



# Stochastic gradient based concurrent criterion learning and parameter-estimation

Rangeet Mitra<sup>a,\*</sup>, Kwonhue Choi<sup>b</sup>, Vimal Bhatia<sup>c,d,e</sup>, Georges Kaddoum<sup>f</sup>

<sup>a</sup> CMR Institute of Technology, AECS Layout, Bengaluru, 560037, Karnataka, India

<sup>b</sup> Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, South Korea

<sup>c</sup> Department of Electrical Engineering and the Centre for Advanced Electronics, IIT Indore, Indore 453552, India

<sup>d</sup> Skoda Auto University, 293 01 Mlada Boleslav, Czech Republic

<sup>e</sup> Faculty of Informatics and Management, University of Hradec Kralove, 500 03 Hradec Kralove, Czechia

<sup>f</sup> Genie Electrique ETS Montreal, 1100-Notre Dame Street Ouest, Montreal, H3C1K3, Quebec, Canada

## ARTICLE INFO

### Keywords:

Criterion learning  
Hyperparameter free learning  
Random Fourier features  
Channel equalization  
Channel estimation

## ABSTRACT

For parameter-estimation over non-Gaussian noise, several learning criteria have emerged, which are known for their hyperparameter dependence. This work proposes a random Fourier feature based online hyperparameter free criterion learning algorithm that comprehensively alleviates dependence on hyperparameter choices and learns the criterion by self-adapting to underlying noise. First, for this joint parameter and criterion estimation, dynamical equations for the proposed hyperparameter-free algorithm are derived. Next, regarding the convergence characteristics of the proposed joint hyperparameter free parameter and criterion learning, rigorous analytical results are presented. Finally, case-studies aligned with classical signal processing applications like channel-estimation and channel-equalization are provided for performance validation of the proposed algorithm, and to verify the derived convergence analysis through computer simulations.

## 1. Introduction

Non-Gaussian noise processes are known to impair several of the next generation communication circuits and systems, internet of things (IoT) applications, and in industrial/manufacturing environments [1]. Apart from communications/IoT, the additive distortion in an analog-digital (A/D) converter/quantizer circuit generally obeys a uniform distribution [2], which is another example of non-Gaussian additive noise process. Generally, the effect of these performance-degradations are especially severe when classical adaptive signal methods, typically designed for Gaussian noise-mitigation, are used for non-Gaussian noise. Works mentioning this performance degradation in non-Gaussian noise are comprehensively reviewed in [3, Sec: 1.2–1.3].

### 1.1. Related works

In this section, a summary of related works is presented in the context of adaptive signal processing over non-Gaussian noise. To mitigate non-Gaussian additive noises, several non-Bussgang learning criteria have emerged recently, such as, maximum correntropy (MCC), generalized maximum correntropy (GMCC) [4], minimum error entropy [3], and the maximum Versoria criterion (MVC) [4–6]. Apart from the

above mentioned works, several information theoretic learning methods are proposed [3]; notably, these learning methods are well known for their consideration of higher order error-statistics, for their motivation from an information theoretic standpoint, and for their robustness to underlying additive non-Gaussian noise processes. However, these criteria are also known for their dependence on hyperparameters whose optimal values depend on the shape and scale of the non-Gaussian noise. Besides, modern signal processing algorithms rely on deployment specific hyperparameter-initializations [7,8]. The best values of these hyperparameters, are well-known to depend on the statistics of the underlying noise process, which are prone to estimation-error in the presence of non-Gaussian outliers. Apart from choosing hyperparameters, the choice of the “best” information potential is also scenario-dependent in practice [3].

To address these drawbacks, the recent works introduce the hyperparameter free criterion learning using sampled random Fourier features (RFFs) [9] that generalizes over diverse additive non-Gaussian noise statistics for a given criterion, namely, minimum error entropy with Fiducial points; however, the work in [9] uses the Shannon information potential (IP) which is not essentially a generalized IP/criterion, and is not guaranteed to generalize over various non-Gaussian noises.

\* Corresponding author.

E-mail addresses: [rangeet.m@cmrit.ac.in](mailto:rangeet.m@cmrit.ac.in) (R. Mitra), [gonew@yu.ac.kr](mailto:gonew@yu.ac.kr) (K. Choi), [vbhatia@iiti.ac.in](mailto:vbhatia@iiti.ac.in) (V. Bhatia), [georges.kaddoum@etsmtl.ca](mailto:georges.kaddoum@etsmtl.ca) (G. Kaddoum).

**Table 1**List of learning criteria and  $g(\cdot)$  pairs

| Learning criterion | $g(\cdot)$                                           | Unknown hyperparameter |
|--------------------|------------------------------------------------------|------------------------|
| MCC                | $\exp(-\gamma e_n^2)e_n$                             | $\gamma$               |
| GMCC               | $\exp(-\gamma e_n^q) e_n ^{q-1}\text{sgn}(e_n)$      | $q, \gamma$            |
| MVC                | $\frac{r e_n ^{r-1}\text{sgn}(e_n)}{(1+r e_n ^r)^2}$ | $r, \tau$              |

To the best of our belief, existing works on stochastic-gradient based *first-order* hyperparameter free learning do not generally address criterion-learning and hyperparameter initialization/learning for arbitrary non-Gaussian noise processes. Therefore, in this paper, we contribute a first-order online algorithm that is capable of learning the parameter and criterion jointly without needing hyperparameters.

