



**HAL**  
open science

## Noise enhancement in robust estimation of location

Yan Pan, Fabing Duan, François Chapeau-Blondeau, Derek Abbott

► **To cite this version:**

Yan Pan, Fabing Duan, François Chapeau-Blondeau, Derek Abbott. Noise enhancement in robust estimation of location. *IEEE Transactions on Signal Processing*, 2018, 66 (8), pp.1953-1966. 10.1109/TSP.2018.2802463 . hal-02170905

**HAL Id: hal-02170905**

**<https://hal.science/hal-02170905>**

Submitted on 21 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Noise Enhancement in Robust Estimation of Location

Yan Pan , Fabing Duan , François Chapeau-Blondeau , and Derek Abbott , *Fellow, IEEE*

**Abstract**—In this paper, we investigate the noise benefits to maximum likelihood type estimators (M-estimator) for the robust estimation of a location parameter. Two distinct noise benefits are shown to be accessible under these conditions. With symmetric heavy-tailed noise distributions, the asymptotic efficiency of the estimation can be enhanced by injecting extra noise into the M-estimators. With an asymmetric contaminated noise model having a convex cumulative distribution function, we demonstrate that addition of noise can reduce the maximum bias of the median estimator. These findings extend the analysis of stochastic resonance effects for noise-enhanced signal and information processing.

**Index Terms**—Noise enhancement, location, M-estimator, asymptotic efficiency, maximum bias.

## I. INTRODUCTION

AN OPTIMAL noise level, obtained by appropriately adding extra noise to a given signal processor or by tuning the existing noise level, can sometimes improve information processing [1]–[6]. This effect was initially called *stochastic resonance* [7] but is now widely referred to as *noise enhancement* [8]–[13], resulting in *noise benefit*—a term coined by Kosko [14]. In recent years, noise-enhanced signal processing theory has received significant attention [4]–[6], [8]–[33]. In the signal estimation field, classical parameter estimation problems [6], [11], [17], [22]–[30] and the Bayesian method [11]–[13], [16]–[21] are of considerable current research interest. The related results of [6], [11]–[13], [16]–[29] confirm that noise enhancement can be exploited to optimize the estimator performance.

Statistical signal processing often relies on strong assumptions, e.g., a Gaussian distribution or a particular signal model, under which optimal estimators or detectors are derived [34]–[38]. Nevertheless, many areas of engineering, for instance,

mobile communication channels, radar systems and neural networks, result in outliers or atypical observations that do not obey these assumptions [38]. This fact can lead to significantly degraded estimator performance, and also highlights the need for robust estimation techniques that are insensitive to outliers [34]–[38].

In this paper, we mainly focus on the noise-enhanced effect in maximum likelihood type estimators (M-estimator) [34]–[38] for estimating a location parameter. Under this scenario, we derive two key results that demonstrate two distinct noise benefits:

(i) For the symmetric heavy-tailed noise distributions, we derive the expression of the asymptotic efficiency for an array of identical M-estimators, and show the possibility of the improved asymptotic efficiency of M-estimators by injecting mutually independent noise components. For a single M-estimator, a discriminant function, elicited from the background noise probability density and the given M-estimator function, provides the condition of existence of the optimal additive noise. It is proven that the optimal additive noise, when it exists, is the symmetrical dichotomous noise given by Chen *et al.* [9]–[11]. When the optimal noise does not exist in a single M-estimator, it is observed that the asymptotic efficiency can still be enhanced to be very close to the upper bound of unity by an array of M-estimators with added noise selected from a parametric class of noise. This fact motivates us to focus on the asymptotic efficiency of an array of identical M-estimators with more than *one* estimator element. We further demonstrate that, for a given M-estimator function and fixed noise levels, the asymptotic efficiency is a monotonically increasing function of the M-estimator number. Thus, the asymptotic efficiency reaches its maximum value for an array with an infinite number of identical M-estimators. Aiming to maximizing the asymptotic efficiency of an infinite number of M-estimators, the optimal probability density of the added noise is proven to be the weighted minimum  $\mathcal{L}^2$ -norm solution of Eq. (44). Specially, by the Cauchy-Schwarz inequality, the upper bound of the asymptotic efficiency of an infinite number of M-estimators is proven to be unity, and the corresponding optimal noise density is the deconvolution of the maximum likelihood estimator and the given M-estimator function. However, this optimal noise density is frequently unattainable, due to absence of a solution to the deconvolution and the fact that the infinite-size array of M-estimators can only be approached in practice. Therefore, as an alternative to such a theoretical optimal distribution, we can select an *a priori* parametric class of noise distributions, of interest, and optimize the parameters and variance of the added noise. The results show that, compared to the method without added noise, the addition of extra noise that obeys the optimized parametric noise distribution does

Y. Pan and F. Duan are with the Institute of Complexity Science, Qingdao University, Qingdao 266071, China (e-mail: yanpn.pan@gmail.com; fabing.duan@gmail.com).

F. Chapeau-Blondeau is with the Laboratoire Angevin de Recherche en Ingénierie des Systèmes, Université d’Angers, Angers 49000, France (e-mail: chapeau@univ-angers.fr).

D. Abbott is with the Centre for Biomedical Engineering and School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia (e-mail: derek.abbott@adelaide.edu.au).

provide an improved asymptotic efficiency. It is also noted that the observed noise-enhanced effect for the robust estimation of a location parameter can also be related to suprathreshold stochastic resonance (SSR) that has been demonstrated in a parallel array of identical elements with the injection of independent noise components [23]–[28], [39], [40]. In the context of SSR, such nonlinear elements were chosen as threshold comparators [27], quantizers [23]–[26], [31], [40] or saturating devices [28] and have been evaluated by various performance measures such as mean square error (MSE) [23], [26], [31], [40], Fisher information [24], [27], [28], and Cramer-Rao bound [25], [26], [40] in the signal estimation field. By contrast here, we mainly focus on the noise-enhanced effect in a parallel array of robust M-estimators evaluated by the measure of the asymptotic efficiency.

(ii) In an asymmetric contaminated noise model, the bias dominates the variance component of M-estimators for a sufficiently large number of observations. We theoretically prove that, under the condition of a convex cumulative distribution function, the maximum bias of M-estimators can be reduced by adding extra noise. Moreover, aiming to minimize the maximum bias of the median estimator, the dichotomous noise is proved to be optimal. Theoretical and numerical results also show that the maximum bias can be distinctly diminished at an optimal dichotomous noise level.

## II. M-ESTIMATOR FOR A LOCATION PARAMETER

Consider a location model of observations [34], [37]

$$x_i = \theta + w_i, \quad i = 1, 2, \dots, n \quad (1)$$

where the location parameter  $\theta$  is unknown and observation errors (noise)  $w_i$  are independent and identically distributed (i.i.d.) random variables with a common probability density  $f_w$ . Then, the observations  $x_i$  have the distribution  $f_w(x - \theta)$ . Here, we assume  $w = \sigma w_0$  is with the scale parameter  $\sigma$  that represents the noise intensity, and  $w_0$  has its probability density  $f_{w_0}$  and unity scale.

