Signal Processing 135 (2017) 17–25

Contents lists available at ScienceDirect

Signal Processing

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

Short communication

Variance analysis of unbiased complex-valued ℓ_p-norm minimizer

Yuan Chen[a,*], Hing Cheung So[b], Ercan Engin Kuruoglu[c], Xiao Long Yang[a]

[a] School of Computer & Communication Engineering, University of Science & Technology Beijing, Beijing, China
[b] Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
[c] Italian National Council of Research (ISTI-CNR), Pisa, Italy

ARTICLE INFO

Keywords:
Variance analysis
Complex-valued signals
Fractional lower-order moment
ℓ_p-norm minimization
Digamma function
Taylor series expansion

ABSTRACT

Parameter estimation from noisy complex-valued measurements is a significant topic in various areas of science and engineering. In this aspect, an important goal is finding an unbiased estimator with minimum variance. Therefore, variance analysis of an estimator is desirable and of practical interest. In this paper, we concentrate on analyzing the complex-valued ℓ_p-norm minimizer with $p \geq 1$. Variance formulas for the resultant nonlinear estimators in the presence of three representative bivariate noise distributions, namely, α-stable, Student's t and mixture of generalized Gaussian models, are derived. To guarantee attaining the minimum variance for each noise process, optimum selection of p is studied, in the case of known noise statistics, such as probability density function and corresponding density parameters. All our results are confirmed by simulations and are compared with the Cramér-Rao lower bound.

1. Introduction

In many areas of science and engineering, such as wireless communications, sensor array signal processing and biomedical sciences [1–3], observations are more conveniently modeled as complex-valued data, which have a simpler analytical form and are easier to deal with than the real-valued model. Parameter estimation for the complex-valued observations [4,5] is an important research topic and has attracted a great deal of attention. For numerous estimators developed in the literature, the goal is to find one which on the average yields the true value and the mean square error (MSE) between the estimate and true value is the smallest. MSE analysis of unbiased estimators is significant to help searching for one with minimum variance. Since the definition of variance relates to expectation, calculating by excessive simulations may be nonconclusive and unrealistic. To obtain the variance elegantly and correctly, approaches such as Taylor series expansion (TSE) on the estimates [6] and on the error function [7] are proposed, which are verified in [8]. However, they only consider estimators in the presence of complex Gaussian noise, which cannot describe other types of disturbances, especially those with impulsive nature, appeared in many fields [9–11]. For example, symmetric α-stable (SαS), Student's t and mixture of generalized Gaussian (MGG) processes are commonly employed to model the complex-valued impulsive noise.

Due to the high computational complexity and difficult implementation in case of lacking analytical probability density function (PDF), the maximum likelihood (ML) estimator may not be the proper choice in the presence of non-Gaussian noise models. Alternatively, a robust estimation methodology, namely, M-estimator is proposed by Huber [12], which generalizes the ML estimator by replacing the logarithm of the likelihood function by an arbitrary ρ-function. Defining the ρ-function as the ℓ_p-norm of residual [13], the least ℓ_p-norm estimator with $p \geq 1$ is widely utilized in the environment with non-Gaussian noise.

In this paper, we investigate the performance of the ℓ_p-norm minimizer using the complex-valued observations. The SαS, Student's t and MGG noise models are considered. To guarantee the unbiasedness of the least ℓ_p-norm estimator, we assume that all discussed models are symmetric with zero location parameter [14]. For the ℓ_p-norm minimizer, since the variance formulas should be a function of p, the selection of optimum p is required to ensure achieving the minimum variance. In [15], the relationship between kurtosis and optimum p is investigated, however, the results cannot be used when the kurtosis does not exist such as in the presence of SαS noise. The optimal selection of p for SαS distribution is discussed in [16], but optimum estimation performance for other noise models is not addressed. Furthermore, although [14] shows that the optimum p can be obtained by a polynomial equation with a very large degree of nonlinearity, it suffers a high computational cost in solving such a high-order polynomial function. To overcome the complexity problem, in this work we simplify the optimal choice of p into finding root of a low-order function using the property of gamma function. It is worth

* Corresponding author.
E-mail address: qchenyuan00@126.com (Y. Chen).

http://dx.doi.org/10.1016/j.sigpro.2016.12.018
Received 20 October 2016; Received in revised form 29 November 2016; Accepted 16 December 2016
Available online 23 December 2016
0165-1684/

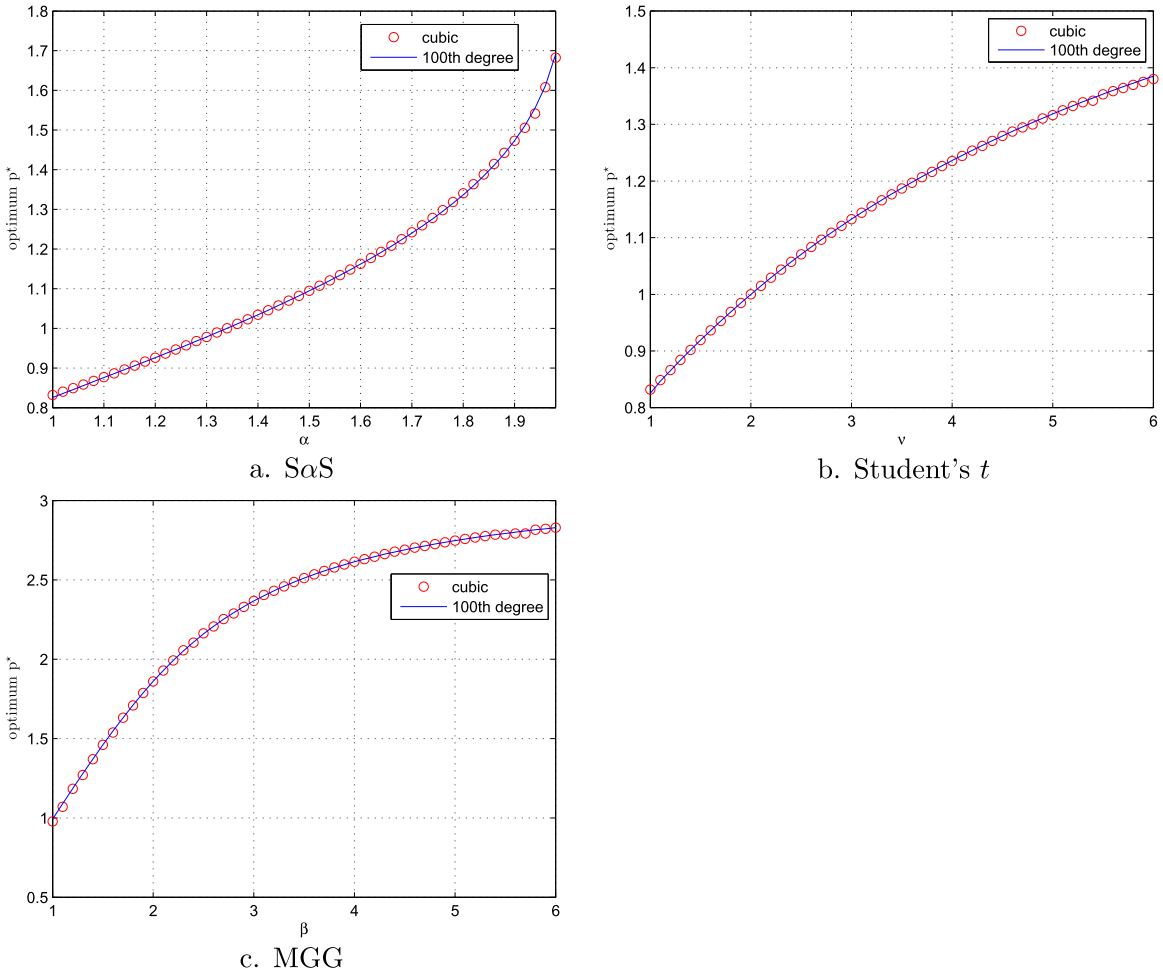

Fig. 1. Optimum $p^\star$ calculated by cubic and 100-th degree polynomials versus density parameters.

mentioning that our investigation on optimality assumes the availability of *a priori* knowledge of the density parameters for each noise model. In the scenario of the unknown distribution parameters, they should be estimated first based on a sufficiently large number of noise-only samples [17–19] and then our method can be applied. Although we focus on the complex-valued estimation problem in this paper, our result can be extended to the real-valued case.