### 1.2. Contributions

Against the literature review in the previous subsection, we outline below the contributions of this paper as follows

- This paper proposes an online first-order RFF based hyperparameter free criterion learning that is capable of adaptively learning the IP/general criterion, without needing insights into the exact statistics of the underlying additive noise process. This is achieved by approximating both the underlying learning criterion, and parameter using RFF; where this representation is guaranteed by the Representer-theorem [10]. This RFF based criterion and the parameter is then jointly optimized by stochastic gradient descent.
- The performance of this stochastic gradient descent based method is evaluated across various case-studies in the context of adaptive signal processing.
- Furthermore, the steady-state convergence of the proposed online criterion learning algorithm is analytically studied in this paper.
- Lastly, the convergence analysis is validated by case-studies comprising of applications of signal processing [3] (like channel-estimation and channel-equalization), over noises arising in practical Internet of things (IoT) scenarios [1].

### 1.3. Organization

This paper is organized as follows: Section 2 proposes the hyperparameter free online criterion learning, Section 3 presents convergence analysis, and Section 4 presents and discusses simulation results. Finally, conclusions are drawn in Section 5.

## 2. Proposed hyperparameter free criterion learning

In this section, we formulate the proposed hyperparameter free criterion learning approach. We consider a training dataset  $\{d_n, \mathbf{x}_n\}_{n=1}^T$  consisting of  $T$  pairs of regressors,  $\mathbf{x}_n$ , and labels,  $d_n$ . From existing works, we find the following general adaptation for parameter weights,  $\mathbf{w}$ , for Bussgang type adaptive algorithms for non-Gaussian noises [4,6]:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta g(e_n) \mathbf{x}_n, \quad (1)$$

with the error,  $e_n = d_n - \mathbf{w}_n^T \mathbf{x}_n$ , and the nature of  $g(\cdot)$  depends on the learning criterion, and  $\eta$  denotes the step-size. A summary of learning criteria and corresponding  $g(e_n)$  is summarized in Table 1, with hyperparameters  $\gamma, \tau, q$ , and  $r$ . However, existing algorithms fail to generalize the best  $g(\cdot)$  and its corresponding hyperparameters for arbitrary distributions of the underlying noise processes. To mitigate this drawback, we formulate an algorithm in this letter that allows for hyperparameter free learning of  $g(\cdot)$ .

For hyperparameter-free RFF based criterion estimation [9,11], we use the following representation for the criterion  $J$  by invoking the

representer theorem [10]:

$$J(e_n; \mathbf{w}_n, \mathbf{\Omega}_n) = \hat{f}(e_n^2) + e_n^2 \approx \mathbf{\Omega}_n^T \mathbf{\Phi}_{\alpha, \beta}(e_n^2) + e_n^2, \quad (2)$$

where  $\mathbf{\Omega}$  denotes the weights for criterion learning,  $\hat{f}$  denotes the estimate of the criterion, and  $\mathbf{\Phi}_{\alpha, \beta} \in \mathbb{R}^{n_G}$  denotes the sampled hyperparameter free RFFs [9–11] with hyperparameters  $\alpha$  and  $\beta$ , and with  $n_G$  denoting the number of RFF.

$$\mathbf{\Phi}_{\alpha, \beta}(\cdot) = \sqrt{\frac{2}{n_G}} \begin{bmatrix} \cos(h^{(1)} \xi^{(1)T}(\cdot) + b^{(1)}) \\ \vdots \\ \cos(h^{(n_G)} \xi^{(n_G)T}(\cdot) + b^{(n_G)}) \end{bmatrix}, \quad (3)$$

where each  $\xi^{(i)} \sim \mathcal{N}(0, \mathbf{I}_n)$ ,  $b^{(i)} \sim U(-\pi, \pi)$ , and each  $h^{(i)} \sim \sqrt{\Gamma(\alpha, \beta)}$ , with  $\Gamma(\alpha, \beta)$  denoting the Gamma distribution with shape-parameter  $\alpha$ , and rate-parameter  $\beta$ .

Physically, when there is an outlier due to non-Gaussian noise processes, it gets mapped to a different feature in RFF-domain  $\mathbf{\Phi}_{\alpha, \beta}$ . Then the transformation  $\mathbf{\Omega}$  adapts itself to a weight-value such that the steady-state error is mitigated by outliers. The following values/initialization methods of  $\alpha$  and  $\beta$  are needed to ensure a unimodal error within an infinitesimal neighborhood  $\epsilon$ , and with a process-noise  $\sigma_v^2$  [9, Eq. (10–11)]:

Method I:

$$\alpha = \frac{1}{2} \quad \beta = \frac{\sigma_v^2}{4\pi^2}, \quad (4)$$

Method II:

$$\alpha = \frac{1}{2} \quad \beta = \frac{\epsilon}{4\pi^2}, \quad (5)$$

As for choice of  $n_G$ , from [12, Eq. (1)], we can infer that for a given approximation-error floor  $\epsilon_1$ ,  $n_G$  can be as large as possible, since,  $n_G = \mathcal{O}(\frac{1}{\epsilon_1} \log(\frac{1}{\epsilon_1}))$ . This is mostly a trade-off between  $n_G$ /computational resources, and the solicited approximation error  $\epsilon_1$  [12]. Existing works find that 50–500  $n_G$  are enough for practical purposes to estimate an unknown function [12,13]. We obtain the following equations for the stochastic gradient based adaptation for the concurrent adaptation for the parameter weights,  $\mathbf{w}$ , and the criterion-learning weights,  $\mathbf{\Omega}$ :

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta \nabla_{\mathbf{w}_n} J, \quad \mathbf{\Omega}_{n+1} = \mathbf{\Omega}_n + \mu \nabla_{\mathbf{\Omega}_n} J, \quad (6)$$

$$\begin{aligned} \mathbf{w}_{n+1} &= \mathbf{w}_n + \eta(e_n + \mathbf{\Omega}_n^T \mathbf{\Phi}'_{\alpha, \beta}(e_n^2) e_n) \mathbf{x}_n, \\ \mathbf{\Omega}_{n+1} &= \mathbf{\Omega}_n + \mu \mathbf{\Phi}'_{\alpha, \beta}(e_n^2) \odot \mathbf{\Phi}_{\alpha, \beta}(e_n^2) e_n, \end{aligned} \quad (7)$$

with  $\odot$  denoting the Hadamard product,  $\nabla$  is a partial projected gradient, and  $(\cdot)'$  denoting the derivative. Physically, this implies that  $\mathbf{\Omega}$  is adapted such that  $\mathbf{\Omega}$  reaches an optimal value by minimization with respect to  $\mathbf{\Omega}$ , and also aligns with optimal  $\mathbf{w}$  via simultaneous partial derivative with respect to  $\mathbf{w}$ . On the other hand,  $\mathbf{w}$  needs to be adapted only to minimize the cost-function  $J$ . We assume a dataset  $D = \{d_n, \mathbf{x}_n\}_{n=0}^{T-1}$ , using which the proposed algorithm is summarized in Algorithm 1.

**Algorithm 1** Proposed hyperparameter free joint parameter and criterion learning

Initialize  $\mathbf{w}_0 = \mathbf{0}$ ,  $\mathbf{\Omega}_0 = -\mathbf{1}$ , train dataset  $D = \{d_n, \mathbf{x}_n\}_{n=0}^{T-1}$ , number of iterations  $T$ .

**for**  $n=0:T-1$  **do**

Calculate error  $e_n = d_n - \mathbf{w}_n^T \mathbf{x}_n$ .

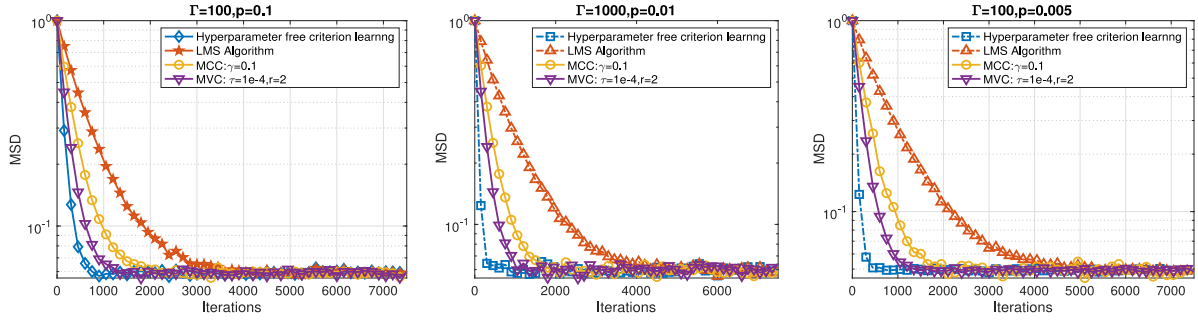
Adapt parameter weights  $\mathbf{w}_n$  as:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \eta(e_n + \mathbf{\Omega}_n^T \mathbf{\Phi}'_{\alpha, \beta}(e_n^2) e_n) \mathbf{x}_n$$

Adapt criterion weights  $\mathbf{\Omega}_n$  as:

$$\mathbf{\Omega}_{n+1} = \mathbf{\Omega}_n + \mu \mathbf{\Phi}'_{\alpha, \beta}(e_n^2) \odot \mathbf{\Phi}_{\alpha, \beta}(e_n^2) e_n$$

**end for**



**Fig. 1.** MSD evolution for the proposed hyperparameter free criterion learning for channel equalization for (a)  $\Gamma = 100, p = 0.1$ , (b)  $\Gamma = 1000, p = 0.01$ , and (c)  $\Gamma = 100, p = 0.005$ .

