Variance Analysis of Unbiased Least ℓ_p-Norm Estimator in Non-Gaussian Noise

Yuan Chen [1] , Hing Cheung So [2] , *EURASIP Member*, *Ercan Engin Kuruoglu* [3] , *EURASIP Member*

Abstract

Modeling time and space series in various areas of science and engineering require the values of parameters of interest to be estimated from the observed data. It is desirable to analyze the performance of estimators in an elegant manner without the need for extensive simulations and/or experiments. Among various performance measures, variance is the most basic one for unbiased estimators. In this paper, we focus on the estimator based on the ℓ_p-norm minimization in the presence of zero-mean symmetric non-Gaussian noise. Four representative noise models, namely, α-stable, generalized Gaussian, Student's t and Gaussian mixture processes, are investigated, and the corresponding variance expressions are derived for linear and nonlinear parameter estimation problems at $p \geq 1$. The optimal choice of p for different noise environments is studied, where the global optimality and sensitivity analyses are also provided. The developed formulas are verified by computer simulations and are compared with the Cramér-Rao lower bound.

Key words: Variance analysis, parameter estimation, non-Gaussian noise, fractional lower-order moment, ℓ_p-norm minimization

1. Introduction

Parameter estimation [1] is a common task required in many areas of science and engineering such as radar, sonar, speech, image analysis, biomedicine, communications and seismology. It refers to accurately finding the values of parameters of interest from the observed data which consist of two components, viz., signal and noise. Typically, a deterministic model is adopted for the signal while a random process model is employed for the noise. Among numerous estimators developed in the literature, least squares (LS) and maximum likelihood (ML) methods have been widely used.

[1] School of Computer & Communication Engineering, University of Science & Technology Beijing, Beijing, China (Email: qchenyuan00@126.com)

[2] Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China (Email: hcso@ee.cityu.edu.hk)

[3] Italian National Council of Research (ISTI-CNR), Pisa, Italy (Email: ercan.kuruoglu@isti.cnr.it)

To assess the quality of an estimator, two fundamental performance measures in the aspect of accuracy are bias and mean square error (MSE). To calculate bias and MSE, approaches such as Taylor series expansion (TSE) of the estimates [2] or TSE on the error function [3] can be used. Although we have only demonstrated the usefulness of the formulas in the presence of Gaussian noise in [4], it is found that the latter has more applicability and may be simpler to derive particularly for nonlinear parameter estimation problems. Despite its theoretical and computational convenience, it is generally understood that the validity of the Gaussian distribution is at best approximate in reality. In fact, the occurrence of non-Gaussian noise has been reported in many fields [5]–[6]. Note that some non-Gaussian models correspond to impulsive noise. In this case, the LS approach which is based on ℓ_2-norm minimization of errors, fails to provide reliable parameter estimation, since its performance is very sensitive to outliers. The ML estimator may also not be a proper choice. In particular, it is hard to implement for the process whose probability density function (PDF) has a complicated analytical form or lacks an analytical expression. One strategy is to detect and discard the suspicious observations but it may not be feasible for large data sets or complex applications [5].

Alternatively, M-estimator [7], which is based on robust statistics, can resist outliers without preprocessing the data. Its key idea is to replace the squared residuals in the LS methodology by another function which emphasizes large samples less than the square. The least ℓ_p-norm estimator with $p < 2$ belongs to the M-estimator family, which is commonly solved by iterative techniques such as the iteratively reweighted least squares (IRLS), Levenberg-Marquardt (LM) and subgradient methods [8]–[10].

As a follow-up to [4], we study the performance of the least ℓ_p-norm estimator for parameter estimation in additive non-Gaussian noise in this work. Four representative models, namely, symmetric α-stable (SαS) [11], generalized Gaussian (GG) [12], Student's t [13] and Gaussian mixture (GM) [14] processes, are investigated. For simplicity but without loss of generality, all noise models are assumed zero-mean and symmetrically distributed, implying that the least ℓ_p-norm estimator is unbiased and we only analyze the variance formulas. Note that this study assumes the availability of the noise statistic information, i.e., PDF and density parameters. Therefore, in the case of unknown noise statistics [15]–[16], they should be estimated [17]–[18] prior to applying our results. The logarithm moment method [17] can be employed for density parameter estimation from the available noise-only samples. Even when we know nothing about the noise statistics, the GM model can be utilized to approximate the impulsive noise [18]. The parameters of the GM process, that is, numbers of components and their variances, can be adjusted adaptively based on the noise-only observations.

The rest of this paper is organized as follows. In Section 2, we briefly present the bias and MSE formulas based on TSE on the estimator cost function, and then the least ℓ_p-norm estimator. The SαS, GG, Student's t and GM models, are then reviewed in Section 3. Linear and nonlinear signal models are studied with illustrative examples in Sections 4 and 5, respectively. Selection of p for different non-Gaussian noise models which result in minimum variance as well as global optimality and sensitivity analysis are examined in Section 6. In Section 7, computer simulation results are provided to validate the derived variance formulas and contrast with the Cramér-Rao lower bound (CRLB). Finally, conclusions are drawn in Section 8.

2. Variance Formula and Least ℓ_p-Norm Estimator

We start with a general signal model:

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{q}, \tag{1}$$

where $\mathbf{y} = [y_1 \ \cdots \ y_N]^T \in \mathbb{R}^N$ is the observation vector with ^{T} being the transpose operator, $\mathbf{g}(\cdot)$ is a known function, $\mathbf{x} = [x_1 \ \cdots \ x_M]^T \in \mathbb{R}^M$ is the deterministic parameter vector of interest with $M \leq N$ and $\mathbf{q} = [q_1 \ \cdots \ q_N]^T \in \mathbb{R}^N$ denotes the additive random noise component with zero location parameter. The task of parameter estimation is to find $\mathbf{x}$ from $\mathbf{y}$.

A common approach for estimating $\mathbf{x}$ is to design a cost function $J(\mathbf{x})$ which is constructed from $\mathbf{y}$, and the estimate of $\mathbf{x}$, denoted by $\hat{\mathbf{x}}$, is computed by minimizing $J(\mathbf{x})$:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} J(\mathbf{x}). \tag{2}$$

Equivalently,

$$\nabla\left(J(\hat{\mathbf{x}})\right) = \mathbf{0}_N, \tag{3}$$

where $\nabla\left(J(\hat{\mathbf{x}})\right)$ denotes the gradient vector of $J(\mathbf{x})$ at $\mathbf{x} = \hat{\mathbf{x}}$ and $\mathbf{0}_N \in \mathbb{R}^N$ is a column vector with all zeros. In this study, we consider a general class of estimators such that $J(\mathbf{x})$ is twice differentiable. When $\hat{\mathbf{x}}$ is located at a reasonable proximity to $\mathbf{x}$, which is valid when the noise is small enough and/or observation number is sufficiently large, the truncated TSE of $\nabla\left(J(\hat{\mathbf{x}})\right)$ around $\mathbf{x}$ is

$$\nabla\left(J(\hat{\mathbf{x}})\right) \approx \nabla\left(J(\mathbf{x})\right) + \mathbf{H}\left(J(\mathbf{x})\right)\left(\hat{\mathbf{x}} - \mathbf{x}\right), \tag{4}$$

where $\mathbf{H}(J(\mathbf{x}))$ is the Hessian matrix. Assuming that $\mathbf{H}(J(\mathbf{x}))$ is smooth enough to have $\mathbf{H}(J(\mathbf{x})) \approx E\{\mathbf{H}(J(\mathbf{x}))\}$, (4) can be utilized to compute the bias and MSE of $\hat{\mathbf{x}}$:

$$\begin{aligned} \text{bias}(\hat{\mathbf{x}}) =& E\{\hat{\mathbf{x}}\} - \mathbf{x} \\ \approx& -\left(E\left\{\mathbf{H}(J(\mathbf{x}))\right\}\right)^{-1} E\left\{\nabla(J(\mathbf{x}))\right\}, \end{aligned} \tag{5}$$

$$\begin{aligned} \mathbf{M}(\hat{\mathbf{x}}) =& E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T\} \\ \approx& \left(E\left\{\mathbf{H}(J(\mathbf{x}))\right\}\right)^{-1} E\left\{\nabla(J(\mathbf{x}))\nabla^T(J(\mathbf{x}))\right\}\left(E\left\{\mathbf{H}(J(\mathbf{x}))\right\}\right)^{-1}, \end{aligned} \tag{6}$$

where E denotes the expectation operator and $(\cdot)^{-1}$ is the matrix inverse. The MSE of $\hat{x}_m$, $m = 1, \cdots, M$, is given by the (m, m) entry of $\mathbf{M}(\hat{\mathbf{x}})$. Note that the expressions in (5)–(6) are exact when $J(\mathbf{x})$ is a quadratic function. Furthermore, the validity of (5) and (6) may not relate to the parameter dimension, namely, M, but depends on whether $\mathbf{x}$ is linear in the observation vector or not. It is because nonlinear estimators nearly always exhibit the threshold effect [1].

For an unbiased estimator where $\text{bias}(\hat{\mathbf{x}}) = \mathbf{0}_M$, the MSE is in fact the variance. Then the covariance matrix for $\hat{\mathbf{x}}$, denoted by $\mathbf{C}(\hat{\mathbf{x}})$, is approximated by (6) while the variance of $\hat{x}_m$, denoted by $\text{var}(\hat{x}_m)$, is provided by the (m, m) entry of $\mathbf{C}(\hat{\mathbf{x}})$.

Denote $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}) \ \cdots \ g_N(\mathbf{x})]^T$. A typical choice for $J(\mathbf{x})$ is the LS cost function:

$$J(\mathbf{x}) = \sum_{n=1}^{N} |y_n - g_n(\mathbf{x})|^2, \tag{7}$$

which corresponds to ℓ_2-norm minimization. It is well known that the LS solution is equivalent to the ML estimate when $\mathbf{q}$ is a zero-mean white Gaussian process. In fact, (5)–(6) have been verified in [4] for LS-based parameter estimation in the presence of white Gaussian noise. Nevertheless, when the noise is non-Gaussian distributed, particularly if $\mathbf{q}$ is impulsive, unreliable parameter estimation will result since the performance of the ℓ_2-norm minimizer is very sensitive to outliers. To achieve robust estimation, ℓ_p-norm minimization with $p < 2$ is widely used since it is less sensitive to outliers than the square function. In this work, we focus on the ℓ_p-norm cost function:

$$J(\mathbf{x}) = \sum_{n=1}^{N} |y_n - g_n(\mathbf{x})|^p, \quad p \in [1, 2), \tag{8}$$

when $\{q_n\}_{n=1}^N$ are independent and identically distributed (IID) zero-mean impulsive non-Gaussian variables with a closed-form PDF or characteristic function (CF) expression. Note that CF and PDF form a Fourier transform pair. In this study, the case of $0 \le p < 1$ is not examined, because the corresponding ℓ_p-norm cost function may not be differentiable.