The rest of this paper is organized as follows. In Section 2, we briefly review the bivariate SαS, Student's t and MGG models. The variance formulas of the least ℓ_p-norm estimator are devised in Section 3. The selection of p for different noise processes with the minimum variance is also examined. In Section 4, simulations are provided to validate the derived variance formulas with comparison to the Cramér-Rao lower bound (CRLB). Finally, conclusions are drawn in Section 5.

2. Review of well-known distributions

According to the analysis in [14], the variance formula needs the fractional lower-order moment (FLOM) of the noise. Since a complex random variable corresponds to a bivariate distribution, in this section, we review the bivariate SαS, Student's t and MGG distributions.

2.1. SαS distribution

Let $q = \Re\{q\} + j\Im\{q\}$ follow a bivariate SαS distribution. Its characteristic function has the form of [20]:

$$\varphi(t) = \exp(j(\delta_r\Re\{t\} + \delta_i\Im\{t\}) - \gamma(\Re\{t\}^2 + \Im\{t\}^2)^{\frac{\alpha}{2}}), \tag{1}$$

where $\alpha \in (1, 2]$ is the characteristic exponent, controlling the impulsiveness of the distribution, $(\delta_r, \delta_i) \in (-\infty, \infty) \times (-\infty, \infty)$ denote the location parameters, and $\gamma > 0$ is the dispersion parameter which determines the spread of the distribution. Note that we set $\delta_r = \delta_i = 0$ here because the location parameter of noise term is assumed zero.

The FLOMs of real and imaginary parts of q, denoted by $E\{|q|^{l-2}\Re\{q\}^2\}$ and $E\{|q|^{l-2}\Im\{q\}^2\}$, are [21]:

$$E\{|q|^{l-2}\Re\{q\}^2\} = D_\alpha(l, \alpha)\gamma^{\frac{l}{\alpha}}, \tag{2}$$

$$E\{|q|^{l-2}\Im\{q\}^2\} = D_\alpha(l, \alpha)\gamma^{\frac{l}{\alpha}}, \quad -2 < l < \alpha, \tag{3}$$

where $E\{\cdot\}$ denotes the expectation operator and

$$D_\alpha(l, \alpha) = \frac{\Gamma\left(\frac{l+2}{2}\right)\Gamma\left(1 - \frac{l}{\alpha}\right)}{\Gamma\left(1 - \frac{l}{2}\right)}2^{l-1}, \tag{4}$$

with $\Gamma(\cdot)$ being the gamma function.

2.2. Student's t distribution

For a zero-mean complex-valued variable $q = \Re\{q\} + j\Im\{q\}$, which follows the bivariate Student's t distribution, the PDF is [10]:

$$f(q) = \frac{1}{2\pi\eta^2}\left(1 + \frac{1}{\nu}\left(\frac{\Re\{q\}^2 + \Im\{q\}^2}{\eta^2}\right)\right)^{-\frac{\nu+2}{2}}, \tag{5}$$

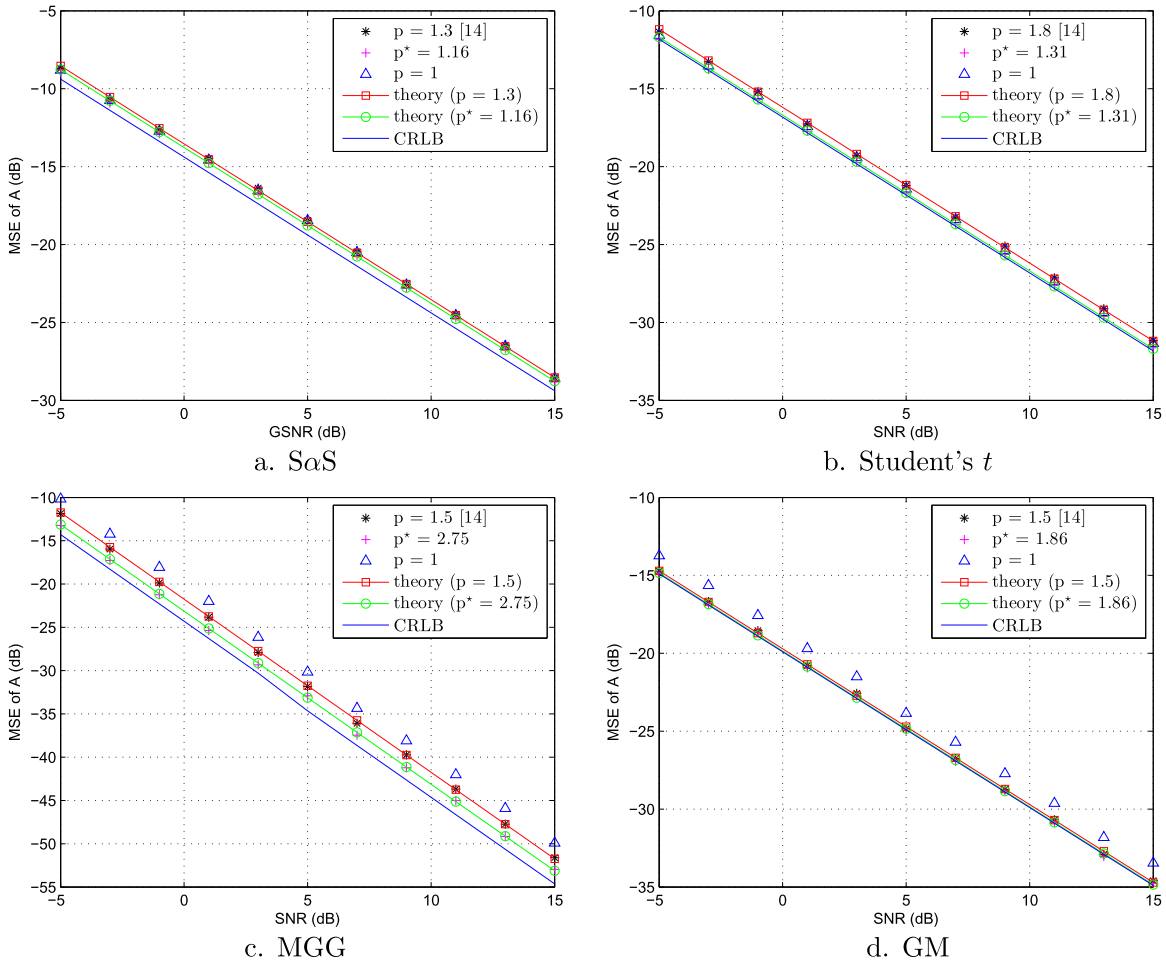

Fig. 2. MSE of sinusoidal amplitude in complex-valued observations.

where $\nu > 0$ is the number of degrees of freedom, and $\eta > 0$ denotes the scaling parameter determining the spread of the PDF.

The FLOMs of $\Re\{q\}$ and $\Im\{q\}$ are:

$$E\{|q|^{l-2}\Re\{q\}^2\} = D_t(l, \nu)\eta^l, \tag{6}$$

$$E\{|q|^{l-2}\Im\{q\}^2\} = D_t(l, \nu)\eta^l, \quad -2 < l < \nu, \tag{7}$$

where

$$D_t(l, \nu) = \begin{cases} 0 & \nu > 2 \ \& \ l \text{ is odd}, \\ \dfrac{\Gamma\left(\dfrac{l+2}{2}\right)\Gamma\left(\dfrac{\nu-l}{2}\right)}{2\Gamma(\nu/2)}\nu^{\frac{l}{2}} & \text{otherwise}. \end{cases} \tag{8}$$

2.3. MGG distribution

MGG process [22] can have different combinations, but we only consider the two-component case with the same shape parameter.