Assuming that the loss function  $\rho$  satisfies certain regularities [34], [37], we define the M-estimator  $\hat{\theta}$  as

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \rho(x_i - \theta). \quad (2)$$

Differentiating Eq. (2) with respect to  $\theta$  yields

$$\sum_{i=1}^n \psi(x_i - \theta) = 0, \quad (3)$$

where  $\psi = d\rho/d\theta = \rho'$  and the root  $\hat{\theta}$  that corresponds to the global minimum of the loss function in Eq. (2) is just the M-estimator or the generalized maximum likelihood estimator [34]–[38]. The maximum likelihood estimator is a special case of M-estimators when  $\psi = -f'_w/f_w \triangleq \psi_M$  and  $\rho = -\log f_w$  [34]–[38].

In order to compare the approximation by  $\hat{\theta}$  with the true value of  $\theta$ , we usually compute the MSE

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \text{var}(\hat{\theta}) + \text{b}^2(\hat{\theta}), \quad (4)$$

with the estimator variance  $\text{var}(\hat{\theta}) = E(\hat{\theta}^2) - E^2(\hat{\theta})$  and the bias  $\text{b}(\hat{\theta}) = E(\hat{\theta}) - \theta$  [34]–[38]. Here, the expectation operator  $E(\cdot) = \int \cdot f_w(x - \theta) dx$  represents the expectation according to the shifted probability density  $f_w(x - \theta)$  that is the density for the noisy random signal  $x_i$  in Eq. (1). According to the central limit theorem and for a sufficiently large observation number  $n$ , a Fisher-consistent M-estimator  $\hat{\theta}$  satisfies  $E[\psi(x - \theta)] = 0$  [34]–[38], and asymptotically converges to the Gaussian distribution with mean  $\theta$  and variance

$$\text{var}(\hat{\theta}) \approx \frac{1}{n} \frac{E[\psi^2(x - \theta)]}{E^2[\psi'(x - \theta)]} = \frac{1}{n} \frac{E_w[\psi^2(x)]}{E_w^2[\psi'(x)]}, \quad (5)$$

by taking the first-order Taylor expansion of  $\psi(x_i - \hat{\theta})$  at  $\theta$  [34]–[38] (also see Appendix B). Here, the expectation operator  $E_w(\cdot) = \int \cdot f_w(x) dx$  is according to the probability density of the random noise  $w_i$ . Also in Eq. (5) the change of variable  $x - \theta \rightarrow x$  has been performed. It is noted that the minimum asymptotic variance of  $\text{var}(\hat{\theta}) = 1/(nJ(\theta))$  is achieved by the maximum likelihood estimator [37], [41], and the Fisher information  $J(\theta)$  with respect to the location parameter  $\theta$  is defined as

$$\begin{aligned} J(\theta) &= E \left[ \frac{(df_w(x - \theta)/d\theta)^2}{f_w^2(x - \theta)} \right] \\ &= E_w \left[ \frac{(df_w(x)/dx)^2}{f_w^2(x)} \right] \\ &= J(f_w), \end{aligned} \quad (6)$$

where  $J(f_w)$  is the Fisher information of the probability density  $f_w$  [41]. Therefore, in order to measure how near the M-estimator  $\hat{\theta}$  is to the optimal estimator, the asymptotic efficiency is defined as the ratio [37]

$$\text{Eff}(\hat{\theta}) = \frac{1}{nJ(f_w)} \frac{1}{\text{var}(\hat{\theta})} = \frac{1}{J(f_w)} \frac{E_w^2[\psi'(x)]}{E_w[\psi^2(x)]}. \quad (7)$$

In the following section, we will use the asymptotic efficiency of Eq. (7) and the bias  $\text{b}(\hat{\theta})$  to assess the occurrence of noise-enhanced effects in M-estimators associated with two noise distribution models.

## III. NOISE-ENHANCED ASYMPTOTIC EFFICIENCY

In a number of areas of engineering, outliers in the measurement cause a heavy-tailed distribution and so a Gaussian assumption no longer holds true [34]–[38]. In this Section, we consider the symmetric heavy-tailed distribution model of noise. Under such scenarios, the function  $\psi$  in Eq. (3) is assumed to be odd, so that there is no bias problem for the unbiased M-estimator  $\hat{\theta}$  derived from Eq. (3). Thus, we mainly discuss the noise-enhanced asymptotic efficiency effect in this situation under consideration.

When the expression for the noise distribution  $f_w$  is known, the maximum likelihood estimator is the optimal option if it exists [34]–[38]. However, in practice, the noise distribution  $f_w$  may be incompletely known, or a complete closed-form of the maximum likelihood estimator is too complex to implement [34]–[38]. Thus, other M-estimators with an easily implemented

feature are frequently employed [34]–[38]. We note that M-estimators are critically connected the  $\psi$ -function in Eq. (3), which can be grouped into the monotone or the redescending classes [37]. Here, we mainly consider the monotone Huber function [34]

$$\psi(x) = \begin{cases} x, & |x| \leq \gamma, \\ \gamma \operatorname{sgn}(x), & |x| > \gamma, \end{cases} \quad (8)$$

and the redescending bisquare function [34], [37]

$$\psi(x) = x [1 - (x/\gamma)^2]^2, \quad (9)$$

for  $|x| \leq \gamma$  and otherwise zero. Here,  $\gamma$  is an adjustable parameter of the M-estimator. We further consider a scale family of the probability density  $f_w(x) = f_{w_0}(x/\sigma)/\sigma$ , where  $f_{w_0}$  has unity scale and the Fisher information of Eq. (6) can be expressed as  $J(f_w) = J(f_{w_0})/\sigma^2$  [41], [42]. Here,  $J(f_{w_0})$  is the Fisher information of the probability density  $f_{w_0}$ . Then, the asymptotic efficiency of Eq. (7) can be calculated as

$$\operatorname{Eff}(\hat{\theta}) = \frac{[1 - 2F_{w_0}(-\frac{\gamma}{\sigma})]^2}{2J(f_{w_0}) \left[ \int_0^{\gamma/\sigma} x^2 f_{w_0}(x) dx + \frac{\gamma^2}{\sigma^2} F_{w_0}(-\frac{\gamma}{\sigma}) \right]} \quad (10)$$

for the Huber estimator of Eq. (8) and

$$\operatorname{Eff}(\hat{\theta}) = \frac{\left[ \int_{-\gamma/\sigma}^{\gamma/\sigma} \left[ 1 - \frac{\sigma^2}{\gamma^2} x^2 \right] \left[ 1 - \frac{5\sigma^2}{\gamma^2} x^2 \right] f_{w_0}(x) dx \right]^2}{J(f_{w_0}) \int_{-\gamma/\sigma}^{\gamma/\sigma} x^2 \left[ 1 - \frac{\sigma^2}{\gamma^2} x^2 \right]^4 f_{w_0}(x) dx} \quad (11)$$

for the bisquare estimator of Eq. (9). Here,  $F_{w_0}$  is the cumulative distribution function of noise  $w_0$ .