According to (8), we have

$$E\{\nabla\left(J(\mathbf{x})\right)\} = p \sum_{n=1}^{N} \frac{\partial g_n(\mathbf{x})}{\partial \mathbf{x}} E\left\{|q_n|^{p-1}\text{sign}(q_n)\right\}, \tag{9}$$

where $\text{sign}(q_n) = 1$ if $q_n > 0$ and -1 otherwise. It will be shown in Section 3 that for our studied noise models, $E\left\{|q_n|^{p-1}\text{sign}(q_n)\right\} = 0$, indicating the unbiasedness of the ℓ_p-norm estimator with the use of (5). Note that the least absolute deviation function of $p = 1$ is non-differentiable, but we will apply the limit $p \to 1$ to obtain the variance expressions. Based on (6) and (8), we also need to compute two fractional lower-order moment (FLOM) terms, that is, $E\{|q_n|^{p-2}\}$ and $E\{|q_n|^{2p-2}\}$.

It is worth noting that the statistical properties of the least ℓ_p-norm estimator have been studied in the literature [9], [19]. Nevertheless, our research results extend these existing works in the following aspects. Firstly, in [9] and [19], the variance formula of linear parameter estimation via ℓ_p-norm minimization is derived analytically but its extension to the nonlinear case is only a conjecture. On the other hand, the development of (6) considers general nonlinear parameter estimation, where the linear model is regarded as a special case. That is, we extend existing knowledge by providing theoretical justification for the variance expression of the nonlinear ℓ_p-norm minimizer. Secondly, their result corresponds to the asymptotic condition where infinite samples are considered. While the derivation of (6) does not require the large sample assumption. Thirdly, the investigated distributions in this paper are frequently encountered in signal processing applications. While in the numerical examples of [19], the studied distributions, such as the parabolic and triangular processes, are much more relevant to the statistics community. Finally, it is assumed in [9] and [19] that the noise should be IID, but (6) can also be employed for non-IID noise environments.

3. Review of Common Non-Gaussian Distributions

In this section, we review four well-known non-Gaussian distributions, namely, SαS, GG, Student's t and GM processes. Note that (6) can also be applied to other noise distributions with either closed-form PDF or CF.

3.1. α-Stable Process

We first consider the univariate SαS distribution. Although there is no closed-form PDF other than the special Cauchy and Gaussian cases, the CF of the SαS process is explicitly expressed as:

$$\varphi(t) = \exp(j\delta t - \gamma|t|^{\alpha}), \tag{10}$$

where $\alpha \in (0, 2]$ is the characteristic exponent parameter which controls the impulsiveness of the distribution. In particular, when $\alpha = 1$ and $\alpha = 2$, this function corresponds to the Cauchy and Gaussian distributions, respectively. The $\delta \in (-\infty, \infty)$ denotes the location parameter and we set $\delta = 0$ here because $\mathbf{q}$ is assumed zero-mean. The $\gamma > 0$ is the dispersion parameter which determines the spread of the distribution around δ. Note that in the case of Gaussian variable, γ is the variance.

For the SαS family, there exists finite moments only for orders less than α, which are known as FLOMs. Let q be a random variable which follows SαS distribution with $\delta = 0$. The FLOM of q is [17]:

$$E\{|q|^s\} = C_{\alpha}(s, \alpha)\gamma^{\frac{s}{\alpha}}, \qquad s \in (-1, 1) \cup (1, \alpha), \tag{11}$$

where $\cup$ denotes the union operator and

$$C_{\alpha}(s, \alpha) = \frac{\Gamma(\frac{s+1}{2})\Gamma(-s/\alpha)}{\alpha\sqrt{\pi}\Gamma(-s/2)} 2^{s+1}, \tag{12}$$

with $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ being the gamma function. Moreover, we can deduce that

$$E\{|q|^{s-1}\text{sign}(q)\} = 0. \tag{13}$$

3.2. Generalized Gaussian Process

Let q be a zero-mean random variable following the univariate symmetric GG distribution. The PDF of q is:

$$f(q) = \frac{\beta}{2\kappa\Gamma(1/\beta)} \exp\left(-\left(\frac{q}{\kappa}\right)^{\beta}\right), \tag{14}$$

where $\beta > 0$ denotes the shape parameter which tunes the decay rate of the density function. In particular, (14) reduces to the Laplace and Gaussian distributions at $\beta = 1$ and $\beta = 2$, respectively. The $\kappa > 0$ is the scale parameter and it becomes the standard deviation of a Gaussian variable when $\beta = 2$.

The sth-order FLOM of q is calculated by integrating $|q|^s$ over $f(q)$ given in (14):

$$E\{|q|^s\} = C_{\text{GG}}(s, \beta)\kappa^s, \qquad s \in (-1, 1) \cup (1, \beta], \tag{15}$$

where

$$C_{\text{GG}}(s, \beta) = \begin{cases} 0 & s = 2l + 1, \forall l \in \mathbb{Z};\ \beta > 2, \\ \frac{\Gamma(\frac{s+1}{\beta})}{\Gamma(1/\beta)} & \text{otherwise.} \end{cases} \tag{16}$$

Note that $E\{|q|^{s-1}\text{sign}(q)\}$ is also shown to be zero by integrating $|q|^{s-1}\text{sign}(q)$ over $f(q)$ in (14).

3.3. Student's t Process

Let q be a zero-mean random variable with the univariate symmetric Student's t distribution. Its PDF is:

$$f(q) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\eta\Gamma(\nu/2)} \left(1 + \frac{1}{\nu}\left(\frac{q}{\eta}\right)^2\right)^{-\frac{\nu+1}{2}}, \tag{17}$$

where $\nu > 0$ is the number of degrees of freedom. Two special cases, namely, $\nu = 1$ and $\nu = \infty$, correspond to the Cauchy and Gaussian distributions, respectively. The $\eta > 0$ is the scaling parameter determining the spread of the PDF.

The sth-order FLOM of q is also derived by integrating $|q|^s$ over $f(q)$ given in (17):

$$E\{|q|^s\} = C_t(s, \nu)\eta^s, \qquad s \in (-1, 1) \cup (1, \nu), \tag{18}$$

where

$$C_t(s, \nu) = \begin{cases} 0 & s = 2l + 1, \forall l \in \mathbb{Z};\ \nu > 2, \\ \frac{\Gamma(\frac{s+1}{2})\Gamma(\frac{\nu-s}{2})}{\sqrt{\pi}\Gamma(\nu/2)}\nu^{\frac{s}{2}} & \text{otherwise.} \end{cases} \tag{19}$$

Furthermore, similar with GG process, the expected values of the derivative of $|q|^s$ with respect to q are zero.

3.4. Gaussian Mixture Process

GM process [14] can have arbitrarily large number of components, but here we only consider the two-component case which is also known as contaminated Gaussian noise [20].

Let q be a zero-mean random variable following the univariate symmetric GM process which is constructed from two independent Gaussian distributions. The PDF of q is:

$$f(q) = (1-\epsilon) f_1(q;\sigma^2) + \epsilon f_1(q;\tau\sigma^2), \tag{20}$$

where $f_1(q;\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{q^2}{2\sigma^2}\right)$ and $\epsilon \in (0,1)$ is the weight parameter. The two Gaussian processes have variances σ^2 and $\tau\sigma^2$ with $\tau > 1$. Apparently, the second term of (20) corresponds to the outlier and we expect that the impulsiveness increases with ϵ and/or τ.

The FLOM of q is:

$$E\{|q|^s\} = C_{\mathrm{GM}}(s,\epsilon,\tau)\sigma^s, \tag{21}$$

where

$$C_{\mathrm{GM}}(s,\epsilon,\tau) = \frac{\left((1-\epsilon)+\epsilon\tau^{\frac{s}{2}}\right)\Gamma(\frac{s+1}{2})}{\sqrt{\pi}} 2^{\frac{s}{2}}. \tag{22}$$

Similarly, the expected values of the derivative of $|q|^s$ with respect to q are also equal to zero.

4. Linear Parameter Estimation

When the parameters are linear in the observations, the signal model of (1) can be rewritten as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{q}, \tag{23}$$

where $\mathbf{A} \in \mathbb{R}^{N\times M}$ is a known full-rank matrix with $N \geq M$ and the elements in $\mathbf{q}$ are now zero-mean IID random variables with a symmetric non-Gaussian distribution.

According to (8), the cost function to be minimized is:

$$J(\mathbf{x}) = \sum_{n=1}^{N}\left| y_n - \sum_{m=1}^{M} a_{n,m}x_m \right|^p, \quad p \in [1,2), \tag{24}$$

where $a_{n,m}$ denotes the (n,m) entry of $\mathbf{A}$.

