



HAL
open science

LLR estimation using machine learning

Latifa Mostari, Alban Goupil, Abdelmalik Taleb-Ahmed

► **To cite this version:**

Latifa Mostari, Alban Goupil, Abdelmalik Taleb-Ahmed. LLR estimation using machine learning. Alexandria Engineering Journal, 2024, 105, pp.230 - 236. 10.1016/j.aej.2024.06.057 . hal-04664447

HAL Id: hal-04664447

<https://hal.science/hal-04664447v1>

Submitted on 30 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Original article

LLR estimation using machine learning

Latifa Mostari^{a,*}, Alban Goupil^b, Abdelmalik Taleb-Ahmed^c

^a LACOSI, Electronics Department, University of Sciences and Technology of Oran, USTO-MB, Algeria

^b CReSTIC, University of Reims Champagne-Ardenne, 51100 Reims, France

^c IEMN DOAE, UMR CNRS 8520, Polytechnic University of Hauts de France, Valenciennes, France

ARTICLE INFO

Keywords:

LLR
High order modulations
BPSK
Estimation
Approximation
Machine learning
Impulsive noise

ABSTRACT

Many decoders of error-correcting codes use the Log-Likelihood Ratio (LLR) as an input, which involves the probability density function (pdf) of the noise. In impulsive noise, the pdf of the noise is not accessible in closed form and is only available through very complex numerical computation. Therefore, the LLR calculation for Binary Phase Shift Keying (BPSK) is too complex. It becomes even more complex for high-order modulations. Moreover, the LLR computational complexity grows as the modulation order increases. The main contribution of our work lies in the LLR approximation for high-order modulations and its estimation using supervised machine learning, without requiring prior knowledge of the noise distribution model. To this end, we propose two approaches to approximate the LLR values using supervised machine learning, for high-order modulated symbols. The first approach can also be used for BPSK modulated symbols. The second approach aims to approximate the LLR for high-order modulated symbols in a more simplified manner compared to the first approach. For both approaches, we estimate the parameters of the approximate LLR under known noise channel conditions using the linear regression algorithm. To estimate these parameters without prior knowledge of the noise distribution model, we use a binary logistic regression algorithm. Our simulations focus on the second proposed approach to estimate the LLR with unknown noise distributions. The results are presented for the 4-ASK (Amplitude Shift Keying) modulation scheme, where the receiver is assumed to suffer from noise ranging from Gaussian to highly impulsive models. The proposed LLR estimation is shown to achieve a comparable performance to the one attained using the exact LLR function.

1. Introduction

Impulsive noise has been measured in various communication environments, including power-line communications [1], shallow underwater acoustic communication [2], and a Poisson field of interference in wireless communications [3]. The parameters of this impulsive noise must be estimated at the channel output to compute the LLR used by many decoders. The LLR computation depends on the pdf of this noise, which is not available in closed form and requires complex numerical computations. This complexity increases further for high-order modulations, and the LLR computational complexity grows as the modulation order increases. For this reason, in our research work, we propose estimating the LLR for high-order modulations with unknown noise parameters.

The impulsive nature can be accurately modeled by Symmetric α -Stable (S α S) distributions [4–7]. The S α S distribution is characterized by two parameters α and γ , where α , $0 < \alpha < 2$, is the characteristic exponent or stability index, and γ , $\gamma > 0$, is the scale factor that determines the impulsiveness and the spread of the noise respectively,

where the impulsiveness of the noise decreases with the parameter α . In wireless context, α is directly associated with the path loss exponent of the radio channel [8].

On the other hand, using high-order modulations is one of the most common methods for improving the bit rate in communications systems without increasing the required bandwidth. However, communication systems using high-order modulations require a high signal-to-noise ratio (SNR). To this end, it is beneficial to combine error correction codes such as LDPC codes [9] and turbo-codes [10] with high-order modulations. The decoder input must utilize the LLR calculation computed through demapping. Since the LLR calculation is contingent upon the chosen modulation and the pdf of the noise, estimating the noise parameters is essential for its evaluation. However, the pdf of impulsive noise is typically not available in closed form. It can be computed through numerical integration of the inverse Fourier transform of the characteristic function, which is computationally intensive. Furthermore, the computational complexity of the LLR increases rapidly with the modulation order. It presents a challenging task to approximate and

* Corresponding author.

E-mail addresses: latifa.mostari@univ-usto.dz (L. Mostari), alban.goupil@univ-reims.fr (A. Goupil), abdelmalik.taleb-ahmed@uphf.fr (A. Taleb-Ahmed).

<https://doi.org/10.1016/j.aej.2024.06.057>

Received 1 April 2024; Received in revised form 28 May 2024; Accepted 20 June 2024

Available online 5 July 2024

1110-0168/© 2024 The Authors. Published by Elsevier B.V. on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

estimate the LLR formula for high-order modulations when the noise parameters are unknown.

Many works have been proposed to approximate the LLR for BPSK, in impulsive noise. In such a scenario, the LLR of the received symbols is a non-linear function and cannot be analytically expressed requiring intensive calculations for its evaluation. It has been previously demonstrated that the LLR for BPSK modulations, under the Gaussian assumption, are not efficient enough and can be prone to performance degradation [5,11]. A possible scenario to circumvent this problem is to use an approximating function for the LLR. For example, a simple approximation using piece-wise linear segments is proposed in [12] to feed a Viterbi decoder. The clipping method presented in [13] also represents a fairly basic approximation. The reader can refer to [11,14,15] for more details on these approximation methods.

