

Robust Frequency Estimation Under Additive Mixture Noise

Yuan Chen¹, Yulu Tian¹, Dingfan Zhang², Longting Huang^{3,*} and Jingxin Xu⁴

¹School of Computer and Communication Engineering, University of Science & Technology Beijing, Beijing, 100083, China

²School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, China

³School of Information Engineering, Wuhan University of Technology, Wuhan, 430070, China

⁴Department of Energy and Public Works, Queensland, 4702, Australia

*Corresponding Author: Longting Huang. Email: huanglt08@whut.edu.cn

Received: 05 August 2021; Accepted: 01 December 2021

Abstract: In many applications such as multiuser radar communications and astrophysical imaging processing, the encountered noise is usually described by the finite sum of α -stable ($1 \leq \alpha < 2$) variables. In this paper, a new parameter estimator is developed, in the presence of this new heavy-tailed noise. Since the closed-form PDF of the α -stable variable does not exist except $\alpha = 1$ and $\alpha = 2$, we take the sum of the Cauchy ($\alpha = 1$) and Gaussian ($\alpha = 2$) noise as an example, namely, additive Cauchy-Gaussian (ACG) noise. The probability density function (PDF) of the mixed random variable, can be calculated by the convolution of the Cauchy's PDF and Gaussian's PDF. Because of the complicated integral in the PDF expression of the ACG noise, traditional estimators, e.g., maximum likelihood, are analytically not tractable. To obtain the optimal estimates, a new robust frequency estimator is devised by employing the Metropolis-Hastings (M-H) algorithm. Meanwhile, to guarantee the fast convergence of the M-H chain, a new proposal covariance criterion is also devised, where the batch of previous samples are utilized to iteratively update the proposal covariance in each sampling process. Computer simulations are carried out to indicate the superiority of the developed scheme, when compared with several conventional estimators and the Cramér-Rao lower bound.

Keywords: Frequency estimation; additive cauchy-gaussian noise; voigt profile; metropolis-hastings algorithm; cramér-rao lower bound

1 Introduction

Heavy-tailed noise is commonly encountered in a variety of area such as wireless communication and image processing [1–8]. Typical models of impulsive noise are α -stable, Student's t and generalized Gaussian distributions [9–12], which cannot represent all kinds of the noise types in the real-world applications. Therefore, the mixture models have been developed including the Gaussian mixture and the Cauchy Gaussian mixture models [13,14], of which the probability density functions (PDFs) is



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

the weighted sum of the corresponding components' PDF. However, these mixture models still cannot describe all impulsive noise types, especially for the case where the interference is caused by both channel and device. In astrophysical imaging processing [15], the observation noise is modelled as the sum of a symmetric α -stable ($S\alpha S$) and a Gaussian noise, caused by the radiation from galaxies and the satellite antenna, respectively. Moreover, in a multiuser radar communication network [16], the multi-access interference and the environmental noise corresponds to $S\alpha S$ distributed and Gaussian distributed variables. Therefore, a new description of the mixture impulsive noise model is proposed, referring to as the sum of $S\alpha S$ and Gaussian random variables in time domain.

In this work, the frequency estimation is considered in the presence of the additive Cauchy-Gaussian (ACG) noise [17,18], which is the sum of two variables with one following Cauchy distribution ($\alpha = 1$) [19] and the other being Gaussian process ($\alpha = 2$). The PDF of ACG noise can be calculated by the convolution of the Gaussian and Cauchy PDFs. According to [20], the PDF of the mixture can be expressed as Voigt profile. Due to the involved form of the Voigt profile, classical frequency estimators [21–24], such as the maximum likelihood estimator (MLE) and M -estimator [25], has convergence problem and cannot provide the optimal estimation in the case of high noise power. To obtain the estimates accurately, Markov chain Monte Carlo (MCMC) [26–28] method is utilized, which samples unknown parameters from a simple proposal distribution instead of from the complicated posterior PDF directly. Among series of MCMC methods, the Metropolis-Hastings (M-H) and Gibbs sampling algorithms are typical ones. The M-H algorithm provides a general sampling framework requiring the computations of an acceptance criterion to judge whether the samples come from the correct posterior or not. While in the case that the posterior PDF is easy to be sampled from, Gibbs sampling is utilized without the calculation of acceptance ratios. It is noted that once the posterior of parameters is known, M-H and Gibbs sampling methods can be utilized in any scenarios.

Because of the Voigt profile in the target PDF of the ACG noise, we choose the M-H algorithm as the sampling method. To improve the performance of the M-H algorithm, an updating criterion of proposal covariance is devised, with the use of the samples in a batch process. Since we assume all unknown parameters are independent, the proposal covariance is in fact a diagonal matrix. Therefore, the square of difference between the neighbor samples is employed as the diagonal elements of proposal covariance, referred to as the proposal variance. Meanwhile, a batch-mode method is utilized to make the proposal variance more accurate. As the proposal variance is updated only according to the samples in each iteration, this criterion can be extended to any other noise type such as the Gaussian and non-Gaussian processes.

The rest of this paper is organized as follows. We review the MCMC and M-H algorithm in Section 2. The main idea of the developed algorithm is provided in Section 3, where the PDF of the additive impulsive noise and posterior PDF of unknown parameters are also included. Then the Cramér-Rao lower bounds (CRLBs) of all unknown parameters are calculated in Section 4. Computer simulations in Section 5 are given to assess the performance of the proposed scheme. Finally, in Section 6, conclusions are drawn. Moreover, a list of symbols is shown in Tab. 1, which are appeared in the following.

