Signal Processing
Manuscript Draft

Manuscript Number: SIGPRO-D-14-00375

Title: Optimum Linear Regression in Additive Cauchy-Gaussian Noise

Article Type: Fast Communication

Keywords: Impulsive noise, Cauchy distribution, Gaussian distribution, mixture noise, Voigt profile, maximum likelihood estimator, pseudo-Voigt function, M-estimator

Abstract: In this paper, we study the estimation problem of linear regression in the presence of a new impulsive noise model, which is a sum of Cauchy and Gaussian random variables in time domain. The probability density function (PDF) of this mixture noise, referred to as the Voigt profile, is derived from the convolution of the Cauchy and Gaussian PDFs. To determine the linear regression parameters, the maximum likelihood estimator is first developed. Since the Voigt profile suffers from a complicated analytical form, an M-estimator with the pseudo-Voigt function is also derived. In our algorithm development, both scenarios of known and unknown density parameters are considered. In the unknown scenario, density parameters need to be estimated prior to proposals, by utilizing the empirical characteristic function and characteristic function. Simulation results show that the performance of both proposed methods can attain the Cram\'{e}r-Rao lower bound.

Highlights:

- An additive mixture noise is studied in this paper and the corresponding noise PDF, i.e., the Voigt function is derived.
- To determine the parameters of a linear regression model, the maximum likelihood estimator (MLE) is developed, where both the scenarios of known and unknown density parameters, are considered.
- To reduce the computational complexity of the MLE, an M-estimator with pseudo-Voigt function is presented.
- Both presented estimators approach the CRLB well.

Optimum Linear Regression in Additive Cauchy-Gaussian Noise

Yuan Chen *,[1], Ercan Engin Kuruoglu[2], Hing Cheung So[1]

[1] Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China

[2] ISTI-CNR (Italian National Council of Research), Pisa, Italy

Abstract: In this paper, we study the estimation problem of linear regression in the presence of a new impulsive noise model, which is a sum of Cauchy and Gaussian random variables in time domain. The probability density function (PDF) of this mixture noise, referred to as the Voigt profile, is derived from the convolution of the Cauchy and Gaussian PDFs. To determine the linear regression parameters, the maximum likelihood estimator is first developed. Since the Voigt profile suffers from a complicated analytical form, an M-estimator with the pseudo-Voigt function is also derived. In our algorithm development, both scenarios of known and unknown density parameters are considered. In the unknown scenario, density parameters need to be estimated prior to proposals, by utilizing the empirical characteristic function and characteristic function. Simulation results show that the performance of both proposed methods can attain the Cramér-Rao lower bound.

Indexing terms: Impulsive noise, Cauchy distribution, Gaussian distribution, mixture noise, Voigt profile, maximum likelihood estimator, pseudo-Voigt function, M-estimator

1 Introduction

Impulsive noise is encountered in a variety of applications such as wireless communications, radar, sonar and image processing [1]. Unlike Gaussian noise, impulsive noise belongs to a family of heavy-tailed noise distributions. Popular models in the literature for impulsive noise are divided into two categories, namely, single process and hybrid process mixed in the probability density function (PDF) domain. Typical single distributions are Student's t-distribution [2], α-stable distribution [3] and generalized Gaussian (GG) process [4], while the mixture models include Gaussian mixture (GM) [5] and Cauchy Gaussian mixture (CGM) [6]. Nevertheless, these models alone may not be able to represent all varieties of impulsive noises in the real world such as the case that the noise measured is the sum of two separate time series: one is an intrinsic Gaussian noise due to the electronic devices in receiver and the other is environmental noise which can be non-Gaussian, in particular impulsive. For example, considering some schemes in frequency-hopping spread spectrum (FH SS) radio communication networks [7], binary transmission systems [8] and multiple-input multiple-output (MIMO) systems [9], we model the multiple access interference as the α-stable distribution and regard the environmental noise as the Gaussian distribution. Similarly, in astrophysical imaging [10], the cosmic microwave background radiation is contaminated with the Gaussian noise from the satellite beam and α-stable distributed radiation

*Corresponding Author (Email: qchenyuan00@126.com; Fax: (852) 2788 7791)

from galaxies and stars. In these potential applications, the disturbance components can be combined into a new mixture model which is a sum of two different random processes in the time domain.

To demonstrate the applicability of this model, we consider the linear regression problem and take the sum of a symmetric Cauchy distribution with dispersion γ and zero-mean Gaussian distribution with variance σ^2 as an illustrative example. This mixture model belongs to the Middletons Class B [11] which is a classical impulsive noise model that has been employed for decades. The PDF of the mixture has an analytical form, known as the Voigt function [12], which is obtained via the convolution of the PDFs of these two processes. When the density parameters, namely, γ and σ^2 are known, the PDF of the mixture is readily determined, and the maximum likelihood estimator (MLE) which is a special case of M-estimator, can be directly applied to find the parameters of interest. The class of M-estimators introduced by Huber [13] generalizes the MLE by replacing the logarithm of the likelihood function by an arbitrary ρ-function. Note that the MLE is in the class of M-estimators by letting $\rho = -\log(f(y))$ with $f(y)$ denoting the likelihood function. However, when γ and σ^2 are unknown, they should be estimated through the relationship between the empirical characteristic function (ECF) and the characteristic function (CF) prior to employing the MLE. Although the MLE has the best performance in the sense of attaining Cramér-Rao lower bound (CRLB), it suffers from having a highly complex analytical form because of the Faddeeva function that appears in the PDF of the mixture noise. Therefore, in order to keep the high accuracy of the MLE and to reduce the computational complexity, a new M-estimator with the loss function chosen as the logarithm of pseudo-Voigt function is employed, which is referred to as the MEPV.