To improve the performance of the approximated LLR with easy implementation, the authors in [8] simplified the LLR formula for BPSK modulations, to a family of functions, $(\min(\frac{\sqrt{2}}{\gamma}y, 2^{\frac{\alpha+1}{\gamma}}))$, with performance close to the true LLR, for a SaS noise distribution. The use of this approximation implies the estimation of the two noise parameters α and γ , by two well-known approaches [8]: the first one proposed by Koutrovelis in [16] and the second one presented in [17] by McCulloch. In [8] the robustness of the proposed approximation in [8] has been studied in other interference models: Middleton class A noise, ϵ -contaminated noise, and a mixture of Gaussian and alpha-stable noise. The authors in [18] used two parameters a and b instead of α and γ in the proposed approximation in [8], where $a = \frac{\sqrt{2}}{\gamma}$ and $b = 2(\alpha + 1)$, and they show that these new parameters can be estimated by direct estimation of α and γ as shown in [8], or by two other methods, using a training sequence: from the estimated pdf of the noise using kernel density estimation, or by maximization of the mutual information. The authors in [19] extended the work of [18], to improve the performance, by approximating the LLR using a function with three parameters $L_{abc} = \text{sign}(y) \min(a|y|, c, b/|y|)$. In [15], the authors also proposed a novel simple approximation for BPSK modulations, which is not a function of γ . The authors in [20] estimated the LLR for BPSK modulations, with unknown impulsive noise distribution, using the concept of maximization of the mutual information proposed in [18].

However, to the best of our knowledge, the approximation and estimation of the LLR for high-order modulations have not been previously addressed in the literature. Given that the noise parameters must be estimated at the channel output to compute the LLR, there is an advantage to estimating the LLR without knowing the noise distribution. This work proposes three main contributions to address this problem, as outlined below:

- **Contribution 1 (Approach 1):** The LLR is approximated by a system of simple linear equations using a linear regression that is supervised machine learning. To approximate the LLR using simple linear regression, the LLR must be linearly related to the channel input y , with fixed channel parameters α and γ . Then the parameters of each equation are estimated via the gradient descent algorithm to minimize the mean squared error between the predicted and exact LLR values. Therefore, since the LLR is a non-linear function, we first divide the interval of y associated with the LLR into sub-intervals based on the shape of the LLR, where the LLR of each sub-interval can be well approximated by a linear equation. Thus, the LLR can be approximated by a system of simple equations. We then estimate the parameters of each equation via the gradient descent algorithm. This approach can be applied to any modulation type and order. However, as the modulation order increases, the number of equations also increases, resulting in a large number of parameters to estimate. Therefore, we propose Approach 2.
- **Contribution 2 (Approach 2):** The LLR for high-order modulation is approximated in a more simplified manner compared to the first approach. This includes 2^{2p} -Quadrature Amplitude

Modulation (QAM) and 2^n -Amplitude Shift Keying (ASK), where p and n are integers. We have demonstrated that LLRs of 2^{2p} -QAM and 2^n -ASK modulated symbols can be obtained from the LLR of BPSK. Then, the exact LLR of BPSK is replaced by its approximation using Approach 1. Therefore, Approach 2 requires the estimation of the parameters for only the approximated LLR of BPSK modulated symbols.

In both approaches, simple linear regression is used to approximate the LLR function, where the LLR is related to the channel input y , with fixed channel parameters α and γ . The noise distribution must be known; therefore, the channel parameters α and γ must be estimated at the channel output.

- **Contribution 3:** The LLR is approximated by a system of simple linear equations, without prior knowledge about the noise distribution model, using binary logistic regression that is supervised machine learning. Therefore, it is not necessary to estimate the channel parameters. We first use the simplified LLR function (Approach 2) to estimate the parameters for only the approximated LLR of BPSK modulated symbols. Then, we estimate the approximation parameters, using binary logistic regression that requires the channel output y , and its corresponding LLR decision. We consider the channel input as the LLR decision of its corresponding channel output y . However, the channel input is unknown at the receiver. Therefore, as used in [20], we generate a random input channel \tilde{x} , and the corresponding channel output is formulated as $\tilde{y} = \tilde{x} + \tilde{w}$, we then extract the noise \tilde{w} from the channel output y to obtain the same parameters estimations since these parameters depend on the noise distribution, as introduced in [20] for BPSK modulation. We have extended this work for other modulations. This is a new method to estimate the LLR for BPSK modulations.

In our simulation, we have used the second proposed approach to estimate the LLR with unknown noise distributions, for a 4-ASK modulation, assuming a receiver affected by noise ranging from Gaussian to highly impulsive models.

The remainder of the paper is organized as follows: Section 2 presents the system model and LLR computation. Section 3 introduces the proposed Approach 1, which uses simple linear regression to estimate the LLR of BPSK and high-order modulations. Section 4 presents the proposed Approach 2, where the simplified LLR of high-order modulations is used to approximate the LLR of 2^{2p} -QAM and 2^n -ASK modulations and ends up getting another LLR approximation. Section 5 presents the proposed method to estimate the LLR, using Approach 2, based on binary logistic regression, with unknown noise distributions. Then, simulation results for a 4-ASK modulation are presented in Section 5. Finally, the paper is concluded with some final remarks in Section 6.

2. System model and LLR computation

In this paper, we assume that the channel output y is modeled by $y = x + w$, where w is an independent and identically distributed noise, following a SaS distribution, and x is the input channel. At the transmitter, a binary codeword of length N is mapped by a Gray mapping into a symbols sequence from a 2^n -ary signal constellation \mathcal{X} of size $|\mathcal{X}| = 2^n$, where each block of n bits, $[b_0, b_1, \dots, b_{n-1}]$, is mapped into a symbol x from \mathcal{X} . At the receiver, LLRs are computed for each bit associated to a transmitted symbol and then used to feed the decoder input. The code used is a (c, r) -regular LDPC code of rate $R = K/N$, where c and r are column and row weights, respectively, of the regular sparse parity check matrix, of size $((N-K) \times N)$, representing this code.

