Channel Parameter Estimation in the Presence of Phase Noise Based on Maximum Correntropy Criterion

Amir Alizadeh, and Ghosheh Abed Hodtani

Abstract—Oscillator output generally has phase noise causing the output power spectral density (PSD) to disperse around a Dirac delta function. In this paper, the AWGN channel is considered, where the sent signal accompanying with phase noise is added to the channel Gaussian noise and received at the receiver. Conventional channel estimation algorithms such as least mean square (LMS) and mean square error (MSE) criterion are not suitable for this channel estimation. We (i) analyze this phase noise channel estimation with information theoretic learning (ITL) criterion, i.e., maximum correntropy criterion (MCC), leading to robustness in the channel estimator’s steady state behavior; and (ii) improve the convergence rate by combining MSE and MCC as a novel mixed-LMS algorithm.

Index Terms—Information theoretic learning, phase noise, least mean squares, maximum correntropy criterion.

I. INTRODUCTION

INFORMATION theory has opened a wide vision for researchers in signal processing. This field of study has been first introduced to machine learning by Principe and Erdogmuz [1]. They introduced information theoretic learning (ITL) criteria such as minimum error entropy (MEE) and maximum correntropy criterion (MCC) to improve machine learning performance in the presence of outliers or heavy-tailed noise distributions. Subsequent works showed the superiority of these ITL criteria compared to conventional Mean Square Error (MSE) criterion [2]. [3] compared MEE and MSE criteria using MLP as an equalizer in both high and low SNR regimes for additive white Gaussian noise (AWGN). Later, [3] made the same comparison for non-Gaussian noise distributions such as exponential and Cauchy and in [4], a mathematical expression was derived for this superiority in the sense of KL divergence. [5] and [6] devised a conjugate gradient (CG) and Levenberg-Marquardt (LM) algorithm based on correntropy criterion, respectively. These two criteria have not been compared to each other theoretically until recently, [6] proposed an information theoretic approach to formulate this difference, which is the Euclidean distance between the noise distribution and a Gaussian distribution. Impulsive noise is present in environments such as underwater communications [7, 8], powerline communications (PLC) [9], digital subscriber lines [10], OFDM [11], etc.

Phase noise (PN) as another case of non-Gaussian noise [12] needs to be mitigated at the receiver. Low-cost oscillators at the receiver cannot produce an ideal dirac delta power spectral density (PSD). Thus, denoising techniques have to be utilized. [13-15] investigated the effect of phase noise in an OFDM symbol. Inner and outer bounds for channel capacity have been calculated in the presence of Wiener phase noise for point-to-point [16] and broadcast channels [17].

Pilot-based channel estimation involves the detection and extraction of channel state information (CSI) using the observed symbols at the receiver while undesirable effects, including AWGN, phase noise, doppler shift, multipath fading, intersymbol interference, etc distort data symbols. Various methods and algorithms have been incorporated in equalizers depending on the assumptions made on channel’s conditions like linearity, Gaussianity, sparsity, infinity of impulse response, modulation techniques, etc.

Least Squares (LS) and Minimum Mean Square Error (MMSE) have been used in OFDM receivers. However, by increasing computation power, adaptive equalizers gained attraction. Reference [18] adopted MCC in sparse channel estimation, [19] used a recursive least square (RLS) algorithm that adopted a generalized form of MCC and [20] added MCC to the cost function of an adaptive filtering problem two fold. Firstly, MCC is the main term in the performance function and secondly, a Correntropy Induced Metric (CIM) mimics l0-norm as the penalization term.

In this paper, we adopted a complex LMS method, as in [21], to estimate complex channel coefficients of a multipath fading channel corrupted by Tikhonof phase noise plus AWGN. The main difference to the mentioned reference is that we incorporated MCC instead of MSE as the learning criteria to combat a non-Gaussian noise distribution, namely, Tikhonof phase noise.

Throughout this paper, uppercase bold letters and lowercase
bold letters denote random variables and deterministic vectors, respectively. \( E[\cdot], \nabla, \ast, \mathbf{H}, \) and \( \dagger \) express expectation, gradient, conjugate, Hermitian, and transpose of matrices, respectively.

The rest of the paper is organized as follows. In section II the basic concepts behind correntropy and its key properties are reviewed. In section III, the complex multipath fading channel model is illustrated and our proposed algorithm is explained based on this model. In section IV, simulation results validate the theoretical findings. Finally, section V concludes the paper.