The rest of this paper is organized as follows. The proposed methods, namely, the MLE and MEPV are presented in Section 2. Both cases of known and unknown density parameters are investigated. Computer simulations are provided in Section 3 to evaluate the accuracy and complexity of the MLE and MEPV. Finally, conclusions are drawn in Section 4.

2 Proposed Algorithms

Without loss of generality, the observed data vector $\mathbf{y} = [y_1 \ \cdots \ y_N]^T$ is modeled as:

$$y_n = s_n(\boldsymbol{\theta}) + e_n, \quad n = 1, 2, \ldots, N, \tag{1}$$

where $s_n(\boldsymbol{\theta})$ denotes the noise-free signal with $\boldsymbol{\theta}$ being the parameter vector of interest, $e_n = p_n + q_n$ is the mixture noise which is a sum of two independent and identically distributed (i.i.d.) processes p_n and q_n, whose PDFs are f_P and f_Q, respectively.

The PDF of e_n can be obtained from the convolution of f_P and f_Q:

$$f_E = f_P * f_Q, \tag{2}$$

where $*$ stands for the convolution operator.

Considering the simplest case of the linear regression model, i.e., $s_n(\boldsymbol{\theta}) = s_n([A \ B]^T) = An + B$, where A and B are the unknown parameters, the data model can be rewritten in vector form as:

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e} \quad n = 1, 2, \ldots, N, \tag{3}$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ \vdots & \vdots \\ N & 1 \end{bmatrix}, \qquad \boldsymbol{\theta} = \begin{bmatrix} A \\ B \end{bmatrix} \tag{4}$$

and $\mathbf{e} = [e_1 \ \cdots \ e_N]^T$ with $e_n = c_n + g_n$ denoting the additive Cauchy Gaussian (ACG) noise which is the sum of i.i.d. Cauchy noise c_n with dispersion γ and the i.i.d. zero-mean Gaussian noise g_n with variance σ^2. Although we only study this simple model, our analysis can be extended to the general linear data model [14], that is, $\mathbf{H} \in R^{N\times M}$ where $N \geq M$ is known and $\boldsymbol{\theta} \in R^M$ is unknown. It is noteworthy that (3)-(4) are also a common signal model for kick detection in oil drilling [15]. The PDFs of Cauchy and Gaussian distributions are:

$$f_C(c_n;\gamma) = \frac{\gamma}{\pi(c_n^2+\gamma^2)}, \tag{5}$$

$$f_G(g_n;\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{g_n^2}{2\sigma^2}\right). \tag{6}$$

Then the PDF of e_n is calculated based on (2):

$$f_E(e_n;\gamma,\sigma^2) = \int_{-\infty}^{\infty} \frac{\gamma}{\pi((e_n-\tau)^2+\gamma^2)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\tau^2}{2\sigma^2}} d\tau. \tag{7}$$

The result of (7) is called the Voigt function which can be represented as [12]

$$f_E(e_n;\gamma,\sigma^2) = \frac{\mathrm{Re}\{w\}}{\sigma\sqrt{2\pi}}, \tag{8}$$

where

$$w = \exp\left(-\left(\frac{e_n+i\gamma}{\sigma\sqrt{2}}\right)^2\right)\left(1+\frac{2i}{\sqrt{\pi}}\int_0^{\frac{e_n+i\gamma}{\sigma\sqrt{2}}} \exp\left(t^2\right) dt\right) \tag{9}$$

and w is called the Faddeeva function with Re denoting the real part.

To estimate the parameter vector $\boldsymbol{\theta}$, we can utilize the M-estimator [13] whose cost function is:

$$J(\boldsymbol{\theta}) = \sum_{n=1}^{N} \rho_n = -\sum_{n=1}^{N} \log\left(f(y_n,\boldsymbol{\theta};\gamma,\sigma^2)\right), \tag{10}$$

where ρ_n is an arbitrary function [13]. Note that the M-estimator coincides with the MLE when $f(y_n,\boldsymbol{\theta};\gamma,\sigma^2)$ is the ACG's PDF $f_E(y_n,\boldsymbol{\theta};\gamma,\sigma^2)$. In the following, we introduce two types of functions, namely, the Voigt function and its approximation which is referred to as the pseudo-Voigt function.

2.1 Maximum Likelihood Estimator

We first use the MLE to find the unknown parameters, assuming the scenario of known γ and σ^2. The study is then extended to the case of unknown distribution parameters.

In the first scenario, the PDF of the mixture noise is known and the PDF of $\mathbf{y}$ is:

$$f_E(\mathbf{y},\boldsymbol{\theta};\gamma,\sigma^2) = \prod_{n=1}^{N} \frac{\mathrm{Re}\{w_n\}}{\sigma\sqrt{2\pi}}, \tag{11}$$

where

$$w_n = \exp\left(-\left(\frac{y_n - \mathbf{h}_n\boldsymbol{\theta} + i\gamma}{\sigma\sqrt{2}}\right)^2\right)\left(1 + \frac{2i}{\sqrt{\pi}}\int_0^{\frac{y_n - \mathbf{h}_n\boldsymbol{\theta} + i\gamma}{\sigma\sqrt{2}}} \exp\left(t^2\right) dt\right) \tag{12}$$

with $\mathbf{h}_n = [n\ 1]$.

