposed algorithm is presented in Section 5. Some simulations to assess the performance of the proposed algorithm are shown in Section 6 and, finally, Section 7 concludes the paper.

## 2 Problem Statement

Let us consider a nonlinear model represented in Figure 1; it consists of a finite impulse response of known order, followed by a nonlinear static function.

\*Corresponding author: [fateh.smi@gmail.com](mailto:fateh.smi@gmail.com)



**Figure 1.** Block of the nonlinear system with binary outputs and noises.

The following equations are used to model the system (Figure 1):

$$\begin{cases} u_k = \sum_{i=0}^{n-1} h_i x_{k-i} \\ d_k = f(u_k) + b_k \end{cases} \quad (1)$$

where,  $x_k$  is the input signal,  $\{h_i\}_{i=0}^{n-1}$ , the channel impulse response,  $n$  is the order of FIR system,  $f(\cdot)$  is the nonlinearity, and  $b_k$  the measurement noise.

The measurement of the system output  $d_k$  is carried out by binary-valued sensor  $I_{[\cdot]}$ , with a finite threshold  $C \in \mathbb{R}$ , such that:

$$s_k = I_{|d_k| \geq C} = \begin{cases} 1 & \text{if } d_k \geq C, \\ -1 & \text{otherwise.} \end{cases} \quad (2)$$

The following system assumptions will be made in this paper:

- The input  $x_k$ , is independent and identically distributed (i.i.d) with zero mean.
- The additive noise  $b_k$  is gaussian and independent of  $x_k$  and  $d_k$ .
- The non-linearity  $f(\cdot)$  is continuous and invertible for any finite  $x$ .
- There is no delay in the system, i.e.  $h_0 \neq 0$ .
- The threshold  $C$  is known.
- $\|\cdot\|$  is the  $l_2$  norm.

### 3 Pouliquen et al Algorithm

A recursive identification algorithm was presented by Pouliquen et al. [18] in 2016. The basic idea of this algorithm is to formulate the identification problem as a bounded error identification problem. It is based on the following assumptions:

- $C = 0$ .
- $\|\theta\| = 1$ .
- At any time  $k$ ,  $x_{k-i}^\top \theta \neq 0$ , where  $i = 0, 1, \dots, n - 1$ .
- The noise sequence  $b_k$  is an i.i.d. sequence of random variables with zero mean, finite covariance and uncorrelated with the input sequence.

The identification is realized such that:

$$|s_k - x_k^\top \theta| < \delta. \quad (3)$$

with  $\delta < 1$  and  $\theta = [h_0, h_1, \dots, h_{n-1}]^\top$ .

The estimation of  $\theta$  is given by:

$$\widehat{\theta}_k = \widehat{\theta}_{k-1} + \Gamma_k \varepsilon_{k/k-1} \quad (4)$$

with:

$$\Gamma_k = \frac{P_{k-1} x_k \sigma_k}{\lambda + x_k^\top P_{k-1} x_k \sigma_k}. \quad (5)$$

$$P_k = \frac{1}{\lambda} (I_{n+1} - \Gamma_k x_k^\top) P_{k-1}. \quad (6)$$

$$\varepsilon_{k/k-1} = s_k - x_k^\top \widehat{\theta}_{k-1}. \quad (7)$$

where:

$$\sigma_k = \begin{cases} \frac{\lambda}{x_k^\top P_{k-1} x_k} (|\frac{\varepsilon_{k/k-1}}{\delta}| - 1), & \text{if } (|\varepsilon_{k/k-1}| > \delta) \text{ and } (x_k^\top P_{k-1} x_k > 0) \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

and  $0 < \lambda < 1$  is the forgetting factor.