For instance, we especially consider the generalized Cauchy noise with the probability density

$$f_w(x) = B(k, v) \left\{ 1 + \frac{1}{v} \left[ \frac{|x|}{A(k)} \right]^k \right\}^{-(v+1/k)} \quad (12)$$

with  $A(k) = \sigma \sqrt{\Gamma(1/k)/\Gamma(3/k)}$ ,  $B(k, v) = kv^{-1/k} \Gamma(v + 1/k) / [2A(k)\Gamma(v)\Gamma(1/k)]$ , and parameters  $\sigma, k, v > 0$  [36], [42]. This generalized Cauchy distribution of Eq. (12) is capable of modelling a wide range of practical noise types. For instance, we obtain a typical heavy-tailed distribution model for  $v = 1/2$  and  $k = 2$ , which is the Cauchy noise [42].

In Fig. 1(a) and (b), we plot the asymptotic efficiencies of Eq. (10) and Eq. (11) as a function of  $\gamma/\sigma$ , respectively. It is clearly seen that, for the generalized Cauchy noise in Eq. (12) with  $v = 0.5, 1, 2, 5$  and  $k = 2$ , the asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  can be enhanced to its maximum as  $\gamma/\sigma$  increases. For instance, in the case of Cauchy noise with  $v = 1/2$  and  $k = 2$  in Eq. (12), the asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  of Huber estimator of Eq. (10) is maximized to the value of 0.88 at  $\gamma/\sigma = 0.39$  and for the bisquare estimator, the asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  of Eq. (11) has the maximum value of 0.90 as  $\gamma/\sigma = 3.28$ . For a given parameter  $\gamma$ , this is the noise-enhanced effect occurring in the location M-estimator, since the asymptotic efficiency is maximized at an optimal noise scale. Of course, if the noise scale  $\sigma$  is given, we can also optimize the estimator parameter  $\gamma$  to obtain the maximum asymptotic efficiency.

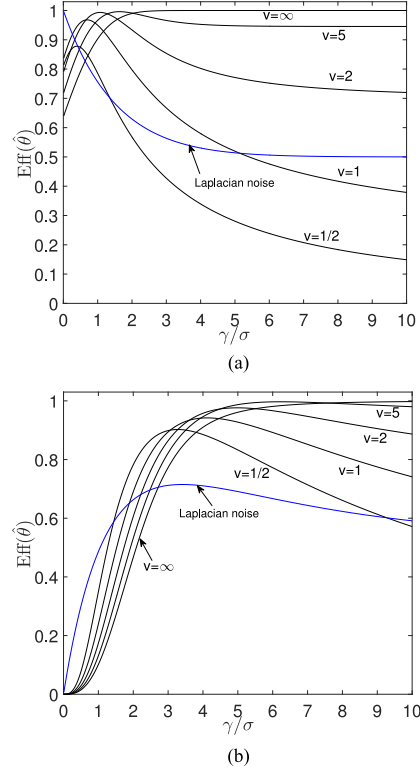


Fig. 1. Asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  of (a) Eq. (10) for the Huber estimator and (b) Eq. (11) for the bisquare estimator as a function of  $\gamma/\sigma$  for the generalized Cauchy noise in Eq. (12) and the Laplacian noise in Eq. (18) ( $\beta = 1$ ). Here, the generalized Cauchy noise is with parameters  $v = 1/2, 1, 2, 5, \infty$  and  $k = 2$ .

It is also noted that, for the Gaussian noise ( $v = \infty$  and  $k = 2$  in Eq. (12)), the asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  approaches unity as  $\gamma/\sigma$  increases, as shown in Fig. 1(a) and (b). The reason is that, as the parameter  $\gamma \rightarrow \infty$ , both Huber and the bisquare functions are linear, which just corresponds to the optimal estimator for the Gaussian noise. For the Laplacian noise in Eq. (18) ( $\beta = 1$ ), Fig. 1(a) shows the asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  monotonically decreases from unity as  $\gamma/\sigma$  increases. The reason is that, as  $\gamma \rightarrow 0$ , the Huber function of Eq. (8) becomes  $\psi(x) = \operatorname{sgn}(x)$  that corresponds to the maximum likelihood estimator for the Laplacian noise, i.e., the median estimator [34]–[38].

Next, two interesting problems arise. The first problem is, when the noise scale  $\sigma$  can be exactly estimated or is known and the asymptotic efficiency is maximized at an optimal parameter  $\gamma_{\text{opt}}$ , whether the addition of extra noise is beneficial or not? The second problem arises because the noise scale  $\sigma$  is often unknown, while the optimization of the asymptotic efficiency  $\operatorname{Eff}(\hat{\theta})$  is closely tied with the ratio  $\gamma/\sigma$ , as shown in Fig. 1. Thus, the M-estimators of Eqs. (8) and (9) often estimate the location away of the maximum asymptotic efficiency obtained by the optimal parameter  $\gamma_{\text{opt}}$ . Under this circumstance, can we add extra noise to the given observations and obtain an improved performance of a robust M-estimator with a fixed parameter  $\gamma$ ? In the following discussion, the answers to these two interesting problems will be presented.

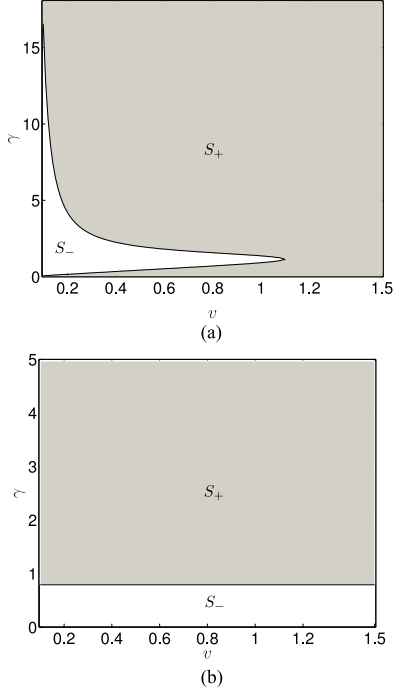


Fig. 2. Region  $S_+$  (gray) of the positive signs of  $D(\psi, f_w)$  in Eq. (46) where addition of noise is beneficial, and the complementary region  $S_-$  of the negative signs of  $D(\psi, f_w)$ , versus the parameter  $v$  of the generalized Cauchy distribution  $f_w$  in Eq. (12) ( $k = 2$ ) and the parameter  $\gamma$  of (a) the Huber and (b) the bisquare function. Here, the existing noise scale  $\sigma = 1$  and the added uniform noise level  $d \rightarrow 0$ .

We purposefully add the extra noise to the original observations

$$y_i = x_i + \eta_i = \theta + z_i, \quad i = 1, 2, \dots, n \quad (13)$$

where the composite noise components  $z_i = w_i + \eta_i$  and the extra noise components  $\eta_i$  are the i.i.d. random variables with the common probability density  $f_\eta$ . Therefore,  $z_i$  have a convolved probability density  $f_z(z) = \int f_w(z - \eta) f_\eta(\eta) d\eta$ . It is proven by the Fisher information convolution inequality in Appendix A that the addition of extra noise to the original observations cannot improve the performance of the maximum likelihood estimator.