The LLR for the i th transmitted bit, $b_i(x)$, associated with the transmitted symbol $x \in \mathcal{X}$, is given by

$$\text{LLR}_i(y) = \log \frac{\sum_{x \in \mathcal{X}_0^i} f_\alpha(\gamma; y - x)}{\sum_{x \in \mathcal{X}_1^i} f_\alpha(\gamma; y - x)}, \quad (1)$$

where $\mathcal{X}_j^i = \{x|b_i(x) = j\}$, $f_\alpha(\cdot)$ is the pdf of the noise w . Such pdf cannot be analytically expressed, except for special cases: Lévy and Cauchy distributions. It can be calculated numerically using the Inverse Fourier Transform of the characteristic function:

$$f_\alpha(\gamma; y) = \frac{1}{\pi} \int_0^{+\infty} \exp(-\gamma t^\alpha |t|^\alpha) \cos(yt) dt. \quad (2)$$

By letting $\alpha = 0.5, 1$, we obtain the special cases: Lévy and Cauchy distributions, respectively. For Gaussian distribution, the exponent $\alpha = 2$ and the scale factor γ is similar to the variance σ^2 , where $\gamma = \frac{\sigma}{\sqrt{2}}$. In our simulation, we involved γ as a measurement of the strength of the noise since the impulsive noise power is not well-defined and the SNR is meaningless [8].

The pdf in (2) requires intensive calculations to evaluate it. Therefore, the computation of the LLR in (1) is too complex, and the complexity increases with the order of the modulation, $|\mathcal{X}| = 2^n$. Thus, approximating the LLR by a simple formula is a significant challenge. In our work, we propose two approaches to approximate the LLR for BPSK and high-order modulations in impulsive noise environments. Our proposed approximations can be applied to any noise type. We also propose a method to estimate the LLR with unknown noise parameters.

To clearly illustrate the principle of our proposal, we explained the case of the 4-ASK modulation ($|\mathcal{X}| = 2^2$), and simulations presented in this research paper consider this modulation. The LLR expressions of 4-ASK are the same as the LLR related to the in-phase and the quadrature components of 16-QAM, where $|\mathcal{X}| = 2^4$. QAM of order 2^{2p} can be considered as two independent 2^p -ASK, modulating two carriers, in-phase and quadrature parts. The p LLR expressions related to the in-phase components (real part) and the other p LLR expressions related to the quadrature components (imaginary part) are identical. From (1), LLRs of 4-ASK, where $\mathcal{X} = \{\pm 1, \pm 3\}$, are given by

$$\text{LLR}_0(y) = \log \frac{f_\alpha(\gamma; y-1) + f_\alpha(\gamma; y-3)}{f_\alpha(\gamma; y+1) + f_\alpha(\gamma; y+3)}, \quad (3)$$

and

$$\text{LLR}_1(y) = \log \frac{f_\alpha(\gamma; y-1) + f_\alpha(\gamma; y+1)}{f_\alpha(\gamma; y-3) + f_\alpha(\gamma; y+3)}. \quad (4)$$

The LLR of BSPK modulations is given by

$$\text{LLR}_b(y) = \log \frac{f_\alpha(\gamma; y-1)}{f_\alpha(\gamma; y+1)}. \quad (5)$$

3. Approach 1: LLR approximation for BPSK and high-order modulations

This section discusses the proposed Approach 1 in detail, which employs simple linear regression to approximate the LLR function from the LLR shape, without simplifying the LLR formula. This approach can be used for any modulation type and order. Here simple linear regression algorithm is presented first. After, we present the LLR approximation for a BPSK modulation. Then, we present the LLR approximation for 4-ASK modulations.

3.1. Simple linear regression algorithm

A simple linear regression algorithm is a supervised machine learning algorithm. It predicts the output, which is linearly related to the input, where the input is a single unknown variable. Hence, this regression technique is set to find a linear relationship between the exact output, $\text{output}_{\text{exact}}$, and the associated input, which is based on known data.

This linear regression equation can be represented as: $\text{output}_{\text{predicted}} = \theta_0 \cdot \text{input} + \theta_1$, where θ_0 and θ_1 are two real variables that are to be estimated. To estimate the optimal parameters of the equation, we construct a cost function and reduce it to minimize the error between $\text{output}_{\text{predicted}}$ and the corresponding value of $\text{output}_{\text{exact}}$, and solve by implementing an optimization algorithm. In our research work, we

Table 1
Estimated parameters for BPSK.

α	γ	θ_0	θ_1	θ_2
1.8	0.682	2.172	3.084	7.383
1.6	0.617	2.527	2.954	6.084
1.4	0.552	2.807	2.790	5.239

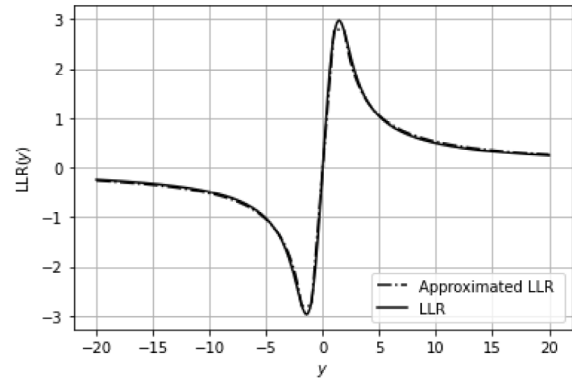


Fig. 1. Comparison between LLR of BPSK and its approximation for $\alpha = 1.4$ and $\gamma = 0.552$.

used the gradient-descent algorithm to reduce the cost function, which represents the mean squared error, given as

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m ((\theta_0 \cdot \text{input}^{(i)} + \theta_1) - \text{output}_{\text{exact}}^{(i)})^2, \quad (6)$$

where m is the length of $\text{output}_{\text{exact}}$.

To approximate the LLR from the LLR shape, using simple linear regression, the LLR must be linearly related to the channel input y , with fixed parameters α and γ . However, the LLR is a non-linear function. Therefore, we divide the interval of y associated to the LLR into sub-intervals, where the LLR of each sub-interval can be well approximated by a linear equation. Thus, the LLR can be approximated by a system of simple equations as we show, in the following, for BPSK and 4-ASK modulations.