Table 1: List of symbols

Notation	Description	Notation	Description
\mathbf{X}	Matrix	T	Transpose
\mathbf{x}	Vector	$^{-1}$	Inverse
$\mathbf{X}(m, n)$	(m, n) entry of \mathbf{X}	$E(x)$	Expectation of x
$ \mathbf{X} $	Determinant of \mathbf{X}	$\frac{\partial}{\partial x}$	partial derivative on x
\hat{x}	Estimates of x	$x_n, \mathbf{x}(n)$	n -th entry of \mathbf{x}
$\theta^{(k)}$	k -th samples	$Re\{x\}$	Real part of x
$\delta(\cdot)$	Dirac delta function	$U(0, 1)$	Standard uniform distribution
$\Gamma(\cdot)$	Gamma function	i	$i = \sqrt{-1}$
$\text{atan}(\cdot)$	Arctangent function	$\exp(\cdot)$	Exponential function
$\lim_{x \rightarrow \infty} f(x)$	Limit of a function	$\int_A^B f(x)dx$	Definite integral of $f(x)$ from A to B
\mathbf{I}	Identity matrix	\mathbf{F}	Fisher information matrix

2 Review of MCMC and M-H Algorithm

Before reviewing the M-H algorithm, some basic concepts, such as Markov chain should be introduced [29,30]. By employing several dependent random variables $\{x^{(l)}\}$ [31], we define a Markov chain as

$$x^{(1)}, x^{(2)}, \dots, x^{(l)}, x^{(l+1)}, \dots \tag{1}$$

where the probability of $x^{(l+1)}$ relies only on $\{x^{(l)}\}$ with the conditional PDF being defined by $\mathcal{P}(x^{(l+1)}|x^{(l)})$. The PDF of $x^{(l+1)}$, denoted by π_{l+1} , can be expressed as

$$\pi_{l+1} = \int \mathcal{P}(x^{(l+1)}|x^{(l)})\pi_l dx^{(l)}. \tag{2}$$

Then the Markov chain is said to be stable if

$$\pi^* = \mathcal{P}(\cdot|\cdot)\pi^* \tag{3}$$

with $\pi^* = \lim_{l \rightarrow \infty} \pi_{l+1}$. To ensure (3), a sufficient but not necessary condition can be written as

$$\pi_l \mathcal{P}(x^{(l+1)}|x^{(l)}) = \pi_{l+1} \mathcal{P}(x^{(l)}|x^{(l+1)}). \tag{4}$$

Typical MCMC algorithms draw samples from the conditional PDF $\mathcal{P}(x^{(l+1)}|x^{(l)})$ with a Markov chain, instead of directly sampling from a target PDF $f(x)$. Therefore, if a proper conditional PDF $\mathcal{P}(\cdot|\cdot)$ is chosen, the stationary distribution can align with the target PDF $f(x)$. In other word, samples drawn from a stable Markov chain will eventually tend to be generated from $f(x)$ accordingly. Furthermore, only samples generated from the stable Markov chain are independent and identically distributed (IID).

In the following, the details of the MCMC method are provided in Tab. 2. It is worth to point out that the burn-in period is a term of an MCMC run before convergent to a stationary distribution.

Table 2: Steps of MCMC method

-
1. Initialize $x^{(1)}$ using an arbitrary value;
 2. Iteratively draw $x^{(l+1)}$ from the conditional PDF $\mathcal{P}(x^{(l+1)}|x^{(l)})$;
 3. Throw out the first P burn-in period samples.
-

Among typical MCMC methods, M-H algorithm is commonly employed, whose main idea [32] is drawing samples from a proposal distribution with a rejection criterion, instead of sampling from $\mathcal{P}(x^{(l+1)}|x^{(l)})$ directly. In this method, a candidate, denoted by x^* is generated from a proposal distribution $q(x^*|x^{(l)})$. Then the acceptance probability is defined as

$$\mathcal{A}(x^{(l)}, x^*) = \min \left\{ 1, \frac{q(x^{(l)}|x^*)f(x^*)}{q(x^*|x^{(l)})f(x^{(l)})} \right\}, \quad (5)$$

which determines whether the candidate is accepted or not. It is noted that the proposal distributions are usually chosen as uniform, Gaussian or Student's t processes, which are easier to be sampled. The details of the M-H algorithm can be seen in Tab. 3.

Table 3: Steps of M-H algorithm

-
1. Initialize $x^{(1)}$;
 2. Sample $u \sim U(0, 1)$;
 3. Sample $x^* \sim q(x^*|x^{(l)})$, $l = 1, 2, \dots$;
 4. Calculate $\mathcal{A}(x^{(l)}, x^*)$ according to (5);
 5. If $u < \mathcal{A}(x^{(l)}, x^*)$
 - $x^{(l+1)} = x^*$
 - else
 - $x^{(l+1)} = x^{(l)}$
 6. Repeat Steps (ii)–(iv);
 7. Discard the first P burn-in samples.
-

In the M-H algorithm, to prove the stationary, we define a transition kernel [33] as

$$\mathcal{P}(x^{(l+1)}|x^{(l)}) = q(x^{(l+1)}|x^{(l)})\mathcal{A}(x^{(l)}, x^{(l+1)}) + \delta(x^{(l+1)} - x^{(l)})\mathcal{B}(x^{(l)}), \quad (6)$$

where $\mathcal{B}(x^{(l)}) = \int q(x^*|x^{(l)})(1 - \mathcal{A}(x^{(l)}, x^*))dx^*$. By employing (4), (6) can be rewritten as

$$\mathcal{P}(x^{(l+1)}|x^{(l)}) = \min \left\{ q(x^{(l+1)}|x^{(l)}), \frac{q(x^{(l)}|x^{(l+1)})f(x^{(l+1)})}{f(x^{(l)})} \right\} + \delta(x^{(l+1)} - x^{(l)})\mathcal{B}(x^{(l)}). \quad (7)$$