II. PRELIMINARIES

A. Correntropy

In this section a brief review of MCC is presented and its main properties are discussed. Correntropy as a similarity measure between two random variables \( X \) and \( Y \) is defined as [1]:

\[
\psi(X, Y) = E[k_{\sigma}(X, Y)] = \int k_{\sigma}(x, y)p_{X, Y}(x, y)dx\,dy, \tag{1}
\]

Where \( E[\cdot] \) is the expectation function, \( p_{X, Y}(x, y) \) is the joint probability density function (PDF) and \( k \) is a Mercer kernel function with a free parameter known as kernel bandwidth (KBW) \( \sigma^2 \). \( k \) is assumed Gaussian in most ITL applications and is defined as \( G_{\sigma}(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-y)^2}{2\sigma^2}\right] \).

Since correntropy is symmetric, positive, bounded, and reaches its maximum if and only if \( x=y \), it can be used as a similarity measure [22].

Expanding the exponential function in \( G_{\sigma}(\cdot) \) by means of Taylor series expansion, it can be deduced that correntropy involves all the even moments of the error random variable \( E = X - Y \):

\[
\psi(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{n=0}^{\infty} (-1)^n/(2^n n!) E[(X-Y)^{2n}/\sigma^{2n}] \tag{2}
\]

This can be viewed as a generalized MSE because setting \( n = 1 \), results in an MSE criterion. As a result, more information is extracted from the data samples in MCC [23].

Since only a limited number of error samples could be observed at the receiver, the above equation is estimated, using sample mean estimator:

\[
\hat{\psi}_{\sigma}(X, Y) = \frac{1}{N} \sum_{i=1}^{N} G_{\sigma}(x(i) - y(i)) \tag{3}
\]

Where \( N \) is the number of error samples. Note that \( k \) is replaced by the special case of a Gaussian kernel.

An interesting property of MCC is the mapping of data samples to a Reproducing Kernel Hilbert Space (RKHS) [23]. Thus, unlike MSE that considers the Euclidean distance between two random variables, MCC maps data to the surface of a hyper-sphere in a high dimensional feature space defined by \( \varphi(x) = k(., x) \). In this space, the inner product between two elements can easily be obtained by the kernel function without even knowing the non-linear mapping function \( \varphi(.) \):

\[
k(x, y) = \varphi(x) \varphi(y). \tag{4}
\]

Consequently, a nonlinear version of MSE is achieved [23].

III. OUR WORK: COMPLEX MAXIMUM CORRENTROPY CRITERION—BASED LEAST MEAN SQUARE EMPLOYED FOR CHANNEL PARAMETER ESTIMATION

A. System Model

Consider a point-to-point multipath channel of length \( M \) impaired by AWGN and phase noise which is described as:

\[
r(n) = h^T(n)x(n)e^{j\Phi(n)} + \vartheta(n), \tag{5}
\]

Where \( x(n) = [x(n), ..., x(n-M+1)]^T \), \( h(n) = [h_0(n), ..., h_{M-1}(n)]^T \), \( \Phi \), \( \vartheta \), and \( r \) are QAM modulated symbols, complex channel vector, phase noise, zero-mean circularly-symmetric complex Gaussian noise, and received data, respectively. If a phase-locked loop eliminates some phase perturbations, \( \Phi(n) \) is small, i.e. \( |\Phi(n)| \ll 1 \). Using a linear approximation (5) is rewritten as [24]:

\[
r(n) = h^T(n)x(n) + jh^T(n)x(n)\Phi(n) + \vartheta(n). \tag{6}
\]

The second and third terms on the right-hand side are complex noise terms. The real part is Gaussian while the imaginary part is non-Gaussian if \( \Phi(n) \) is assumed non-Gaussian. For a Von Mises distributed phase noise, the PDF of the imaginary part could be estimated through a kernel density estimation (KDE) as shown in Fig. 1. In this figure, the imaginary noise term displays a more heavy-tailed behavior than the standard Gaussian PDF. The KBW is chosen according to Silverman’s law [25] and kappa parameter for Von Mises distribution is set to \( \kappa = 8 \). As a result, the heavy-tailed assumption for phase noise is valid.

B. Proposed Algorithm

We aim at obtaining an estimation of \( h \), namely, \( \hat{h} \) via a training algorithm which minimizes the error:

\[
e(n) = r(n) - \hat{r}(n) = r(n) - \hat{h}^T(n)x(n) = r(n) - x^T(n)\hat{h}(n)
\]

\[
= r(n) - x^T(n)(\hat{h}_R(n) + j\hat{h}_I(n)), \tag{7}
\]

Where \( \hat{r}(n) \), \( \hat{h}_R(n) \), and \( \hat{h}_I(n) \) are the estimated received symbol, real part of the estimated channel coefficient vector, and imaginary part of the estimated channel coefficient vector at time instant \( n \). The conjugate of the error symbol is expressed as:

\[
e^*(n) = r^*(n) - x'^T(n)(\hat{h}_R(n) + j\hat{h}_I(n)). \tag{8}
\]