3.2. LLR approximation for BPSK modulations

As illustrated in Fig. 1, $\text{LLR}_b(y)$ can be divided into three parts to obtain a system of simple equations. Thus, $\text{LLR}_b(y)$ can be approximated by

$$L_b(y) = \begin{cases} \theta_0 y & \text{if } |y| \leq 1.2, |y| \leq \sqrt{\frac{\theta_2}{\theta_0}}, \\ \theta_1 & \text{if } 1.2 < |y| \leq 2.3, \frac{\theta_1}{\theta_0} < |y| \leq \frac{\theta_2}{\theta_1}, \\ \text{sign}(y) \frac{\theta_2}{|y|} & \text{otherwise.} \end{cases} \quad (7)$$

Here, the estimated parameters $\theta_j, j \in \{0, 1, 2\}$, are given in Table 1. To find the last equation's parameter for $L_b(y)$, we convert this equation to a linear form. A comparison between the LLR of BPSK symbols in (5) and its approximation from (7), using the estimated parameters in Table 1 for $\alpha = 1.4$ and $\gamma = 0.552$, is presented in Fig. 1. The proposed approximation is very close to the exact computation of the LLR.

In the case of noise with a Gaussian distribution, the approximated LLR is: $1.852 \times y$, where $\sigma = 1$. The parameters are determined offline for fixed values of α and γ (σ in a Gaussian distribution).

3.3. LLR approximation for 4-ASK modulations

As we can see on Fig. 2, $\text{LLR}_0(y)$ and $\text{LLR}_1(y)$, for 4-ASK modulations, can be divided into five and four parts respectively to obtain a system

Table 2
Estimated parameters for $L_0(y)$.

α	γ	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
1.8	0.682	2.25	1.65	0.60	1.03	1.89	-0.521	6.99	20.05	-0.60
1.6	0.617	2.61	0.99	1.54	0.97	1.59	-0.82	7.44	15.69	-0.35
1.4	0.552	2.89	0.39	2.34	1.14	0.88	-1.06	7.91	13.15	-0.23

Table 3
Estimated parameters for $L_1(y)$.

α	γ	θ_a	θ_b	θ_c	θ_d	θ_e	θ_f	θ_g	θ_h
1.8	0.682	-1.11	3.47	-2.15	4.32	-0.45	-1.18	-14.75	0.83
1.6	0.617	-0.49	3.09	-2.51	5.04	0.19	-3.39	-10.84	0.57
1.4	0.552	0.061	-2.71	2.89	5.82	0.62	-4.81	-8.66	0.44

of simple equations. Thus, $LLR_0(y)$ and $LLR_1(y)$ can be approximated respectively by

$$L_0(y) = \begin{cases} \theta_0 y & \text{if } |y| \leq 1, \\ \text{sign}(y) \cdot (\theta_1 |y| + \theta_2) & \text{if } 1 < |y| \leq 2, \\ \text{sign}(y) \cdot (\theta_3 |y| + \theta_4) & \text{if } 2 < |y| \leq 3.2, \\ \text{sign}(y) \cdot (\theta_5 |y| + \theta_6) & \text{if } 3.2 < |y| \leq 4.8, \\ \text{sign}(y) \cdot \left(\frac{\theta_7}{|y|} + \theta_8\right) & \text{otherwise,} \end{cases} \quad (8)$$

and

$$L_1(y) = \begin{cases} \theta_a |y| + \theta_b & \text{if } 0 \leq |y| \leq 1, \\ \theta_c |y| + \theta_d & \text{if } 1 < |y| \leq 3, \\ \theta_e |y| + \theta_f & \text{if } 3 < |y| \leq 4.8, \\ \frac{\theta_g}{|y|} + \theta_h & \text{otherwise,} \end{cases} \quad (9)$$

Here, the estimated parameters $\theta_j, j \in \{0, 1, \dots, 8\}$, and $\theta_k, k \in \{a, b, \dots, h\}$ are given in Tables 2 and 3 respectively. Fig. 2 shows that the LLRs of 4-ASK, $LLR_0(y)$ in (3) and $LLR_1(y)$ in (4), are very close to their approximations, $L_0(y)$ in (8) and $L_1(y)$ in (9).

Moreover, as the modulation order increases, the number of equations also increases (from the LLR shape), giving rise to enormous parameters. Therefore, we propose Approach 2, which simplifies the approximation of the LLR for high-order modulations. It requires the parameter estimation for only the approximated LLR of BPSK modulated symbols. This methodology is explained in the following section:

4. Approach 2: LLR approximation for high-order modulations

This section discusses the proposed Approach 2 in detail. We first introduce the LLR simplification for high-order modulations. The simplified LLR is then used to extract the LLR of 2^{2p} -QAM and 2^n -ASK modulated symbols from the LLR of BPSK. Following that, the approximated LLR of the 4-ASK modulation, as a function of the approximated LLR of BPSK modulations, is shown.

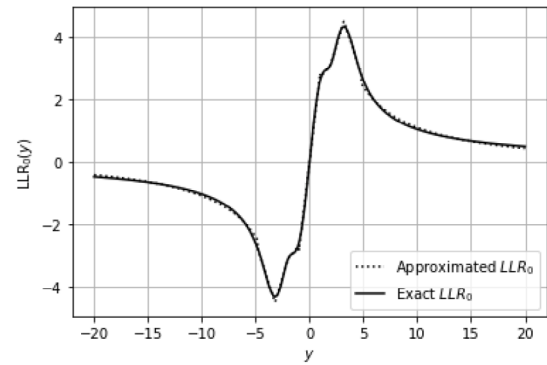
4.1. LLR simplification for high-order modulations

The LLR of high-order modulations in (1), can be reformulated as

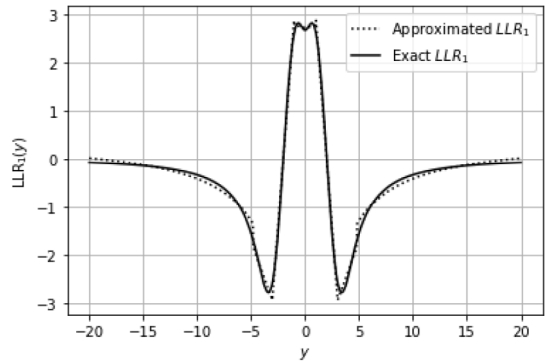
$$LLR_i(y) = \log \frac{\Re \int_0^{+\infty} \sum_{x \in \mathcal{X}_0^i} \exp(-\gamma^\alpha |t|^\alpha - jt(y-x)) dt}{\Re \int_0^{+\infty} \sum_{x \in \mathcal{X}_1^i} \exp(-\gamma^\alpha |t|^\alpha - jt(y-x)) dt} \quad (10)$$