Then we have

$$\mathcal{P}(x^{(l+1)}|x^{(l)})f(x^{(l)}) = \min\{q(x^{(l+1)}|x^{(l)})f(x^{(l)}), q(x^{(l)}|x^{(l+1)})f(x^{(l+1)})\} + \delta(x^{(l+1)} - x^{(l)})\mathcal{B}(x^{(l)})f(x^{(l)}), \quad (8)$$

$$\mathcal{P}(x^{(l)}|x^{(l+1)})f(x^{(l+1)}) = \min\{q(x^{(l)}|x^{(l+1)})f(x^{(l)}), q(x^{(l+1)}|x^{(l)})f(x^{(l+1)})\} + \delta(x^{(l)} - x^{(l+1)})\mathcal{B}(x^{(l+1)})f(x^{(l+1)}). \quad (9)$$

According to [33], it can be proven that $\delta(x^{(l+1)} - x^{(l)})\mathcal{B}(x^{(l)})f(x^{(l)}) = \delta(x^{(l)} - x^{(l+1)})\mathcal{B}(x^{(l+1)})f(x^{(l+1)})$. Therefore, with the use of (8)-(9) as well as (4), the balance condition of the M-H algorithm can easily to be hold.

In this algorithm, samples obtained in each iteration are closing to each other and can be highly correlated since M-H moves tend to be local moves. Asymptotically, the samples drawn from the Markov chain are all unbiased and all come from the target distribution.

3 Proposed Method

Without loss of generality, the observed data $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ is modeled as:

$$\begin{aligned} y_n &= A\cos(\omega n + \varphi) + q_n \\ &= a_1\cos(\omega n) + a_2\sin(\omega n) + q_n, \end{aligned} \tag{10}$$

where $a_1 = A\cos(\varphi), a_2 = -A\sin(\varphi)$ with A, ω, φ denoting amplitude, frequency and phase, respectively. The $q_n = c_n + g_n$ is the IID ACG noise, where c_n is the Cauchy noise with unknown median γ and g_n is the zero-mean Gaussian distributed with unknown variance σ^2 . Here our task is estimating A, ω and φ from observations $\{y_n\}_{n=1}^N$.

3.1 Posterior of Unknown Parameters

Here we investigate the posterior of unknown parameters. Before that, we first express the PDFs of noise terms c_n and g_n as:

$$f_C(c_n|\gamma) = \frac{\gamma}{\pi(c_n^2 + \gamma^2)}, \tag{11}$$

$$f_G(g_n|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{g_n^2}{2\sigma^2}\right). \tag{12}$$

Then the PDF of the mixture noise q_n , known as the Voigt profile [23], can be computed according to convolution of (9) and (10), which is

$$\begin{aligned} f_Q(q_n|\gamma, \sigma^2) &= \int_{-\infty}^{\infty} \frac{\gamma}{\pi((q_n - \tau)^2 + \gamma^2)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\tau^2}{2\sigma^2}} d\tau \\ &= \frac{\text{Re}\{w_n\}}{\sigma\sqrt{2\pi}}, \end{aligned} \tag{13}$$

where

$$w_n = \exp\left(-\left(\frac{q_n + i\gamma}{\sigma\sqrt{2}}\right)^2\right) \left(1 + \frac{2i}{\sqrt{\pi}} \int_0^{\frac{q_n + i\gamma}{\sigma\sqrt{2}}} \exp(t^2) dt\right). \tag{14}$$

Let $\theta = [a_1, a_2, \omega, \gamma, \sigma^2]^T$ being unknown parameter vector. According to the investigation in [34], the unknown parameters a_1 and a_2 are usually assumed to be following the IID Gaussian distribution with variance δ^2 and zero mean. While ω is the continuous uniform distributed between 0 and π , respectively.

Furthermore, it is also assumed in [34] that both γ and σ^2 follows the conjugate inverse-gamma distributions. Therefore, the priors for all unknown parameters a_1, a_2, ω, γ and σ^2 can be written as

$$f(a_1, a_2) = f(a_1)f(a_2) = \frac{1}{2\pi\delta^2} \exp\left(-\frac{a_1^2 + a_2^2}{2\delta^2}\right), \tag{15}$$

$$f(\omega) = \frac{1}{\pi}, \omega \in [0, \pi], \tag{16}$$

$$f(\gamma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma^{-\alpha-1} \exp\left(-\frac{\beta}{\gamma}\right), \quad (17)$$

$$f(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-2\alpha-2} \exp\left(-\frac{\beta}{\sigma^2}\right), \quad (18)$$

where α and β are set to $\alpha=10^{-10}$ and $\beta=0.01$, respectively.

According to the PDF expression of ACG noise in (13), the conditional PDF of the observation vector \mathbf{y} has the form of

$$\begin{aligned} f(\mathbf{y}|\theta) &= \prod_{n=1}^N f_{\mathcal{Q}}(e_n|\gamma, \sigma^2) \\ &= \prod_{n=1}^N \frac{\operatorname{Re}\{v_n\}}{\sigma\sqrt{2\pi}}, \end{aligned} \quad (19)$$

where $e_n = y_n - a_1 \cos(\omega n) - a_2 \sin(\omega n)$ denotes the residual between the observed data and the noise-free signal, and

$$v_n = \exp\left(-\left(\frac{e_n + i\gamma}{\sigma\sqrt{2}}\right)^2\right) \left(1 + \frac{2i}{\sqrt{\pi}} \int_0^{\frac{e_n + i\gamma}{\sigma\sqrt{2}}} \exp(t^2) dt\right). \quad (20)$$

Assume that the priors for all unknown parameters θ and the observations \mathbf{y} are statistically independent. With use of (15)–(18) and (19)–(20) as well as Bayes' theorem [25], the posterior expression of all unknown parameters θ can be

$$\begin{aligned} f(\theta|\mathbf{y}) &= f(\mathbf{y}|\theta) f(a_1, a_2) f(\omega) f(\gamma) \\ &= \frac{C_1}{C_2} \prod_{n=1}^N \operatorname{Re}\{v_n\}, \end{aligned} \quad (21)$$

where $C_1 = \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2}}{(2\pi)^{\frac{3}{2}} \pi \delta^2 \sigma \Gamma(\alpha_1) \Gamma(\alpha_2)}$ and $C_2 = \exp\left(\frac{a_1^2 + a_2^2}{2\delta^2} + \frac{\beta_1}{\gamma} + \frac{\beta_2}{\sigma^2}\right)$.