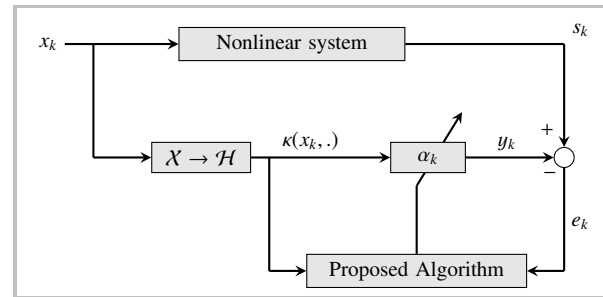
### 4 Theoretical Framework of Kernel

Nonlinear adaptive filtering problems can be formulated as linear finite order problems, where the original input data  $\mathcal{X}$  has been mapped to a nonlinear Hilbert space  $\mathcal{H}$  (infinite-dimensional) with a Mercer kernel [19]:

$$\kappa(x_i, x_j) = \langle \psi(x_i), \psi(x_j) \rangle_{\mathcal{H}}, \quad \forall (x_i, x_j) \in \mathcal{X}^2 \quad (9)$$

where  $\psi$  maps  $\mathcal{X}$  into a higher dimensions space  $\mathcal{H}$  with an inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ .

Figure 2 shows a functional diagram illustrating an adaptive filter based on the kernel, where,  $\{x_k\}_1^N$  is the vector of the  $N$  most recent input signal samples,  $\{y_k\}_1^N$  is the estimated desired response and  $\{e_k\}_1^N$  is the estimation error.



**Figure 2.** Kernel-based adaptive system identification.

At each instant  $k$ , the learning procedure is performed in two consecutive phases:

- First of all, we get the desired output  $s_k$  from the nonlinear system with binary output and noise.
- In the second phase, after the transformation of the data to a nonlinear Hilbert space with a Mercer kernel, the model coefficients  $\alpha_k$  are updated while minimizing the functional cost.

Let us begin with some definitions in order to correctly determine the existence of a functional space  $\mathcal{H}$ .

**Definition 4.1 (Positive Definite Kernel).** A kernel is called positive definite, if, for each input data point  $\{x_i\}_{i=1}^N \in \mathcal{X}$  satisfies the following condition:

$$\sum_{i,j=1}^N \alpha_i \alpha_j \kappa(x_i, x_j) \geq 0, \quad (10)$$

for all  $N \in \mathbb{N}$ ,  $(x_1, \dots, x_N) \in \mathcal{X}^N$  and  $(\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$ .

To represent the attributes of  $\mathcal{H}$  from their coordinates, we can allocate an orthonormal framework to a Hilbert space  $\mathcal{H}$ . The associated kernel must be continuous, symmetric, normalized and positive definite function  $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ . The input domain,  $\mathcal{X}$ , is a subset of  $\mathbb{R}^N$  that is compact.

**Definition 4.2** (Reproducing kernels and Hilbert spaces). Let  $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$  represent a Hilbert space constituted by functions of  $\mathcal{X}$  in  $\mathbb{R}$ . The function  $\kappa(x_i, x_j)$  of  $\mathcal{X} \times \mathcal{X}$  in  $\mathbb{R}$  is the reproducing kernel of  $\mathcal{H}$ , provided that  $\mathcal{H}$  admits one, if there exists a function  $\kappa(x, \cdot) : x_i \rightarrow \kappa(x, x_j)$  belongs to  $\mathcal{H}$ , for any  $x \in \mathcal{X}$ .

## 5 Proposed Algorithm

In this section, we present the proposed algorithm. The basic idea is to run the algorithm presented in [18] in the kernel feature space, which is connected with a positive definite kernel  $\kappa$ , through the characteristic map  $\Psi(\cdot)$  as defined in (11). The sample sequence is transformed by using a feature map called  $\Psi$ :

$$\Psi : \mathcal{X} \rightarrow \mathcal{H} \\ x_i \rightarrow \kappa(x_i, \cdot), \quad 0 \leq i \leq N \quad (11)$$

To create the Hilbert space model of the reproducing kernel ( $\kappa$ ), we use the Gaussian kernel, which is generally a default choice due to its universal approximation capability:

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad \forall x_i, x_j \in \mathcal{X}. \quad (12)$$

where  $\sigma > 0$  is the kernel width.