First, we consider the case that $\mathbf{q}$ is IID SαS distributed. Employing (11) and (13) yields:

$$E\left\{\nabla(J(\mathbf{x}))\nabla^T(J(\mathbf{x}))\right\} = p^2 C_\alpha(2p-2,\alpha)\gamma^{\frac{2p-2}{\alpha}}\left(\mathbf{A}^T\mathbf{A}\right), \tag{25}$$

$$E\left\{\mathbf{H}(J(\mathbf{x}))\right\} = p(p-1)C_\alpha(p-2,\alpha)\gamma^{\frac{p-2}{\alpha}}\left(\mathbf{A}^T\mathbf{A}\right). \tag{26}$$

Substituting (25)–(26) into (6), we obtain the covariance matrix, denoted by $\mathbf{C}_\alpha(\hat{\mathbf{x}})$:

$$
\begin{aligned}
\mathbf{C}_\alpha(\hat{\mathbf{x}}) &= \frac{C_\alpha(2p-2,\alpha)}{(p-1)^2 C_\alpha^2(p-2,\alpha)}\gamma^{\frac{2}{\alpha}}\left(\mathbf{A}^T\mathbf{A}\right)^{-1} \\
&= \frac{2\alpha\sqrt{\pi}\Gamma^2(-\frac{p-2}{2})\Gamma(\frac{2p-1}{2})\Gamma(-\frac{2p-2}{\alpha})}{(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(-\frac{p-2}{\alpha})\Gamma(1-p)}\gamma^{\frac{2}{\alpha}}\left(\mathbf{A}^T\mathbf{A}\right)^{-1},
\end{aligned}
\tag{27}
$$

where $1 < p < \alpha$. Although (27) does not hold for $p = 1$, we apply the limit $p \to 1$ to handle this special case. Taking the limit of $p \to 1$ and employing $\Gamma(z+1) = z\Gamma(z)$, the covariance matrix at $p = 1$ becomes:

$$
\begin{aligned}
\lim_{p\to1}\mathbf{C}_\alpha(\hat{\mathbf{x}}) &= \lim_{p\to1}\left\{\frac{\sqrt{\pi}\Gamma^2(2-\frac{p}{2})\Gamma(p-\frac{1}{2})\Gamma(1-\frac{2p-2}{\alpha})}{\Gamma^2(\frac{p+1}{2})\Gamma^2(1-\frac{p-2}{\alpha})\Gamma(2-p)}\right\}\gamma^{\frac{2}{\alpha}}\left(\mathbf{A}^T\mathbf{A}\right)^{-1} \\
&= \frac{\pi^2}{4\Gamma^2(1+\frac{1}{\alpha})}\gamma^{\frac{2}{\alpha}}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}.
\end{aligned}
\tag{28}
$$

In a similar manner, when $\mathbf{q}$ is GG distributed, the corresponding covariance matrix, denoted by $\mathbf{C}_{\mathrm{GG}}(\hat{\mathbf{x}})$, can be derived as:

$$
\mathbf{C}_{\mathrm{GG}}(\hat{\mathbf{x}}) = \frac{\Gamma(\frac{2p-1}{\beta})\Gamma(\frac{1}{\beta})}{(p-1)^2\Gamma^2(\frac{p-1}{\beta})}\kappa^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1},\ p>1. \tag{29}
$$

While the limiting case of $p \to 1$ is shown to be

$$
\lim_{p\to1}\mathbf{C}_{\mathrm{GG}}(\hat{\mathbf{x}}) = \Gamma^2\left(1+\frac{1}{\beta}\right)\kappa^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{30}
$$

For Student's t noise, the corresponding covariance matrix, denoted by $\mathbf{C}_t(\hat{\mathbf{x}})$, is computed as:

$$
\mathbf{C}_t(\hat{\mathbf{x}}) = \frac{\nu\sqrt{\pi}\Gamma(\frac{\nu}{2})\Gamma(\frac{2p-1}{2})\Gamma(\frac{\nu-2p+2}{2})}{(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(\frac{\nu-p+2}{2})}\eta^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1},\ p>1. \tag{31}
$$

While for $p \to 1$, we obtain:

$$
\lim_{p\to1}\mathbf{C}_t(\hat{\mathbf{x}}) = \frac{\nu\pi\Gamma^2(\frac{\nu}{2})}{4\Gamma^2(\frac{\nu+1}{2})}\eta^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}. \tag{32}
$$

Finally, the development when the noise is a symmetric GM noise is given as follows:

$$
\mathbf{C}_{\mathrm{GM}}(\hat{\mathbf{x}}) = \frac{2\sqrt{\pi}\Gamma(\frac{2p-1}{2})\left((1-\epsilon)+\epsilon\tau^{p-1}\right)\sigma^2}{(p-1)^2\Gamma^2(\frac{p-1}{2})\left((1-\epsilon)+\epsilon\tau^{\frac{p-2}{2}}\right)^2}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}, \tag{33}
$$

where $p > 1$ and for $p \to 1$,

$$
\lim_{p\to1}\mathbf{C}_{\mathrm{GM}}(\hat{\mathbf{x}}) = \frac{\pi}{2\left((1-\epsilon)+\epsilon\tau^{-\frac{1}{2}}\right)^2}\sigma^2\left(\mathbf{A}^T\mathbf{A}\right)^{-1}, \tag{34}
$$

where $\mathbf{C}_{\mathrm{GM}}(\hat{\mathbf{x}})$ denotes the corresponding covariance matrix.

Two simple yet representative examples are presented for illustration as follows.

Example 1

We consider estimating an unknown constant c from

$$\mathbf{y} = \mathbf{1}_N c + \mathbf{q}, \tag{35}$$

where $\mathbf{A} = \mathbf{1}_N \in \mathbb{R}^N$ is now a column vector with all ones. Applying (27)–(34), we obtain the variance expressions for $\hat{c}$ in the presence of SαS, GG, Student's t and GM noise models:

$$\mathrm{var}_\alpha(\hat{c}) = \begin{cases} \frac{2\alpha\sqrt{\pi}\Gamma^2(-\frac{p-2}{2})\Gamma(\frac{2p-1}{2})\Gamma(-\frac{2p-2}{\alpha})}{N(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(-\frac{p-2}{\alpha})\Gamma(1-p)}\gamma^{\frac{2}{\alpha}} & p > 1, \\ \frac{\pi^2}{4N\Gamma^2(1+\frac{1}{\alpha})}\gamma^{\frac{2}{\alpha}} & p = 1, \end{cases} \tag{36}$$

$$\mathrm{var}_{\mathrm{GG}}(\hat{c}) = \begin{cases} \frac{\Gamma(\frac{2p-1}{\beta})\Gamma(\frac{1}{\beta})}{N(p-1)^2\Gamma^2(\frac{p-1}{\beta})}\kappa^2 & p > 1, \\ \frac{\Gamma^2(1+\frac{1}{\beta})}{N}\kappa^2 & p = 1, \end{cases} \tag{37}$$

$$\mathrm{var}_t(\hat{c}) = \begin{cases} \frac{\nu\sqrt{\pi}\Gamma(\frac{\nu}{2})\Gamma(\frac{2p-1}{2})\Gamma(\frac{\nu-2p+2}{2})}{N(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(\frac{\nu-p+2}{2})}\eta^2 & p > 1, \\ \frac{\nu\pi\Gamma^2(\frac{\nu}{2})}{4N\Gamma^2(\frac{\nu+1}{2})}\eta^2 & p = 1, \end{cases} \tag{38}$$

$$\mathrm{var}_{\mathrm{GM}}(\hat{c}) = \begin{cases} \frac{2\sigma^2\sqrt{\pi}\Gamma(\frac{2p-1}{2})\left((1-\epsilon)+\epsilon\tau^{p-1}\right)}{N(p-1)^2\Gamma^2(\frac{p-1}{2})\left((1-\epsilon)+\epsilon\tau^{\frac{p-2}{2}}\right)^2} & p > 1, \\ \frac{\pi}{2N\left((1-\epsilon)+\epsilon\tau^{-\frac{1}{2}}\right)^2}\sigma^2 & p = 1. \end{cases} \tag{39}$$

Notice that (37) and (38) are invalid for the conditions specified in (16) and (19) when $C_{\mathrm{GG}}(s,\beta) = C_t(s,\nu) = 0$. It is also worth mentioning that the result of $p = 1$ in the GM noise is in agreement with the variance analysis of the least ℓ_1-norm estimator in [5].

Example 2

Now we consider the single-tone signal model where the observations are:

$$y_n = A\sin(\omega n + \theta) + q_n, \quad n = 1, \cdots, N. \tag{40}$$

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

Noting that $A\sin(\omega n+\theta) = A\cos(\theta)\sin(\omega n) + A\sin(\theta)\cos(\omega n)$, we first transform the parameters of interest as $\mathbf{x} = [A\cos(\theta) \quad A\sin(\theta)]^T$. Comparing (40) with (23), we easily deduce that:

$$\mathbf{A} = \begin{bmatrix} \sin(\omega) & \cdots & \sin(N\omega) \\ \cos(\omega) & \cdots & \cos(N\omega) \end{bmatrix}^T. \tag{41}$$

The $\mathbf{C}(\hat{\mathbf{x}})$ can be computed using (27)–(34) for different noise models. Subsequently, the variances of $\hat{A}$ and $\hat{\theta}$ are calculated as [21]:

$$\mathrm{var}(\hat{A}) \approx \cos^2(\theta)c_{1,1} + \sin^2(\theta)c_{2,2} + 2\sin(\theta)\cos(\theta)c_{1,2}, \tag{42}$$

$$\mathrm{var}(\hat{\theta}) \approx A^2\left(\sin^2(\theta)c_{1,1} + \cos^2(\theta)c_{2,2} - 2\sin(\theta)\cos(\theta)c_{1,2}\right). \tag{43}$$

where $c_{k,l}$ denotes the (k,l) entry of $\mathbf{C}(\hat{\mathbf{x}})$ with $k,l = 1,2$.

5. Nonlinear Model

We now consider the general model of (1) as a nonlinear one and the ℓ_p-norm minimizer has been given in (8).

To compute the covariance matrix, we need the first-order and second-order derivatives of $J(\mathbf{x})$. Their expected values are:

$$E\left\{\nabla(J(\mathbf{x}))\nabla^T(J(\mathbf{x}))\right\} = p^2\left(\nabla^T(\mathbf{g}(\mathbf{x}))\right)\mathbf{V}\left(\nabla(\mathbf{g}(\mathbf{x}))\right), \tag{44}$$

$$E\left\{\mathbf{H}(J(\mathbf{x}))\right\} = p(p-1)\left(\nabla^T(\mathbf{g}(\mathbf{x}))\right)\mathbf{W}\left(\nabla(\mathbf{g}(\mathbf{x}))\right), \tag{45}$$

where $\mathbf{V} = \mathrm{diag}(E\{|e_1|^{2p-2} \ \cdots \ E\{|e_N|^{2p-2}\})$ and $\mathbf{W} = \mathrm{diag}(E\{|e_1|^{p-2}\} \ \cdots \ E\{|e_N|^{p-2}\})$ with $e_n = y_n - g_n(\mathbf{x})$.

As the exact expressions depend on $\mathbf{g}(\mathbf{x})$, we illustrate the variance formula using a sinusoidal parameter estimation model as follows.

<u>Example 3</u>

The data model is identical to (40) but now A is known while ω and θ are the unknown parameters to be estimated. Here we have $\mathbf{x} = [\omega \ \ \theta]^T$.