3.2 Proposed M-H Algorithm

Due to the multimodality of the likelihood function, the maximum likelihood estimator cannot be employed and the high computational complexity of the grid search. Furthermore, other typical robust estimators, such as the ℓ_1 -norm minimizer [35], cannot provide optimum estimation for the mixture noise. Moreover, even when the posterior PDFs of each unknown parameters $f(\theta|\mathbf{y})$ are known, the Gibbs sampling algorithm cannot be applied because of the complicated expression in (21).

Therefore, to estimate parameters accurately, the M-H algorithm is utilized, whose details are provided in Tab. 3. To simplify the sampling process, the multivariate Gaussian distribution is selected as the proposal distribution $q(\cdot|\cdot)$. In l -th sampling iteration, $q(\mathbf{x}^*|\mathbf{x}^{(l-1)})$ can be written as

$$q(\mathbf{x}^*|\theta^{(l-1)}) = \frac{1}{2\pi \sqrt{|\Sigma^{(l)}|}} \exp\left(-\frac{1}{2}(\mathbf{x}^* - \theta^{(l-1)})^T (\Sigma^{(l)})^{-1} (\mathbf{x}^* - \theta^{(l-1)})\right), \quad (22)$$

where $\mathbf{x}^* = [x_1 x_2 x_3 x_4 x_5]^T$ is the candidate vector with x_1, x_2, x_3, x_4 and x_5 corresponding to a_1, a_2, ω, γ and σ^2 , respectively, $\theta^{(l-1)} = [\theta_1^{(l-1)} \theta_2^{(l-1)} \theta_3^{(l-1)} \theta_4^{(l-1)} \theta_5^{(l-1)}]^T$ denotes the proposal mean vector with $\theta^{(l-1)}$ being the samples in $(l-1)$ th iteration and the $\Sigma^{(l)}$ is the 5×5 proposal covariance matrix. With

the assumption that all unknown parameters are independent, the $\Sigma^{(l)}$ is regarded as a diagonal matrix. Furthermore, the main diagonal entries of the proposal covariance are also called the proposal variances.

It is noted that the larger proposal variance will cause a faster convergence but possible oscillation around the correct value. While the smaller values of proposal variance lead to slower convergence but small fluctuation. Therefore, the choice of the proposal variance will significantly influence the performance of the estimator. In this paper, a batch-mode proposal variance selection criterion is developed.

To estimate $\Sigma^{(l)}$, we define two new batch-mode vector using the former L sampling vectors, which are

$$\begin{aligned} \Phi_1 &= [\theta^{(l-L)} \theta^{(l-L-1)} \dots \theta^{(l-1)}]^T \\ \Phi_2 &= [\theta^{(l-L-1)} \theta^{(l-L-2)} \dots \theta^{(l-2)}]^T, \end{aligned} \tag{23}$$

where L is also called the length of the batch-mode window. Then the proposal covariance $\Sigma^{(l)}$ can be defined by the empirical covariance of Φ_1 and Φ_2 , which is

$$\Sigma^{(l)} = (\Phi_1 - \Phi_2)(\Phi_1 - \Phi_2)^T. \tag{24}$$

To state the criterion clearly, the details are also shown in Fig. 1.

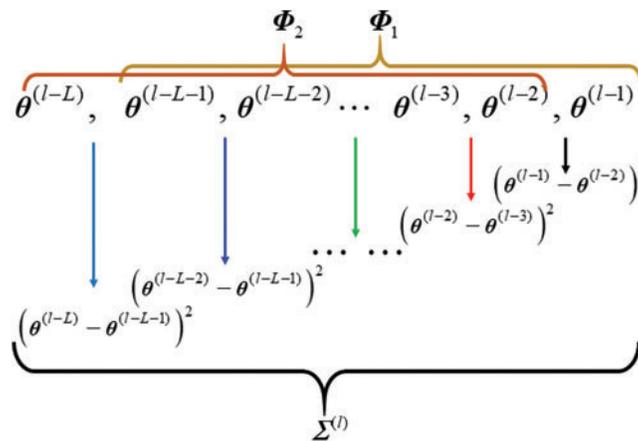


Figure 1: The construction of the proposal covariance

To start the algorithm, the initial estimate of θ and the burn-in period P should be determined. As it is discussed before, $\theta^{(1)}$ can be chosen arbitrarily because the initialization of the M-H method only affects the convergence rate. To guarantee the enough samples in the batch-mode criterion, the first P burn-in samples are generated with a fixed proposal covariance. From the l -th iteration ($l = P + 1, P + 2, \dots$), $\theta^{(l)}$ is calculated by the M-H algorithm using an adaptive $\Sigma^{(l)}$. After K numbers of sampling, the estimates \hat{a}_1, \hat{a}_2 and $\hat{\omega}$ are obtained from the mean of the first three row of the samples, which are

$$\begin{aligned}
\hat{a}_1 &= \frac{1}{K-P} \sum_{l=P+1}^K \theta^{(l)}(1), \\
\hat{a}_2 &= \frac{1}{K-P} \sum_{l=P+1}^K \theta^{(l)}(2), \\
\hat{\omega} &= \frac{1}{K-P} \sum_{l=P+1}^K \theta^{(l)}(3),
\end{aligned} \tag{25}$$

where $\theta^{(l)}(1)$, $\theta^{(l)}(2)$ and $\theta^{(l)}(3)$ are the first three elements of l -th sampling vector, respectively. The details of the proposed algorithm are shown in [Tab. 4](#).