The proposed algorithm is composed of the four following steps:

- **Step 1:** In the first step, we transform the input data from the observation space ( $\mathcal{X}$ ) into a Hilbert space ( $\mathcal{H}$ ) to generate the input data:

$$\{(\Psi(x_1), s_1), (\Psi(x_2), s_2), \dots, (\Psi(x_N), s_N), \dots)\}. \quad (13)$$

- **Step 2:** In this second step, we apply the logic of the Pouliquen et al. algorithm to the sequence defined in (13) to minimize the cost function:

$$E[|s_k - \langle (\Psi(x_k)), \alpha \rangle_{\mathcal{H}}|^2]$$

where the term  $\alpha$  represents the weight vector in reproducing kernel Hilbert space.

- **Step 3:** In this third step, we operate immediately in the reproducing kernel Hilbert space, assuming the data has been mapped in the RKHS  $\mathcal{H}$  using the  $\Psi$  function map, i.e.,

$$\mathcal{X} \ni x \rightarrow \Psi(x_k) := \kappa(x, \cdot) \in \mathcal{H}. \quad (14)$$

The proposed algorithm update equations in kernel Hilbert space results in:

$$\widehat{\alpha}_k = \widehat{\alpha}_{k-1} + \Gamma_k^{\mathcal{H}} e_k, \quad (15)$$

$$\Gamma_k^{\mathcal{H}} = \frac{P_{k-1}^{\mathcal{H}} \kappa(x_k, \cdot) \sigma_k}{\lambda + \kappa(x_k, \cdot)^{\top} P_{k-1}^{\mathcal{H}} \kappa(x_k, \cdot) \sigma_k}, \quad (16)$$

$$P_k^{\mathcal{H}} = \frac{1}{\lambda} (I_{n+1} - \Gamma_k^{\mathcal{H}} \kappa(x_k, \cdot)^{\top}) P_{k-1}^{\mathcal{H}}, \quad (17)$$

and

$$e_k = s_k - \kappa(x_k, \cdot)^{\top} \widehat{\alpha}_{k-1}. \quad (18)$$

where:

$$\sigma_k = \begin{cases} \frac{\lambda}{\kappa(x_k, \cdot)^{\top} P_{k-1}^{\mathcal{H}} \kappa(x_k, \cdot)} (| \frac{e_k}{\delta} | - 1), \\ \text{if } (|e_k| > \delta) \text{ and } (\kappa(x_k, \cdot)^{\top} P_{k-1}^{\mathcal{H}} \kappa(x_k, \cdot) > 0) \\ 0, \text{ otherwise.} \end{cases} \quad (19)$$

## 6 Comparison and numerical simulation

In this section, our aim was to evaluate the performance of the proposed algorithm. We simulated the system, in which the linear dynamics is a BRAN channel and the output nonlinearity is a hyperbolic function ( $\tanh(2x)$ ). The input sequence was randomly generated with a uniform distribution in the range  $[1; -1]$  using 50 Monte Carlo runs.

### 6.1 BRAN C radio channel

The measurements for the BRAN C channel are described in this paragraph (see Table 1). The BRAN radio channel impulse response is described as follows:

$$h(n) = \sum_{i=0}^{p-1} A_i \delta(n - \tau_i), \quad p = 18. \quad (20)$$

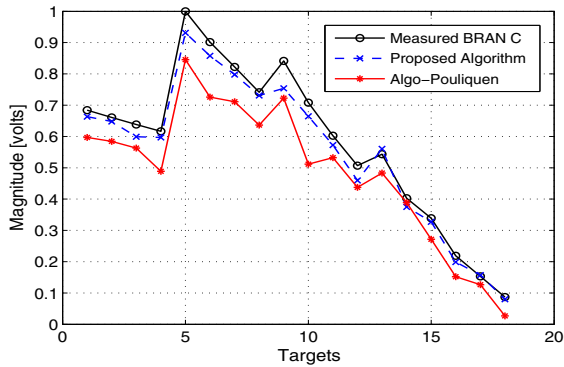
where  $\delta(n)$ ,  $\tau_i$  and  $A_i \in N(0, 1)$  denote, the Dirac function, the paths  $i$  time delay and path  $i$  magnitude respectively.