First, we examine the case when $\mathbf{q}$ is α-stable distributed. According to (6) and with the use of (11) and (13), (27)–(28) as well as (44)–(45), the corresponding covariance of $\mathbf{x}$, namely, $\mathbf{C}_\alpha(\hat{\mathbf{x}})$, is

$$\mathbf{C}_\alpha(\hat{\mathbf{x}}) = \begin{cases} \frac{2\alpha\sqrt{\pi}\Gamma^2(-\frac{p-2}{2})\Gamma(\frac{2p-1}{2})\Gamma(1-\frac{2p-2}{\alpha})}{(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(-\frac{p-2}{\alpha})\Gamma(1-p)}\gamma^{\frac{2}{\alpha}}\mathbf{S}^{-1} & p > 1, \\ \frac{\pi^2}{4\Gamma^2(1+\frac{1}{\alpha})}\gamma^{\frac{2}{\alpha}}\mathbf{S}^{-1} & p = 1. \end{cases} \tag{46}$$

where

$$\mathbf{S} = A^2\begin{bmatrix} \sum_{n=1}^N n^2\cos^2(\omega n+\theta) & \sum_{n=1}^N n\cos^2(\omega n+\theta) \\ \sum_{n=1}^N n\cos^2(\omega n+\theta) & \sum_{n=1}^N \cos^2(\omega n+\theta) \end{bmatrix}. \tag{47}$$

In a similar manner, the covariance expressions for GG, Student's t and GM noise models are evaluated as:

$$\mathbf{C}_{\mathrm{GG}}(\hat{\mathbf{x}}) = \begin{cases} \frac{\Gamma(\frac{2p-1}{\beta})\Gamma(\frac{1}{\beta})}{(p-1)^2\Gamma^2(\frac{p-1}{\beta})}\kappa^2\mathbf{S}^{-1} & p > 1, \\ \Gamma^2(1+\frac{1}{\beta})\kappa^2\mathbf{S}^{-1} & p = 1, \end{cases} \tag{48}$$

$$\mathbf{C}_t(\hat{\mathbf{x}}) = \begin{cases} \frac{\nu\sqrt{\pi}\Gamma(\frac{\nu}{2})\Gamma(\frac{2p-1}{2})\Gamma(\frac{\nu-2p+2}{2})}{(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(\frac{\nu-p+2}{2})}\eta^2\mathbf{S}^{-1} & p > 1, \\ \frac{\nu\pi\Gamma^2(\frac{\nu}{2})}{4\Gamma^2(\frac{\nu+1}{2})}\eta^2\mathbf{S}^{-1} & p = 1, \end{cases} \tag{49}$$

and

$$\mathbf{C}_{\mathrm{GM}}(\hat{\mathbf{x}}) = \begin{cases} \frac{2\sqrt{\pi}\Gamma(\frac{2p-1}{2})\left((1-\epsilon)+\epsilon\tau^{p-1}\right)}{(p-1)^2\Gamma^2(\frac{p-1}{2})\left((1-\epsilon)+\epsilon\tau^{\frac{p-2}{2}}\right)^2}\sigma^2\mathbf{S}^{-1} & p > 1, \\ \frac{\pi}{2\left((1-\epsilon)+\epsilon\tau^{-\frac{1}{2}}\right)^2}\sigma^2\mathbf{S}^{-1} & p = 1. \end{cases} \tag{50}$$

6. Choice of p

It is apparent that the choice of p is important because it affects the performance of the ℓ_p-norm estimator. In the literature, [19] and [22]–[25] have mentioned that p can be chosen according to its empirical relationship with the residual kurtosis. However, these suggestions cannot be applied if the kurtosis does not exist, such as in the SαS and Student's t processes with $\nu \leq 4$. The work of [25] overviews the three *ad hoc* schemes in [22]–[24] and designs an adaptive algorithm for determining p, but the simulation results indicate that the optimum p cannot be obtained in all these methods. For SαS distribution with $1 < \alpha < 2$, [26] suggests $p = \alpha - o$ with $o \ll \alpha$ being a user-defined parameter. Nevertheless, this criterion cannot guarantee the minimum MSE because its design motivation is to increase the convergence speed of the IRLS instead of minimizing the variance. Furthermore, [19] and [27] propose selecting p by finding the minimum variance expression, but only numerical results are provided. In this section, we determine the optimal value of p, which is achieved by minimizing the variance in an analytic manner. Our method is able to find the best p and is not dependent on the kurtosis of the noise distributions. Moreover, the global optimality and sensitivity analysis are included.

6.1. α-Stable Process

6.1.1. Optimum of p

We first study least ℓ_p-norm estimation in the presence of SαS noise. For the covariance matrices of (27) and (46), the scalar term characterized by p, denoted by $\Phi_\alpha(p)$, is:

$$\begin{aligned}\Phi_\alpha(p) &= \frac{2\alpha\sqrt{\pi}\Gamma^2(-\frac{p-2}{2})\Gamma(\frac{2p-1}{2})\Gamma(-\frac{2p-2}{\alpha})}{(p-1)^2\Gamma^2(\frac{p-1}{2})\Gamma^2(-\frac{p-2}{\alpha})\Gamma(1-p)} \\ &= \frac{\sqrt{\pi}\Gamma^2(1-\frac{p-2}{2})\Gamma(\frac{2p-1}{2})\Gamma(1-\frac{2p-2}{\alpha})}{\Gamma^2(\frac{p+1}{2})\Gamma^2(1-\frac{p-2}{\alpha})\Gamma(2-p)}.\end{aligned} \tag{51}$$

To find the value of p which minimizes $\Phi_\alpha(p)$, a standard technique is to set its derivative, namely, $\Phi'_\alpha(p)$, to zero. To solve the problem in an efficient manner, here we express $\Phi'_\alpha(p)$ as:

$$\Phi'_\alpha(p) = \Phi_\alpha(p) h_\alpha(p), \tag{52}$$

where

$$h_\alpha(p) = \psi\left(2-p\right) - \psi\left(2-\frac{p}{2}\right) + \psi\left(p-\frac{1}{2}\right) - \psi\left(\frac{p+1}{2}\right) + \frac{2}{\alpha}\left(\psi\left(1-\frac{p-2}{\alpha}\right) - \psi\left(1-\frac{2p-2}{\alpha}\right)\right) \tag{53}$$

with $\psi(\cdot)$ being the digamma function [28] which is defined as $\psi(\cdot) = \frac{\Gamma'(\cdot)}{\Gamma(\cdot)}$. Since $\Phi_\alpha(p) > 0$, we only need to solve the root of $h_\alpha(p) = 0$. Employing TSE on (53) and using the property that

$\psi(x) = -a + \sum_{n=0}^{\infty}\left(\frac{1}{n+1} - \frac{1}{n+1+x}\right)$ where a is the Euler-Mascheroni constant, $h_\alpha(p)$ can be written as

$$h_\alpha(p) = \sum_{k=0}^{\infty}\left(a_k p^k + b_k p^{-k-1}\right), \tag{54}$$

where

$$a_k = \frac{(-1)^k}{k!}\left[\frac{2-2^{k+1}}{\alpha^{k+1}}\psi^{(k)}\left(\frac{2+\alpha}{\alpha}\right) + \frac{2^k-1}{2^k}\psi^{(k)}(2) - \left(\left(-\frac{1}{2}\right)^k - (-1)^k\right)\psi^{(k)}\left(\frac{3}{2}\right)\right], \tag{55}$$

$$b_k = 2(-1)^k - \frac{(-1)^k+1}{2^k}, \tag{56}$$

with $\psi^{(k)}(\cdot)$ being the polygamma function [28] and $k!$ denoting factorial of k. For SαS noise, the admissible value of p is in $[1,\alpha)$. To determine its optimal value practically, we can solve $\sum_{k=0}^{K}(a_k p^k + b_k p^{-k-1}) = 0$ where K is chosen sufficiently large, using a numerical method with the initial value of $p = \frac{\alpha+1}{2}$. It is because according to extensive simulation results, the variance expressions based on $p = \frac{\alpha+1}{2}$ are indeed very close to those with the optimal p.

6.1.2. Global Optimality Analysis

Let $p^\star$ be the root of $h(p) = 0$ when $p \in [1,\alpha)$. To prove that $p^\star$ is the global optimal value, two conditions should be satisfied simultaneously: $h_\alpha(p)$ is a monotonically function concerning p and $h_\alpha(1)h_\alpha(\alpha) < 0$. Note that if $h_\alpha(p)$ is differentiable, the monotonicity of $h_\alpha(p)$ is determined by the sign of its first-order derivative, denoted by $h'_\alpha(p)$.

If α is fixed, then according to (53), $h'_\alpha(p)$ has the form of

$$\begin{aligned} h'_\alpha(p) =& \frac{4}{\alpha^2}\left(\psi^{(1)}\left(1 - \frac{2p-2}{\alpha}\right) - \frac{1}{2}\psi^{(1)}\left(1 - \frac{p-2}{\alpha}\right)\right) + \frac{1}{2}\psi^{(1)}\left(2 - \frac{p}{2}\right) \\ &- \psi^{(1)}(2-p) + \psi^{(1)}\left(p - \frac{1}{2}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{p+1}{2}\right), \quad p \in [1,\alpha). \end{aligned} \tag{57}$$

Since $\psi^{(1)}(\cdot)$ is decreasing, we know that $\psi^{(1)}\left(1 - \frac{2p-2}{\alpha}\right) - \frac{1}{2}\psi^{(1)}\left(1 - \frac{p-2}{\alpha}\right) > 0$, and (57) becomes

$$\begin{aligned} h'_\alpha(p) >& \frac{1}{2}\psi^{(1)}\left(2 - \frac{p}{2}\right) - \psi^{(1)}(2-p) + \psi^{(1)}\left(p - \frac{1}{2}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{p+1}{2}\right) \\ =& \psi^{(1)}\left(p - \frac{1}{2}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{p+1}{2}\right) > 0. \end{aligned} \tag{58}$$

Accordingly, with a fixed α, $h_\alpha(p)$ is monotonically increasing.