Using $\sum_j \exp(-x_j) \approx \exp(-\min_j |x_j|)$, the LLR in (10) can be simplified as

$$\begin{aligned} LLR_i(y) &\approx \log \frac{\int_0^{+\infty} \exp(-\gamma^\alpha |t|^\alpha) \cos(\min_{x \in \mathcal{X}_0^i} |y-x| t) dt}{\int_0^{+\infty} \exp(-\gamma^\alpha |t|^\alpha) \cos(\min_{x \in \mathcal{X}_1^i} |y-x| t) dt} \\ &\approx \log \frac{f(\min_j |y - a_{i,j}^0|)}{f(\min_j |y - a_{i,j}^1|)}. \end{aligned} \quad (11)$$



(a)



(b)

Fig. 2. Comparisons between LLRs of 4-ASK and their approximations (Approach 1) for $\alpha = 1.4$ and $\gamma = 0.552$: (a) $LLR_0(y)$, (b) $LLR_1(y)$.

Therefore, the simplified LLR can be written as

$$LLR_i(y) \approx \log \frac{f_\alpha(\gamma; \min_{x \in \mathcal{X}_0^i} |y-x|)}{f_\alpha(\gamma; \min_{x \in \mathcal{X}_1^i} |y-x|)}. \quad (12)$$

In the following, we use this simplified version to obtain the LLR of 2^{2p} -QAM and 2^n -ASK modulated symbols from the LLR of BPSK. We explain the case of 4-ASK modulations.

4.2. LLR approximation of 2^{2p} -QAM and 2^n -ASK modulations

Using the simplified LLR in (12), $LLR_0(y)$ in (3) and $LLR_1(y)$ in (4) for 4-ASK can be simplified respectively as

$$LLR_0(y) \approx \begin{cases} \log \frac{f_\alpha(\gamma; y-1)}{f_\alpha(\gamma; y+1)} & \text{if } |y| \leq 2, \\ \text{sign}(y) \cdot \log \frac{f_\alpha(\gamma; |y|-3)}{f_\alpha(\gamma; |y|+1)} & \text{if } |y| > 2, \end{cases} \quad (13)$$

and

$$LLR_1(y) \approx \log \frac{f_\alpha(\gamma; |y|-1)}{f_\alpha(\gamma; |y|-3)}, \quad (14)$$

where $\log \frac{f_\alpha(\gamma; y-1)}{f_\alpha(\gamma; y-3)} = LLR_b(y)$ in (5), $\log \frac{f_\alpha(\gamma; |y|-1)}{f_\alpha(\gamma; |y|-3)} = -LLR_b(|y|-2)$, and $\log \frac{f_\alpha(\gamma; |y|-3)}{f_\alpha(\gamma; |y|+1)}$ can be written as

$$\begin{aligned} \log \frac{f_\alpha(\gamma; |y|-3)}{f_\alpha(\gamma; |y|+1)} &= \log \frac{f_\alpha(\gamma; |y|-3) \cdot f_\alpha(\gamma; |y|-1)}{f_\alpha(\gamma; |y|-1) \cdot f_\alpha(\gamma; |y|+1)} \\ &= \log \frac{f_\alpha(\gamma; |y|-3)}{f_\alpha(\gamma; |y|-1)} + \log \frac{f_\alpha(\gamma; |y|-1)}{f_\alpha(\gamma; |y|+1)} \\ &= LLR_b(|y|-2) + LLR_b(|y|). \end{aligned} \quad (15)$$

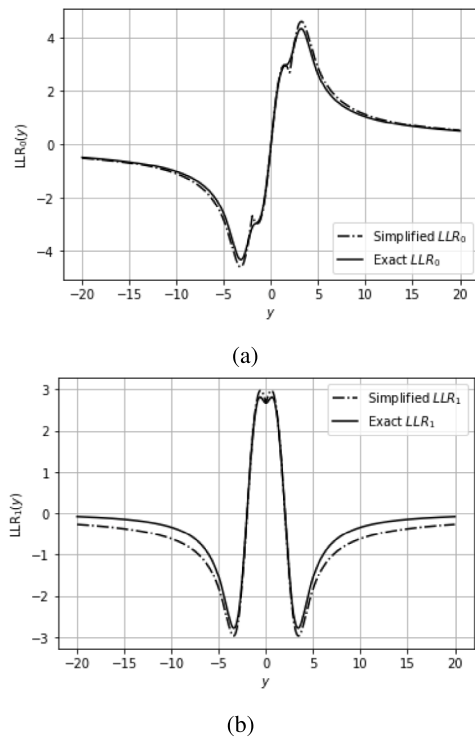


Fig. 3. Comparisons between LLRs of 4-ASK and their approximations (Approach 2) for $\alpha = 1.4$ and $\gamma = 0.552$: (a) $LLR_0(y)$, (b) $LLR_1(y)$.

Therefore, the simplified LLR expressions, for 4-ASK, becomes:

$$LLR_0(y) \simeq \begin{cases} LLR_b(y) & \text{if } |y| \leq 2, \\ \text{sign}(y) \cdot (LLR_b(|y| - 2) + LLR_b(|y|)) & \text{if } |y| > 2, \end{cases} \quad (16)$$

and

$$LLR_1(y) \simeq -LLR_b(|y| - 2). \quad (17)$$

Relations (16) and (17) show that LLRs of 4-ASK can be obtained from the LLR of BPSK. In the same way, we simplify the LLRs of the other modulations. Fig. 3 show that the simplified LLRs are close to the exact LLRs.