The performance function of the algorithm is described by:
The gradient of \( (e(n)e^*(n)) \) with respect to \( h_R \) is expressed as:

\[
\mathbf{V}_{h_R}(e(n)e^*(n)) = \begin{bmatrix} \frac{\partial e(n)e^*(n)}{\partial h_{R,0}(n)} \\ \frac{\partial e(n)e^*(n)}{\partial h_{R,1}(n)} \\
\frac{\partial e(n)e^*(n)}{\partial h_{R,2}(n)} \\
\frac{\partial e(n)e^*(n)}{\partial h_{R,3}(n)} \end{bmatrix}
\]

(10)

One can obtain the gradient of \( (e(n)e^*(n)) \) with respect to \( h_i \) in a similar fashion:

\[
\mathbf{V}_{h_i}(e(n)e^*(n)) = e(n)i\mathbf{x}^*(n) + e^*(n)\mathbf{y}^*(n).
\]

(11)

Using the conventional gradient descent (GD) algorithm, the real and imaginary parts for the estimated channel coefficient is updated by the following formulas:

\[
\mathbf{h}_R(n+1) = \mathbf{h}_R(n) + \mu \mathbf{V}_{h_R}(f(n))
\]

(12)

\[
\mathbf{h}_I(n+1) = \mathbf{h}_I(n) + \mu \mathbf{V}_{h_I}(f(n)).
\]

(13)

Substituting (9) in (12) and (13), and utilizing the gradients derived in (10) and (11), we obtain:

\[
\mathbf{h}_R(n+1) = \mathbf{h}_R(n) - \frac{\mu}{2\sigma^2} e^{-\frac{(e(n)x^*(n))}{2\sigma^2}} (e(n)(-x^*(n)) + e^*(n)(-x(n))).
\]

(14)

\[
\mathbf{h}_I(n+1) = \mathbf{h}_I(n) - \frac{\mu}{2\sigma^2} e^{-\frac{(e(n)y^*(n))}{2\sigma^2}} (e(n)(jx^*(n)) + e^*(n)(-jx(n))).
\]

(15)

Combining (14) and (15) using \( \mathbf{h}(n) = \mathbf{h}_R(n) + j\mathbf{h}_I(n) \), results in:

\[
\mathbf{h}(n+1) = \mathbf{h}(n) - \frac{\mu}{\sigma^2} e^{-\frac{(e(n)x^*(n))}{2\sigma^2}} e(n)x^*(n).
\]

(16)

We name it complex MCC-based LMS (MCC-LMS) to distinguish it from the standard complex LMS (MSE-LMS) algorithm which is characterized by:

\[
\mathbf{h}(n+1) = \mathbf{h}(n) + \mu e(n)x^*(n).
\]

(17)

Comparing (16) and (17), it is obvious that an adaptive step size \( \mu = \frac{\mu}{\sigma^2}G_0(e) \) improves MCC-LMS’s performance in the presence of heavy-tailed distributions. As [22] puts it, it is a localized similarity measure in contrast to global MSE. Locality means MCC puts more emphasis on the points near the line \( x=y \), suppressing the outliers that are away from this bissector in the joint space. This emphasis is defined by the free parameter \( \sigma \) (KBW) in the Gaussian kernel function \( G_0(e) \). On the other hand, MSE sees all the points in the joint space through the same lens and each error sample contributes linearly to the overall performance function’s output in respect to its distance to \( x=y \).

IV. SIMULATION RESULTS

In this section, simulation results are presented for a multipath fading (\( M=5 \) paths) channel with complex coefficients:

\[
\mathbf{h} = [1 + j 0 1 + j 0 1]^T / \sqrt{5},
\]

(18)

Where the constant factor is used to normalize the channel vector power. KBW is set to \( \sigma^2 = 2 \) through trial and error and all simulations are carried out over 50 Monte Carlo iterations. Fig 2. Illustrates the performance of the proposed MCC-LMS and standard MSE-LMS in a high SNR AWGN and PN-free regime. In order to indicate a time-variant channel, the coefficients are altered to \( h = [1 - j 0 1 - j 0 1]^T / \sqrt{5} \) in iteration 2000. As can be seen, applying MCC criterion provides us with no significant improvement. In other words, it hiders convergence rate badly. Besides, the improvement in the steady state MSE is only 4dB.

It is worth mentioning that the slow rate of convergence in MCC roots in the exponential factor in (16) because large errors at the beginning contribute to small \( \mu' \). Therefore, the coefficients are updated slightly but in the correct direction.

Adding PN turns the table in favor of MCC. Fig 3. shows the algorithms’ behavior in the presence of moderate PN, where \( \kappa = 4 \) and SNR=60. Obviously, the steady states of both algorithms get worse. Moreover, The difference between the two rises to 8dB while it was about 4dB in the lack of PN.