Suppose $q = \Re\{q\} + j\Im\{q\}$ following the zero-mean symmetric MGG process, its PDF is:

$$f(q) = (1 - \epsilon)f_1(q; \beta, \sigma) + \epsilon f_1(q; \beta, \tau\sigma), \tag{9}$$

where $f_1(q; \beta, \sigma) = \frac{\beta}{2\pi\sigma^2\Gamma(2/\beta)}\exp\left(-\frac{(\Re\{q\}^2 + \Im\{q\}^2)^{\frac{\beta}{2}}}{\sigma^\beta}\right)$, $\epsilon \in (0, 1)$ is the weight parameter, $\beta > 0$ denotes the shape parameter tuning the decay rate of the density function, σ and $\tau\sigma$ with $\tau > 0$ are the scaling parameters of the two components. Note that when $\beta = 2$ or/and $\epsilon = 0, 1$, the MGG processes are reduced to Gaussian mixture (GM) and generalized Gaussian (GG) models, respectively.

The FLOMs of real and imaginary parts are:

$$E\{|q|^{l-2}\Re\{q\}^2\} = D_{\mathrm{MGG}}(l, \beta, \epsilon, \tau)\sigma^l, \tag{10}$$

$$E\{|q|^{l-2}\Im\{q\}^2\} = D_{\mathrm{MGG}}(l, \beta, \epsilon, \tau)\sigma^l, \tag{11}$$

where

$$D_{\mathrm{MGG}}(l, \beta, \epsilon, \tau) = \begin{cases} 0 & \beta > 2 \ \& \ l \text{ is odd}, \\ \dfrac{\Gamma\left(\dfrac{l+2}{\beta}\right)(1 - \epsilon + \epsilon\tau^l)}{2\Gamma(2/\beta)} & \text{otherwise}. \end{cases} \tag{12}$$

3. Variance analysis of least ℓ_p-norm estimator

Without loss of generality, the complex-valued observed sequence $\mathbf{y} = [y_1 \cdots y_N]^T$ is modeled as

$$y_n = g_n(\mathbf{x}) + q_n, \tag{13}$$

where ^{T} is the transpose operator, $g_n(\mathbf{x}) = \Re\{g_n(\mathbf{x})\} + j\Im\{g_n(\mathbf{x})\}$ denotes a differentiable function of $\mathbf{x}$ with $\mathbf{x} = [x_1 \cdots x_M]^T$ being the deterministic parameter vector of interest, and q_n is the identically independent distributed (IID) noise component with zero location parameter. Our task is to find $\mathbf{x}$ from $\mathbf{y}$ in the presence of one of the non-Gaussian noise models, namely, SαS, Student's t and MGG.

To solve the problem, we concentrate on the complex-valued ℓ_p-norm with $p \geq 1$:

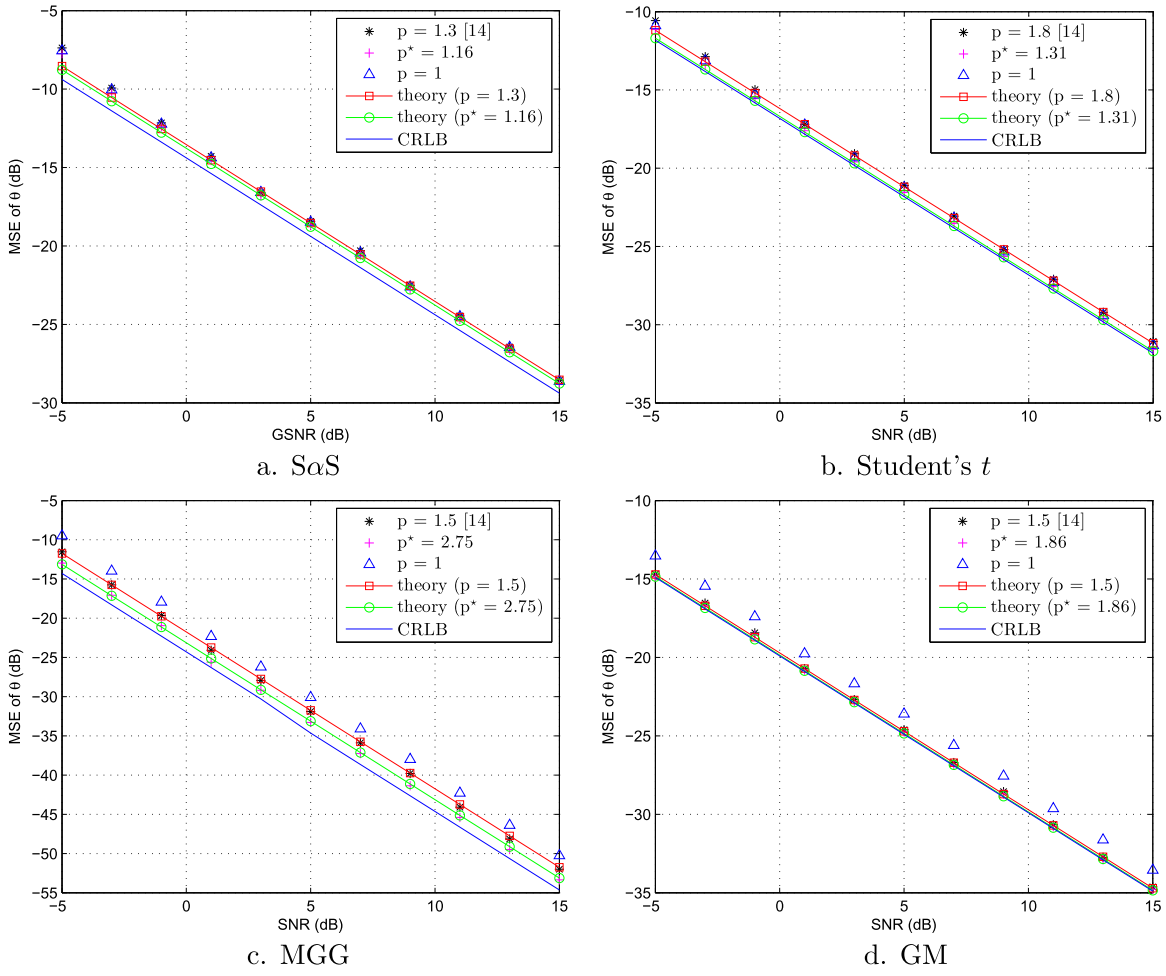

Fig. 3. MSE of sinusoidal phase in complex-valued observations.

$$J(\mathbf{x}) = \sum_{n=1}^{N} |y_n - g_n(\mathbf{x})|^p = \sum_{n=1}^{N} ((\Re\{y_n\} - \Re\{g_n(\mathbf{x})\})^2 + (\Im\{y_n\} - \Im\{g_n(\mathbf{x})\})^2)^{\frac{p}{2}}. \tag{14}$$

3.1. Variance formula derivation

For the symmetric noise model, the ℓ_p-norm minimizer is always unbiased and the covariance matrix for $\hat{\mathbf{x}}$, denoted by $\mathbf{C}(\hat{\mathbf{x}})$, is [8]:

$$\mathbf{C}(\hat{\mathbf{x}}) = (E\{\mathbf{H}(J(\mathbf{x}))\})^{-1} E\{\nabla(J(\mathbf{x}))\nabla^H(J(\mathbf{x}))\}(E\{\mathbf{H}(J(\mathbf{x}))\})^{-1}, \tag{15}$$

where $\mathbf{H}(J(\mathbf{x}))$ and $\nabla(J(\mathbf{x}))$ are the Hessian matrix and gradient vector at the true value of $\mathbf{x}$, $^{-1}$ and ^{H} denote the matrix inverse and conjugate transpose operator, respectively. The variances of the estimates in $\hat{\mathbf{x}}$ are the diagonal elements of $\mathbf{C}(\hat{\mathbf{x}})$.