Thus, the approximated LLRs of 2^{2p} -QAM and 2^n -ASK modulated symbols can be obtained from the approximated LLR of BPSK, where the LLR of BPSK in (5) is replaced by its approximation, as follows

$$LL_0(y) \simeq \begin{cases} L_b(y) & \text{if } |y| \leq 2, \\ \text{sign}(y) \cdot (L_b(|y| - 2) + L_b(|y|)) & \text{if } |y| > 2, \end{cases} \quad (18)$$

and

$$LL_1(y) \simeq -L_b(|y| - 2). \quad (19)$$

Therefore, this approach requires the estimation parameters for only the approximated LLR of BPSK modulated symbols. As explained before (simple linear regression algorithm), The parameters approximation of the approximated LLR, for two approaches, can be estimated by reducing the cost function in (6), using gradient descent algorithm, to minimize the error between the approximated LLR and the corresponding exact LLR, with fixed parameters α and γ (noise distribution is known). In the following section, we show the parameter estimation, using Approach 2, with unknown noise distributions.

5. Parameters estimation for LLR approximation with unknown noise distributions

This section discusses the parameters estimation for LLR approximation, using Approach 2, with no prior knowledge about the noise

distribution model, we know only the channel output. This estimation can also be used for BPSK modulation. Here binary logistic regression algorithm is presented first. After, we present the method to extract the data from the channel output, for binary logistic regression algorithm, to estimate the parameters of the approximated LLR.

5.1. Binary logistic regression algorithm

Binary logistic regression is a supervised machine learning classification algorithm that predicts the outcome, d , of a categorical dependent variable, based on the concept of probability. Therefore, the outcome must be a binary value, which means that there are only two possible classes. In simple terms, the dependent variable is binary, having data coded as either 1 or 0.

The adopted logistic regression algorithm uses the Sigmoid function, S , to convert the outcome into a categorical value. It maps the predicted values to probabilities, as follows:

$$\Pr(y) = \begin{cases} S(y) & \text{if } d = 1 \\ 1 - S(y), & \text{if } d = 0 \end{cases} \quad (20)$$

Therefore, $\Pr(y) = S(y)^d \times (1 - S(y))^{(1-d)}$, so that, $0 \leq S(y) \leq 1$ and the Sigmoid function is given by:

$$S(y) = \frac{1}{1 + e^{-(\theta_a y + \theta_b)}}, \quad (21)$$

where $\theta_a y + \theta_b$ is a linear equation representing the approximated LLR, such that, $LLR(y) = \theta_a y + \theta_b$, where θ_a and θ_b are the parameters to be estimated. Hence, the pdf in (2) can be estimated by $\Pr(y)$ in (20).

To estimate the optimal parameters, we define a cost function and aim to find the minimum error between the predicted outcome and the corresponding exact value. The cost function is defined as shown below,

$$J(\theta_a, \theta_b) = \frac{1}{2m} \sum_{i=1}^m [d^{(i)} \log(S(y^{(i)})) + (1 - d^{(i)}) \log(S(1 - y^{(i)}))]. \quad (22)$$

where m is the length of the outcome. In this work, we apply the logistic regression from the sklearn library of Python. To apply logistic regression in our work, it requires the channel output y and its corresponding LLR decision. In the following section, we demonstrate how to find the LLR decision given the channel output.

Algorithm 1: Parameters estimation for LLR approximation

Input: Channel output (Received symbols) y .

Output: Estimation of the LLR function.

- 1: $\tilde{w} = y - (3 \times \text{sign}(y))$, where $|y| \geq 3$.
 - 2: Generate a random BPSK symbol \tilde{x} .
 - 3: $\tilde{y} = \tilde{x} + \tilde{w}$.
 - 4: Find the optimal parameters, θ_a and θ_b , using binary logistic regression, where the data are \tilde{y} and \tilde{x} (decision of the LLR): $\tilde{LLR}(y) = \theta_a \tilde{y} + \theta_b$.
 - 5: Replace estimated parameters in equations of the approximated LLR that corresponds to y : $LLR(y) = -\theta_a y - \theta_b$.
-

5.2. LLR estimation under unknown noise distributions

Estimating the approximation parameters, using binary logistic regression, requires the channel output y , where $y = x + w$, and its corresponding LLR decision. We can consider the channel input x as the LLR decision of its corresponding channel output y . However, the channel input is unknown at the receiver. Therefore, we generate a random input channel \tilde{x} , and the corresponding channel output is formulated as $\tilde{y} = \tilde{x} + \tilde{w}$, where the noises \tilde{w} and w have the same distribution to obtain the same parameters estimations since these parameters depend on the noise distribution. However, the noise distribution w is also unknown at the receiver.

To apply a noise \tilde{w} , where \tilde{w} and w have the same distribution, we extract the noise from the channel output y as introduced in [20] for BPSK modulation, and extend this work for other modulations. The authors in [20] show that, for BPSK modulation, the noise obtained from the channel output is $\tilde{w} = y - \text{sign}(y)$, where $\text{sign}(y)$ corresponds to the channel input $x = \pm 1$. In our work, we extract the noise as follows:

For 4-ASK modulations, $\mathcal{X} = \{\pm 1, \pm 3\}$, where $x = \pm 1$ corresponds to $\text{sign}(y)$, and $x = \pm 3$ corresponds to $3 \times \text{sign}(y)$. We thus take $|y| \geq 3$ to reduce the probability of having y correspond to $x = \pm 1$, and to close extracted noise, \tilde{w} , to the channel noise, w . How good is the resulting LLR approximation depends on how close is the extracted noise, \tilde{w} , to the channel noise, w , since the estimation parameters depend on the noise distribution. We thus extract the noise for 4-ASK modulation, as follows:

$$\tilde{w} = y - (3 \times \text{sign}(y)) \quad |y| \geq 3. \tag{23}$$

To estimate the approximation parameters, we consider the channel input \tilde{x} as the LLR decision of its corresponding channel output \tilde{y} . Following that, since the noises \tilde{w} and w follow the same distribution, we apply the estimated parameters in the approximated LLR, which corresponds to the channel output y .