### 3. Convergence analysis

In this section, we derive convergence analysis for the proposed hyperparameter free criterion learning. We denote the optimal value of a parameter as  $(\cdot)^o$ , and the deviation from the optimal value as  $\tilde{\cdot}$ . This allows for re-expression of (7) as follows:

$$\begin{aligned}\tilde{\mathbf{w}}_{n+1} &= \tilde{\mathbf{w}}_n - \eta(\tilde{\mathbf{w}}_n^T \mathbf{x}_n + v_n)(1 + \mathbf{\Omega}_n^T \Phi'_{\alpha,\beta}(e_n^2))\mathbf{x}_n, \\ \tilde{\mathbf{\Omega}}_{n+1} &= \tilde{\mathbf{\Omega}}_n - \frac{\mu}{2} \Phi_{\alpha,\beta}(2e_n^2)(\tilde{\mathbf{w}}_n^T \mathbf{x}_n + v_n),\end{aligned}\quad (8)$$

with  $v_n$  denoting the underlying noise with variance  $\sigma_v^2$ . For further analysis, we denote

$$K(\mathbf{\Omega}_n) = \mathbb{E}[(1 + \mathbf{\Omega}_n^T \Phi'_{\alpha,\beta}(e_n^2))^2] \approx 1 + \|\mathbf{\Omega}_n\|^2, \quad M(\mathbf{\Omega}_n) = \mathbb{E}[(1 + \mathbf{\Omega}_n^T \Phi'_{\alpha,\beta}(e_n^2))] \quad (9)$$

with  $\mathbb{E}[\cdot]$  denoting the statistical expectation operator, and  $\sigma_e^2$  denoting the excess mean squared error.

We present the following justifications for derivation of  $K(\mathbf{\Omega}_n)$  and  $M(\mathbf{\Omega}_n)$ :

- *Justification for  $K(\mathbf{\Omega}_n)$ :* We note that

$$K(\mathbf{\Omega}_n) = \mathbb{E}[(1 + \mathbf{\Omega}_n^T \Phi'_{\alpha,\beta}(e_n^2))^2].$$

Expanding and noting  $\mathbf{m}_{\alpha,\beta}' = 0$ , and  $\mathbf{R}_{\alpha,\beta}' = \mathbb{I}_{n_G \times n_G}$ , where  $\mathbf{m}_{\alpha,\beta}'$  is the mean of  $\Phi'_{\alpha,\beta}(e_n^2)$ , and  $\mathbf{R}_{\alpha,\beta}'$  denotes the autocorrelation matrix of  $\Phi'_{\alpha,\beta}$ , we get:

$$K(\mathbf{\Omega}_n) = 1 + \|\mathbf{\Omega}_n\|^2.$$

- *Justification for  $M(\mathbf{\Omega}_n)$ :* We note that  $[\mathbf{1}^T \Phi'_{\alpha,\beta}(e_n^2)] = 0$ , which implies  $\mathbf{\Omega}^o = \delta \mathbf{1} - \mathbf{\Omega}$ , where  $\delta$  is an arbitrary scaling factor, and  $\mathbf{1}$  is the all ones vector. At convergence, we approximate for  $\delta = -1$ :

$$M(\mathbf{\Omega}_\infty) \approx \mathbb{E}[1 + \sum_{i=1}^{n_G} \sqrt{\frac{2}{n_G}} h^{(i)} \sin(h^{(i)} e_n^2 + b^{(i)})] \quad (10)$$

Noting that  $h^{(i)}$  and  $b^{(i)}$  are drawn from Gamma and uniform distributions respectively, with parameter choices  $\frac{1}{2}$  and  $\frac{\epsilon}{4\pi^2}$  (as given in Method II), we get:

$$\mathbb{E}[M(\mathbf{\Omega}_\infty)] \approx 1 + \sqrt{\frac{n_G}{2}} \frac{\epsilon}{2\pi^2} \mathbb{E}\left[\cos\left(\frac{\pi}{2} - [h^{(i)} e_n^2 + b^{(i)}]\right)\right], \quad (11)$$

or, by taking series at  $\frac{\pi}{2}$ ,

$$\mathbb{E}[M(\mathbf{\Omega}_\infty)] \approx 1 + \sqrt{\frac{n_G}{2}} \frac{\epsilon}{2\pi^2} \mathbb{E}\left[1 - \frac{(\pi/2 - h^{(i)} e_n^2 - b^{(i)})^2}{2}\right], \quad (12)$$

Assuming small error at convergence, and noting that  $h^{(i)}$  and  $b^{(i)}$  are drawn from Gamma distribution,  $\Gamma[\alpha, \beta]$ , and uniform distribution,  $U[-\pi, \pi]$ , respectively, we find:

$$\mathbb{E}[M(\mathbf{\Omega}_\infty)] \leq 1 + \sqrt{\frac{n_G}{2}} \frac{\epsilon}{2\pi^2} \mathbb{E}\left[1 - \frac{(\pi/2 - \alpha^{(i)} e_n^2 - \beta^{(i)})^2}{2}\right] \quad (13)$$

$$\leq 1 + \sqrt{\frac{n_G}{2}} \frac{\epsilon}{2} \times \frac{17}{24}$$

Squaring both sides of (7), we get the following:

$$\begin{aligned}\mathbb{E}[\|\tilde{\mathbf{w}}_{n+1}\|^2] &= (1 - 2\eta M(\mathbf{\Omega}_n) \mathbb{E}[\|\mathbf{x}_n\|^2] + \eta^2 \mathbb{E}[\|\mathbf{x}_n\|^4] K(\mathbf{\Omega}_n)) \mathbb{E}[\|\tilde{\mathbf{w}}_n\|^2] \\ &\quad + \eta^2 \sigma_v^2 K(\mathbf{\Omega}_n) \mathbb{E}[\|\mathbf{x}_n\|^2].\end{aligned}\quad (14)$$