Furthermore, inspired by the suprathreshold stochastic resonance model [39], [40], we collect  $L$  statistics  $\hat{\theta}_l$  generated by identical M-estimators for  $l = 1, 2, \dots, L$ , and average them as the M-estimation of the location parameter in the context of noise benefits. It is emphasized that each M-estimator is subject to the same input  $x_i$  but mutually independent noise components  $\eta_{li}$  with the common probability density  $f_\eta$ . Therefore, the input of each M-estimator is

$$y_{li} = \theta + w_i + \eta_{li} = \theta + z_{li}, \quad (14)$$

where the composite noise  $z_{li} = w_i + \eta_{li}$  have the probability density  $f_z$ . The estimate  $\hat{\theta}_l$  satisfies  $\sum_{i=1}^n \psi(y_{li} - \hat{\theta}_l) = 0$  of Eq. (3), and the devised M-estimator is given by

$$\hat{\theta} = \frac{1}{L} \sum_{l=1}^L \hat{\theta}_l. \quad (15)$$

Due to the independent characteristic of noise  $\eta_{li}$  and using the Taylor expansion of  $\hat{\theta}_l$ , the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of  $\hat{\theta}$  in Eq. (15) is theoretically derived in Appendix B as

$$\text{Eff}_a(\hat{\theta}) = \frac{1}{J(f_w)} \frac{L \text{E}_z[\psi'(z)]}{\text{E}_z[\psi^2(z)] + (L-1) \text{E}_w\{\text{E}_\eta^2[\psi(w+\eta)]\}}, \quad (16)$$

with the expectations  $\text{E}_z(\cdot) = \int \cdot f_z(x) dx$  and  $\text{E}_\eta(\cdot) = \int \cdot f_\eta(x) dx$ .

For a single M-estimator ( $L = 1$ ), the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) reduces to Eq. (32), and the corresponding optimal noise probability density  $f_\eta^{\text{opt}}$  that maximizes the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) is discussed in Appendix C. This optimization problem is similar to the issue of finding the optimal additive noise addressed by Chen *et al.* [9]–[11] based on Carathéodory's theorem. It is proven [11] that, for a general estimation parameter, the optimal noise, if one exists, can be chosen as a suitable randomization of no more than two constant vectors to reduce of the estimation variance without increasing the bias of the estimator. In Appendix C, for a location parameter under the Fisher-consistent constraint  $\text{E}_z[\psi(z)] = 0$ , we provide the existence condition of the optimal noise implied by the function  $g(\eta)$  in Eq. (34b) and prove the symmetrical dichotomous noise in Eq. (25) to be the optimal one on the basis of a convex real-value vector function and the Jensen's inequality [17], [43], [44]. The demonstration of Appendix C is in particular complementary to the approach of Ref. [11] for location estimation.

However, the derived optimal additive noise may not exist for a single M-estimator in some situations where the function of  $g(\eta)$  in Eq. (34b) has its maximum at  $\eta = 0$ , e.g., the Cauchy background noise  $f_w$  of Eq. (12) and the Huber M-estimator of Eq. (8). It is interesting to note in Fig. 3 that, even if the added noise is not optimal, the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  can be enhanced by injecting extra noise components into the M-estimator of Eq. (15) with the estimator number  $L > 1$ . This noise-enhanced effect motivates us to further consider the situation of  $L > 1$ .

Furthermore, for a given M-estimator function  $\psi$  and at a given extra noise level, we demonstrate in Appendix D that the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  in Eq. (16) is a monotonically increasing function of the estimator number  $L$ , as illustratively shown in Fig. 3. Here, the additive noise distribution  $f_\eta$  is arbitrary. We recognize that, due to the nonlinear functional term  $\text{E}_w\{\text{E}_\eta^2[\psi(w+\eta)]\}$  in Eq. (16), the theoretical determination of the optimal probability density that maximizes the asymptotic efficiency of Eq. (16) remains an open question for a finite M-estimator number  $1 < L < \infty$ . In the limiting case of  $L \rightarrow \infty$ , the asymptotic efficiency of Eq. (16) attains its maximum with respect to  $L$  and asymptotically approaches to

$$\text{Eff}_\infty(\hat{\theta}) = \lim_{L \rightarrow \infty} \text{Eff}_a(\hat{\theta}) = \frac{1}{J(f_w)} \frac{\text{E}_z[\psi'(z)]}{\text{E}_w\{\text{E}_\eta^2[\psi(w+\eta)]\}}, \quad (17)$$

where  $\text{E}_z[\psi^2(z)] < \infty$  and  $\lim_{L \rightarrow \infty} \text{E}_z[\psi^2(z)]/L \rightarrow 0$ . In Appendix E, aiming to maximize the asymptotic efficiency of

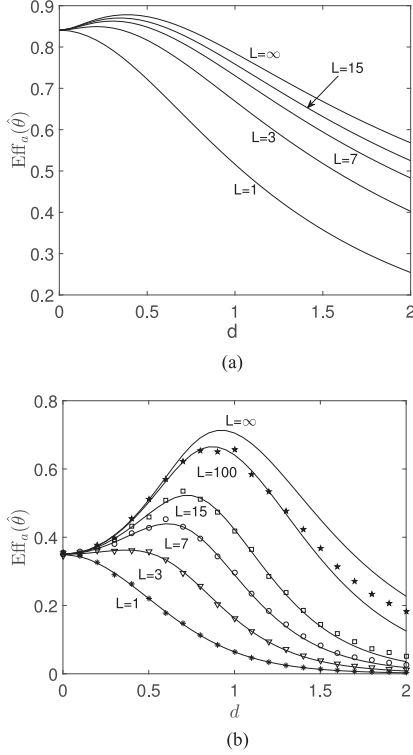


Fig. 3. Asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) for an array of (a) Huber estimators with  $\gamma = 0.1$  and (b) bisquare estimators with  $\gamma = 1$  as a function of uniform noise level  $d$  and the estimator number  $L$ . Here, the existing Cauchy noise is with the fixed scale  $\sigma = 1$ . Using the iterative reweighting algorithm (see Appendix G-A), the numerical results computed by  $10^4$  Monte Carlo trails are plotted in (b) for the M-estimator number  $L = 1$  (\*),  $L = 3$  ( $\nabla$ ),  $L = 7$  ( $\circ$ ),  $L = 15$  ( $\square$ ) and  $L = 100$  ( $\times$ ). Here, the observation size takes  $n = 7000$ , the number of simulation trails is  $10^4$  and the tolerance parameter  $\zeta = 10^{-6}$ .

Eq. (17), we find the optimal noise density  $f_\eta^{\text{opt}}$  is the solution of the weighted minimum  $\mathcal{L}^2$ -norm of Eq. (44), and the asymptotic efficiency  $\text{Eff}_\infty(\hat{\theta})$  is theoretically proven to be not more than unity by the Cauchy-Schwarz inequality. Then, for achieving the upper bound of  $\text{Eff}_\infty(\hat{\theta}) = 1$ , it is demonstrated that the optimal noise probability density  $f_\eta^{\text{opt}}$  of Eq. (42) is the deconvolution of  $\psi_M(w)$  and  $\psi(x)$ . Here, the condition of existence of this optimal probability density  $f_\eta^{\text{opt}}$  is that the inverse of the Fourier transform in Eq. (42) must exist and be nonnegative. In Appendix E, an achievable optimum as a logistic noise is worked out as an example. Unfortunately, this kind of optimal noise often does not exist. For instance, there is no solution of Eq. (42) for the Huber function in Eq. (8) (or the bisquare function in Eq. (9)) and the Cauchy noise distribution  $f_w$  in Eq. (18) ( $v = 1/2$  and  $k = 2$ ). In addition, the infinite M-estimator number  $L = \infty$  is also inaccessible and can only be approached in practice.