**Table 1.** Delay and magnitudes of 18 targets of BRAN C channel.

Del. $\tau_i(ns)$	Mag. $A_i(dB)$	Del. $\tau_i(ns)$	Mag. $A_i(dB)$
0	-3.3	230	-3.0
10	-3.6	280	-4.4
20	-3.9	330	-5.9
30	-4.2	400	-5.3
50	0	490	-7.9
80	-0.9	600	-9.7
110	-1.7	730	-13.2
140	-2.6	880	-16.3
180	-1.5	1050	-21.2

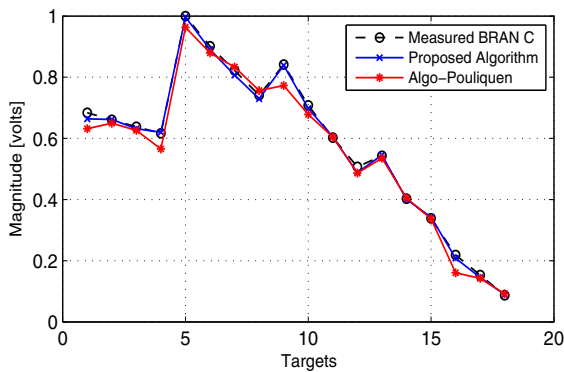
### 6.2 BRAN C Channel Identification

The estimated parameters of the BRAN C radio channel impulse response, using the proposed algorithm and Pouliquen et al. algorithm is presented in the figure (Fig. 3), for a data length  $N = 1000$  and for an  $SNR = 10dB$ . From this figure (Fig. 3) we remark that, when estimating the BRAN C channel impulse response using the proposed algorithm, there is a slight difference between the estimated values and the measured values, and an apparent difference when using Pouliquen et al. algorithm.



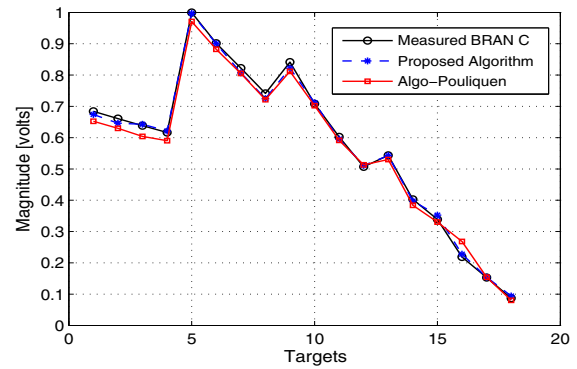
**Figure 3.** Estimation of the BRAN C channel impulse response, for a data length  $N = 1000$  and  $SNR = 10dB$ .

Figure 4 shows the estimation of the BRAN C channel impulse response parameters as a function of targets, for a data length  $N = 1000$  and for an  $SNR = 20dB$ . The impulse response is estimated with good precision using the proposed algorithm, but represents some fluctuations if we use Pouliquen et al. algorithm.



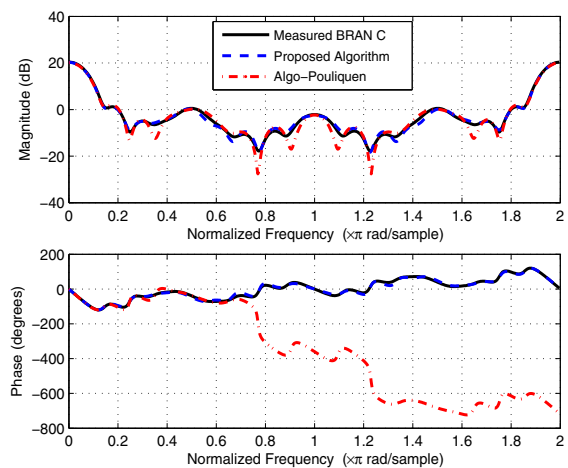
**Figure 4.** Estimation of the BRAN C channel impulse response, for a data length  $N = 1000$  and  $SNR = 20dB$ .