In the case that γ and σ^2 are unknown, an exact expression of the PDF cannot be derived readily, therefore, γ and σ^2 are estimated using the ECF first. For the ACG noise, the CF of the observed data y_n is:

$$\phi(t) = E\{\exp(iy_n t)\} = \exp\left(it(An + B) - \gamma|t| - \frac{t^2}{2}\sigma^2\right), \tag{13}$$

where E stands for expectation and the magnitude of $\phi(t)$ is:

$$|\phi(t)| = \exp\left(-\gamma|t| - \frac{t^2}{2}\sigma^2\right). \tag{14}$$

Taking the logarithm on both sides of (14) yields:

$$\Phi(t) = -\log\left(|\phi(t)|\right) = \gamma|t| + \frac{t^2}{2}\sigma^2. \tag{15}$$

On the other hand, the ECF, denoted by $\psi(t)$, is

$$\psi(t) = \frac{1}{N}\sum_{n=1}^{N} e^{iy_n t}. \tag{16}$$

The error distribution between the ECF and CF is unknown. Here ℓ_1-norm estimator is employed. Let $\Psi(t) = -\log(|\psi(t)|)$, γ and σ^2 can be estimated if t is chosen in a grid $t \in [t_1, t_K]$ [16]:

$$\left(\hat{\gamma}, \hat{\sigma}^2\right) = \arg\min_{\tilde{\gamma}, \tilde{\sigma}^2} ||\boldsymbol{\Psi} - \mathbf{F}\mathbf{x}||_1, \tag{17}$$

where

$$\mathbf{F} = \begin{bmatrix} |t_1| & \frac{t_1^2}{2} \\ |t_2| & \frac{t_2^2}{2} \\ \vdots & \vdots \\ |t_K| & \frac{t_K^2}{2} \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \gamma \\ \sigma^2 \end{bmatrix}, \quad \boldsymbol{\Psi} = [\Psi(t_1)\ \cdots\ \Psi(t_K)]^T. \tag{18}$$

Since (17) is not differentiable, subgradient method [17] is employed to update $\hat{\mathbf{x}}$:

$$\hat{\mathbf{x}}^{(\ell+1)} = \hat{\mathbf{x}}^{(\ell)} - \eta_\ell \mathbf{g}^{(\ell)}, \tag{19}$$

where $^{(\ell)}$ stands for the ℓth iteration, $\mathbf{g}^{(\ell)} = -\mathbf{F}^T\text{sign}(\boldsymbol{\Psi} - \mathbf{F}\mathbf{x}^{(\ell)})$ and $\eta_\ell = 1/||\mathbf{g}^{(\ell)}||_2$. We employ the least squares solution $||\boldsymbol{\Psi} - \mathbf{F}\mathbf{x}||_2^2$ as $\hat{\mathbf{x}}^{(0)}$ and update (19) until the relative error $\left|\frac{\hat{\mathbf{x}}^{(\ell+1)} - \hat{\mathbf{x}}^{(\ell)}}{\hat{\mathbf{x}}^{(\ell+1)}}\right| < \epsilon$ is reached, where $\epsilon > 0$ is the tolerance. After γ and σ^2 have been estimated by (19), the PDF of $\mathbf{y}$, namely, $f_E(\mathbf{y}, \boldsymbol{\theta}; \hat{\gamma}, \hat{\sigma}^2)$ is calculated by (11).

The MLE of $\boldsymbol{\theta}$ is the minimum of the cost function in (10):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \{J_1(\boldsymbol{\theta})\}, \tag{20}$$

where $J_1(\boldsymbol{\theta}) = -\sum_{n=1}^{N} \log\left(\mathrm{Re}\{w_n\}\right)$.

The MLE cost function in (20) is multimodal for the linear regression model. In our study, the optimal estimator is realized by the Newton's method which is a local search algorithm with quadratic rate of convergence [18]. As a result, it is clear that global convergence depends on the initialization.

The updated procedure of the Newton's method is:

$$\hat{\boldsymbol{\theta}}^{(\ell+1)} = \hat{\boldsymbol{\theta}}^{(\ell)} - \left(\nabla^2\left(J_1(\hat{\boldsymbol{\theta}}^{(\ell)})\right)\right)^{-1} \nabla\left(J_1(\hat{\boldsymbol{\theta}}^{(\ell)})\right), \tag{21}$$

where

$$\nabla\left(J_1(\boldsymbol{\theta})\right) = \begin{bmatrix} \sum_{n=1}^{N} n v_n \\ \sum_{n=1}^{N} v_n \end{bmatrix}, \tag{22}$$

$$\nabla^2\left(J_1(\boldsymbol{\theta})\right) = \begin{bmatrix} \sum_{n=1}^{N} n^2\left(u_n - v_n^2\right) & \sum_{n=1}^{N} n\left(u_n - v_n^2\right) \\ \sum_{n=1}^{N} n\left(u_n - v_n^2\right) & \sum_{n=1}^{N}\left(u_n - v_n^2\right) \end{bmatrix} \tag{23}$$