To obtain good estimation parameters with an extracted noise, the shape of the $\text{LLR}_b(y)$ is divided into three segments to obtain a system of equations with a good approximation. Thus, $\text{LLR}_b(y)$ can be approximated by

$$L_b(y) = \begin{cases} \theta_0 y & \text{if } |y| \leq 1.2, \\ \text{sign}(y)(\theta_1 |y| + \theta_2) & \text{if } 1.2 < |y| \leq 2.3, \\ \text{sign}(y) \frac{\theta_3}{|y|} & \text{otherwise,} \end{cases} \tag{24}$$

where $\theta_j, j \in \{0, 1, 2, 3\}$, are the estimation parameters.

Algorithm 1 shows how to estimate the parameters of the LLR approximation, for all modulation orders, given uncertainty about the noise distribution. In this Algorithm, we emphasize that the approximated LLR uses the opposite values of the estimated parameters since the LLR used in the logistic regression algorithm is the inverse of the LLR used in our work.

6. Simulations

In this section, assuming 4-ASK modulated symbols, we use the bit error rate (BER) to compare the developed LLR function, with the exact LLR expression. The BER is considered as a function of the α -stable noise dispersion, γ . In our results, γ is interpreted as a measurement of the noise strength since the impulsive noise power is not well-defined and the signal-to-noise ratio (SNR) is meaningless [8].

The used code in our simulations is a (3, 6)-regular LDPC of length 20,000 bits using the Belief propagation algorithm. The channel is assumed to be SaS, where $\alpha \in \{1.8, 1.6, 1.4, 1.2\}$. Comparisons are made between the BER obtained using the estimated LLR functions in (18) and (19), where the approximated LLR of BPSK modulation is given in (24), and the BER obtained using the exact LLR computed via the numerical integration in (3) and (4). Performance comparisons for various values of α , shown in Fig. 4, demonstrate that the estimated LLR expression for 4-ASK modulated symbols, yields a comparable BER performance to that achieved by the exact LLR function. This result comes with the advantage of reduced complexity brought about by the proposed LLR approximation method.

It is necessary to show that the LLR approximation, given in (18) and (19), generalizes to all noise cases, including Gaussian noise. Hence, in Fig. 5, the BER versus the SNR per bit is plotted for a receiver under Gaussian noise. It is clear that the proposed technique yields comparable results to the exact LLR when Gaussian noise is present. Hence, our proposed techniques are effective under different types of receiver noise distributions. Also, as mentioned earlier, the proposed approximation requires estimation parameters only once (for BPSK modulated symbols).

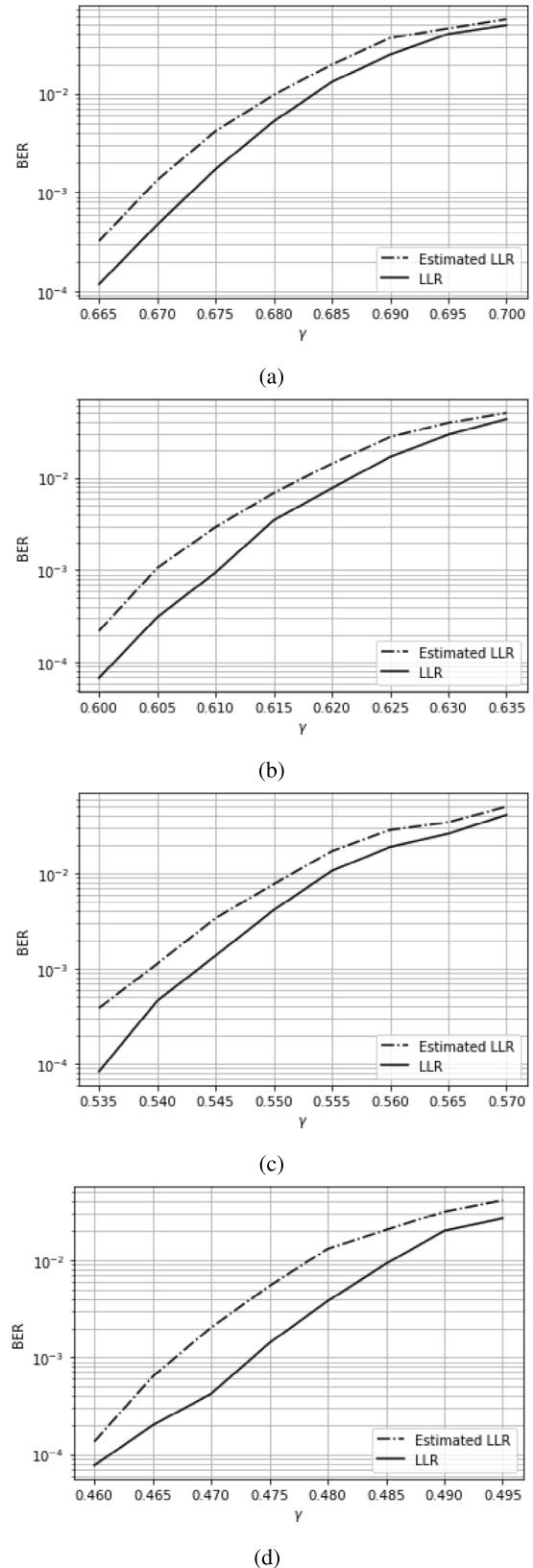


Fig. 4. Comparison between the BER obtained using the exact and estimated LLR of 4-ASK for: (a) $\alpha = 1.8$, (b) $\alpha = 1.6$, (c) $\alpha = 1.4$, (d) $\alpha = 1.2$.