Further, assuming  $\tilde{\mathbf{\Omega}}_n$  and  $\tilde{\mathbf{w}}_n$  uncorrelated, we have,

$$\mathbb{E}[\|\tilde{\mathbf{\Omega}}_{n+1}\|^2] = \sum_{i=0}^n \frac{\mu^2}{4} \mathbb{E}[(\|\mathbf{x}_i\|^2 \|\tilde{\mathbf{w}}_i\|^2 + \sigma_v^2)].$$

This implies that the step-size range to guarantee convergence of the proposed hyperparameter free criterion learning is  $0 < \eta < \frac{2M(\mathbf{\Omega}_\infty)}{\mathbb{E}[\|\mathbf{x}_n\|^2]}$ . Further, from (14), the mean squared deviation (MSD) of the proposed approach can be quantified as below:

$$\mathbb{E}[\|\tilde{\mathbf{w}}_\infty\|^2] = \frac{\eta \sigma_v^2 (1 + \mathbb{E}[\|\mathbf{\Omega}_\infty\|^2])}{2M(\mathbf{\Omega}_\infty) - \eta \mathbb{E}[\|\mathbf{x}_n\|^2]} \approx \frac{\eta \sigma_v^2}{2M(\mathbf{\Omega}_\infty) - \eta \mathbb{E}[\|\mathbf{x}_n\|^2]}, \quad (15)$$

where  $M(\mathbf{\Omega}_\infty)$  is derived in (13). For a general noise distribution with variance  $\sigma_v^2$ , the above equation is a max-entropy approximation for steady state MSD [14]. Note that for large number of RFF,  $n_G$ , the proposed algorithm converges to a lower MSD compared to conventional LMS algorithm.

This result indicates that the proposed approach converges to a lower MSE floor as compared to the least mean squares (LMS) algorithm [3] for a given step-size  $\eta$ . Lastly, from (14), it is trivial to note that the time constant,  $\tau$ , for the proposed approach is given as  $\tau \leq \frac{1}{2\eta \mathbb{E}[M(\mathbf{\Omega}_\infty)] \lambda_{\min}}$ , with  $\lambda_{\min}$  denoting the least eigenvalue of the autocorrelation matrix corresponding to the regressors  $\mathbf{x}_n$ . The assumption of infinitesimal steady-state error assures faster convergence than the LMS algorithm for scenarios impaired by non-Gaussian noise processes without requiring scenario-dependent hyperparameter initializations. We can note that MCC and MVC can also attain the performance of the proposed algorithm; however it will require precise hyperparameter initialization to achieve this.

As for the computational complexity of this approach, we infer from Eq. (7), that the overall computational complexity of the proposed approach is  $\mathcal{O}(n_G)$ , where  $n_G$  is the number of RFF and  $\mathcal{O}(\cdot)$  denotes the big O notation. In the next section, we present case-studies to validate the proposed hyperparameter free algorithm.

### 4. Simulation results

In this section, we provide simulation based case studies to validate the proposed algorithm. We compare our algorithms with respect to MVC as it is the “best” algorithm compared to MCC and GMCC as reported in [4]. We use the MSD,  $\mathbb{E}[\|\tilde{\mathbf{w}}\|^2]$ , to evaluate all the proposed algorithms. We also notice from simulations presented in this section, that MVC needs hyperparameters  $r, \tau$ , and its convergence severely depends on its proper choice. Below, we split our results into three

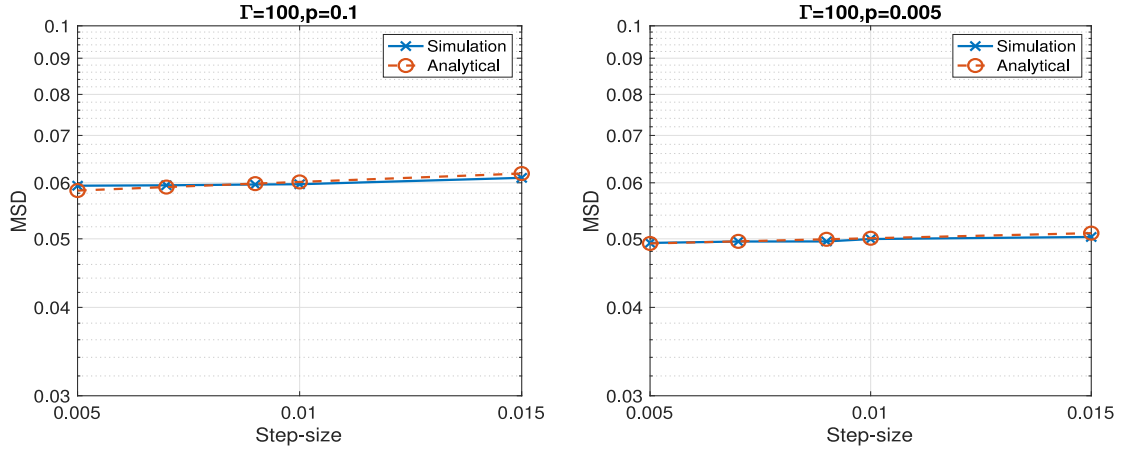


Fig. 2. Validation of MSD vs.  $\eta$  for various values of  $\Gamma, p$  denoting various impulsive noise scenarios.