On the other hand, if p equals 1 or α, $h_\alpha(1)$ and $h_\alpha(\alpha)$ are expressed as:

$$h_\alpha(1) = \frac{2}{\alpha}\left(\psi\left(\frac{1}{\alpha}\right) - \psi(1)\right) < 0, \tag{59}$$

$$h_\alpha(\alpha) = \psi(2-\alpha) - \psi\left(1 - \frac{\alpha}{2}\right) + \psi\left(\alpha - \frac{1}{2}\right) - \psi\left(\frac{\alpha+1}{2}\right). \tag{60}$$

To see if $h_\alpha(1)h_\alpha(\alpha) < 0$, we investigate the sign of $h_\alpha(\alpha)$ in (60). Let $l(\alpha) = h_\alpha(\alpha)$ be a function of $\alpha \in (1,2)$ and employing the multiplication theorem [28] of $\psi^{(1)}(\cdot)$, the first-order derivative of $l(\alpha)$ is computed as

$$l^{'}(\alpha) = \frac{1}{4}\left(\psi^{(1)}\left(1-\frac{\alpha}{2}\right) - \psi^{(1)}\left(\frac{3-\alpha}{2}\right)\right) + \frac{1}{2}\psi^{(1)}\left(\alpha-\frac{1}{2}\right) + \frac{1}{2}\left(\psi^{(1)}\left(\alpha-\frac{1}{2}\right) - \psi^{(1)}\left(\frac{\alpha+1}{2}\right)\right)$$
$$>0. \tag{61}$$

Then $l(\alpha)$ is proved to be an increasing function, and the lower bound of $h_\alpha(\alpha)$ is

$$h_\alpha(\alpha) > l(1) = \psi(1) - \psi\left(\frac{1}{2}\right) + \psi\left(\frac{1}{2}\right) - \psi(1) = 0. \tag{62}$$

Combining (58), (59) and (62), it is easy to show that $p^\star$ results in the global optimum.

6.1.3. Sensitivity Analysis

We now investigate how the variation on $p^\star$ affects $\Phi(p)$. Let Δp be the deviation from $p^\star$, the sensitivity is measured by the difference quotient:

$$s_\alpha(\Delta p) = \frac{\Phi_\alpha(p^\star + \Delta p) - \Phi_\alpha(p^\star)}{\Delta p}, \tag{63}$$

where $1 \leq p^\star + \Delta p < \alpha$.

Employing the TSE, we expand $\Phi_\alpha(p^\star + \Delta p)$ around $p^\star$:

$$\Phi_\alpha(p^\star + \Delta p) \approx \Phi_\alpha(p^\star) + \Phi^{'}_\alpha(p^\star)\Delta p + \frac{1}{2}\Phi^{''}_\alpha(p^\star)(\Delta p)^2, \tag{64}$$

Since $\Phi^{'}_\alpha(p^\star) = 0$, we obtain from (63)–(64):

$$s_\alpha(\Delta p) \approx \frac{1}{2}\Phi^{''}_\alpha(p^\star)\Delta p, \tag{65}$$

where $\Phi^{''}_\alpha(p^\star) = h^{'}_\alpha(p^\star)\Phi_\alpha(p^\star)$. Equation (65) indicates that the variance sensitivity is a function of α.

To investigate the relationship between the variance formula and α, we employ the approximation of $p^\star \approx \frac{1+\alpha}{2}$. Let $\Upsilon_\alpha(\alpha) = h^{'}_\alpha(p^\star)$ which has the form of:

$$\Upsilon_\alpha(\alpha) = \frac{1}{2}\psi^{(1)}\left(\frac{7-\alpha}{4}\right) - \psi^{(1)}\left(\frac{3-\alpha}{2}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{3+\alpha}{4}\right) + \psi^{(1)}\left(\frac{\alpha}{2}\right)$$
$$+ \frac{4}{\alpha^2}\left(\psi^{(1)}\left(\frac{1}{\alpha}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{\alpha+3}{4}\right)\right). \tag{66}$$

Utilizing the multiplication theorem of $\psi^{(1)}(\cdot)$, the first-order derivative of (66) is

$$\Upsilon^{'}_\alpha(\alpha) = \frac{1}{16}\psi^{(2)}\left(\frac{3-\alpha}{4}\right) - \frac{1}{8}\psi^{(2)}\left(\frac{7-\alpha}{4}\right) + \frac{1}{16}\psi^{(2)}\left(\frac{\alpha}{4}\right) + \frac{1}{16}\psi^{(2)}\left(\frac{5-\alpha}{4}\right) + \frac{1}{16}\psi^{(2)}\left(\frac{\alpha+2}{4}\right)$$
$$- \frac{1}{8}\psi^{(2)}\left(\frac{3+\alpha}{4}\right) + \frac{4}{\alpha^4}\left(\psi^{(2)}\left(\frac{\alpha+3}{4}\right) - \psi^{(2)}\left(\frac{1}{\alpha}\right)\right) - \frac{8}{\alpha^3}\left(\psi^{(1)}\left(\frac{1}{\alpha}\right) - \frac{1}{2}\psi^{(1)}\left(\frac{\alpha+3}{4}\right)\right). \tag{67}$$

Employing the series representation of $\psi^{(1)}(\cdot)$, it is easy to prove that $\Upsilon^{'}_\alpha(\alpha) < 0$. Then we know that $\Phi^{''}_\alpha(p^\star)$ is larger for smaller α. This indicates that the variance formula is more sensitive to the variation of $p^\star$ if α is smaller.

6.2. Generalized Gaussian Process

For GG noise, the corresponding scalar term, denoted by $\Phi_{\rm GG}(p)$, is extracted from (29) or (48) as

$$\Phi_{\rm GG}(p) = \frac{\Gamma(\frac{2p-1}{\beta})}{(p-1)^2\Gamma^2(\frac{p-1}{\beta})} = \frac{\Gamma(\frac{2p-1}{\beta})}{\beta^2\Gamma^2(\frac{p-1}{\beta}+1)}. \tag{68}$$

As in (52), the derivative of $\Phi_{\rm GG}(p)$, namely, $\Phi^{'}_{\rm GG}(p)$, can also be factorized as:

$$\Phi^{'}_{\rm GG}(p) = \Phi_{\rm GG}(p) h_{\rm GG}(p), \quad \Phi_{\rm GG}(p) > 0, \tag{69}$$

where

$$h_{\rm GG}(p) = \frac{2}{\beta}\left(\psi\left(\frac{2p-1}{\beta}\right) - \psi\left(\frac{p-1}{\beta}+1\right)\right). \tag{70}$$

Since $\psi(x)$ is monotonically increasing for $x > 0$, it is easily deduced that $h_{\rm GG}(p) = 0$ if and only if $p = \beta$. Note that this is not surprising since the ML estimator is the ℓ_β-norm minimizer.

Similar to the α-stable distribution, it is easy to prove that $p^\star = \beta$ is the global optimum. While for the sensitivity analysis, if β is smaller, $\Phi_{\rm GG}(p)$ is more sensitive to the deviation from $p^\star$.

6.3. Student's t Process

From (31) or (49), the scalar term with Student's t noise, denoted by $\Phi_t(p)$, is

$$\Phi_t(p) = \frac{\sqrt{\pi}\Gamma(\frac{\nu}{2})\Gamma(\frac{2p-1}{2})\Gamma(\frac{\nu-2p+2}{2})}{4\Gamma^2(\frac{p+1}{2})\Gamma^2(\frac{\nu-p+2}{2})}. \tag{71}$$

The derivative of $\Phi_t(p)$ can be expressed as

$$\Phi^{'}_t(p) = \Phi_t(p) h_t(p), \quad \Phi_t(p) > 0, \tag{72}$$

where

$$h_t(p) = \psi\left(p - \frac{1}{2}\right) - \psi\left(\frac{p+1}{2}\right) + \psi\left(\frac{\nu+2-p}{2}\right) - \psi\left(\frac{\nu+2}{2} - p\right). \tag{73}$$

Applying TSE on (73) yields the same expression as in (54) except that a_k and b_k are

$$a_k = \frac{(-1)^k}{k!}\left[\left((-1)^k - \left(\frac{-1}{2}\right)^k\right)\psi^{(k)}\left(\frac{3}{2}\right) + \left(\left(\frac{1}{2}\right)^k - 1\right)\psi^{(k)}\left(1+\frac{\nu}{2}\right)\right], \tag{74}$$

$$b_k = 2(-1)^k + \frac{(-1)^{k+1}-1}{2^k}. \tag{75}$$

with $1 \le p < \nu$.

To determine the optimum value of p, we follow the same procedure as in the case of α-stable process. Nevertheless, the variance expressions based on $p = \frac{\nu+4}{5}$ are very close to those with the best p according to our simulation results.

Assume $p^\star$ is the solution of $h_t(p) = 0$ when $p \in [1, \nu)$. Similarly to the SαS distribution, it is easy to show that $p^\star$ is the globally optimal value, while $\Phi_t(p)$ is more sensitive to Δp for the smaller ν.

6.4. Gaussian Mixture Process

For GM noise, the corresponding scalar term, denoted by $\Phi_{\mathrm{GM}}(p)$, is extracted from (33) or (50) as

$$\Phi_{\mathrm{GM}}(p) = \frac{\sqrt{\pi}\Gamma(\frac{2p-1}{2})\left((1-\epsilon)+\epsilon\tau^{p-1}\right)}{2\Gamma^2(\frac{p+1}{2})\left((1-\epsilon)+\epsilon\tau^{\frac{p-2}{2}}\right)^2}. \tag{76}$$

The derivative of $\Phi_{\mathrm{GM}}(p)$ is factorized as:

$$\Phi'_{\mathrm{GM}}(p) = \Phi_{\mathrm{GM}}(p)h_{\mathrm{GM}}(p), \quad \Phi_{\mathrm{GM}}(p) > 0, \tag{77}$$

where

$$h_{\mathrm{GM}}(p) = \psi\left(p-\frac{1}{2}\right) - \psi\left(\frac{p+1}{2}\right) - \frac{(1-\epsilon)\ln(\tau)}{(1-\epsilon)+\epsilon\tau^{p-1}} + \frac{(1-\epsilon)\ln(\tau)}{(1-\epsilon)+\epsilon\tau^{\frac{p}{2}-1}}. \tag{78}$$

To solve $h_{\mathrm{GM}}(p) = 0$, we apply TSE on (78) as in the case of SαS process, which is identical to (54) except that a_k and b_k are now

$$a_k = \frac{\ln(\tau)}{k!}\left[I^{(k)}\left(\frac{\epsilon\sqrt{\tau}}{(1-\epsilon)}, \sqrt{\tau}\right) - I^{(k)}\left(\frac{\epsilon}{(1-\epsilon)}, \tau\right)\right] + \frac{(-1)^k}{k!}\left[\left((-1)^k - \left(\frac{-1}{2}\right)^k\right)\psi^{(k)}\left(\frac{3}{2}\right)\right], \tag{79}$$

$$b_k = 2(-1)^k - \frac{(-1)^k + 1}{2^k}, \tag{80}$$

where $1 \le p < 2$ and $I^{(k)}(c, x)$ denotes kth-order derivative of $\frac{1}{1+cx^{p-1}}$ with respect to p when $p = 0$.