In the figure 5, we represent the estimation of the BRAN C radio channel impulse response, using both algorithms, for a data length  $N = 1000$ , and for an  $SNR = 30dB$ . From this result, we notice that the estimated amplitude has the same form as those of the BRAN C measured data. The impulse response parameters are estimated with a high accuracy using a proposed algorithm. For the Pouliquen, et al. algorithm we have a difference in some paths.



**Figure 5.** Estimation of the BRAN C channel impulse response, for a data length  $N = 1000$  and  $SNR = 30dB$ .

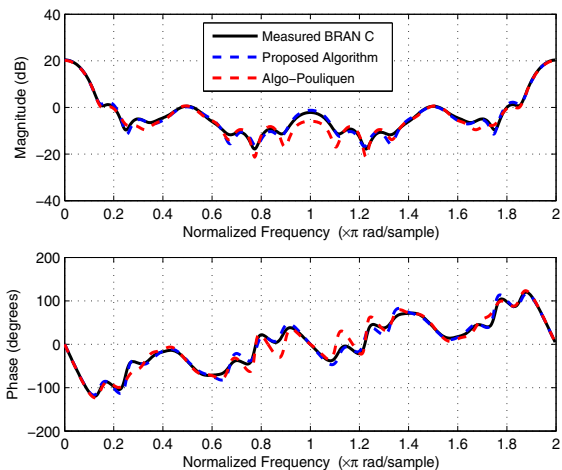
According to the figures 3, 4 and 5, we find that the proposed algorithm has good results despite the noise environment (i.e.  $SNR = 10dB$ ) compared to the Pouliquen et al. algorithm. In order to test the robustness of the proposed algorithm in the frequency domain, we show the estimated magnitude and phase of the BRAN C radio channel impulse response, in the cases of  $N = 1000$ ,  $N = 2000$  and  $N = 3000$ , in Figures 6, 7 and 8, respectively, for an  $SNR = 20dB$ , as compared to the Pouliquen et al. algorithm. From Figure 6, in the case where  $N = 1000$ , we can conclude that the estimated phase of the BRAN C channel impulse response, via the proposed algorithm, is closer to the true one. However, using the Pouliquen et al. algorithm, there is a significantly larger difference between the measured data and the estimated phases.



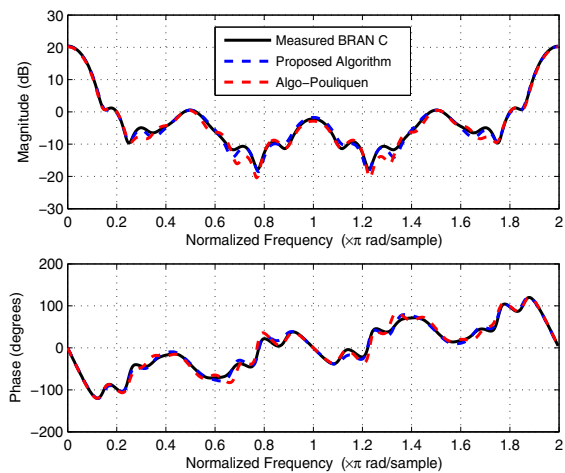
**Figure 6.** Estimated BRAN C magnitude and phase, for a data length  $N = 1000$  and  $SNR = 20dB$ .

When  $N$  is equal to 2000, we can observe that the estimated magnitude and phase of the BRAN C channel, have the same appearance and are very close to the true values when using the proposed algorithm; whereas, when using the Pouliquen et al. algorithm, there is a minor difference between the estimated and the measured ones (Fig. 7). Moreover, we note that the estimated values of

the BRANC channel using the proposed algorithm and the Pouliquen et al. algorithm are in excellent agreement with the real model and closer to the true one, respectively, when the data length is large (e.g. when we take  $N = 3000$ ) (Fig. 8). In conclusion, the input data length ( $N$ ) and the signal-to-noise ratio ( $SNR$ ) have a great impact on the estimated phase and a small impact on the estimated magnitude.