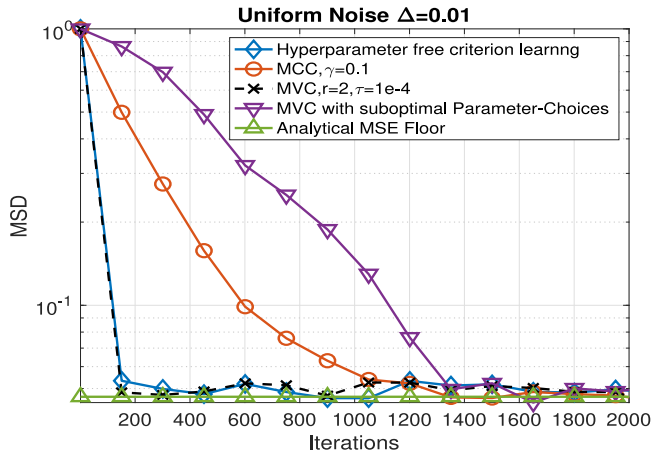


Fig. 3. MSD comparison for channel-estimation over uniform noise with  $\Delta = 0.01$ .

case case-studies: (a) Channel Equalization under impulsive noise, (b) Channel Estimation in the Presence of Quantization Noise, and (c) Channel Estimation in the Presence of  $\alpha$ -stable noise.

#### 4.1. Evaluation of convergence-speed and validation of steady-state convergence-analysis for channel equalization in the presence of impulsive noise

In this case study, we consider a two-tap channel impulse response,  $\mathbf{h} = [1, 0.5]$ . Further, we denote the sequence of training symbols as  $\mathbf{s}$ , and the impulsive noise-process as  $\mathbf{v}$ , and express the output sequence  $\mathbf{y}$  as,  $\mathbf{y} = \mathbf{h} \otimes \mathbf{s} + \mathbf{v}$ , where  $\otimes$  denotes linear convolution, and  $\mathbf{v}$  is drawn from a bimodal Bernoulli Gaussian noise model whose severity is parameterized by its parameters  $\Gamma, p$  [1]. At the receiver, regressor for the T-tap equalizer,  $\mathbf{x}_n$  is denoted as  $\mathbf{x}_n = [y_n, y_{n-1}, \dots, y_{n-T+1}]^T$ . For this scenario, case-studies are presented in Fig. 1 for various levels of impulsive noise corresponding to different  $\Gamma, p$ . From Fig. 1, it is observed that the proposed hyperparameter free criterion learning exhibits similar convergence characteristics for different levels of impulsive noise. From Fig. 1, it is further observed that the proposed algorithm exhibits faster convergence than the LMS algorithm, MCC and the MVC for various impulse noise conditions (parameterized by  $\Gamma, p$ ); it is to be noted that the step-sizes for respective algorithms are chosen such that all the algorithms achieve the same error-floor, and convergence speed is compared for all the considered algorithms.

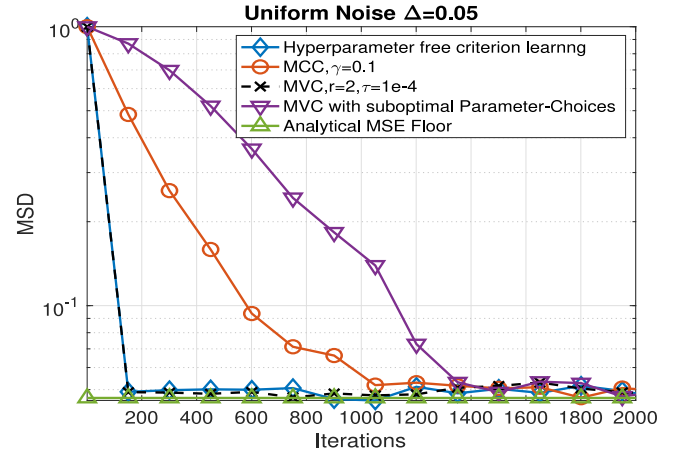


Fig. 4. MSD comparison for channel-estimation over uniform noise with  $\Delta = 0.05$ .