Assuming that τ is large enough, say, $\tau > 100$, which corresponds to an impulsive noise, $h_{\mathrm{GM}}(p)$ is a monotonically increasing function, that is,

$$h_{\mathrm{GM}}(p) > h_{\mathrm{GM}}(1) = -2\ln(2) - (1-\epsilon)\ln(\tau) + \frac{(1-\epsilon)\ln(\tau)}{1-\epsilon+\epsilon\tau^{-\frac{1}{2}}}. \tag{81}$$

As a rule of thumb, for $\epsilon \ge 0.35$, $h_{\mathrm{GM}}(p) > 0$ and we will choose $p = 1$ as the optimal value because of monotonicity of $h_{\mathrm{GM}}(p)$. While for $0 < \epsilon < 0.35$, the empirically optimum value is $p = 1 + \frac{2}{100\epsilon+3}$.

Let $p^\star$ be the root of $h_{\mathrm{GM}}(p) = 0$ among $p \in [1, 2)$. The $p^\star$ can be proved to be the global optimum in a similar way to the analysis of SαS process. Furthermore, from the monotonicity study of $\Phi_{\mathrm{GM}}(p^\star + \Delta p)$, we can know that if ϵ and τ are large, $\Phi_{\mathrm{GM}}(p)$ is more sensitive to the variation on $p^\star$.

Our analysis for sensitivity under four distributions has been verified in Figure 1. Analogous to $\Upsilon_\alpha(\alpha)$ in (66), we define $\Upsilon_{\mathrm{GG}}(\beta)$, $\Upsilon_t(\nu)$ and $\Upsilon_{\mathrm{GM}}(\epsilon, \tau)$ to measure sensitivity for other three distributions. In GM process, to show the variation of $\Upsilon_{\mathrm{GM}}(\epsilon, \tau)$ for different ϵ, we fix $\tau = 100$ while in varying τ, $\epsilon = 0.05$ is selected. It is shown that when the density parameter increases, the sensitivity decreases for SαS, GG and Student's t processes while it increases for GM distribution.

7. Simulation Results

Computer simulations have been carried out to corroborate the derived variance formulas for least ℓ_p-norm estimation under four zero-mean IID symmetric non-Gaussian noise models. The density parameters are chosen as $\alpha = \nu = 1.8$, $\beta = 1.5$, $\epsilon = 0.05$ and $\tau = 100$. According to Section 6, the nearly optimum values are $p = 1.4$, $p = 1.16$ and $p = 1.25$ for the SαS, Student's t and GM models, respectively. It is found that $K = 65$ is sufficiently large to compute the corresponding $p^\star$, which are 1.43, 1.14 and 1.27, and they are very close to their empirical counterparts. While it is apparent that $p^\star = 1.5$ for the GG disturbance. We scale the noise sequences to produce different signal-to-noise ratio (SNR) conditions. It is worth noting that second-order power diverges for the α-stable and Student's t models with $\nu \leq 2$, and hence the geometric power is employed instead, corresponding to the geometric SNR (GSNR) [29]. The CRLB [30] is included as the benchmark while the results of classical approaches, namely, the ℓ_1-norm and ℓ_2-norm estimators, are also provided for comparison. Here, the ℓ_1-norm estimator is realized by the subgradient method with 200 iterations [10]. The signal models in Examples 1 to 3 are studied. The ℓ_p-norm minimization with $1 < p < 2$ in the linear estimation problem is realized by the IRLS approach while the nonlinear estimation problem is solved by the LM algorithm [9]. Unless stated otherwise, the data length is $N = 50$. All results are based on 1000 independent runs.

First, linear estimation of a scalar is investigated, which corresponds to Example 1. The signal is generated according to (35) where $c = 1$. Figure 2 shows the MSE of $\hat{c}$ versus SNR/GSNR. It is observed that the variance formulas align with the simulation results for all noise models and the ℓ_p-norm estimator is superior to the ℓ_1-norm and ℓ_2-norm minimizers. In all cases, there is no threshold performance because the parameter of interest is linear in the observations. The ℓ_p-norm estimator provides the optimum performance for GG process when $p = \beta$, since its variance can attain CRLB. Moreover, the MSE performance is nearly optimum for the SαS, Student's t and GM disturbances. The validity of (36), (37) and (39) versus N is studied in Figure 3. It is seen that the variance formulas may not be valid for a very short data length, say, $N \leq 3$. Moreover, when N increases, the corresponding MSE decreases.

Second, we follow Examples 2 and 3 to investigate the sinusoidal parameter estimation performance. The signal is generated according to (40) and the sinusoidal parameters are $A = 1$, $\omega = 1.25$ and $\theta = 0.5$. Figures 4 and 5 show the MSE performance versus SNR/GSNR for the linear and nonlinear estimation scenarios, respectively. In both figures, we see the optimality of the least ℓ_p-norm estimator in the GG noise model and its suboptimal performance in the SαS, Student's t and GM disturbance. Most importantly, the variance formulas align with the simulation results for all cases when the SNR or GSNR is sufficiently high. Note that threshold effect occurs when the noise power is large enough because the sinusoidal phase and frequency are nonlinear in nature, although for the former, parameter transformation is utilized in linearization.

In summary, the MSE decreases with the number of samples and (6) may not be accurate for a very small value of N. For linear parameter estimation, the derived variance formulas are corroborated for all SNR conditions while they may be invalid for nonlinear estimators when there are large estimation errors. When the PDF or CF of the noise is exactly known, we have shown that the optimum value of p depends on the density parameters via studying the α-stable, GG, Student's t and GM processes. Note that $p = 1$ is an appropriate choice for a heavier-tailed distribution particularly when the density parameters are unknown. It is already indicated in Figures 2 to 5 that accurate parameter estimation can be achieved with $p = 1$ for the α-stable, Student's t and GM processes. On the other hand, we also demonstrate that the performance of the ℓ_1-norm estimator is inferior to that with $p = 2$ for a less impulsive process, namely, the GG noise. For a heavier-tailed α-stable noise with $\alpha = 1$, the nearly optimal value of p is $(\alpha+1)/2 = 1$, which also supports the use of the ℓ_1-norm minimizer. Nevertheless, when the α-stable noise is more impulsive with $\alpha < 1$, we may need $p < 1$, which corresponds to an interesting and challenging future research.

8. Conclusion

In this work, we focus on parameter estimation in the presence of non-Gaussian impulsive noise. Variance of robust estimator based on the least ℓ_p-norm technique with $1 \leq p < 2$ in four representative disturbances has been analyzed. The accuracy of the derived variance formulas is validated using linear and nonlinear estimation examples. It is demonstrated when an appropriate value of p is chosen, the variances of the least ℓ_p-norm estimator for SαS, Student's t and GM noise models can be close to corresponding CRLB while optimality is achieved for the GG process. Moreover, it is worth mentioning that the bias and MSE analysis can also be employed for constrained optimization problems. For example, the time-of-arrival based localization problem is formulated as minimizing an objective function subject to a quadratic constraint in [31]. To apply (5)-(6), this constrained problem is converted to an unconstrained one by substituting the constraint into the objective function. Nevertheless, its extension to general constrained optimization problems, which include ridge regression [32] and ℓ_1-regularization [33], will be one of our future works.

References

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

[2] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 1991.

[3] V.H. MacDonald and P.M. Schultheiss, "Optimum passive bearing estimation in a spatially incoherent noise environment," *Journal of the Acoustical Society of America*, vol. 46, no. 1, pt. 1, pp. 37-43, Jul. 1969.

[4] H.C. So, Y.T. Chan, K.C. Ho and Y. Chen, "Simple formulas for bias and mean square error computation," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 162-165, Jul. 2013.

[5] R.A. Maronna, R.D Martin and V.J. Yohai, *Robust Statistics: Theory and Methods*, Chichester, England: Wiley, 2006.

[6] 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, Jul. 2012.

[7] P.J. Huber and E.M. Ronchetti, *Robust Statistics*, Hoboken, N.J.: Wiley, 2009.

[8] R. Byrd and D. Payne, "Convergence of the IRLS algorithm for robust regression," John Hopkins University, Baltimore, Maryland, United States, Technical Report 313, 1979.

[9] R. Gonin and A.H. Money, *Nonlinear L_p-Norm Estimation*, New York: Marcel Dekker, 1989.

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

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

[12] G.R. Arce, *Nonlinear Signal Processing: A Statistical Approach*, New York: John Wiley & Sons Inc., 2005.

[13] 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.

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

[15] M. Nassar, K. Gulati, Y. Mortazavi and B.L. Evans, "Statistical modeling of asynchronous impulsive noise in powerline communication networks," *in Proceedings of 2011 IEEE Global Telecommunications Conference (GLOBECOM)*, Houston, TX, Dec. 2011, pp. 1-6.

[16] P.G. Georgiou, P. Tsakalides and C. Kyriakakis, "Alpha-stable modeling of noise and robust time delay estimation in the presence of impulsive noise," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 291-301, Sep. 1999.

[17] E.E. Kuruoglu, "Density parameter estimation of skewed α-stable distributions," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2192-2220, Oct. 2001.

[18] J. Vila-Valls, Q. Wei, P. Closas and C. Fernandez-Prades, "Robust Gaussian sum filtering with unknown noise statistics: Application to target tracking," *in Proceedings of 2014 IEEE Workshop on Statistical Signal Processing (SSP)*, Gold Coast, Australia, Jun. 2014, pp. 416-419.

[19] R. Gonin and A.H. Money, "Nonlinear L_p-norm estimation: part I-on the choice of the exponent, p, where the errors are additive," *Communications in Statistics-Theory and Methods*, vol. 14, no. 4, pp. 827-840, 1985.

[20] B. Seyfe and S. Valaee, "A new choice of penalty function for robust multiuser detection based on M-estimation," *IEEE Transactions on Communications*, vol. 53, no. 2, pp. 224-227, Feb. 2005.

[21] E. Anarim and Y. Istefanopulos, "Statistical analysis of Pisarenko type tone frequency estimator," *Signal Processing*, vol. 24, no. 1, pp. 291-298, 1991.

[22] A.H. Money, J.F. Affleck-Graves, M.L. Hart and G.D.I. Barr, "The linear regression model: L_p-norm estimation and the choice of p," *Communications in Statistics-Simulation and Computation*, vol. 11, no. 1, pp. 89-109, 1982.

[23] V.A. Sposito, M.L. Hand and B. Skarpness, "On the efficiency of using sample kurtosis in selecting optimal L_p estimators," *Communications in Statistics-Simulation and Computation*, vol. 12, no. 3, pp. 265-272, Jan. 1983.