As an alternative, some prior interesting parametric classes of noise distributions can be selected and optimized to improve the asymptotic efficiency of M-estimators. For example, consider the generalized Gaussian noise  $\eta_i$  with its probability density

$$f_\eta(x) = \frac{c_1(\beta)}{\sigma_\eta} \exp\left(-c_2(\beta) \left|\frac{x}{\sigma_\eta}\right|^{2/(1+\beta)}\right), \quad (18)$$

where  $c_1(\beta) = \Gamma^{1/2}(3(1+\beta)/2)/(1+\beta)\Gamma^{3/2}((1+\beta)/2)$ ,  $c_2(\beta) = [\Gamma(3(1+\beta)/2)/\Gamma((1+\beta)/2)]^{1/(1+\beta)}$ , the decay exponent  $\beta \geq -1$  and the scale parameter  $\sigma_\eta > 0$  [45]. This generalized Gaussian noise model of Eq. (18) describes many noise types encountered in real-world systems, ranging from Gaussian noise ( $\beta = 0$ ), Laplacian noise ( $\beta = 1$ ) to uniform noise ( $\beta = -1$ ) [42], [45]. For  $\beta = -1$  in Eq. (18), the probability density of uniform noise can be expressed as  $f_\eta(x) = 1/(2d)$  for  $-d \leq x \leq d$  ( $d = \sqrt{3}\sigma_\eta$ ) and zero otherwise.

The theoretical asymptotic efficiency of Eq. (17) is then applicable to evaluate the benefit of added noise for various background noise types. For instance, with the addition of uniform noise, the asymptotic efficiency in Eq. (17) can be expressed as

$$\begin{aligned} \text{Eff}_\infty(\hat{\theta}, d) &= \frac{1}{J(f_w)} \frac{E_w^2\left(\int_{-d}^d \frac{1}{2d} \psi'(w+\eta) d\eta\right)}{E_w\left[\left(\int_{-d}^d \frac{1}{2d} \psi(w+\eta) d\eta\right)^2\right]} \\ &= \frac{1}{J(f_w)} \frac{E_w^2[\psi(w+d) - \psi(w-d)]}{E_w\{[\rho(w+d) - \rho(w-d)]^2\}}. \end{aligned} \quad (19)$$

In Appendix F, we assume the uniform noise level  $d \rightarrow 0$ , and derive the condition of Eq. (46) to determine whether the addition of uniform noise is beneficial or not. For the generalized Cauchy noise in Eq. (12) ( $k = 2$ ) with the fixed scale  $\sigma = 1$ , the signs of  $D(\psi, f_w)$  in Eq. (46) versus the parameter  $\gamma$  of the  $\psi$  function and the parameter  $v$  of the generalized Cauchy distribution  $f_w$  in Eq. (12) are shown in Fig. 2(a) and (b). It is seen in Fig. 2 that, using the discriminant inequality of Eq. (46), the regions  $S_+$  of the positive signs of  $D(\psi, f_w)$  clearly manifest that the benefit of adding uniform noise exists for substantial ranges of the parameter pair  $(\gamma, v)$ . In addition, the benefit of added uniform noise occurs for various noise distributions characterised by the parameter  $v$  of the generalized Cauchy distribution  $f_w$  in Eq. (12).

The discriminant inequality of Eq. (46) only theoretically demonstrates the noise benefit for a very small noise level and an infinite M-estimator number. Furthermore, for the background Cauchy noise and the estimator parameter  $\gamma$  indicated in Fig. 2 that elicits the benefit of adding noise, the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  as a function of the uniform noise level  $d$  is shown in Fig. 3 for different M-estimator numbers. Here, the Huber function in Fig. 3(a) has a given estimator parameter  $\gamma = 0.1$ , and the bisquare function in Fig. 3(b) is with  $\gamma = 1$ . The fixed scale of the Cauchy noise is  $\sigma = 1$ . It is seen in Fig. 3 that, for a single M-estimator ( $L = 1$ ), the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  monotonically decreases with the increase of noise level  $d$ . While, for the estimator number  $L > 1$ , the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  can be improved to a maximum as the noise level  $d$  increases. This kind of noise enhancement is distinct from the beneficial effect of noise in a single processor in that it critically relies on the collective role of the estimator number  $L$  [39]. Then, the noise-enhanced effect shown in Fig. 3 can be regarded as a form of SSR [39] for location estimation in essence.

We also numerically demonstrate the noise-enhanced asymptotic efficiency effect for the bisquare M-estimator by the iterative reweighting method [37]. The procedure of the

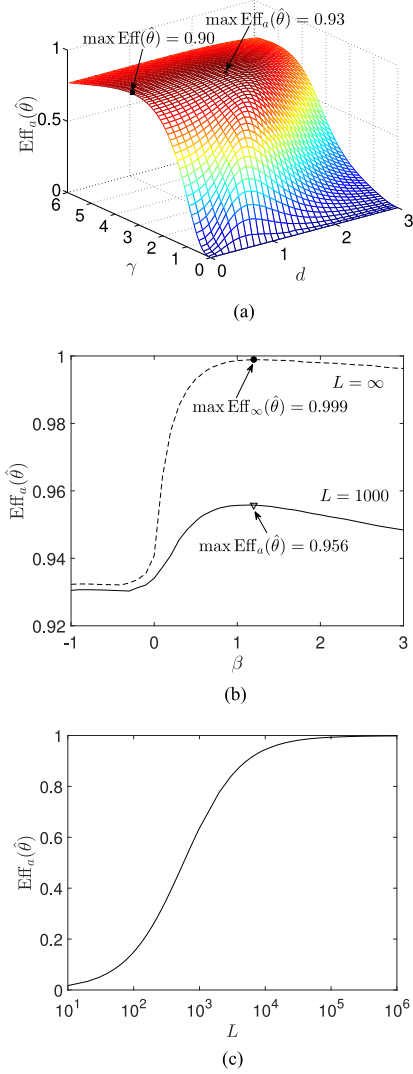


Fig. 4. (a) Asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) for an array of bisquare estimators versus the estimator parameter  $\gamma$  and uniform noise level  $d$ . Here, the M-estimator number  $L = 1000$  and the background Cauchy noise scale  $\sigma = 1$ ; (b) Maximum asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) as a function of the decay exponents  $\beta$  of generalized Gaussian distributions via the optimized parameter pair of  $(\sigma_\eta, \gamma)$  for  $L = 1000$  (solid line) and  $L = \infty$  (dashed line). (c) Asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  as a monotonically increasing function of the M-estimator number  $L$  with the parameters  $\gamma_{\text{opt}}, \sigma_\eta^{\text{opt}}$  and  $\beta$  that correspond to the maximum asymptotic efficiency  $\text{Eff}_\infty(\hat{\theta}) = 0.999$  (●) in Fig. 4(b).

numerical computation is given in Appendix G-A. In Fig. 3(b), the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  is numerically evaluated for  $10^4$  simulation trails and the tolerance parameter  $\zeta = 10^{-6}$ . Here, the observation size  $n = 7000$  is large enough but finite. The numerical results of the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  are plotted in Fig. 3(b) for the M-estimator numbers  $L = 1$  (\*),  $L = 3$  (▽),  $L = 7$  (○),  $L = 15$  (□) and  $L = 100$  (★). It is shown in Fig. 3(b) that the numerical results of  $\text{Eff}_a(\hat{\theta})$  obtained by this method are well consistent with the theoretical values of  $\text{Eff}_a(\hat{\theta})$  in Eq. (16).