Next, we present some meaningful results about the significance of KBW’s impact on MCC-LMS. Large KBW worsens convergence rate while it improves MCC-LMS’s robustness against outliers. This trade-off is a design parameter and the decision is dependent on the application. Fig 4. compares MCC with different values and MSE.
In order to have both good convergence rate and desirable steady state, we combine both algorithms and introduce mixed-LMS. More precisely, for the few initial iterations after a change in the channel coefficients (i.e., a rise in MSE of error), MSE-LMS is utilized and for the following iterations after the change, MCC is replaced in the algorithm. This change can be triggered by a condition such as, the number of passed iterations or MSE value falling beneath a predefined threshold, say \( \textit{MSE} = -40 dB \). In Fig. 5, We used the latter.

V. CONCLUSION

With regard to higher order statistics (HOS), we presented a complex correntropy-based learning algorithm, named MCC-LMS, that could achieve a lower MSE for channel coefficients estimation. This algorithm is robust against non-Gaussian distributions. Therefore, the parameters of a channel corrupted by phase noise can be well estimated by our proposed algorithm. Although this algorithm illustrated good performance in the steady state, the convergence rate was quite low. To address this issue, we proposed a mixed-LMS algorithm that exploited both fast convergence rate of standard LMS and good steady state of MCC-LMS.

REFERENCES

[1] B. W. Silverman, “The kernel method for univariate data,” in \textit{Density estimation for statistics and data analysis}, 1st ed. London, UK: Chapman and Hall, 1986, pp. 75-93.

[2] I. Santamaria, D. Erdogmus, and J. C. Principe, "Entropy minimization for supervised digital communications channel equalization," \textit{IEEE Trans. Signal Process}, vol. 50, no. 5, pp. 1184-1192, Aug. 2002, doi: 10.1109/78.995074.

[3] A. R. Heravi and G. A. Hodtani, "Additive non-Gaussian noise channel estimation by using minimum error entropy criterion," in \textit{2017 7th International Conference on Computer and Knowledge Engineering (ICCKE)}, Mashhad, Iran, 2017, pp. 181–186.

[4] A. R. Heravi and G. A. Hodtani, "Where Does Minimum Error Entropy Outperform Minimum Mean Square Error? A New and Closer Look," \textit{IEEE Access}, vol. 6, pp. 5856-5864, Jan. 2018, doi: 10.1109/ACCESS.2018.2792329.

[5] A. R. Heravi and G. A. Hodtani, "A New Correntropy-Based Conjugate Gradient Backpropagation Algorithm for Improving Training in Neural Networks," \textit{IEEE Trans. Neural. Netw. Learn. Syst.}, vol. 29, no. 12, pp. 6252-6263, May 2018, doi: 10.1109/TNNLS.2018.2827778.

[6] A. R. Heravi and G. A. Hodtani, "A new robust correntropy based Levenberg-Marquardt algorithm," in \textit{2016 Iran Workshop on Communication and Information Theory (ICWIT)}, Tehran, Iran, 2016, pp. 1-6.

[7] M. A. Chitre, J. R. Potter, and S. Ong, "Optimal and Near-Optimal Signal Detection in Snapping Shrimp Dominated Ambient Noise," \textit{IEEE Journal of Oceanic Engineering}, vol. 31, no. 2, pp. 497-503, Apr. 2006, doi: 10.1109/JOE.2006.875272.

[8] M. Chitre, J. Potter, and O. S. Heng, "Underwater acoustic channel characterisation for medium-range shallow water communications," in \textit{Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37609)}, Kobe, Japan, 2004, pp. 40-45.

[9] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," \textit{IEEE Transactions on..."
Amir Alizadeh received the B.Sc. in electronics engineering from Guilan University, Rasht, Iran in 2018 and the M.Sc. in communications engineering from Ferdowsi University of Mashhad, Mashhad, Iran in 2021.

His research interests include information theoretic learning, machine learning, deep learning, information theory, and signal processing.

Ghosheh Abed Hodtani received the B.Sc. degree in electronics engineering and the M.Sc. degree in communications engineering, both from Isfahan University of Technology, Isfahan, Iran, in 1985, 1987, respectively. He joined Electrical Engineering Dept., at Ferdowsi University of Mashhad, Mashhad, Iran, in 1987. He decided to pursue his studies in 2005 and received the Ph.D. degree (with excellent grade) from Sharif University of Technology, Tehran, Iran, in 2008; and he has been a full professor in electrical engineering since 2016, selected as distinguished engineering professor by Ferdowsi Academic Foundation in 2016. His research interests are in multi-user information theory, communication theory, information-theoretic learning and signal processing. Prof. Hodtani is the author of a textbook on electrical circuits, the winner of the best paper award at IEEE ICT-2010 and a member of technical program and steering committees of Iran workshop on communication and information theory.