Utilizing (14), the expected values of the required derivatives in (15) are:

$$E\{\nabla(J(\mathbf{x}))\nabla^H(J(\mathbf{x}))\} = p^2\{\nabla_{\mathbf{x}}^H(\Re\{\mathbf{g}(\mathbf{x})\})E\{\Re\{\mathbf{V}\}\}\nabla_{\mathbf{x}}(\Re\{\mathbf{g}(\mathbf{x})\}) + \nabla_{\mathbf{x}}^H(\Im\{\mathbf{g}(\mathbf{x})\}E\{\Im\{\mathbf{V}\}\}\nabla_{\mathbf{x}}(\Im\{\mathbf{g}(\mathbf{x})\})\}, \tag{16}$$

$$E\{\mathbf{H}(J(\mathbf{x}))\} = p^2\{\nabla_{\mathbf{x}}^H(\Re\{\mathbf{g}(\mathbf{x})\})E\{\Re\{\mathbf{W}\}\}\nabla_{\mathbf{x}}(\Re\{\mathbf{g}(\mathbf{x})\}) + \nabla_{\mathbf{x}}^H(\Im\{\mathbf{g}(\mathbf{x})\}E\{\Im\{\mathbf{W}\}\}\nabla_{\mathbf{x}}(\Im\{\mathbf{g}(\mathbf{x})\})\}, \tag{17}$$

where $\Re\{\mathbf{V}\}$, $\Im\{\mathbf{V}\}$, $\Re\{\mathbf{W}\}$ and $\Im\{\mathbf{w}\}$ are $N \times N$ diagonal matrices with n-th elements $\Re\{V_n\} = E\{|q_n|^{2p-4}\Re\{q_n\}^2\}$, $\Im\{V_n\} = E\{|q_n|^{2p-4}\Im\{q_n\}^2\}$, $\Re\{W_n\} = E\{|q_n|^{p-4}\Re\{q_n\}^2\}$ and $\Im\{W_n\} = E\{|q_n|^{p-4}\Im\{q_n\}^2\}$, respectively. Here we consider two scenarios, which are real-valued $\mathbf{x}$ and complex-valued $\mathbf{x}$. For the real-valued case, $\nabla_{\mathbf{x}}(\cdot)$ denotes the gradient vector at $\mathbf{x}$. While for the complex-valued $\mathbf{x}$, employing the Wirtinger derivatives [23] yields $\nabla_{\mathbf{x}}(\cdot) = \frac{1}{2}\left(\frac{\partial}{\partial\Re\{\mathbf{x}\}} - i\frac{\partial}{\partial\Im\{\mathbf{x}\}}\right)$.

Using (15)–(17) as well as FLOM expressions in Section 2, the covariance matrix expression of the ℓ_p-norm minimizer in the presence of three distributions, namely, SαS, Student's t and MGG processes, can be derived as:

$$\mathbf{C}_{\alpha}(\hat{\mathbf{x}}) = \frac{2\Gamma^2\left(2-\frac{p}{2}\right)\Gamma(p)\Gamma\left(1-\frac{2p-2}{\alpha}\right)\gamma^{\frac{2}{\alpha}}}{\Gamma^2\left(\frac{p}{2}+1\right)\Gamma(2-p)\Gamma^2\left(1-\frac{p-2}{\alpha}\right)}\mathbf{G}^{-1}, \quad 1 \le p < \alpha, \tag{18}$$

$$\mathbf{C}_{t}(\hat{\mathbf{x}}) = \frac{\nu\Gamma\left(\frac{\nu}{2}\right)\Gamma(p)\Gamma\left(\frac{\nu-2p+2}{2}\right)\eta^2}{\Gamma^2\left(1+\frac{p}{2}\right)\Gamma^2\left(\frac{\nu-p+2}{2}\right)}\mathbf{G}^{-1}, \quad p \ge 1, \tag{19}$$

$$\mathbf{C}_{\mathrm{MGG}}(\hat{\mathbf{x}}) = \frac{\Gamma\left(\frac{2p}{\beta}\right)\Gamma\left(\frac{2}{\beta}\right)((1-\epsilon)+\epsilon\tau^{2p-2})\sigma^2}{\beta^2\Gamma^2\left(1+\frac{p}{\beta}\right)((1-\epsilon)+\epsilon\tau^{p-2})^2}\mathbf{G}^{-1}, \quad p \ge 1. \tag{20}$$

where

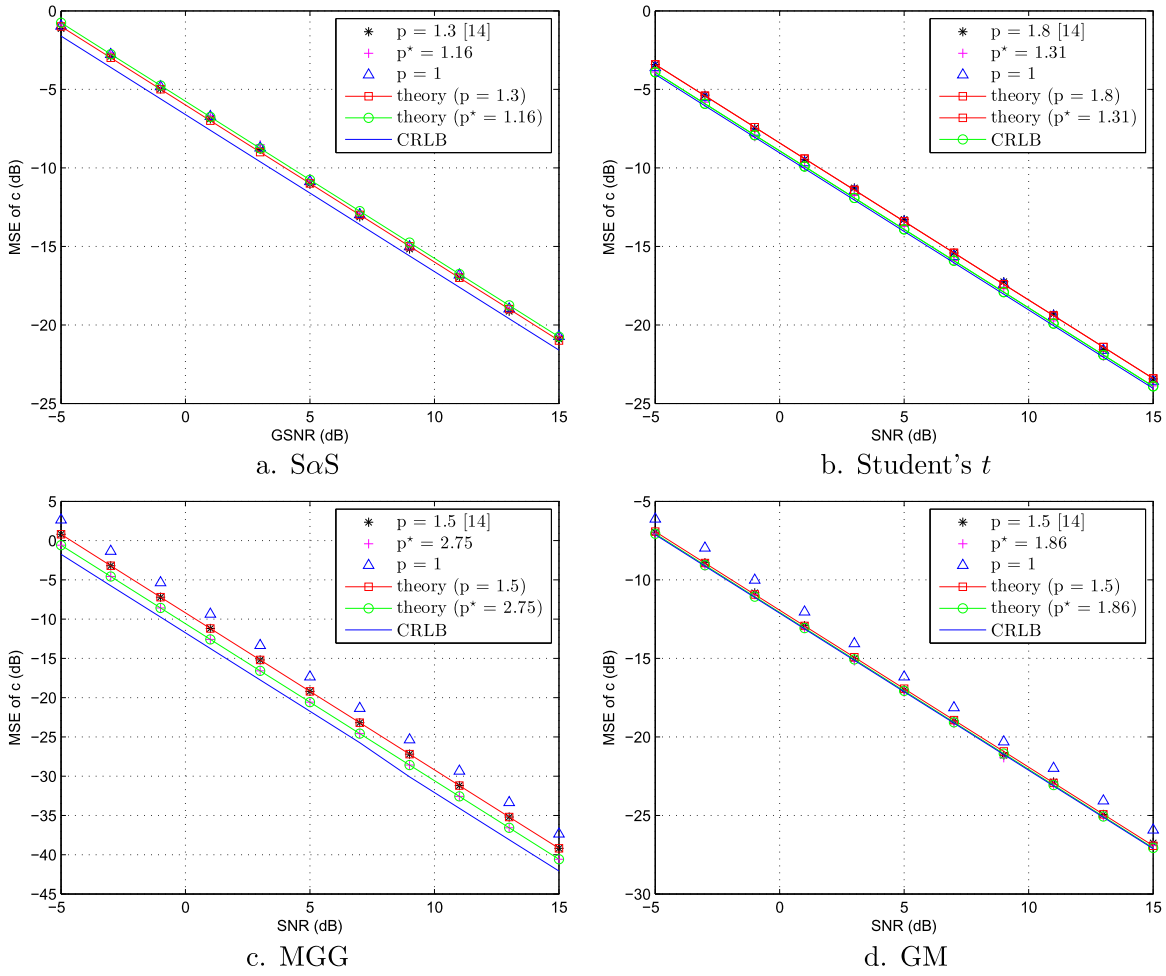

Fig. 4. MSE of scalar in complex-valued observations.