Now we return to the question of whether, for an optimized M-estimator with its optimal parameter  $\gamma_{\text{opt}}$ , the addition of extra noise is beneficial or not. For a fixed input Cauchy noise

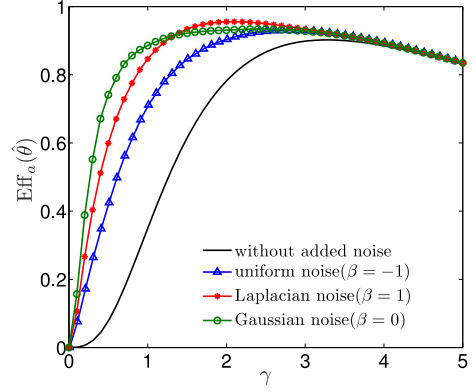


Fig. 5. Asymptotic efficiency  $\text{Eff}(\hat{\theta})$  of Eq. (11) without added noise (solid line), and the maximum asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) versus the estimator parameter  $\gamma$  for Gaussian (○), Laplacian (\*) and uniform (△) types of the added noise from Eq. (18). For each point of the curves, the level  $\sigma_\eta$  of added noise is optimized to achieve the maximum asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$ . Here, the background Cauchy noise scale  $\sigma = 1$  and the M-estimator number  $L = 1000$ .

scale  $\sigma = 1$  and the M-estimator number  $L = 1000$ , we plot the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of the bisquare M-estimator array as a function of both the estimator parameter  $\gamma$  and the added noise level  $d$ , as shown in Fig. 4(a). It is found in Fig. 4(a) that the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) attains its maximum value of 0.93 (\*) at the optimized parameter pair of  $(d_{\text{opt}}, \gamma_{\text{opt}}) = (1.3, 2.8)$ . Without the help of added noise ( $d = 0$ ), the maximum  $\text{Eff}(\hat{\theta})$  is 0.90 (■) with the optimal parameter  $\gamma_{\text{opt}} = 3.28$ , as shown in Fig. 4(a). The maximum asymptotic efficiency  $\text{Eff}_a(\hat{\theta}) = 0.93$  via optimizing the parameter pair  $(\gamma, d)$  presents a slight improvement over the maximum  $\text{Eff}(\hat{\theta})$  of 0.90 without the help of added noise.

Furthermore, we can consider the possibility of adding extra noise from Eq. (18) to the M-estimator. For different decay exponents  $\beta$  of the generalized Gaussian distributions of Eq. (18), Fig. 4(b) shows the maximum asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  in Eq. (16) via optimizing the parameter pair  $(\gamma, \sigma_\eta)$ . It is interesting to note in Fig. 4(b) that, for a finite M-estimator number  $L = 1000$ , the maximum asymptotic efficiency of Eq. (16) can reach 0.956 (▽) for the decay exponent  $\beta = 1$  (Laplacian noise). More interestingly, based on Eq. (17) and for  $L = \infty$ , Fig. 4(b) also shows the the maximum asymptotic efficiency  $\text{Eff}_\infty(\hat{\theta})$  can achieve as high as 0.999 (●) that is very close to unity. Here, the decay exponent  $\beta = 1.2$  and the optimal parameter pair  $(\gamma_{\text{opt}}, \sigma_\eta^{\text{opt}}) = (1.4, 8.1)$ . Of course, this limit value of  $\text{Eff}_\infty(\hat{\theta}) = 0.999$  can only be approached in practice. Under this circumstance, Fig. 4(c) illustrates the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  as a monotonically increasing function of the M-estimator number  $L$  with the parameters  $\gamma_{\text{opt}}, \sigma_\eta^{\text{opt}}$  and  $\beta$  that correspond to the maximum asymptotic efficiency  $\text{Eff}_\infty(\hat{\theta}) = 0.999$ . It is shown in Fig. 4(c) that the asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  can gradually approach the limit value of  $\text{Eff}_\infty(\hat{\theta}) = 0.999$ , but at the cost of the very heavy computation complexity induced by the large number  $L$ .

Finally, we answer the second question of the possibility of a noise benefit to an array of robust M-estimators with a

fixed parameter  $\gamma$ . Fig. 5 shows the asymptotic efficiency  $\text{Eff}(\hat{\theta})$  (solid line) of Eq. (11) without added noise. Furthermore, for each given parameter  $\gamma$  and the estimator number  $L = 1000$ , Fig. 5 also depicts the maximum asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) for Gaussian ( $\circ$ ), Laplacian ( $*$ ) and uniform ( $\triangle$ ) types of the added noise via optimizing the noise level  $\sigma_\eta$  ( $d = \sqrt{3}\sigma_\eta$  for uniform noise). From Fig. 5, it is visible that the asymptotic efficiencies  $\text{Eff}_a(\hat{\theta})$  obtained by optimizing the level of added noise are always superior to the efficiency with no added noise (solid line). These results show that, for a fixed parameter  $\gamma$ , the addition of extra noise is effectively beneficial to the estimation process.

#### IV. ADDING NOISE TO REDUCE THE MAXIMUM BIAS

In Section III, we discussed the noise-enhanced asymptotic efficiency in symmetric heavy-tailed noise models. In this Section, we further consider the case of  $\epsilon$ -contaminated noise model  $h$  ranging in the  $\epsilon$ -neighborhood [34], [37]

$$h_\epsilon(x, \theta) = \{h \mid h = (1 - \epsilon)f_w(x, \theta) + \epsilon q(x), q \in \mathcal{Q}\}, \quad (20)$$

of an assumed parametric distribution  $f_w(x, \theta)$ , where  $q$  that represents the probability model of outliers belongs to a family of probability density functions  $\mathcal{Q}$  and  $0 \leq \epsilon < 1$  indicates the proportion of the observations generated by the contamination model of  $q$  [34], [37]. Here, the contamination model of  $q$  is usually not exactly known, and depicts the statistical characteristics of the outliers produced by some unknown mechanisms [34], [36]–[38]. In order to analyze some extreme contaminated cases,  $q$  is chosen from the set of point mass distribution  $\delta(x_0)$  such that the probability of  $x = x_0$  is unity [34], [37]. Here,  $\delta$  is the Dirac function. It has been pointed out [34], [37] that, for the location M-estimation in neighborhoods of Eq. (20) of a symmetric  $f_w$  with a symmetric probability density  $q$  of contamination, it is easy to derive an unbiased M-estimator  $\hat{\theta}$  with  $E(\hat{\theta}) = \theta$  and we still care about the asymptotic efficiency of M-estimators. Thus, in this Section, we mainly focus on the asymmetric contaminated noise model of  $h_\epsilon(x, \theta)$ . Under this condition, the bias of the estimators to the true value becomes the main factor of MSE as a measurement of estimators rather than the variance for a sufficiently large observation size  $n$ . This is because the estimation variance  $\text{var}(\hat{\theta})$  of Eq. (4) tends to zero with the order of  $n^{-1}$ , while the bias does not [34]–[38].