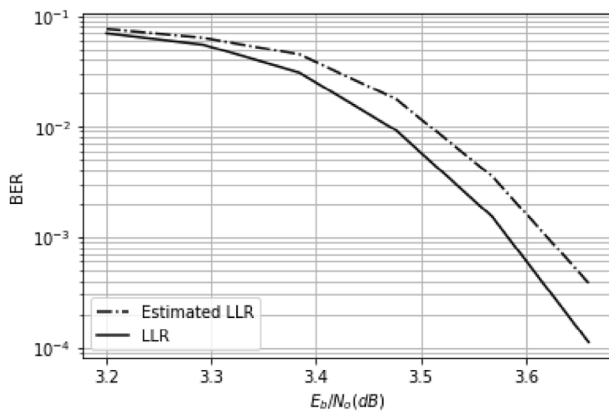


Fig. 5. Comparison between the BER obtained using the exact and approximated LLR of 4-ASK and its approximation, for Gaussian channel ($\alpha = 2$).

7. Conclusion

In this work, we have proposed a method to estimate the LLR for high-order modulated symbols and BPSK modulations, with no prior knowledge of the noise distribution model. Two approaches were proposed to approximate the LLR values using supervised machine learning, for high-order modulated symbols. The second approach is proposed to approximate the LLR for high-order modulated symbols in a simple way. We have used the second proposed approach to estimate the LLR with unknown noise distributions, using supervised machine learning. Our results show that the proposed estimation of the LLR yields comparable results to those obtained using the exact LLR function. The perspective of this work includes extending the proposed approach to estimate the LLR with no prior knowledge about the noise distribution model under channels with attenuation. Additionally, future research could focus on estimation without prior knowledge about the modulation type and order. This could significantly enhance the flexibility and applicability of the proposed methods in various communication scenarios.

CRedit authorship contribution statement

Latifa Mostari: Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Writing – review & editing. **Alban Goupil:** Writing – original draft, Supervision, Resources, Methodology, Investigation. **Abdelmalik Taleb-Ahmed:** Writing – original draft, Supervision, Resources, Investigation, Methodology, Writing – review & editing.

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] D. Ankit, K. Ranjan, S. Robert, Performance analysis of a multi-hop power line communication system over log-normal fading in presence of impulsive noise, *IET Commun.* 9 (1) (2015) 1–9.
- [2] K. Xiaoyan, S. Haixin, Z. Shengli, C. En, Impulsive noise mitigation in underwater acoustic OFDM systems, *IEEE Trans. Veh. Technol.* 65 (10) (2016) 8190–8202.
- [3] S. Fabien, L. Fabrice, L. Basile, Wide band time-correlated model for wireless communications under impulsive noise within power substation, *IEEE Trans. Wireless Commun.* 13 (3) (2014) 1449–1461.
- [4] A. Gorge, L. Chysostomos, Fast estimation of the parameters of alpha-stable impulsive interference, *IEEE Trans. Signal Process.* 44 (6) (1996) 1492–1503.
- [5] P.C. Pinto, M.Z. Win, Communication in a Poisson field of interferers—part I: Interference distribution and error probability, *IEEE Trans. Inform. Theory* 9 (7) (2010) 2176–2186.
- [6] H.E. Ghannudi, L. Clavier, N. Azzaoui, F. Septier, P.A. Rolland, α -stable interference modeling and Cauchy receiver for an IR-UWB Ad Hoc network, *IEEE Trans. Commun.* 58 (6) (2010) 1748–1757.
- [7] M.G. Sanchez, A.V. Alejos, I. Cuinas, Urban wide-band measurement of the UMTS electromagnetic environment, *IEEE Trans. Veh. Technol.* 53 (4) (2004) 1014–1022.
- [8] V. Dimanche, A. Goupil, L. Clavier, G. Gellé, On detection method for soft iterative decoding in the presence of impulsive interference, *IEEE Commun. Lett.* 18 (6) (2014) 945–948.
- [9] R. Gallager, Low-density parity-check codes, *IEEE Trans. Inform. Theory* 8 (1) (1962) 21–28.
- [10] C. Berrou, A. Glavieux, P. Thitimajshima, Near Shannon limit error correcting coding and decoding: Turbo-codes, *IEEE ICC* (1993) 1064–1070.
- [11] L. Clavier, G. Peters, F. Septier, I. Nevat., Impulsive noise modeling and robust receiver design, *EURASIP J. Wirel. Commun. Netw.* 13 (2021).
- [12] T.S. Saleh, I. Marsland, M. El-Tanany, Simplified LLR-based Viterbi decoder for convolutional codes in symmetric alpha-stable noise, in: 25th IEEE Canadian Conference on Electrical and Computer Engineering, CCECE, 2012, pp. 1–4.
- [13] H.B. Mâad, A. Goupil, L. Clavier, G. Gellé, Clipping demapper for LDPC decoding in impulsive channel, *IEEE Commun. Lett.* 17 (5) (2013) 968–971.
- [14] Y. Hou, R. Liu, L. Zhao, A non-linear LLR approximation for LDPC decoding over impulsive noise channels, in: IEEE/CIC International Conference on Communications in China (ICCC), China, 2014.
- [15] B. Dai, R. Liu, Z. Mei, Novel non-linear demapper for soft decision decoder of LDPC codes in impulsive noise, *IET Commun.* 13 (6) (2019) 796–801.
- [16] I.A. Koutrouvelis, An iterative procedure for the estimation of the parameters of stable laws, *Comm. Statist. Simulation Comput.* 10 (1) (1981) 17–38.
- [17] J.H. McCulloch, Simple consistent estimators of stable distribution parameters, *Comm. Statist. Simulation Comput.* 15 (4) (1986) 1109–1136.
- [18] V. Dimanche, A. Goupil, L. Clavier, G. Gellé, Estimation of an approximated likelihood ratio for iterative decoding in impulsive environment, in: IEEE Wireless Communications and Networking Conference, 2016, pp. 1–6.
- [19] Y. Mestrah, A. Savard, A. Goupil, G. Gellé, L. Clavier, Robust and simple log-likelihood approximation for receiver Design, in: IEEE Wireless Communications and Networking Conference, 2019.
- [20] Y. Mestrah, A. Savard, A. Goupil, G. Gellé, L. Clavier, An unsupervised LLR estimation with unknown noise distribution, *EURASIP J. Wireless Commun. Networking* (2020) 1–11.