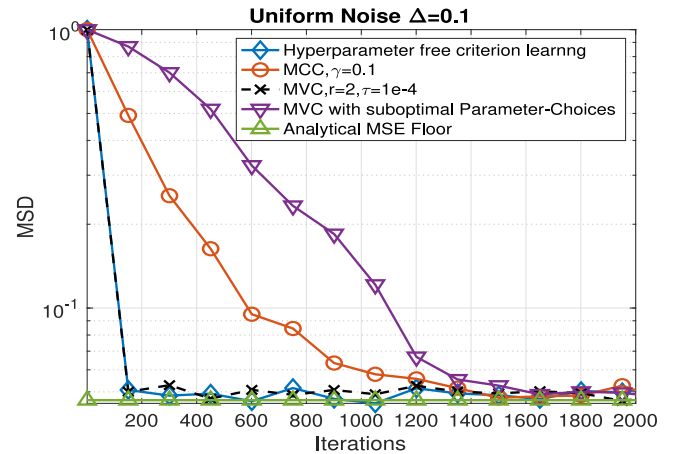


Fig. 5. MSD comparison for channel-estimation over uniform noise with  $\Delta = 0.1$ .

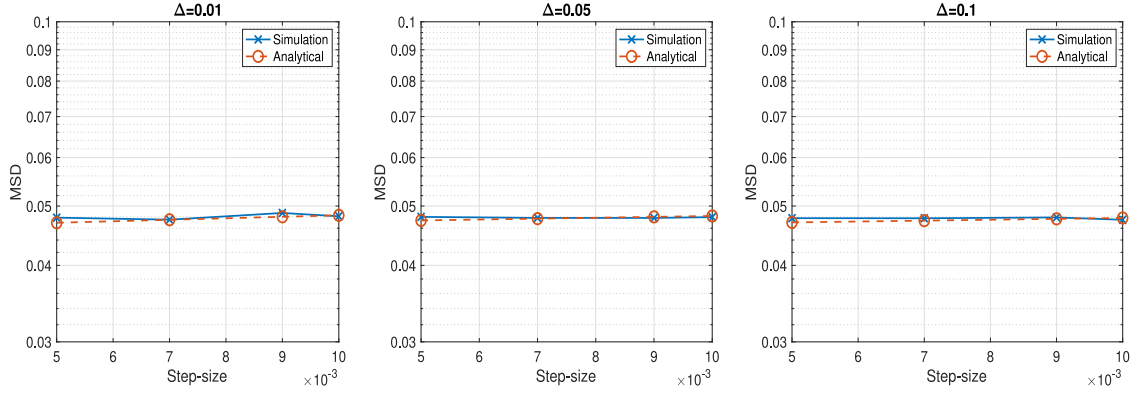


Fig. 6. MSD validation for uniform quantization noise for various step-sizes.

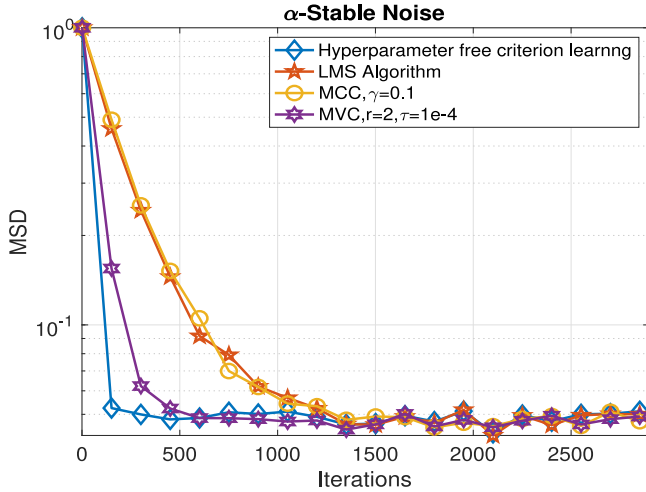


Fig. 7. MSD comparison for  $\alpha$ -stable noise with  $\alpha = 1.4$ .

Furthermore, from Fig. 2, it is found that the analytical expression for steady state MSD derived in Eq. (15) bears a close match with the simulated steady-state error for various impulsive noise levels.

#### 4.2. Evaluation of convergence-speed and validation of steady-state convergence-analysis for channel estimation in the presence of quantization noise

In this section, we demonstrate computer simulations for convergence of the proposed hyperparameter free criterion learning in the presence of quantization noise in the context of channel-estimation. We use the well-known uniform distribution of quantization noise which we denote as  $U(-\Delta, \Delta)$  [2]. Simulations are shown for value of  $\Delta = 0.01$  in Fig. 3,  $\Delta = 0.05$  in Fig. 4, and  $\Delta = 0.1$  in Fig. 5. From these figures, it is observed that MVC relies on appropriate hyperparameter-initialization for their convergence. Upon inaccurate hyperparameter-choices, the convergence characteristics of MVC degrades as seen from Figs. 3–5. However, with appropriate hyperparameter-choices, the convergence characteristics of the proposed hyperparameter free algorithm aligns with the convergence characteristics of MVC with its best  $\tau, r$  initialization. This makes the proposed hyperparameter free criterion learning viable for channel-estimation in the presence of quantization noises arising in A/D converters.

Next, in Fig. 6 we plot steady-state error-floors as a function of step-size and study its agreement with the analytical expression derived in Eq. (15). From Fig. 6, we find that the presented analytical equation for

steady-state error in Eq. (15) agrees well under uniform Quantization noise for a wide range of  $\Delta = 0.01, 0.05, 0.1$ , which validates the analysis.