$$\mathbf{G}=\begin{cases}\frac{1}{4}\left[\left(\frac{\partial\Re\{\mathbf{g}(\mathbf{x})\}}{\partial\Re\{\mathbf{x}\}}\right)^T\frac{\partial\Re\{\mathbf{g}(\mathbf{x})\}}{\partial\Re\{\mathbf{x}\}}+\left(\frac{\partial\Re\{\mathbf{g}(\mathbf{x})\}}{\partial\Im\{\mathbf{x}\}}\right)^T\frac{\partial\Re\{\mathbf{g}(\mathbf{x})\}}{\partial\Im\{\mathbf{x}\}}\right. & \text{complex}\\ \left.+\left(\frac{\partial\Im\{\mathbf{g}(\mathbf{x})\}}{\partial\Re\{\mathbf{x}\}}\right)^T\frac{\partial\Im\{\mathbf{g}(\mathbf{x})\}}{\partial\Re\{\mathbf{x}\}}+\left(\frac{\partial\Im\{\mathbf{g}(\mathbf{x})\}}{\partial\Im\{\mathbf{x}\}}\right)^T\frac{\partial\Im\{\mathbf{g}(\mathbf{x})\}}{\partial\Im\{\mathbf{x}\}}\right],\\ \nabla^T(\Re\{\mathbf{g}(\mathbf{x})\})\nabla(\Re\{\mathbf{g}(\mathbf{x})\})+\nabla^T(\Im\{\mathbf{g}(\mathbf{x})\})\nabla(\Im\{\mathbf{g}(\mathbf{x})\}), & \text{real}\end{cases}\tag{21}$$

with $\Re\{\mathbf{g}(\mathbf{x})\}=[\Re\{g_1(\mathbf{x})\}\cdots\Re\{g_N(\mathbf{x})\}]^T$ and $\Im\{\mathbf{g}(\mathbf{x})\}=[\Im\{g_1(\mathbf{x})\}\cdots\Im\{g_N(\mathbf{x})\}]^T$. Notice that when ν and β are larger than 2, (19) and (20) are invalid for odd p.

In the following, two illustrations including the linear and nonlinear models are provided to elaborate the covariance expression.

Scalar estimation

The simplest complex constant estimation is investigated, whose model is

$$\mathbf{y}=\mathbf{1}_N c+\mathbf{q},\tag{22}$$

where $\mathbf{1}_N$ denotes the $N\times 1$ vector with all elements being 1 and $c=\Re\{c\}+j\Im\{c\}$ is the complex constant to be estimated.

Since the unknown parameter $\mathbf{x}=c$ is complex-valued, we utilize the complex scenario in (21). Taking $\Re\{g(\mathbf{x})\}=\mathbf{1}_N\Re\{c\}$ and $\Im\{g(\mathbf{x})\}=\mathbf{1}_N\Im\{c\}$ on (21) yields,

$$\mathbf{G}=\frac{N}{2}.\tag{23}$$

Then the variance of $\hat{c}$ can be derived by (18)–(20) for different noise models.

Linear estimation

Here we consider the linear estimation problem with the observed signal:

$$y_n=A\exp(j(\omega n+\theta))+q_n,\quad n=1,\ldots,N,\tag{24}$$

where $A>0$, $\omega\in(-\pi,\pi)$ and $\theta\in[0,2\pi]$ denote the amplitude, frequency and phase, respectively. Assuming that ω is known, our task is to estimate A and θ.

Let $x=A\exp(j\theta)$. We easily get:

$$\mathbf{G}=N.\tag{25}$$

According to [24], the variances of $\hat{A}$ and $\hat{\theta}$ are:

$$\mathrm{var}(\hat{A})\approx C(\hat{x}),\tag{26}$$

$$\mathrm{var}(\hat{\theta})\approx A^2C(\hat{x}),\tag{27}$$

where $C(\hat{x})$ is calculated using (18)–(20) according to the respective noise model.

Nonlinear estimation

We now consider the nonlinear estimation problem. The data model is identical to (24). However, here the unknown parameters are ω and θ while A is known. In this case, $\mathbf{x}=[\omega\ \theta]^T$. Then we obtain [25]

$$\mathbf{G}=A^2\begin{bmatrix}\frac{N(N+1)(2N+1)}{6} & \frac{N(N+1)}{2}\\ \frac{N(N+1)}{2} & \frac{N(N+1)(2N+1)}{6}\end{bmatrix}.\tag{28}$$

The corresponding covariance expressions for SαS, Student's t and MGG noise models are computed by (18)–(20), respectively.

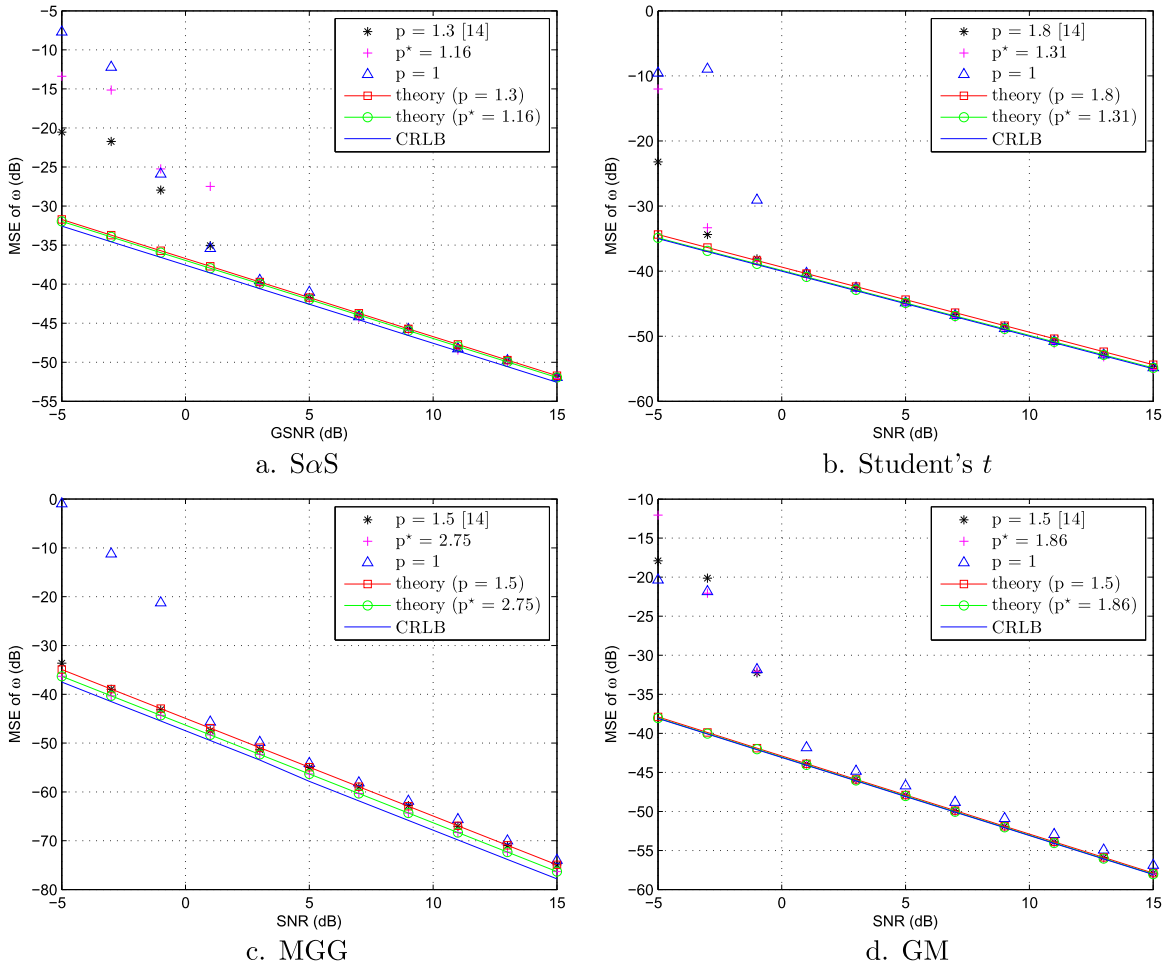

Fig. 5. MSE of sinusoidal frequency in complex-valued observations.

3.2. Minimum variance analysis

We investigate the minimum achievable variance for each studied noise model using the ℓ_p-norm minimizer, via selection of optimum p. This is achieved by minimizing the scalar terms in (18)–(20) because the matrix components are independent of p.