For the asymmetric contaminated noise model of  $h_\epsilon(x, \theta)$  of Eq. (20), there are two cases: either a symmetric central probability density  $f_w$  but an asymmetric point mass distribution  $q$ , or an asymmetric  $f_w$  plus an asymmetric (or symmetric)  $q$ . In the case of the noise components  $w_i$  with the asymmetric probability density  $f_w$ , we can add the i.i.d. noise components  $w'_i$  with the probability density of  $-w_i$ , i.e.,  $f_{w'}(x) = f_w(-x)$ , to the original noise components  $w_i$ . Here,  $w'_i$  are independent with  $w_i$ . Then, the updated random variables  $\tilde{w}_i = w_i + w'_i$  have a symmetric probability density  $f_{\tilde{w}}(x) = \int f_w(x - u)f_{w'}(u)du = \int f_{w'}(-x - u)f_w(u)du = f_{\tilde{w}}(-x)$ . Therefore, we here only concentrate on the maximum bias induced by the contaminated noise model of  $h_\epsilon(x, \theta)$  with a symmetric  $f_w$  but an asymmetric  $q$  in the following parts.

Consider the estimate  $\hat{\theta}_n$  depending on the observation data  $(x_1, x_2, \dots, x_n)$  and for a location M-estimator with the monotonic  $\psi$  in Eq. (3), the asymptotic value  $\hat{\theta}_\infty = \theta$  under  $f_w$ , i.e., the solution of  $E[\psi(x - \theta)] = 0$  [34], [37]. Then, the bias for an arbitrary distribution  $h$  is  $b_{\hat{\theta}} = \hat{\theta}_\infty(h) - \theta$ . For a given proportion  $\epsilon$  and the assumed parametric probability density  $f_w$ , the maximum bias is defined as [34], [37]

$$b_\epsilon = \max\{|b_{\hat{\theta}}(h, \theta)| : h \in h_\epsilon\}, \quad (21)$$

over an arbitrary contamination model of  $q \in \mathcal{Q}$ .

Let  $\psi$  be a nondecreasing function, instead of a redescending function, with a bound  $s = \psi(\infty)$ , the fraction of contamination  $0 < \epsilon < 1/2$  and  $\theta = 0$  (without loss of generality). It can be shown [34], [37] that the maximum bias  $b_\epsilon$  of a location M-estimator is the solution of

$$E_w[\psi(x + b)] = \frac{s\epsilon}{1 - \epsilon}, \quad (22)$$

which is attained in the case of the point mass distribution  $q = \delta(x_0)$  in Eq. (20) with  $x_0 \rightarrow \pm\infty$ .

The median estimator, as a special Huber estimator in Eq. (8) with its parameter  $\gamma \rightarrow 0$ , has the minimax bias property for a symmetric unimodal distribution  $f_w$  of noise [34], [37], [38]. Since the median estimator with  $\psi(x) = \text{sgn}(x)$  is bounded by  $s = 1$ , the maximum bias  $b_\epsilon$  in Eq. (22) can be calculated as [34], [37]

$$b_\epsilon = F_w^{-1}[1/2(1 - \epsilon)], \quad (23)$$

where  $F_w^{-1}$  is the inverse function of the cumulative distribution  $F_w$ . When we add noise  $\eta_i$  to the median estimator, the observations in Eq. (13) are now buried in the composite noise  $z_i = w_i + \eta_i$  ( $i = 1, 2, \dots, n$ ). Then, the maximum bias of Eq. (23) is updated as

$$\tilde{b}_\epsilon = F_z^{-1}[1/2(1 - \epsilon)], \quad (24)$$

with the inverse function  $F_z^{-1}$  of the cumulative distribution  $F_z$ .

In Appendix H, we derive the condition for reducing maximum bias  $b_\epsilon$  of the median estimator in Eq. (23) with additive noise. It is demonstrated that if there exists a positive interval where the cumulative distribution function  $F_w$  is strictly convex, then the maximum bias  $b_\epsilon$  of the median estimator in Eq. (23) can be reduced. Furthermore, under the condition of convexity of  $F_w$ , we also demonstrate in Appendix H that the optimal noise probability density  $f_\eta^{\text{opt}}$  is given by

$$f_\eta^{\text{opt}}(\eta) = \frac{1}{2}[\delta(\eta - a^{\text{opt}}) + \delta(\eta + a^{\text{opt}})], \quad (25)$$

with the optimal noise level

$$a^{\text{opt}} = \arg \max_a \frac{1}{2}[F_w(b_\epsilon - a) + F_w(b_\epsilon + a)]. \quad (26)$$

The noise with the distribution in Eq. (25) is also referred to as the dichotomous noise [2]. This condition applies for bimodal noise distribution models arising in diverse problems [46]–[51], where the cumulative distribution function  $F_w$  satisfies the convexity condition given in Appendix H.



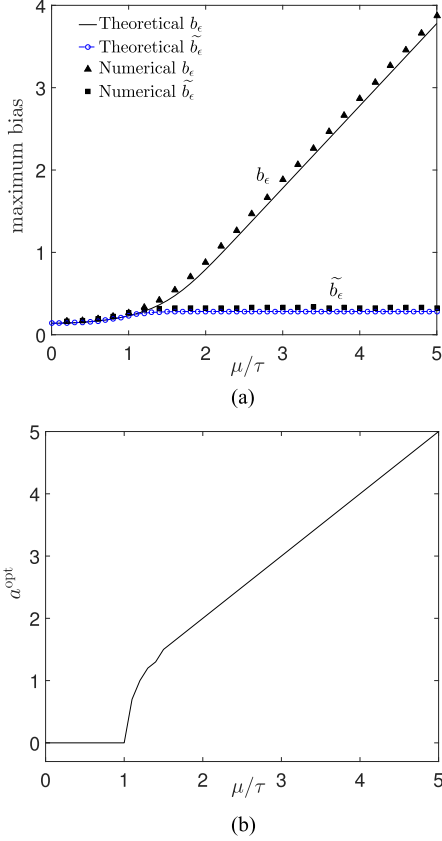


Fig. 6. (a) Maximum bias  $\tilde{b}_\epsilon$  ( $\circ$ ) of Eq. (24) obtained by optimizing the added noise level  $a$ , and maximum bias  $b_\epsilon$  (solid line) of Eq. (23) without the added noise as a function of  $\mu/\tau$ . The numerical results of maximum bias  $\tilde{b}_\epsilon$  ( $\blacktriangle$ ) and  $b_\epsilon$  ( $\blacksquare$ ) are produced by Monte Carlo simulation (see Appendix G-B). (b) The optimal noise level  $a^{\text{opt}}$  corresponding to (a) versus  $\mu/\tau$ . Here, the fraction of contamination  $\epsilon = 0.1$ .