Table 4: The proposed algorithm

-
1. Initialize θ as all ones;
 2. generate P sample vector using M-H algorithm in [Tab. 3](#) with the fixed $\Sigma = \mathbf{I}_5 \times 5$, where $\mathbf{I}_5 \times 5$ is an identity matrix;
 3. For $l = P + 1, P + 2, \dots$
 - 3.1 compute $\Sigma^{(l)}$ using [\(24\)](#);
 - 3.2 obtain $\theta^{(l)}$ using M-H algorithm in [Tab. 3](#) and $\Sigma^{(l)}$.
 4. Estimate the unknown parameters with [\(25\)](#).
-

Finally, utilize the definition of $a_1 = A \cos(\varphi)$ and $a_2 = -A \sin(\varphi)$, we can obtain the estimates of amplitudes and phase, denoted by \hat{A} and $\hat{\varphi}$,

$$\hat{A} = \sqrt{\hat{a}_1^2 + \hat{a}_2^2}, \tag{26}$$

$$\hat{\varphi} = -\text{atan} \left(\frac{\hat{a}_2}{\hat{a}_1} \right). \tag{27}$$

4 Derivation of CRLB

Let $\psi = [A \omega \phi \gamma \sigma^2]^T$. According to [\[21\]](#), the CRLB of ψ is usually obtained by the diagonal elements of \mathbf{F}^{-1} . Then we have

$$\begin{aligned}
\mathbf{F}(m, k) &= -E \left\{ \frac{\partial \log f(\mathbf{y}|\psi)}{\partial \psi} \left(\frac{\partial \log f(\mathbf{y}|\psi)}{\partial \psi} \right)^T \right\} \\
&= -E \left\{ \sum_{n=1}^N \frac{\partial \log f(y_n|\psi)}{\partial \psi} \left(\frac{\partial \log f(y_n|\psi)}{\partial \psi} \right)^T \right\},
\end{aligned} \tag{28}$$

where $m, k = 1, \dots, 5$ and

$$\frac{\partial \log f(y_n|\psi)}{\partial \psi} = \begin{bmatrix} \frac{1}{\sigma^2} \frac{\cos(\omega n + \phi) \text{Re}\{(y_n - A \cos(\omega n + \phi) + i\gamma)v_n\}}{\text{Re}\{w_n\}} \\ \frac{1}{\sigma^2} \frac{An \sin(\omega n + \phi) \text{Re}\{(y_n - A \cos(\omega n + \phi) + i\gamma)v_n\}}{\text{Re}\{v_n\}} \\ \frac{1}{\sigma^2} \frac{A \sin(\omega n + \phi) \text{Re}\{(y_n - A \cos(\omega n + \phi) + i\gamma)v_n\}}{\text{Re}\{v_n\}} \\ \frac{\text{Re}\{i(y_n - A \cos(\omega n + \phi) + i\gamma)w_n\} + \frac{2}{\sqrt{2\pi\sigma^2}}}{\sigma^2 \text{Re}\{v_n\}} \\ \frac{\text{Re}\{(y_n - A \cos(\omega n + \phi) + i\gamma)^2 v_n\} + \frac{\gamma}{\sqrt{2\pi\sigma^2\sigma^2}} - \frac{1}{2\sigma^2}}{\sigma^2 \text{Re}\{v_n\}} \end{bmatrix}. \tag{29}$$

Due to the complicated integral of v_n in (20), the closed-form expression of (29) is difficult to be obtained. As a result, with M Monte Carlo trials, (28) is calculated as

$$\hat{\mathbf{F}}(k, l) \approx -\frac{1}{M} \sum_{m=1}^M \sum_{n=1}^N \frac{\partial \log f(y_n^m|\psi)}{\partial \psi} \left(\frac{\partial \log f(y_n^m|\psi)}{\partial \psi} \right)^T, \tag{30}$$

where y_n^m denotes the n -th observed data in m -th trial. Apparently, (30) is only an approximation of the expectation. Therefore, a sufficiently large value of M will make (30) approaching (28).

5 Simulation Results

To assess the proposed algorithm, several simulations have been conducted. The mean square frequency error (MSFE), referred to as $E\{(\hat{\omega} - \omega)^2\}$, is employed as the performance measure. Then the noise-free signal s_n is generated with the use of (10) where $A = 9.33, \omega = 0.76$ and $\phi = 0.5$. In the M-H algorithm, the initial estimate is set to as $[11111]^T$, while the number of iterations is $K = 8000$. Here comparison with conventional estimators, such as the ℓ_1 -norm estimator, MLE and M -estimator [24] are provided due to its robust and suboptimal for the Cauchy noise, while the CRLB are also included as a benchmark. It is noted that the ℓ_1 -norm minizer is solved by the least absolute deviation [36], while the initial values of MLE and M -estimator are defined using fast Fourier transform. Furthermore, the stopping criterion of these three methods are relative error smaller than 10^{-8} . Simulations are obtained by using Matlab on Intel(R) Core(TM) i7-4790 CPU@3.60 GHz in Windows 7 operation system. While all results are based on 500 Monte Carlo runs and a data length of $N = 100$.