with

$$v_n = \frac{1}{\sigma^2}\frac{1}{\mathrm{Re}\{w_n\}}\mathrm{Re}\Big\{\left(y_n - \mathbf{h}_n\boldsymbol{\theta} + i\gamma\right) w_n\Big\}, \tag{24}$$

$$u_n = \frac{1}{\sigma^2}\frac{1}{\mathrm{Re}\{w_n\}}\mathrm{Re}\Big\{\left(2\left(\frac{y_n - \mathbf{h}_n\boldsymbol{\theta} + i\gamma}{\sigma\sqrt{2}}\right)^2 - 1\right) w_n + \frac{2\gamma}{\sqrt{2\pi}\sigma}\Big\}. \tag{25}$$

In this study, the weighted median method is utilized for algorithm initialization, that is, $\hat{\boldsymbol{\theta}}^{(0)} = \mathrm{MED}\{\mathbf{y}\}$ [19] and the stopping criterion follows that of (19). According to our simulation results, the MLE is able to find the global solution by utilizing this initialization.

2.2 M-estimator with Pseudo-Voigt Function

Although the MLE is maximally efficient, in the sense that its variance asymptotically achieves the CRLB, it suffers from high computational complexity because of the integral in the Faddeeva function, i.e., likelihood function. To reduce the computational cost, we consider that ρ_n is the logarithm of the pseudo-Voigt function and the scenarios of known and unknown density parameters are also discussed.

When σ^2 and γ are known, it has been proved that the Voigt function can be approximated by the sum of the PDF of Cauchy and Gaussian distributions, which is called pseudo-Voigt function [20]:

$$f(e_n;\gamma,\sigma^2) = \mu_a f_1(e_n;\gamma) + (1-\mu_a) f_2(e_n;\sigma^2), \tag{26}$$

where f_1 and f_2 are the PDFs of Cauchy and Gaussian distributions, respectively, which is totally different from f_C and f_G:

$$f_1(e_n;\gamma) = \frac{C_a \xi_a}{\sqrt{\pi}(e_n^2 + \xi_a^2)}, \tag{27}$$

$$f_2(e_n;\sigma^2) = \frac{C_a}{\sqrt{\pi}\xi_a}\exp\left(-\log(2)\left(\frac{e_n}{\xi_a}\right)^2\right) \tag{28}$$

with

$$\mu_a = \frac{C_a - \sqrt{\log(2)}}{C_a(1 - \sqrt{\pi \log(2)})}, \quad C_a = b_{1/2}(a)e^{a^2}\left(1 - \mathrm{erf}(a)\right), \quad \xi_a = \sqrt{2}\sigma b_{1/2}(a), \tag{29}$$

$$b_{1/2}(a) = a + \sqrt{\log(2)}\exp(-0.6055a + 0.0718a^2 - 0.0049a^3 + 0.000136a^4) \tag{30}$$

and $a = \frac{\gamma}{\sqrt{2}\sigma}$, with $\mathrm{erf}(\cdot)$ denoting the error function.

The model in equation (26) is totally different from Swamis model [6]. In his paper, the density parameters are γ and σ^2, which are apparently shown in the PDF model. However, the density parameter in equation (26) is ξ_a, which is a nonlinear function of γ and σ^2. That is, in Swamis paper, the problem is to minimize $g(\gamma, \sigma^2)$. While in our work, we deal with $g(h(\gamma, \sigma^2))$ which is more complicated.

In the unknown density parameter scenario, similar to the MLE, we estimate the density parameters γ and σ^2 first, which can be derived from (19). After $\hat{\gamma}$ and $\hat{\sigma}^2$ are obtained, $f(\mathbf{y}, \boldsymbol{\theta}; \hat{\gamma}, \hat{\sigma}^2)$ can be constructed.

Then the $\boldsymbol{\theta}$ is estimated by minimizing the cost function according to (10):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\tilde{\boldsymbol{\theta}}} \left\{ J_2(\boldsymbol{\theta}) \right\}, \tag{31}$$

where $J_2(\boldsymbol{\theta}) = -\sum_{n=1}^{N} \log\left(f(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)\right)$ with

$$f(y_n, \boldsymbol{\theta}; \gamma, \sigma^2) = \mu_a f_1(y_n, \boldsymbol{\theta}; \gamma) + (1 - \mu_a) f_2(y_n, \boldsymbol{\theta}; \sigma^2) \tag{32}$$

$$= \mu_a \frac{C_a \xi_a}{\sqrt{\pi}\left((y_n - \mathbf{h}_n\boldsymbol{\theta})^2 + \xi_a^2\right)} + (1 - \mu_a)\frac{C_a}{\sqrt{\pi}\xi_a} \exp\left(-\log(2)\left(\frac{y_n - \mathbf{h}_n\boldsymbol{\theta}}{\xi_a}\right)^2\right). \tag{33}$$