3.2.1. SαS

We first investigate the ℓ_p-norm minimizer under SαS noise. For (18), the scalar term related to p, denoted by $H_\alpha(p)$, is:

$$H_\alpha(p) = \frac{\Gamma^2\left(2-\frac{p}{2}\right)\Gamma(p)\Gamma\left(1-\frac{2p-2}{\alpha}\right)}{\Gamma^2\left(\frac{p}{2}+1\right)\Gamma(2-p)\Gamma^2\left(1-\frac{p-2}{\alpha}\right)}. \tag{29}$$

To find the optimal value of p, denoted by $p^\star$, we set the derivative of $\log(H_\alpha(p))$, denoted by $h_\alpha(p)$, to zero. Then $h_\alpha(p)$ has the form of

$$h_\alpha(p) = \psi(2-p) - \psi\left(2-\frac{p}{2}\right) + \psi(p) - \psi\left(\frac{p}{2}\right) + \frac{2}{\alpha}\left(\psi\left(1-\frac{p-2}{\alpha}\right) - \psi\left(1-\frac{2p-2}{\alpha}\right)\right) - \frac{2}{p}, \tag{30}$$

where $\psi(\cdot)$ denotes the digamma function [26].

Let $u = p-1 \in [0, 1)$. Employing the property that $\psi(x+1) = \psi(x) + \frac{1}{x}$, (30) is reduced to

$$h_\alpha(u) = \psi(1-u) - \psi\left(\frac{1-u}{2}\right) + \psi(1+u) - \psi\left(\frac{u+1}{2}\right) + \frac{2}{\alpha}\left(\psi\left(\frac{1-u}{\alpha}\right) - \psi\left(1-\frac{2u}{\alpha}\right)\right) - \frac{2}{u+1}. \tag{31}$$

Utilizing $2\psi(2x) = 2\log(2) + \psi(x) + \psi(x+\frac{1}{2})$, we further simplify (31) as

$$h_\alpha(u) = h_1(u) + h_2(u) + h_3(u) + 4\log(2), \tag{32}$$

where

$$h_1(u) = -\psi(1-u) - \psi(u+1) + \psi\left(1-\frac{u}{2}\right) + \psi\left(1+\frac{u}{2}\right), \tag{33}$$

$$h_2(u) = \frac{2}{\alpha}\left(\psi\left(\frac{1-u}{\alpha}\right) - \psi\left(1-\frac{2u}{\alpha}\right)\right), \tag{34}$$

$$h_3(u) = -\frac{2}{u+1}. \tag{35}$$

Taking the rational zeta series [25] on $h_1(u)$ yields

$$h_1(u) = \sum_{k=1}^{\infty}\zeta(k+1)\left\{(-1)^k + 1 - \left(-\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^k\right\}u^k, \tag{36}$$

where $\zeta(\cdot)$ is the Riemann zeta function. Furthermore, based on the fact that $\psi(x) = -a + \sum_{k=0}^{\infty}\left(\frac{1}{k+1} - \frac{1}{k+x}\right)$ where a is the Euler-Mascheroni constant, as well as TSE at u=0, $h_2(u)$ is rewritten as

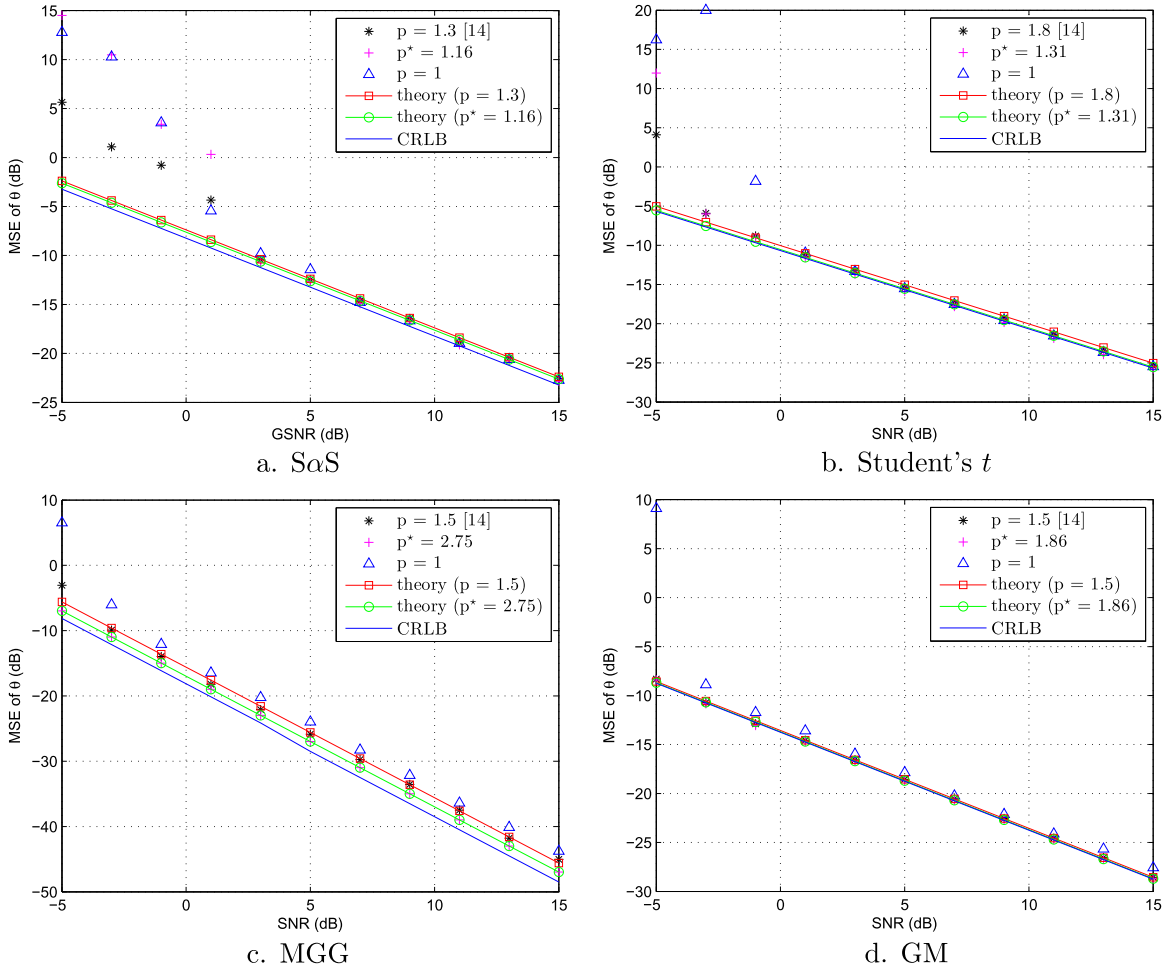

Fig. 6. MSE of sinusoidal phase in complex-valued observations.

$$h_2(u) = \frac{2}{\alpha}\left(\psi\left(\frac{1}{\alpha}\right) - \psi(1)\right) + \sum_{k=1}^{\infty} \frac{2u^k}{\alpha^{k+1}}\left\{2^k\zeta(k+1,\,1) - \zeta\left(k+1,\,\frac{1}{\alpha}\right)\right\}, \tag{37}$$

where $\zeta(\cdot,\cdot)$ denotes the Hurwitz zeta function. Utilizing the TSE at u=0, $h_3(u)$ is

$$h_3(u) = -2 - 2\sum_{k=1}^{\infty}(-u)^k. \tag{38}$$

Combining (32)–(38), $h_\alpha(u)$ has the form of

$$h_\alpha(u) = b(0) + \sum_{k=1}^{\infty} b(k)u^k, \tag{39}$$

where $b(0) = 4\log(2) - 2 + \frac{2}{\alpha}(\psi\left(\frac{1}{\alpha}\right) - \psi(1))$ and

$$b(k) = \left((-1)^k + 1 - \left(\frac{1}{2}\right)^k - \left(-\frac{1}{2}\right)^k\right)\zeta(k+1) - 2(-1)^k + \left(\frac{1}{\alpha}\right)^k\left(2^k\zeta(k+1,\,1) - \zeta\left(k+1,\,\frac{1}{\alpha}\right)\right). \tag{40}$$

According to extensive simulations, we find that the third-order polynomial in (39) can describe $h_\alpha(u)$ well. Therefore, we can obtain $u^\star$ by finding the real-valued root of the cubic function (k=3) [27], which is

$$u^\star = \sqrt[3]{-\frac{y}{2} + \sqrt{\left(\frac{y}{2}\right)^2 + \left(\frac{z}{3}\right)^3}} - \sqrt[3]{\frac{y}{2} + \sqrt{\left(\frac{y}{2}\right)^2 + \left(\frac{z}{3}\right)^3}}, \tag{41}$$

with $y = -2\left(\frac{b(2)}{3b(3)}\right)^3 - \frac{b(1)b(2)}{3(b(3))^2} + \frac{b(0)}{b(3)}$ and $z = -\frac{1}{3}\left(\frac{b(2)}{b(3)}\right)^2 + \frac{b(1)}{b(3)}$.