First of all, the choice of the batch-mode window length L for the proposal covariance matrix is studied. Here the density parameters of ACG noise are set to $\gamma = 0.05$ and $\sigma^2 = 0.5$. Figs. 2 and 3 shows the MSFE vs. L and the corresponding computational time. Here the computation time is measured with stopwatch timer in the simulator. It is shown that when $L \leq 1000$, the MSFE can be aligned with CRLB, while the computational cost of the proposed algorithm becomes higher when L increases. Therefore, in the following test, we choose L as 1000.

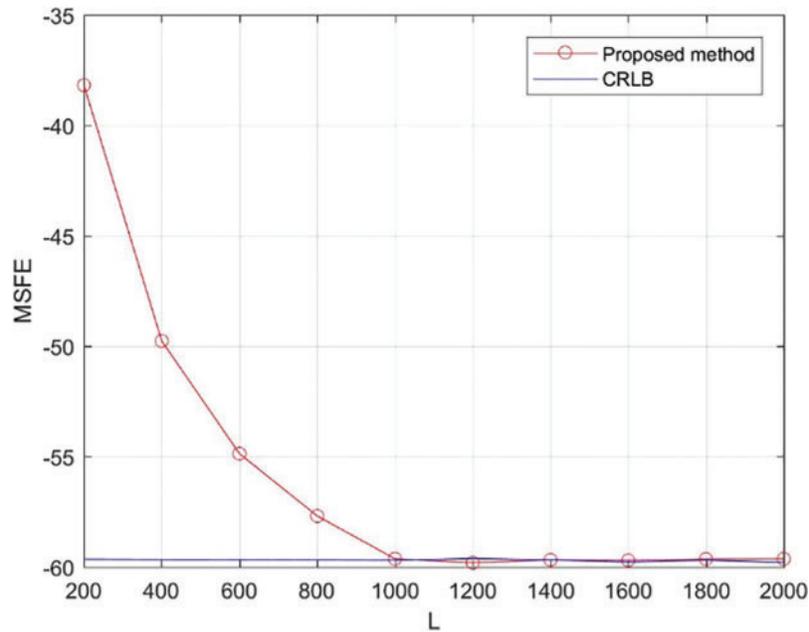


Figure 2: MSFE vs. L

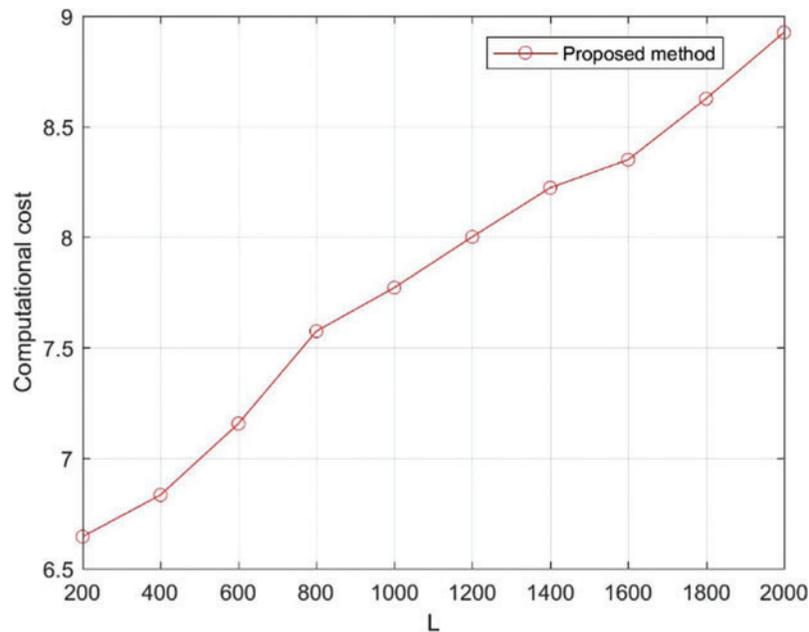


Figure 3: The computational cost vs. L

Second, the convergence rate of the unknown parameters is investigated. Meanwhile, the burn-in period P can be determined, accordingly. In this test, the density parameters are identical to the previous test. Figs. 4 and 5 indicates the estimates of all unknown parameters in different iteration number l , which are ω , A , ϕ , γ and σ^2 . It can be seen in these figures that after the first 2000 samples,

the sampled data approaches the true values of unknown parameters. Therefore, the burn-in period P can be chosen as 2000 in this parameter setting.