To find the minimum of (31), we apply the Newton's method:

$$\hat{\boldsymbol{\theta}}^{(\ell+1)} = \hat{\boldsymbol{\theta}}^{(\ell)} - \left(\nabla^2\left(J_2(\hat{\boldsymbol{\theta}}^{(\ell)})\right)\right)^{-1} \nabla\left(J_2(\hat{\boldsymbol{\theta}}^{(\ell)})\right), \tag{34}$$

where

$$\nabla\left(J_2(\boldsymbol{\theta})\right) = \begin{bmatrix} \sum_{n=1}^{N} nV_n \\ \sum_{n=1}^{N} V_n \end{bmatrix}, \tag{35}$$

$$\nabla^2\left(J_2(\boldsymbol{\theta})\right) = \begin{bmatrix} \sum_{n=1}^{N} n^2\left(U_n + W_n - V_n^2\right) & \sum_{n=1}^{N} n\left(U_n + W_n - V_n^2\right) \\ \sum_{n=1}^{N} n\left(U_n + W_n - V_n^2\right) & \sum_{n=1}^{N}\left(U_n + W_n - V_n^2\right) \end{bmatrix} \tag{36}$$

with

$$V_n = \frac{2}{f(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}\left(\mu_a\left(\frac{y_n - \mathbf{h}_n\boldsymbol{\theta}}{(y_n - \mathbf{h}_n\boldsymbol{\theta})^2 + \xi_a^2}\right) f_1(y_n, \boldsymbol{\theta}; \gamma) + (1 - \mu_a)\left(\frac{\log(2)(y_n - \mathbf{h}_n\boldsymbol{\theta})}{\xi_a^2}\right) f_2(y_n, \boldsymbol{\theta}; \sigma^2)\right), \tag{37}$$

$$U_n = \frac{-2}{f(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}\left(\mu_a\left(\frac{(y_n - \mathbf{h}_n\boldsymbol{\theta})^2 - \xi_a^2}{\left((y_n - \mathbf{h}_n\boldsymbol{\theta})^2 + \xi_a^2\right)^2}\right) f_1(y_n, \boldsymbol{\theta}; \gamma) + \frac{(1 - \mu_a)\log(2)}{\xi_a^2} f_2(y_n, \boldsymbol{\theta}; \sigma^2)\right), \tag{38}$$

$$W_n = \frac{4}{f(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}\left(\mu_a\left(\frac{(y_n - \mathbf{h}_n\boldsymbol{\theta})}{(y_n - \mathbf{h}_n\boldsymbol{\theta})^2 + \xi_a^2}\right)^2 f_1(y_n, \boldsymbol{\theta}; \gamma) + (1 - \mu_a)\left(\frac{\log(2)(y_n - \mathbf{h}_n\boldsymbol{\theta})}{\xi_a^2}\right)^2 f_2(y_n, \boldsymbol{\theta}; \sigma^2)\right). \tag{39}$$

In this method, $\hat{\boldsymbol{\theta}}$ is updated by (34) and the initialization and stopping criterion are the same as for the MLE in (21).

2.3 Complexity of Proposed Methods

The computational complexities of the proposed methods are roughly examined. The Faddeeva function in the PDF can be realized according to [21]. At each iteration of the Newton's method, the numbers of flops in the MLE and MEPV required are $\mathcal{O}\left(N^3 + N^2 \log_2 N\right)$ and $\mathcal{O}(N)$, respectively.

3 Cramér-Rao Lower Bound

In the scenario that the density parameters are known, the CRLB [14] of $\hat{\boldsymbol{\theta}}$ can be calculated by the diagonal elements of the inverse of the Fisher information matrix $\mathbf{I}$:

$$\mathbf{I} = -E\left\{ \sum_{n=1}^{N} \frac{\partial^2 \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \boldsymbol{\theta}^2} \right\} = E\left\{ \sum_{n=1}^{N} \left(\frac{\partial \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \boldsymbol{\theta}} \right)^T \right\}. \tag{40}$$

Based on (11), we have

$$\frac{\partial \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \boldsymbol{\theta}} = \frac{1}{\sigma^2} \begin{bmatrix} \frac{n\mathrm{Re}\{(y_n - An - B + i\gamma) w_n\}}{\mathrm{Re}\{w_n\}} \\ \frac{\mathrm{Re}\{(y_n - An - B + i\gamma) w_n\}}{\mathrm{Re}\{w_n\}} \end{bmatrix}. \tag{41}$$

It is hard to derive a closed form expression for (41), therefore, we use the average of sufficient number of independent runs to replace the expectation.

For unknown γ and σ^2, the parameter vector is $\boldsymbol{\alpha} = [A \ B \ \gamma \ \sigma^2]^T$. Then the CRLB of $\hat{\boldsymbol{\theta}}$ in this case corresponds to the $(1,1)$ and $(2,2)$ entries in the inverse of $\mathbf{I}$. The (k,l) element of $\mathbf{I}$ is written as:

$$\mathbf{I}_{k,l} = -E\left\{ \sum_{n=1}^{N} \frac{\partial \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \alpha_k} \frac{\partial \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \alpha_l} \right\}, \quad k, l = 1, 2, 3, 4, \tag{42}$$

where

$$\frac{\partial \log f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)}{\partial \boldsymbol{\alpha}} = \begin{bmatrix} \frac{1}{\sigma^2} \frac{n\mathrm{Re}\{(y_n - An - B + i\gamma) w_n\}}{\mathrm{Re}\{w_n\}} \\ \frac{1}{\sigma^2} \frac{\mathrm{Re}\{(y_n - An - B + i\gamma) w_n\}}{\mathrm{Re}\{w_n\}} \\ -\frac{\frac{1}{\sigma^2}\mathrm{Re}\{i(y_n - An - B + i\gamma) w_n\} + \frac{2}{\sqrt{2\pi\sigma^2}}}{\mathrm{Re}\{w_n\}} \\ \frac{\frac{1}{\sigma^2}\mathrm{Re}\{(y_n - An - B + i\gamma)^2 w_n\} + \frac{\gamma}{\sqrt{2\pi\sigma^2}\sigma^2}}{\mathrm{Re}\{w_n\}} - \frac{1}{2\sigma^2} \end{bmatrix}. \tag{43}$$