[24] G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, 1973.

[25] R. Gonin and A.H. Money, "Nonlinear L_p-norm estimation: Part II: The asymptotic distribution of the exponent, p, as a function of the sample kurtosis," *Communications in Statistics-Theory and Methods*, vol. 14, no. 4, pp. 841-849, 1985.

[26] E.E. Kuruoglu, P.J.W. Rayner and W.J. Fitzgerald, "Least L_p-norm impulsive noise cancellation with polynomial filters," *Signal Processing*, vol. 69, no. 1, pp. 1-14, Aug. 1998.

[27] H. Nyquist, "The optimal $\mathrm{L_p}$ norm estimator in linear regression models," *Communication in Statistics-Theory and Methods*, vol. 12, no. 21, pp. 2511-2524, Jan. 1983.

[28] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, New York: Dover Publications, 1964.

[29] J.G. Gonzalez, J.L. Paredes and G.R. Arce, "Zero-order statistics: A mathematical framework for the processing and characterization of very impulsive signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3839-3851, Oct. 2006.

[30] W.-J. Zeng, H.C. So and L. Huang, "ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Transactions on Signal Processing*, vol. 61, no. 17, pp. 4296-4308, Sep. 2013.

[31] K.W. Cheung, H.C. So, W.-K. Ma and Y.T. Chan, "Least squares algorithms for time-of-arrival based mobile location," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1121-1128, Apr. 2004.

[32] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, Academic Press/Elsevier, 2015.

[33] M. Figueiredo, R. Nowak and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics on Signal Processing*, vol. 1, no. 4, pp. 586-597, 2007

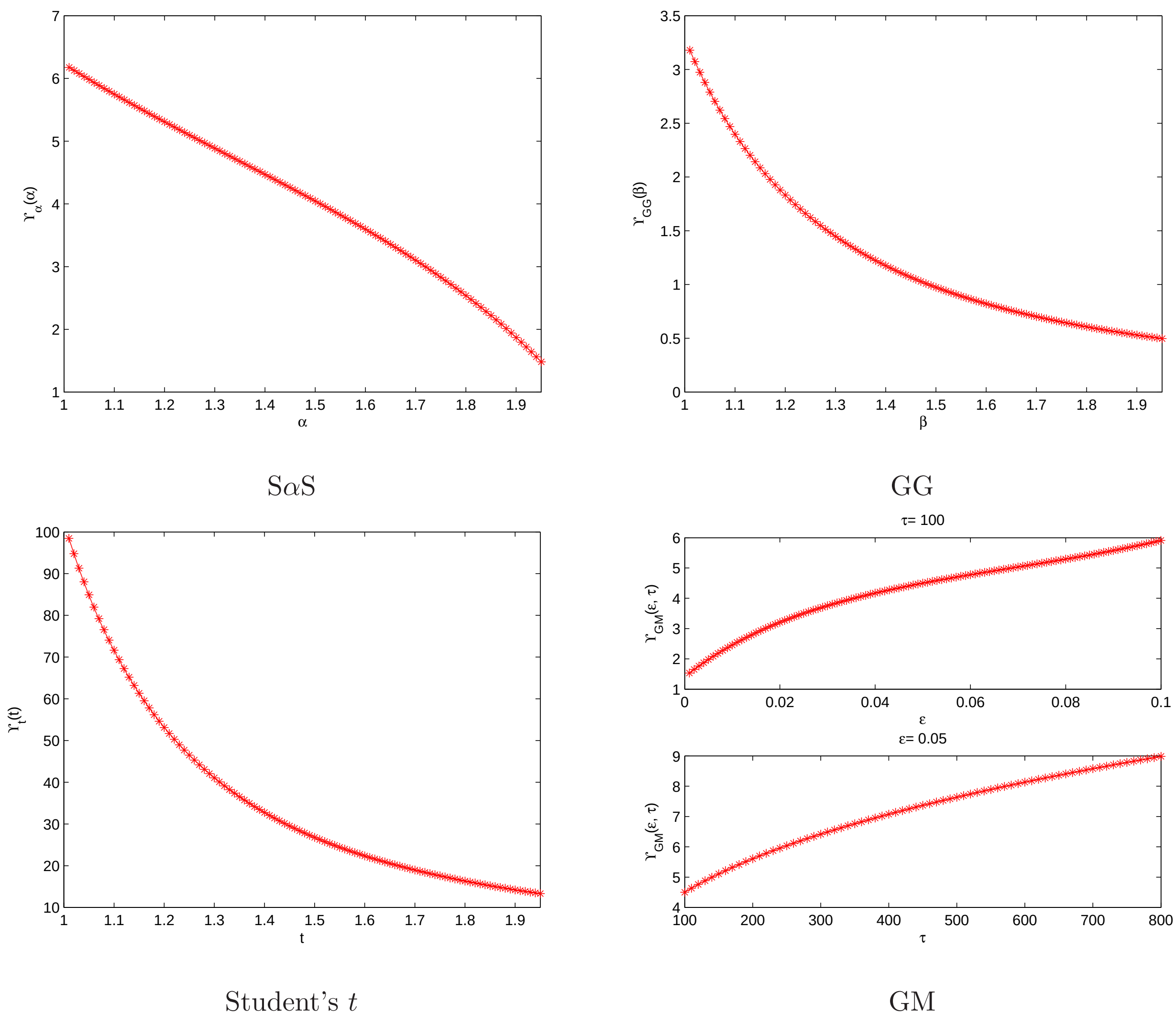
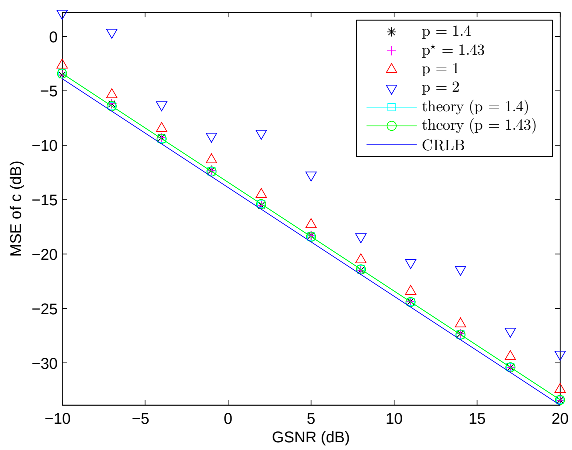
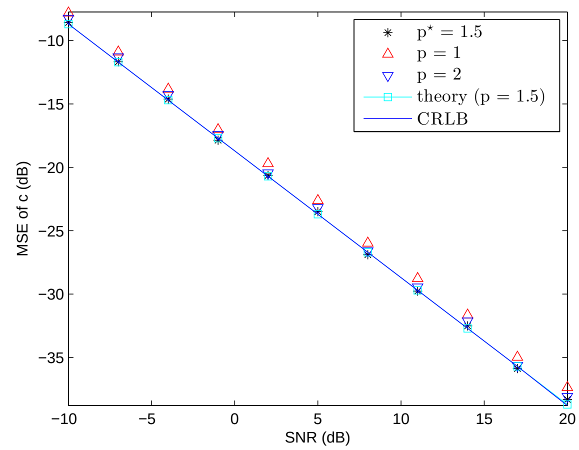


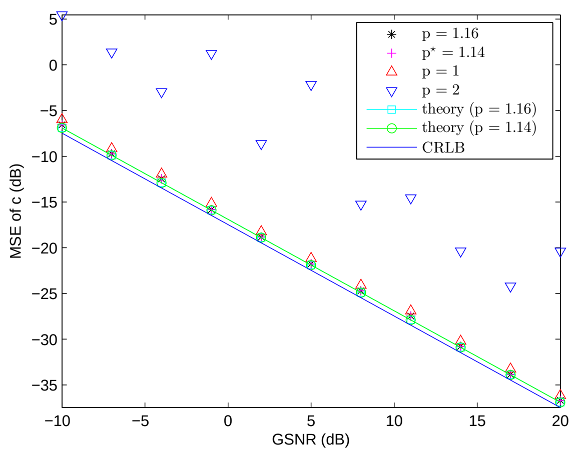
Figure 1: Sensitivity of variance to p versus density parameters. Top left: α-stable noise. Top right: GG noise. Bottom left: Student's t noise. Bottom right: GM noise.



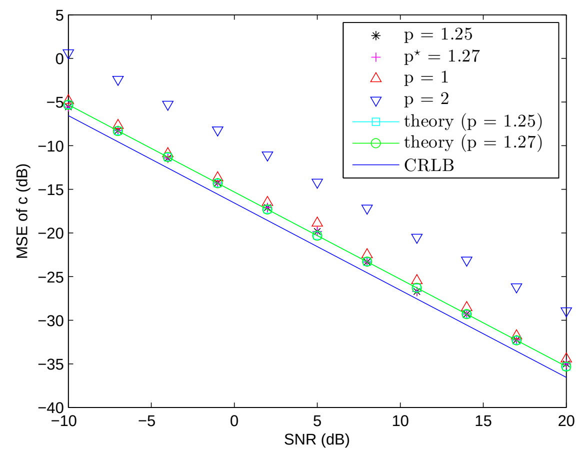
SαS



GG

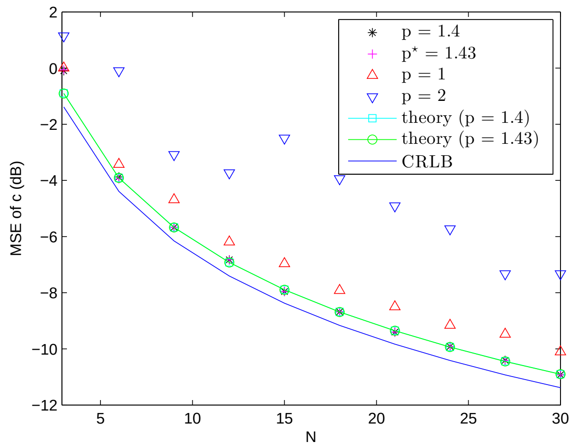


Student's t

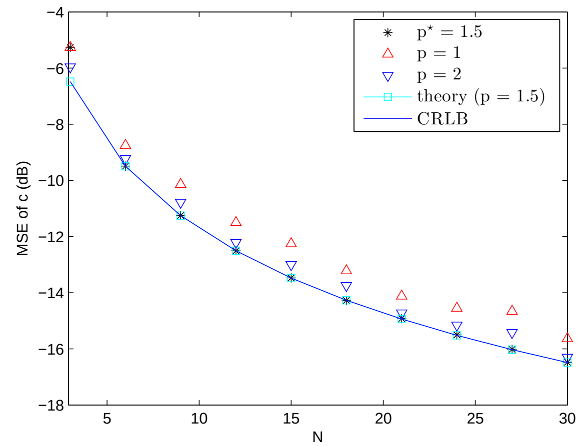


GM

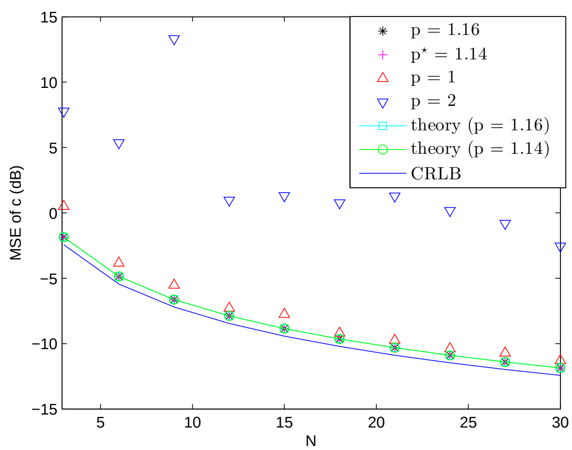
Figure 2: MSE of scalar versus SNR/GSNR. Top left: α-stable noise. Top right: GG noise. Bottom left: Student's t noise. Bottom right: GM noise.