Utilizing the definition of u, $p^\star$ is computed by $u^\star + 1$. Note that for $\alpha \in [1, 2)$, the term $(\frac{y}{2})^2 + (\frac{z}{3})^3$ is larger than 0 and $u^\star$ is the only real root of (39). Since $h_\alpha(p)$ is monotonic, $p^\star$ is in the global optimum.

3.2.2. *Student's* t

From (19), the scalar term corresponding to Student's t noise, denoted by $H_t(p)$, is

$$H_t(p) = \frac{\Gamma(p)\Gamma\left(\frac{\nu - 2p + 2}{2}\right)}{\Gamma^2\left(1 + \frac{p}{2}\right)\Gamma^2\left(\frac{\nu - p + 2}{2}\right)}. \tag{42}$$

In a similar manner, the derivative of $\log(H_t(p))$ can be expressed as

$$h_t(p) = \psi(p) - \psi\left(1 + \frac{p}{2}\right) - \psi\left(\frac{\nu - 2p}{2} + 1\right) + \psi\left(\frac{\nu - p}{2} + 1\right). \tag{43}$$

Here we consider two cases: $\nu < 2$ and $\nu \geq 2$. In the scenario of $\nu < 2$, we have

$$h_t{}'(p) = \psi^{(1)}(p) - \frac{1}{2}\psi^{(1)}\left(1 + \frac{p}{2}\right) + \psi^{(1)}\left(\frac{\nu - 2p}{2} + 1\right) - \frac{1}{2}\psi^{(1)}\left(\frac{\nu - p}{2} + 1\right), \tag{44}$$

where $\psi^{(1)}(\cdot)$ is the trigamma function. Based on the multiplication theorem of trigamma function, (44) is rewritten as

$$h_t{}'(p) = \frac{1}{4}\left(\psi^{(1)}\left(\frac{p}{2}\right) + \psi^{(1)}\left(\frac{1}{2} + \frac{p}{2}\right)\right) - \frac{1}{2}\psi^{(1)}\left(1 + \frac{p}{2}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{\nu}{2} + 1 - \frac{p}{2}\right) + \frac{1}{4}\left(\psi^{(1)}\left(\frac{\nu}{4} + \frac{1}{2} - \frac{p}{2}\right) + \psi^{(1)}\left(\frac{\nu}{4} + 1 - \frac{p}{2}\right)\right), \tag{45}$$

which is greater than 0. Therefore, we can deduce that $h_t(p) > h_t(1) > 0$. In this case, the optimum p is 1.

In the case of $\nu \geq 2$, assigning $u = p - 1 \in [0, 1)$, (43) is

$$h_t(u) = \psi(u+1) - \psi\left(1 + \frac{u+1}{2}\right) - \psi\left(\frac{\nu - 2u}{2}\right) + \psi\left(\frac{\nu - u - 1}{2} + 1\right). \tag{46}$$

Similarly to the SαS distribution, (46) can be written as the same expression as in (39) except that

$$b(0) = 2\log(2) - 2 - \frac{2}{\nu}, \tag{47}$$

$$b(k) = \zeta(k+1)\left((-1)^k - \left(-\frac{1}{2}\right)^k\right) - 2(-1)^k + \left\{\zeta\left(k, \frac{\nu}{2}\right)\left(\left(\frac{1}{2}\right)^k - 1\right) - \frac{2}{\nu^{k+1}}\right\}. \tag{48}$$

It can also be easily shown that the series in Student's t process converges when $k \geq 3$. The optimum value is $p^\star = u^\star + 1$ with $u^\star$ being calculated by (41).

3.2.3. MGG

For MGG noise, the corresponding scalar term, which is extracted from (20), has the form of

$$H_{\mathrm{MGG}}(p) = \frac{\Gamma\left(\frac{2p}{\beta}\right)((1-\epsilon) + \epsilon\tau^{2p-2})}{\Gamma^2\left(1 + \frac{p}{\beta}\right)((1-\epsilon) + \epsilon\tau^{p-2})^2}. \tag{49}$$

Let $u = \frac{p}{\beta} - \frac{1}{2} \in (-1/2, 1/2)$, the derivative of $\log(H_{\mathrm{MGG}}(p))$ is factorized as:

$$h_{\mathrm{MGG}}(u) = \frac{1}{\beta}\left\{\psi(2u+1) - \psi\left(\frac{3}{2} + u\right)\right\} + \ln(\tau)\left\{\frac{1}{1 + \frac{\epsilon}{1-\epsilon}\tau^{\beta u + \beta/2 - 2}} - \frac{1}{1 + \frac{\epsilon}{1-\epsilon}\tau^{2\beta u + \beta - 2}}\right\}. \tag{50}$$

To solve $h_{\mathrm{MGG}}(u) = 0$, we express (50) as (39), where $b(0)$ and $b(k)$ are now

$$b(0) = \frac{2\log(2) - 2}{\beta}, \tag{51}$$

$$b(k) = \frac{1}{\beta}\{\zeta(k+1)((-2)^k - (-1)^k) + (-2)^{k+1}\} - \ln(\tau)\left\{I^{(k)}\left(\frac{\epsilon\tau^{\beta/2-2}}{(1-\epsilon)}, \tau^\beta\right) - I^{(k)}\left(\frac{\epsilon\tau^{\beta-2}}{(1-\epsilon)}, \tau^{2\beta}\right)\right\}, \tag{52}$$

where $I^{(k)}(c, x)$ denotes kth-order derivative of $\frac{1}{1+cx^u}$ with respect to u when u=0.

According to extensive simulations, the series here approximates $h_{\mathrm{MGG}}(u)$ well when $k \geq 3$. Therefore, we derive the optimum $p^\star$ as $p^\star = \beta(u^\star + 1/2)$ with $u^\star$ having the same form in (41).

4. Simulation results

To verify the derived variance formulas in the case of three symmetric impulsive noise distributions, computer simulations have been conducted. The signal is generated according to (24) and the corresponding parameters are A=1, $\omega = 1.25$ and $\theta = 0.5$. For SαS, Student's t and MGG models, the density parameters are chosen as $\alpha = 1.6$ [28], $\nu = 5$ [29], $\beta = 5$, $\epsilon = 0.01$ and $\tau = \sqrt{5}$ [11]. The special case of MGG distribution, referred to as GM process, is also shown here with $\beta = 2$, $\epsilon = 0.01$ and $\tau = \sqrt{5}$. The GG model is not considered here because the optimal ℓ_p-norm minimizer is the ML estimator [14]. According to (41), the optimal value $p^\star$ for these four settings, are 1.16, 1.31, 2.75 and 1.86, respectively. To provide a comparison, the empirical optimum results in [14], referring to as $p = (1+\alpha)/2$, $p = (\nu+4)/5$, and $p = 1 + \frac{2}{100\epsilon + 3}$, are investigated here. Furthermore, comparisons with the least ℓ_1-norm estimator and the CRLB for complex-valued estimation [30] are also included. Since the second-order power diverges for the SαS model, we utilize the geometric signal-to-noise ratio (GSNR) to produce different noise conditions [31]. All results are based on 5000 Monte Carlo simulation trials with a data length of N=50.

First of all, to investigate the sufficiency of using the cubic polynomial, we study the value of optimum p obtained by (39) and (41). According to [14], we choose degree 100 replacing the infinity order in (39). Fig. 1 shows the comparison of $p^\star$ calculation between the cubic function and polynomial (39) with degree 100 versus α, ν, β, respectively. Other density parameters align with the previous set. It is experimentally observed in SαS, Student's t and MGG models that results by cubic function attain the true values well, which corroborates our claim in Section 3.