4 Simulation Results

To evaluate the performance of the MLE and MEPV, computer simulations have been conducted. The mean square errors (MSEs), $E\{(\hat{A} - A)^2\}$ and $E\{(\hat{B} - B)^2\}$, are employed as the performance measure. The signal is generated according to (5) with $A = 1$, $B = 0.5$, the noise e_n is generated as the sum of i.i.d. Cauchy distribution with dispersion γ and i.i.d. zero-mean Gaussian distribution with variance σ^2. Following the setup in [22], the interval of t in (18) is $[0.1, 1]$ with 1000 uniform grid points. Under such mixture noise, the signal-to-noise ratio is hard to define, therefore, we set $\sigma^2 = 10$ and scale $\frac{\sigma^2}{\gamma}$ to produce different noise conditions. Comparison with ℓ_1-norm estimator solved by the least absolute deviation [23], MM-estimator with breakdown point 0.85 [24] and the CRLB are provided. The ℓ_1-norm estimator is included because it is robust and suboptimum estimator for Cauchy noise. While the MM-estimator, which is a double-stage M-estimator, is considered since this estimator

is robust for the linear regression model. All results are based on 1500 independent Monte Carlo runs with a data length of $N = 60$.

First, we examine the difference between the Voigt and pseudo-Voigt functions. The γ is set to 10 and Figure 1 shows the comparison with the logarithmic scale. This approximation has been studied in detail in [20] and we have also verified also with our experiment a good match between the two curves. The closeness of the match depends on the $\frac{\sigma^2}{\gamma}$. For normalized curves in Figure 1, the mismatch measured by the area between the two curves can be less than 0.5%, which agrees with the analysis in [20].

Next, we examine the scenario of the known distribution parameters. It is seen in Figures 2 and 3 that when γ and σ^2 are available, the MSEs of both the MLE and MEPV can attain the CRLB for $\frac{\sigma^2}{\gamma} \in [0, 30]$ dB case. The performance of the MLE and MEPV are almost the same because the main idea of them is to optimize two highly similar functions. Furthermore, they are superior to the ℓ_1-norm estimator and MM-estimator in the case of ACG noise. Figure 4 shows the average computational cost versus the data length N for γ being set to 10. A stopwatch timer is utilized to measure the operation time of both methods. It is demonstrated that the complexity of the MEPV is significantly lower with respect to the MLE. The computational cost of the MLE increases exponentially, while for the MEPV, the computational complexity grows approximately sublinearly for small N ($N \in [50, 2550]$) and linearly for large N ($N > 2550$) since it requires less iterations to converge.

Finally, we investigate the scenario of the unknown γ and σ^2. The ECF is employed to estimate the density parameters γ and σ^2 first. Figures 5 and 6 show that the MSEs of the MLE and MEPV can achieve the CRLB and they significantly outperform the ℓ_1-norm estimator and MM-estimator.

5 Conclusions

We investigate the ACG process in this paper, which is modeled as the sum of Cauchy and Gaussian variables in time domain. The PDF of the ACG, known as the Voigt profile, is calculated by the convolution of PDFs of these two components. To estimate the parameters in the linear regression in the environment of the ACG noise, the MLE and MEPV are developed for both known and unknown density parameters cases. Computer simulation results are provided to show that the MSE performance of both estimators can achieve the CRLB for $\frac{\sigma^2}{\gamma} \in [0, 30]$ dB and are superior to ℓ_1-norm and MM-estimator. It is also seen in simulation results that the MEPV has much lower the computational cost than the MLE because of the complicated analytical form in the PDF. Moreover, this study on the ACG process can be extended to autoregressive model and nonlinear parameter estimation problem.

References

[1] A.M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, "Robust estimation in signal proxcessing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61-80, Jul. 2012.

[2] J. Pfanzagl and O. Sheynin, "Studies in the history of probability and statistics XLIV: A forerunner of the t-distribution," *IEEE Transactions on Signal Processing*, vol. 83, no. 4, pp. 891-898, Apr. 1996.

[3] C.L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distribution and Applications*, John Wiley & Sons Inc., New York, 1995.

[4] J.J. Shynk, *Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications*, Hoboken, N.J.: Wiley, 2013.

[5] D.A. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics*, pp. 659-663, 2009.

[6] A. Swami,"Non-Gaussian mixture models for detection and estimation in heavy tailed noise," *in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, Jun. 2000, vol. 6, pp. 3802-3805.

[7] J. Ilow, D. Hatzinakos and A.N. Venetsanopoulos, "Performance of FH SS radio networks with interference modeled as a mixture of Gaussian and alpha-stable noise," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 509-520, Apr. 1998.

