Signal Processing 106 (2015) 312–318

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

Optimum linear regression in additive Cauchy–Gaussian noise

Yuan Chen [a,*], Ercan Engin Kuruoglu [b], Hing Cheung So [a]

[a] Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China
[b] ISTI-CNR (Italian National Council of Research), Pisa, Italy

ARTICLE INFO

Article history:
Received 17 March 2014
Received in revised form
18 July 2014
Accepted 30 July 2014
Available online 12 August 2014

Keywords:
Impulsive noise
Cauchy distribution
Gaussian distribution
Mixture noise
Voigt profile
Maximum likelihood estimator
Pseudo-Voigt function
M-estimator

ABSTRACT

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 (MLE) is developed first. 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. For the latter case, we estimate the density parameters by utilizing the empirical characteristic function prior to applying the MLE. Simulation results show that the performance of both proposed methods can attain the Cramér–Rao lower bound.

1. Introduction

Impulsive noise appears 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 distributions and hybrid distributions mixed in the probability density function (PDF) domain. Typical single distributions are Student's t-distribution [2], α-stable distribution [3] and generalized Gaussian distribution [4], while the mixture models include Gaussian mixture [5] and Cauchy Gaussian mixture (CGM) [6]. Nevertheless, these models alone may not be able to represent all varieties of impulsive noise in the real world, particularly when the noise measured is the sum of two time series: one is an intrinsic Gaussian noise due to the electronic devices in receiver and the other is interference from the environment which is non-Gaussian distributed. For example, in frequency-hopping spread spectrum radio communication networks [7], binary transmission systems [8] and multiple-input multiple-output systems [9], we may model the multiple access interference as α-stable distribution and regard the environmental noise as Gaussian distribution. Similarly, in astrophysical image processing [10], the cosmic microwave background radiation is contaminated with Gaussian noise from the satellite beam and α-stable distributed radiation from galaxies and stars. In these potential applications, the disturbance components correspond to 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 distributed variable with dispersion γ and zero-mean Gaussian distributed variable with variance σ^2 as an illustrative example. This mixture model belongs to the Middletons Class B [11] model which is a

* Corresponding author. Fax: +852 3442 0562.
E-mail address: qchenyuan00@126.com (Y. Chen).

http://dx.doi.org/10.1016/j.sigpro.2014.07.028
0165-1684/

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 via some statistical means, e.g., through the relationship between the empirical characteristic function (ECF) and 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 while reducing the computational complexity, a new M-estimator with ρ being 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 $\mathbf{y} = [y_1\ y_2\ \cdots\ y_N]^T$ 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

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e}, \tag{3}$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ \vdots & \vdots \\ N & 1 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} A \\ B \end{bmatrix}, \quad \mathbf{e} = [e_1\ e_2\ \cdots\ e_N]^T \tag{4}$$

and $e_n = c_n + g_n$ denotes the additive Cauchy Gaussian (ACG) noise which is the sum of an 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. It is noteworthy that (3) and (4) are also a common signal model for kick detection in oil drilling [14]. Although we only study this simple model, our analysis can be extended to the general linear data model [15], that is, $\mathbf{H} \in R^{N \times M}$ with $N \geq M$ is known and $\boldsymbol{\theta} \in R^M$ is unknown. The PDFs of Cauchy and Gaussian distributions are

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

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

According to (2), the PDF of e_n is calculated as

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

The expression of (7) can be represented as the so-called Voigt function [12]:

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

where

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

is called the Faddeeva function and $\mathrm{Re}\{\cdot\}$ denotes the real part.

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

$$J(\boldsymbol{\theta}) = \sum_{n=1}^{N} \rho_n, \tag{10}$$

where ρ_n is an arbitrary function [13] which can be chosen, e.g., as $-\log(f_n)$. Note that the M-estimator coincides with the MLE when f_n is the ACG's PDF $f_E(y_n, \boldsymbol{\theta}; \gamma, \sigma^2)$. In the following, we work on 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

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