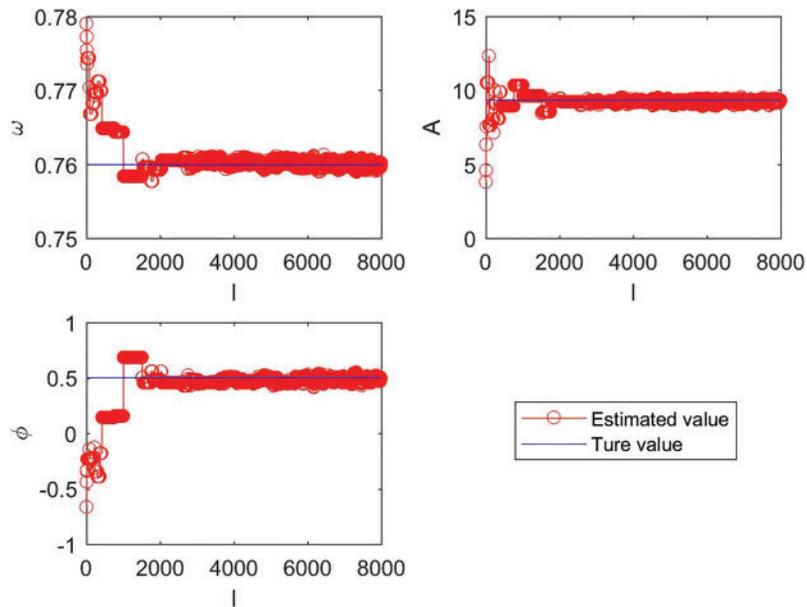


Figure 4: Estimates of unknown parameters vs. iteration number k

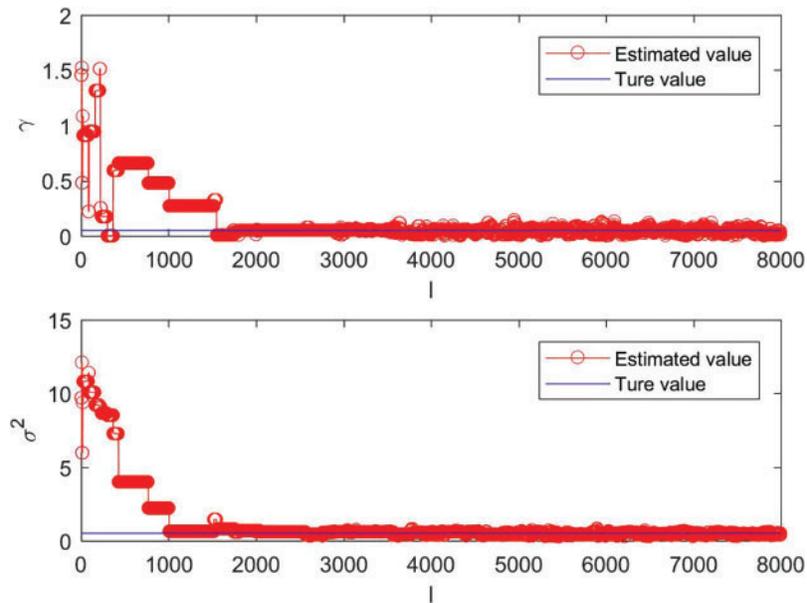


Figure 5: Estimates of density parameters vs. iteration number l

Finally, the MSFE of the proposed estimator is considered. In this test, all parameters are chosen as the same with the previous test. As there is no finite variance for ACG noise [18], the signal-to-noise is difficult to be defined. Therefore, here γ is scaled to produce different noise conditions. According to the study in the previous test, we throw away first 2000 samples to guarantee the stationary of the

Markov chain. It is observed in Fig. 6 that the MSFE of the proposed method can attain the CRLB when $\gamma \in [-20, 10]$ dB. Furthermore, the proposed algorithm is superior to the other three estimators, since it still can work well in higher γ .

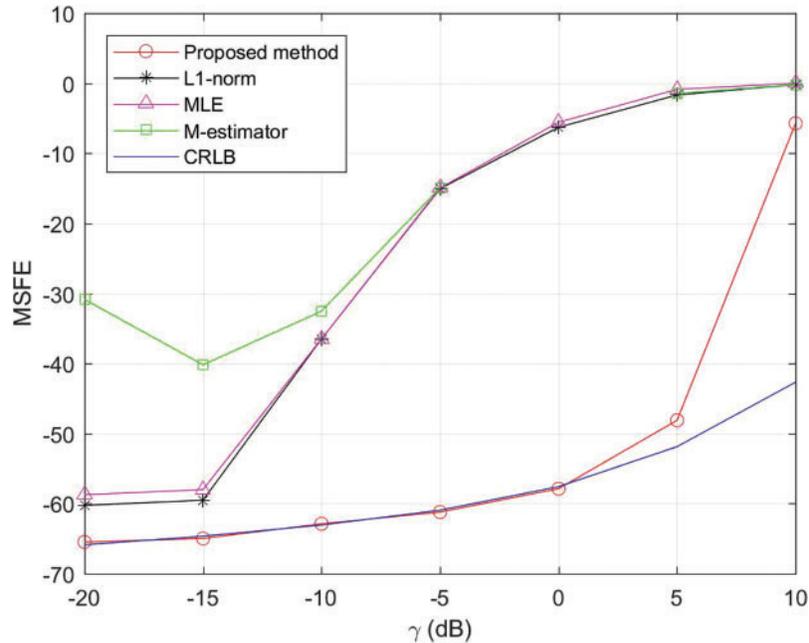


Figure 6: Mean square frequency error of ω vs. γ

6 Conclusion

In this paper, with the use of the M-H algorithm, a robust parameter estimator of a single sinusoid has been developed, in the presence of additive Cauchy-Gaussian noise. Meanwhile, a new proposal covariance updating criterion is also devised by employing the squared error of the batch-mode M-H samples. It is shown in simulation results that the developed estimator can attain the CRLB with a stationary M-H chain, indicating the accurate of our scheme. In the future work, the method can be extended to the signals with more complicated models.

Funding Statement: The work was supported by National Natural Science Foundation of China (Grant No. 52075397, 61905184, 61701021) and Fundamental Research Funds for the Central Universities (Grant No. FRF-TP-19-006A3).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.
- [2] Z. T. Li, W. Wei, T. Z. Zhang, M. Wang, S. J. Hou *et al.*, "Online multi-expert learning for visual tracking," *IEEE Transactions on Image Processing*, vol. 29, pp. 934–946, 2020.