[8] S. Ambike, J. Ilow and D. Hatzinakos, "Detection for binary transmission in a mixture of Gaussian noise and impulsive noise modeled as an alpha-stable process," *IEEE Signal Processing Letters*, vol. 1, no. 3, pp. 55-57, Mar. 1994.

[9] A.X. Li, Y.Z. Wang, W.Y. Xu and Z.C. Zhou, "Receiver design of MIMO systems in a mixture of Gaussian noise and impulsive noise," *in Proceedings of the 60th IEEE Vehicular Technology Conference*, Los Angeles, CA, Sept. 2004, vol. 24, pp.1493-1497.

[10] D.Herranz, E.E. Kuruoglu and L. Toffolatti, "An α-stable approach to the study of the P(D) distribution of unresolved point sources in CMB sky maps," *Astronomy and Astrophysics*, vol. 424, no. 3, pp. 1081-1096, 2004.

[11] Y. Kim and G.T. Zhou, "The Middleton class B model and its mixture representation," Center for Signal and Image Processing, Technical Report No. CSIP TR-98-01, Georgia Institute of Technology, Atlanta, GA, 303320250, May 1998.

[12] F.W.J. Olver, D.M. Lozier and R.F. Boisvert, *NIST Handbook of Mathematical Functions*, Cambridge University Press, Cambridge, pp. 167-168, 2010.

[13] P.J. Huber, *Robust Statistics*, 2nd ed., New York: Wiley, 2009.

[14] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, NJ: Englewood Cliffs, 1993.

[15] T. Burgess, A.A. Starkey and D. White, "Improvements for kick detection," *Oil Review*, vol. 2, no. 1, pp. 43-51, 1990.

[16] R. Brcich and A.M. Zoubir, "Estimation and detection in a mixture of symmetric alpha stable and Gaussian interference," *in Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, Ceasarea, Jun. 1999, pp. 219-223.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, New York. Cambridge University Press, 2004.

[18] J. Nocedal and S.J. Wright, *Numerical Optimization*, 2nd ed., New York: Springer, c2006.

[19] R.W. Hawley and N.C. Gallagher Jr., "On edgeworths method for minimum absolute error linear regression," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2045-2054, Aug. 1994.

[20] H.O. Dirocco and A. Cruzado, "The Voigt profile as a sum of a Gaussian and a Lorentzian functions, when the coefficient depends only on the widths ratio," *Acta Physica Polonica A*, vol. 122, no. 4, pp. 666-669, 2012.

[21] J.A.C. Weideman, "Computation of the complex error function," *SIMAM Journal on Numerical Analysis*, vol. 31, no. 5, pp. 1497-1518, Oct. 1994.

[22] S.M. Kogon and D.B. Williams, "On the characterization of impulsive noise with α-stable distributions using Fourier techniques," *in Proceedings of the 29th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1995, vol. 2, pp. 787-791.

[23] Y. Li and G. Arce, "A maximum likelihood approach to least absolute deviation regression," *EURASIP Journal on Applied Signal Processing*, no. 12, pp. 1762-1769, Sept. 2004.

[24] C. Croux, G. Dhaene and D. Hoorelbeke, "Robust Standard Errors for Robust Estimators," *DTEW Research Report 0367, K.U. Leuven*, Jan. 2004.

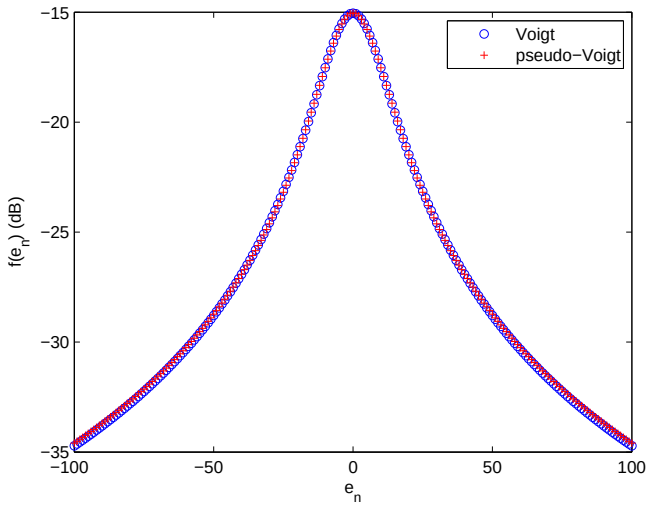

Figure 1: Comparison between Voigt and pseudo-Voigt functions

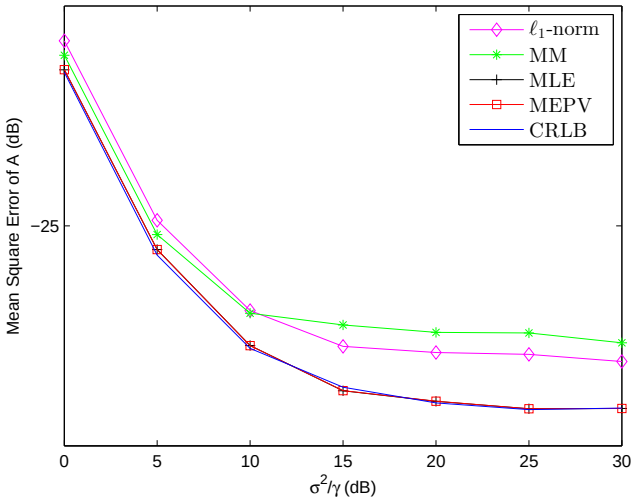

Figure 2: Mean square error of A versus $\frac{\sigma^2}{\gamma}$ with known γ and σ^2