Secondly, we address linear estimation, namely, studying the amplitude and phase estimation performance using the complex-valued observations in (24). Figs. 2 and 3 show MSE performance of amplitude and phase versus SNR/GSNR, respectively. In these figures, we see the near optimality of the least ℓ_p-norm estimator in the Student's t and GM noise models and its suboptimal performance in the SαS and MGG disturbances. Most importantly, in Figs. 2 and 3, the gap between the optimal value and result in [14] indicates that our method is more accurate. Note that the gap between our proposal and [14] becomes bigger as β increases, verifying the importance of our method. Then in Fig. 4, the MSE of the complex-valued scalar estimation is investigated. The observations are generated according to (22) where $c = 1 + j\sqrt{2}$. The findings are similar to those of Figs. 2 and 3.

Thirdly, nonlinear estimation of exponential signal frequency and phase is studied. The MSE results are plotted in Figs. 5 and 6. We again see that the variance formulas are validated and the other findings are similar to those of Figs. 2 and 3. Furthermore, when SNR/GSNR $\geq$5 dB, our proposals perform better than the ℓ_1-norm minimizer and results in [14].

In summary, in the presence of SαS, Student's t and MGG noises, our proposed method is superior to that in [14]. It has also been discussed in [14] that in the real applications that the density parameters of noise are unknown, the least ℓ_1-norm estimator is a good choice. However, according to the simulations on the MGG distribution, the gap between the p=1 and $p^\star$ is not less than 3 dB, indicating the inferiority of the ℓ_1-norm minimizer. In this case, the shape parameter should be estimated firstly.

5. Conclusion

In this work, we focus on the variance expression deviation of the least ℓ_p-norm technique with $p \geq 1$, where three representative bivariate symmetric disturbances have been studied. The optimal choice of p, corresponding to the minimum variance, is discussed, which can be obtained by solving a cubic equation. It is worth pointing out that for

the bivariate Student's t process with $\nu < 2$, the optimum p should be chosen as 1. Simulation results validate the accuracy of the derived variance formulas using linear and nonlinear estimation examples with complex-valued observations. It is also demonstrated that variances of the least ℓ_p-norm estimator for SαS and Student's t noise models are very close to the CRLB. This result indicates that the estimator can provide optimum or nearly-optimum performance for these three noise models if an appropriate value of p is chosen. Note that all results in this paper can be applied to the real-valued scenario.

Acknowledgements

The work described in this paper was supported by grants from the Fundamental Research Funds for the Central Universities (Grant No. 2302015FRF-TP-15-033A1) and China Postdoctoral Science Foundation (Grant No. 2016M590047).

References

[1] P.J. Schreier, L.L. Scharf, Statistical Signal Processing of Complex-Valued Data: the Theory of Improper and Noncircular Signals, Cambridge University Press, Cambridge, U.K., 2010.
[2] V. Madisetti, Wireless, Networking, Radar, Sensor Array Processing, and Nonlinear Signal Processing, Chapman & Hall/CRC Press, 2009.
[3] E. Ollila, V. Koivunen, H.V. Poor, Complex-valued signal processing-essential models, tools and statistics, in: Proceedings of 2011 Information Theory and Applications Workshop (ITA), La Jolla, CA, Feb. 2011, pp. 1–10.
[4] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, NJ: Englewood Cliffs, 1993.
[5] P. Stoica, R. Moses, Introduction of Spectral Analysis, Prentice-Hall, Upper Saddle River, NJ, 1997.
[6] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, 1991.
[7] V.H. MacDonald, P.M. Schultheiss, Optimum passive bearing estimation in a spatially incoherent noise environment, J. Acoust. Soc. Am. 46 (1) (1969) 37–43 (Pt. 1).
[8] H.C. So, Y.T. Chan, K.C. Ho, Y. Chen, Simple formulas for bias and mean square error computation, IEEE Signal Process. Mag. 30 (4) (2013) 162–165.
[9] C.L. Nikias, M. Shao, Signal Processing with Alpha-Stable Distribution and Applications, John Wiley & Sons Inc., New York, 1995.
[10] K. Samuel, N. Saralees, Multivariate t-Distributions and their Applications, Cambridge University Press, 2004.
[11] M.O.M. Mahmoud, M. Jaidane-Saidane, N. Hizaoui, The use of mixture of generalized Gaussian for trend analysis of the load duration curve: Summer and winter load variability in Tunisia, in: Proceedings of the 43rd Universities Power Engineering Conference, Padova, Italy, Sept. 2008, pp. 1–5.
[12] P.J. Huber, E.M. Ronchetti, Robust Statistics, Wiley, Hoboken, N.J., 2009.
[13] T.H. Li, A nonlinear method for robust spectral analysis, IEEE Trans. Signal Process. 58 (5) (2010) 2466–2474.
[14] Y. Chen, H.C. So, E.E. Kuruoglu, Variance analysis of unbiased least ℓ_p-norm estimator in non-Gaussian noise, Signal Process. 122 (2016) 190–203.
[15] A.H. Money, J.F. Affleck-Graves, M.L. Hart, G.D.I. Barr, The linear regression model: ℓ_p-norm estimation and the choice of p, Commun. Stat.-Simul. Comput. 11 (1) (1982) 89–109.
[16] E.E. Kuruoglu, P.J.W. Rayner, W.J. Fitzgerald, Least ℓ_p-norm impulsive noise cancellation with polynomial filters, Signal Process. 69 (1) (1998) 1–14.
[17] E.E. Kuruoglu, Density parameter estimation of skewed α-stable distributions, IEEE Trans. Signal Process. 49 (10) (2001) 2192–2220.
[18] S. Kogon, D.B. Williams, On the characterization of impulsive noise with alpha-stable distributions using Fourier techniques, in: Proceedings of the 29th Asilomar Conference on Signals, System and Computer, Washington, USA, Oct. 1995, pp. 787–791.
[19] O.M.M. Mohamed, M. Jaidane-Saidane, On the parameter estimation of the generalized Gaussian mixture model, in: Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, Scotland, Aug. 2009, pp. 2273–2277.
[20] E. Ollila, D.E. Tyler, V. Koivunen, H.V. Poor, Complex elliptically symmetric distributions: Survey, new results and applications, IEEE Trans. Signal Process. 60 (11) (2012) 5597–5625.
[21] X.Y. Ma, C.L. Nikias, Parameter estimation and blind channel identification in impulsive signal environments, IEEE Trans. Signal Process. 43 (12) (1995) 2884–2897.
[22] M.S. Allili, Wavelet modeling using finite mixtures of generalized Gaussian distributions: Application to texture discrimination and retrieval, IEEE Trans. Image Process. 21 (4) (2012) 1452–1464.
[23] A. van den Bos, Complex gradient and Hessian, IEE Proc. - Vis., Image Signal Process. 141 (6) (1994) 380–383.
[24] E. Anarim, Y. Istefanopulos, Statistical analysis of Pisarenko type tone frequency estimator, Signal Process. 24 (1) (1991) 291–298.
[25] I.D. Gradshteyn, I.M. Ryzhik, Tables of Integrals, Series and Products, Academic, New York, 2007.
[26] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, New York: Dover Publications, 1964.
[27] J. Nathan, Basic Algebra I, Dover Publications, New York, 2009.
[28] P.G. Georgiou, P. Tsakalides, C. Kyriakakis, Alpha-stable modeling of noise and robust time delay estimation in the presence of impulsive noise, IEEE Trans. Multimed. 1 (3) (1999) 291–301.
[29] G.J. Lauprete, A.M. Samarov, R.E. Welsch, Robust portfolio optimization, Metrika 55 (1) (2002) 139–149.
[30] R.J. Kozick, B.M. Sadler, Maximum-Likelihood array processing in non-Gaussian noise with Gaussian mixtures, IEEE Trans. Signal Process. 48 (12) (2000) 3520–3535.
[31] J.G. Gonzalez, J.L. Paredes, G.R. Arce, Zero-order statistics: a mathematical framework for the processing and characterization of very impulsive signals, IEEE Trans. Signal Process. 54 (10) (2006) 3839–3851.