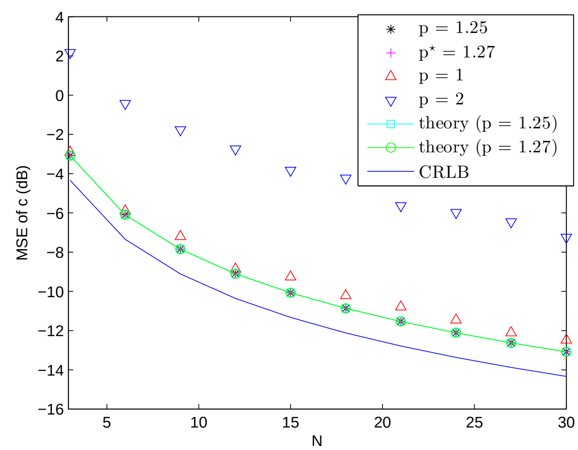
SαS



GG



Student's t



GM

Figure 3: MSE of scalar versus N. Top left: α-stable noise. Top right: GG noise. Bottom left: Student's t noise. Bottom right: GM noise.

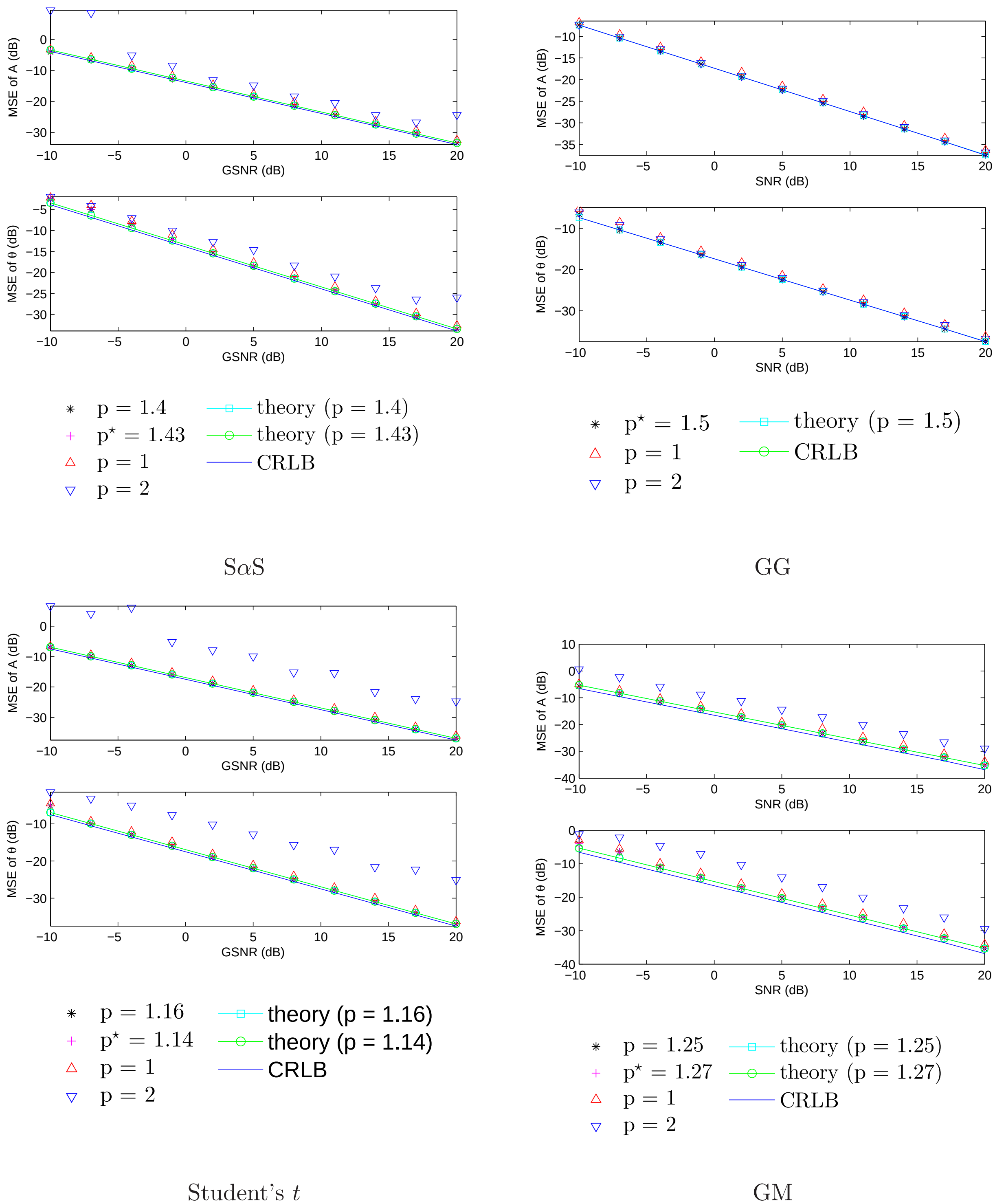


Figure 4: MSE of sinusoidal amplitude and phase versus SNR/GSNR. Top left: α-stable noise. Top right: GG noise. Bottom left: Student's t noise. Bottom right: GM noise.

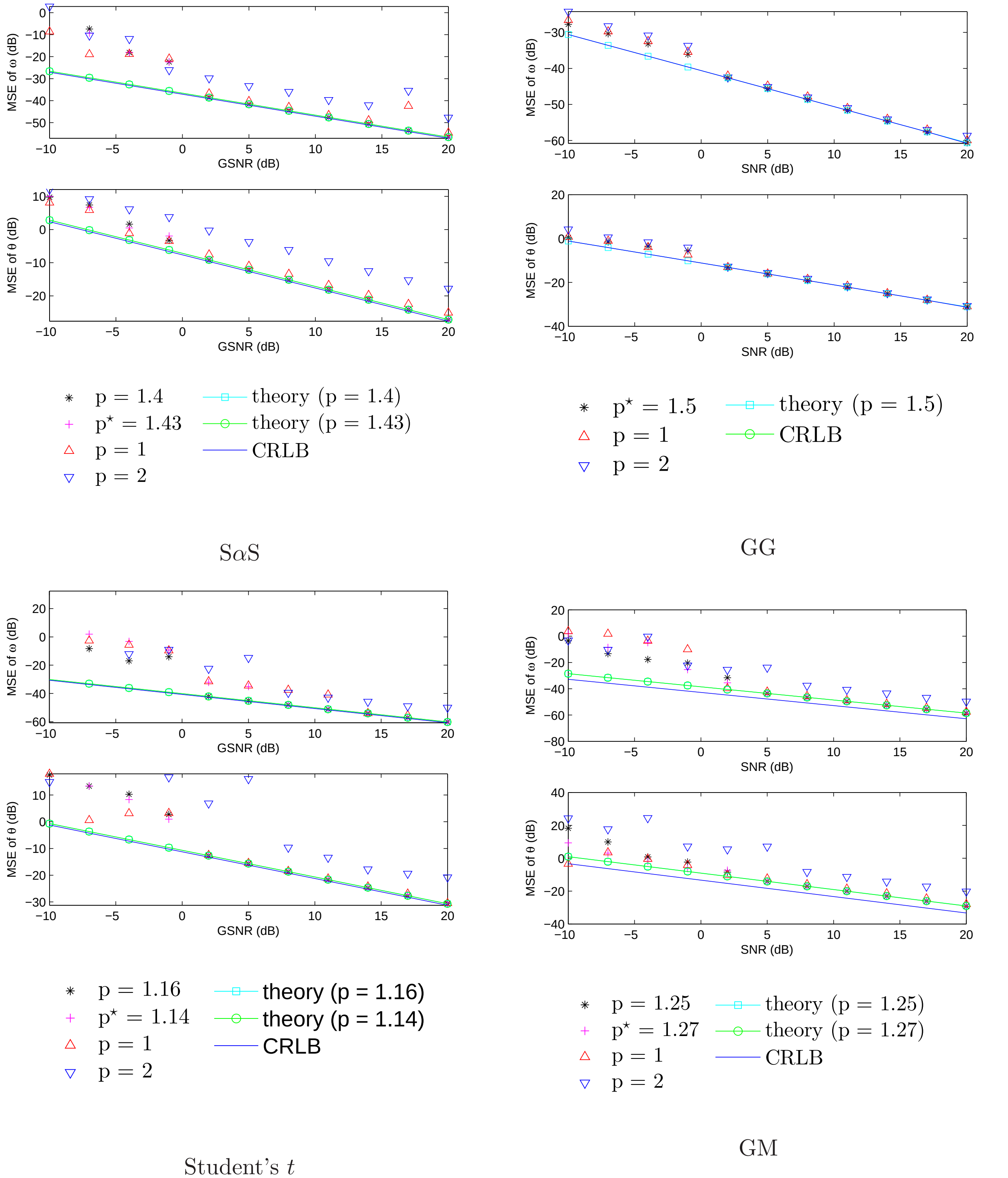


Figure 5: MSE of sinusoidal frequency and phase versus SNR/GSNR. Top left: α-stable noise. Top right: GG noise. Bottom left: Student's t noise. Bottom right: GM noise.