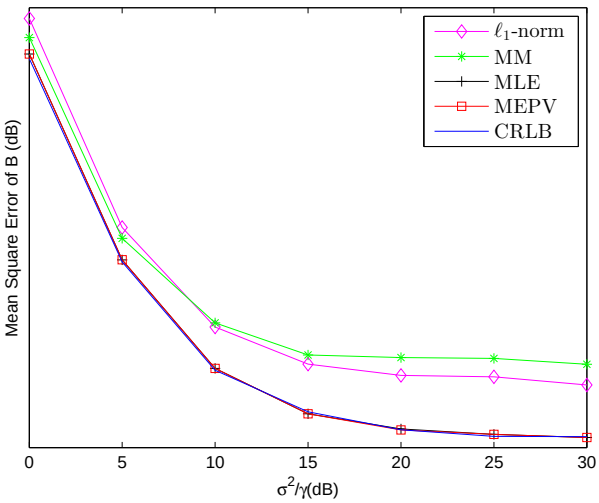

Figure 3: Mean square error of B versus $\frac{\sigma^2}{\gamma}$ with known γ and σ^2

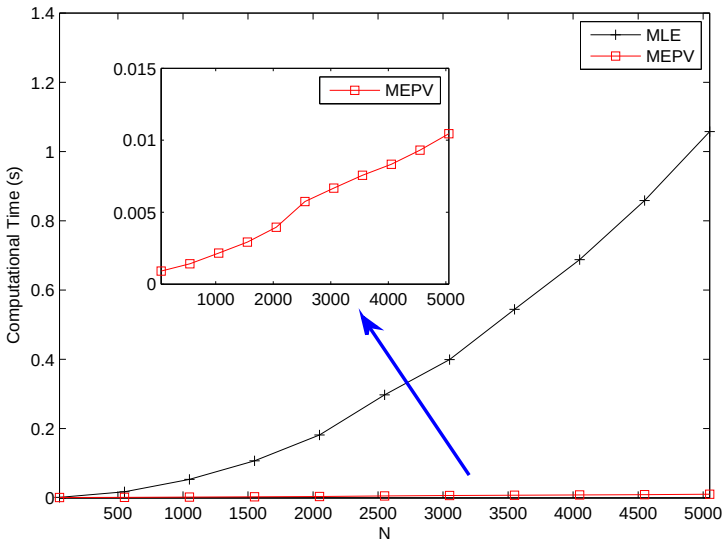

Figure 4: Computational complexity versus N

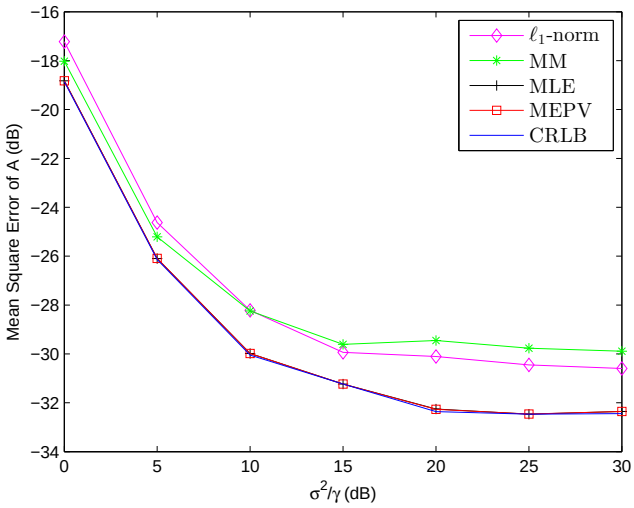

Figure 5: Mean square error of A versus $\frac{\sigma^2}{\gamma}$ with unknown γ and σ^2

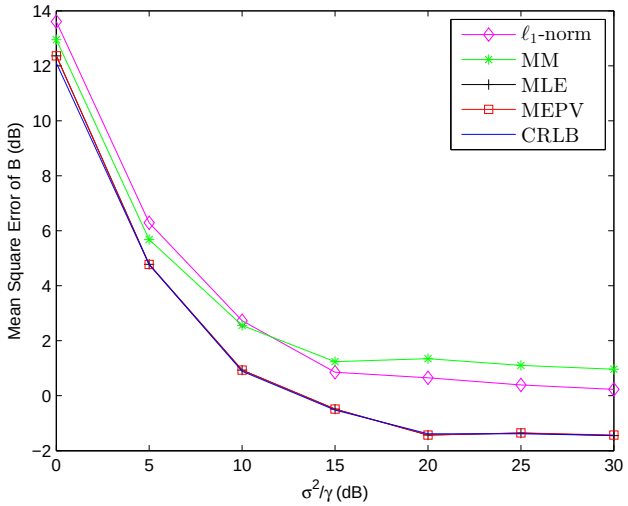

Figure 6: Mean square error of B versus $\frac{\sigma^2}{\gamma}$ with unknown γ and σ^2