- [3] Z. T. Li, J. Zhang, K. H. Zhang and Z. Y. Li, "Visual tracking with weighted adaptive local sparse appearance model via spatio-temporal context learning," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4478–4489, 2018.
- [4] Y. Fei, "A 2.7 GHz low-phase-noise LC-QVCO using the gate-modulated coupling technique," *Wireless Personal Communications*, vol. 86, no. 2, pp. 1–11, 2015.
- [5] Y. Xue, Q. Li and F. Ling, "Teensensor: Gaussian processes for micro-blog based teen's acute and chronic stress detection," *Computer Systems Science and Engineering*, vol. 34, no. 3, pp. 151–164, 2019.
- [6] F. Mallouli, "Robust em algorithm for iris segmentation based on mixture of Gaussian distribution," *Intelligent Automation & Soft Computing*, vol. 25, no. 2, pp. 243–248, 2019.
- [7] J. Wang, Y. Yang, T. Wang, R. Sherratt and J. Zhang, "Big data service architecture: A survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393–405, 2020.
- [8] J. Zhang, S. Zhong, T. Wang, H. -C. Chao and J. Wang, "Block chain-based systems and applications: A survey," *Journal of Internet Technology*, vol. 21, no. 1, pp. 1–4, 2020.
- [9] T. Zhang, A. Wiesel and M. S. Greco, "Multivariate generalized Gaussian distribution: Convexity and graphical models," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4141–4148, 2013.
- [10] C. L. Nikias and M. Shao, "Signal processing with fractional lower order moments: Stable processes and their applications," in *Proc. of the IEEE*, vol. 81, no. 7, pp. 986–1010, 1993.
- [11] K. Aas and I. H. Haff, "The generalized hyperbolic skew student's t-distribution," *Journal of Financial Econometrics*, vol. 4, no. 2, pp. 275–309, 2006.
- [12] G. R. Arce, "NonGaussian models," in *Nonlinear Signal Processing: A Statistical Approach*, 1st ed., New York, USA: John Wiley & Sons Inc., pp. 17–42, 2005.
- [13] D. A. Reynolds, "Gaussian mixture models," in *Encyclopedia of Biometrics*, 1st ed., New York, USA: Springer-Verlag, pp. 659–663, 2009.
- [14] A. Swami, "Non-Gaussian mixture models for detection and estimation in heavy tailed noise," in *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, pp. 3802–3805, 2000.
- [15] 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.
- [16] 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, 1998.
- [17] Y. Chen, E. E. Kuruoglu and H. C. So, "Estimation under additive Cauchy-Gaussian noise using markov chain monte carlo," in *Proc. of IEEE Statistical Signal Processing Workshop (SSP)*, Gold Coast, Australia, pp. 356–359, 2014.
- [18] Y. Chen, E. E. Kuruoglu, H. C. So, L. -T. Huang and W. -Q. Wang, "Density parameter estimation for additive Cauchy-Gaussian mixture," in *Proc. of IEEE Statistical Signal Processing Workshop (SSP)*, Gold Coast, Australia, pp. 205–208, 2014.
- [19] F. Kahrari, M. Rezaei and F. Yousefzadeh, "On the multivariate skew-normal Cauchy distribution," *Statistics & Probability Letters*, vol. 117, pp. 80–88, 2016.
- [20] F. W. J. Olver, D. M. Lozier and R. F. Boisvert, "Parabolic Cylinder functions," in *NIST Handbook of Mathematical Functions*, 1st ed., Cambridge, UK: Cambridge University Press, pp. 167–168, 2010.
- [21] S. M. Kay, "Maximum likelihood estimation," in *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed., New Jersey: Prentice-Hall, pp. 157–218, 1993.
- [22] S. Fang, L. Huang, Y. Wan, W. Sun and J. Xu, "Outlier detection for water supply data based on joint auto-encoder," *Computers, Materials & Continua*, vol. 64, no. 1, pp. 541–555, 2020.
- [23] F. Wang, Z. Wei and X. Zuo, "Anomaly iot node detection based on local outlier factor and time series," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 1063–1073, 2020.
- [24] M. Subzar, A. I. Al-Omari and A. R. A. Alanzi, "The robust regression methods for estimating of finite population mean based on srswor in case of outliers," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 125–138, 2020.

- [25] Y. Chen, E. E. Kuruoglu and H. C. So, "Optimum linear regression in additive Cauchy-Gaussian noise," *Signal Processing*, vol. 106, no. 1, pp. 312–318, 2015.
- [26] G. L. Bretthorst, "Estimating the parameters," in *Bayesian Spectrum Analysis and Parameter Estimation*, 1st ed., New York, USA: Springer-Verlag, pp. 43–54, 1988.
- [27] C. Bishop, "Sampling methods," in *Pattern Recognition and Machine Learning*, 1st ed., New York, USA: Springer, pp. 523–558, 2006.
- [28] C. Andrieu, N. D. Freitas, A. Doucet and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, no. 1–2, pp. 5–43, 2003.
- [29] J. C. Spall, "Estimation via markov chain monte carlo," *IEEE Control Systems*, vol. 23, no. 2, pp. 34–45, 2003.
- [30] C. M. Grinstead and J. L. Snell, "Markov chains," in *Introduction to Probability*, 2nd ed., Rhode Island, USA: American Mathematical Society, pp. 405–470, 2012.
- [31] C. Robert and G. Casella, "Metropolis-Hastings algorithms," in *Introducing Monte Carlo Methods with R*, 1st ed., New York, USA: Springer Verlag, pp. 166–198, 2009.
- [32] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.
- [33] A. T. Cemgil, "A tutorial introduction to Monte Carlo methods, Markov chain Monte Carlo and particle filtering," in *Academic Press Library in Signal Processing, Volume 1: Signal Processing Theory and Machine Learning*, 1st ed., Oxford, UK: Elsevier Science & Technology, pp. 1065–1114, 2013.
- [34] R. Kohn, M. Smith and D. Chan, "Nonparametric regression using linear combinations of basis functions," *Statistics and Computing*, vol. 11, no. 4, pp. 313–322, 2001.
- [35] T. H. Li, "A nonlinear method for robust spectral analysis," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2466–2474, 2010.
- [36] Y. Li and G. Arce, "A maximum likelihood approach to least absolute deviation regression," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 12, pp. 1762–1769, 2004.