**Figure 7.** Estimated BRANC magnitude and phase, for a data length  $N = 2000$  and  $SNR = 20dB$ .



**Figure 8.** Estimated BRANC magnitude and phase, for a data length  $N = 3000$  and  $SNR = 20dB$ .

## 7 Conclusion

In this paper, we have proposed a new algorithm based on positive definite kernel for nonlinear system identification from binary measurements on the output, compared with the Pouliquen et al. algorithm. Both algorithms are used to estimate the parameters of the BRANC radio channel impulse response. In this work, experimental results shows that our proposed algorithm is effective and efficient

for identifying the amplitude and phase of the frequency selective channel impulse response (BRANC). In future work, we will develop new methods to identify communication channels (i.e. using no-trick kernel methods).

## References

- [1] A. Chiuso and G. Pillinetto, “System identification: A machine learning perspective,” Annual Review of Control, Robotics, and Autonomous Systems, vol. 2, p. 281–304 (2019).
- [2] U. Soverini and T. Söderström, “Frequency domain identification of FIR models in the presence of additive input–output noise,” Automatica, vol. 115, p. 108879 (2020).
- [3] H. Zhang, T. Wang, and Y. Zhao, “Asymptotically Efficient Recursive Identification of FIR Systems With Binary-Valued Observations,” IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019.
- [4] J. Guo, Y. Zhao, C. Y. Sun and Y. Yu, “Recursive identification of FIR systems with binary-valued outputs and communication channels,” Automatica, vol. 60, p. 165–172 (2015).
- [5] T. Yuan, Q. Liu, and J. Guo, “Identification of FIR Systems with Quantized Input and Binary-Valued Observations Under A Priori Parameter Constraint,” In : 2020 39th Chinese Control Conference (CCC). IEEE, p. 1099–1104 (2020).
- [6] C. Cortes and V. Vapnik, “Support-vector networks. Machine learning,” vol. 20, no. 3, p. 273–297 (1995).
- [7] B. Scholkopf and A. J. Smola, “Learning with kernels: support vector machines, regularization, optimization, and beyond,” Adaptive Computation and Machine Learning series, 2018.
- [8] W. Liu, P. P. Pokharel and S. Haykin, “Kernel adaptive filtering: a comprehensive introduction,” Vol. 57. John Wiley & Sons (2011).
- [9] W. Liu, P. P. Pokharel and J. C. Principe, “The kernel least-mean-square algorithm,” IEEE Transactions on Signal Processing, vol. 56, no. 2, p. 543–554 (2008).
- [10] W. Liu and J. C. Principe, “Kernel affine projection algorithms,” EURASIP Journal on Advances in Signal Processing, vol. 2008, p. 1–12 (2008).
- [11] R. Castro Garcia, “Structured nonlinear system identification using kernel-based methods,” 2017.
- [12] B. Chen, S. Zhao, P. Zhu and J. C. Principe, “Quantized kernel least mean square algorithm,” IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 1, p. 22–32 (2011).
- [13] S. Wang, Y. Zheng, and C. Ling, “Regularized kernel least mean square algorithm with multiple-delay feedback,” IEEE Signal Processing Letters, vol. 23, no. 1, p. 98–101 (2015).
- [14] B. Chen, J. Liang, N. Zheng, and J. C. Principe, “Kernel least mean square with adaptive kernel size,” Neurocomputing, vol. 191, p. 95–106 (2016).
- [15] R. Fateh and A. Darif, “Mean Square Convergence of Reproducing Kernel for Channel Identification: Ap-

- plication to Bran D Channel Impulse Response,” In : International Conference on Business Intelligence. Springer, Cham, 2021. p. 284–293.
- [16] ETSI, “Broadband radio acces network (BRAN); hyperlan type 2; physical layer,” Technical report, December 2001.
- [17] ETSI, “Broadband Radio Acces Network (BRAN); hyperlan type 2; Requirements and archtiectures for wireless broadband access,” Janvier 1999.
- [18] M. Pouliquen, T. Menard, E. Pigeon, O. Gehan, and A. Goudjil, “Recursive system identification algorithm using binary measurements,” European Control Conference, Aalborg, 2016.
- [19] N. Aronszajn, “Theory of reproducing kernels,” Transactions of the American mathematical society, vol. 68, no. 3, p. 337–404 (1950).