#### 4.3. Evaluation of convergence-speed for channel estimation in the presence of $\alpha$ -stable noise

In this case-study, we present performance-evaluation of the proposed hyperparameter-free approach against the LMS, MCC, and MVC, in the presence of  $\alpha$ -stable noise [15]. By setting  $\alpha = 1.4$ , the said comparison is illustrated in Fig. 7, where we find that MCC and LMS exhibit the same performance due to improper initialization of spread-parameter of MCC for the considered  $\alpha$ -stable noise-process. Furthermore, the proposed hyperparameter-free criterion learning algorithm is found to outperform LMS, MCC, and MVC, which makes it generalized for unknown additive non-Gaussian noise-processes.

To summarize, we can infer from these simulations, and from the Representer theorem [10], that the proposed hyperparameter-free algorithm can generalize over arbitrary unknown noise distributions, e.g., Gauss-Bernoulli, Uniform, and  $\alpha$ -stable noise-distributions.

## 5. Conclusion

In this paper, a first order online hyperparameter free concurrent criterion learning and parameter estimation was proposed to alleviate dependence on hyperparameter-choices in existing cost-functions for adaptive filtering in the presence of unknown non-Gaussian noise. Relevant analytical results were derived for convergence analysis of the proposed algorithm. The derived analysis was validated via simulations considering diverse additive noises arising in practical IoT systems. The distribution-independent generalization of the proposed approach to unknown noise distributions makes it better suited for estimation in the presence of non-Gaussian noises.

#### CRedit authorship contribution statement

**Rangeet Mitra:** Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Kwonhue Choi:** Writing – review & editing. **Vimal Bhatia:** Writing – review & editing, Supervision, Project administration. **Georges Kaddoum:** Writing – review & editing, Supervision, Project administration.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] B. Selim, M.S. Alam, G. Kaddoum, B.L. Agba, Effect of impulsive noise on uplink NOMA systems, *IEEE Trans. Veh. Technol.* 69 (3) (2020) 3454–3458.
- [2] C.W. Ling, C.T. Li, Vector quantization with error uniformly distributed over an arbitrary set, *IEEE Trans. Inform. Theory* 70 (7) (2024) 5392–5407.
- [3] B. Chen, Y. Zhu, J. Hu, J.C. Principe, *System Parameter Identification: Information Criteria and Algorithms*, Newnes, 2013.
- [4] F. Huang, J. Zhang, S. Zhang, Maximum versoria criterion-based robust adaptive filtering algorithm, *IEEE Trans. Circuits Syst. II Express Briefs* 64 (10) (2017) 1252–1256.
- [5] B. Chen, X. Wang, Y. Li, J.C. Principe, Maximum correntropy criterion with variable center, *IEEE Signal Process. Lett.* 26 (8) (2019) 1212–1216.
- [6] B.-X. Su, F.-Y. Wu, K.-D. Yang, T. Tian, Y.-Y. Ni, Steady-state mean-square performance analysis of the block-sparse maximum versoria criterion, *Signal Process.* 213 (2023) 109186.
- [7] X. Cheng, D. Liu, C. Wang, S. Yan, Z. Zhu, Deep learning-based channel estimation and equalization scheme for FBMC/OQAM systems, *IEEE Wirel. Commun. Lett.* 8 (3) (2019) 881–884.
- [8] A.S.M. Mohammed, A.I.A. Taman, A.M. Hassan, A. Zekry, Deep learning channel estimation for OFDM 5G systems with different channel models, *Wirel. Pers. Commun.* 128 (4) (2023) 2891–2912.
- [9] R. Mitra, G. Kaddoum, D.B.d. Costa, Hyperparameter free MEEF-based learning for next generation communication systems, *IEEE Trans. Commun.* 70 (3) (2022) 1682–1696.
- [10] B. Schölkopf, R. Herbrich, A.J. Smola, A generalized representer theorem, in: *International Conference on Computational Learning Theory*, Springer, 2001, pp. 416–426.
- [11] R. Mitra, G. Kaddoum, V. Bhatia, Hyperparameter-free transmit-nonlinearity mitigation using a kernel-width sampling technique, *IEEE Trans. Commun.* 69 (4) (2020) 2613–2627.
- [12] A. Rahimi, B. Recht, Random features for large-scale kernel machines, *Adv. Neural Inf. Process. Syst.* 20 (2007).
- [13] P. Bouboulis, S. Pougkakiotis, S. Theodoridis, Efficient KLMS and KRLS algorithms: A random Fourier feature perspective, in: *2016 IEEE Statistical Signal Processing Workshop, SSP, IEEE*, 2016, pp. 1–5.
- [14] T.M. Cover, *Elements of Information Theory*, John Wiley & Sons, 1999.
- [15] W. Ma, J. Duan, W. Man, H. Zhao, B. Chen, Robust kernel adaptive filters based on mean p-power error for noisy chaotic time series prediction, *Eng. Appl. Artif. Intell.* 58 (2017) 101–110.