For example, consider the Gaussian mixture noise with density

$$f_w(x) = \frac{1}{2} [\kappa(x; -\mu, \tau) + \kappa(x; \mu, \tau)], \quad (27)$$

where  $\kappa(x; \mu, \tau) = (1/\sqrt{2\pi\tau^2}) \exp[-(x - \mu)^2/(2\tau^2)]$  with parameters  $\mu, \tau \geq 0$ . The dichotomous noise of Eq. (25) with the optimal noise level  $a^{\text{opt}}$  of Eq. (26) is added to the observations. For the fraction of contamination  $\epsilon = 0.1$ , Fig. 6(a) presents that the maximum bias  $b_\epsilon$  (solid line) of Eq. (23) without added noise increases as the ratio of  $\mu/\tau$  grows, while the maximum bias  $\tilde{b}_\epsilon$  ( $\circ$ ) of Eq. (24) obtained by optimizing the added noise level  $a^{\text{opt}}$  maintains a lower value with the increase of  $\mu/\tau$ . The corresponding optimal noise level  $a^{\text{opt}}$  is also plotted in Fig. 6(b). The results clearly show the noise-reduced bias effect in the median estimator for the asymmetric  $\epsilon$ -contaminated noise model.

## V. DISCUSSION

In this paper, we study the enhancement of the asymptotic efficiency and the reduction of the maximum bias by adding noise in robust location M-estimators. For symmetric

heavy-tailed noise models, we show that the asymptotic efficiencies of two commonly used M-estimators with the Huber function and the bisquare function are non-monotonic functions of the ratio of the noise scale and the estimator parameter. With a summing array of M-estimators, by injecting extra noise into the array, we show the possibility of noise enhancement of the asymptotic efficiency that is maximized by an optimal nonzero amount of extra noise.

Furthermore, in order to maximize the asymptotic efficiency of a single M-estimator with added noise, a discriminant function is proposed to establish whether an optimal noise condition exists or not. When the optimal additive noise exists, its probability density function is proven to be the same one given by Chen *et al.* [9]–[11]. Nevertheless, even when the optimal additive noise does not exist, the noise-enhanced effect occurs for an array of M-estimators with estimator number larger than one. Moreover, it is proven that the asymptotic efficiency of an array of M-estimators is a monotonically increasing function of the M-estimator number, and achieves its maximum for an infinite M-estimator number. Then, aiming to maximize the asymptotic efficiency of an infinite number of M-estimators, we find that the optimal noise distribution is the weighted minimum  $\mathcal{L}^2$ -norm solution of Eq. (44). It is also theoretically demonstrated that the upper bound of the asymptotic efficiency of an infinite number of M-estimators is unity, and the corresponding optimal noise is analytically derived. Unfortunately, this optimal noise distribution often does not exist. Therefore, we select a class of parametric noise distributions, and optimize the noise distribution parameters as well as the noise variance. Specially, we present a condition for a benefit by adding uniform noise, and show the noise benefit is effective for a number of background heavy-tailed noise models.

It is also shown that, compared with the method without extra noise, the addition of noise can provide an improved asymptotic efficiency with a local optimal probability density chosen from the given parametric noise distributions. In addition, upon increasing the M-estimator number to a sufficiently large value, we can gradually obtain an excellent asymptotic efficiency close to unity. The asymmetric contaminated model is also considered, which contains a proportion of the observations generated by the true distribution, while another proportion of outliers is mixed by an unknown mechanism. In this case, we consider the extreme situation of contamination with an infinite point-mass distribution, and characterize the noise type for which the maximum bias of the median estimator can be minimized by adding an optimal amount of the dichotomous noise.

Several questions of interest remain open: Although the optimal noise density  $f_\eta^{\text{opt}}$  that maximizes the asymptotic efficiency of an infinite number of M-estimators is proven to be the weighted minimum  $\mathcal{L}^2$ -norm solution of Eq. (44), an explicit form of  $f_\eta^{\text{opt}}$  remains an interesting open question for future study. We also note that, for achieving the upper bound of unity of the asymptotic efficiency, the optimal additive noise exists only if the inverse of the Fourier transform of Eq. (42) is nonnegative over all the probability space. Otherwise, the optimal noise does not exist. In cases where the inverse of the Fourier transform of Eq. (42) exists but has

negative parts, one can envisage to clip the negative parts to zero and renormalize the positive parts as a suboptimal solution for the noise density. We test this approach with the following example. Consider a zero-mean Student distribution  $f_w(x) = c_v(1 + x^2/v)^{-(v+1)/2}$  with the degree of freedom  $v > 0$  and the constant  $c_v = \Gamma((v+1)/2)/(\sqrt{v\pi}\Gamma(v/2))$ . The corresponding maximum likelihood estimator  $\psi_M = -f'_w(x)/f_w(x) = (v+1)x/(v+x^2)$ . When the median estimator  $\psi(x) = \text{sgn}(x)$  is employed, the inverse of the Fourier transform of Eq. (42) is given by  $I(x) = \sqrt{\pi/2}(v+1)(v-x^2)/(v+x^2)^2$ , which has negative parts as  $|x| > \sqrt{v}$  and is actually not a probability density function. We can clip these negative parts of  $I(x)$  to zero and renormalize the nonnegative parts as a probability density function  $f_\eta^c(x) = \sqrt{v}(v-x^2)/(v+x^2)^2$  for  $|x| \leq \sqrt{v}$  and zero otherwise. It is very interesting to note that, with this clipped probability density function  $f_\eta^c$  and the degree of freedom  $v = 4$ , the asymptotic efficiency of Eq. (17) is 0.995 that is very close to unity. Thus, this suboptimal solution  $f_\eta^c(x)$  provides a suitable noise density to improve the asymptotic efficiency of an array of M-estimators. Adding noise in an array as we propose is another alternative to improve the efficiency. However, even if this optimal noise distribution exists, the infinite-size array can only be approached in practice. As an alternative, it is meaningful to further explore the optimal noise type that achieves the maximum asymptotic efficiency for a finite M-estimator number, and ideally with a noise type which is not restricted to a prior class of parametric noise distributions. In addition, for the special case of the median estimator, the optimal dichotomous noise is demonstrated to minimize the maximum bias for estimating a location buried in bimodal noise models with convex cumulative distributions. It is interesting to analyze the reduction of maximum bias with addition of noise for other robust M-estimators, and prove whether there are other contaminated models exhibiting the noise-enhanced effects or not. We here consider the asymmetrical contaminated model of Eq. (20) for the point mass distribution of extreme cases, and derive the optimal symmetrical noise to reduce the maximum bias at a given proportion of the contamination. In this contaminated model, the contaminating probability density  $q$  is actually arbitrary and unknown. But, with sufficient knowledge of the contamination distribution in Eq. (20), it is very meaningful to further investigate the optimal asymmetrical noise obeying the distribution of  $h_c(-x)$ . Then, the asymmetrical contaminated model of Eq. (20) can be simplified as an unbiased location estimation. For robust statistics of multi-parameter problems and M-estimators for regression models [34]–[38], the benefits of noise also deserve to be further studied. These questions are of interest for robust estimation where addition of noise properly optimized may reveal, as in the present study, to be a beneficial option.

APPENDIX A  
NO NOISE-ENHANCED EFFECT FOR THE MAXIMUM  
LIKELIHOOD ESTIMATOR

In this Appendix, we show that adding extra noise to observations cannot reduce the variance of the maximum likelihood

estimator. Based on Eqs. (5) and (6), the asymptotic variance of the maximum likelihood estimator is  $1/(nJ(f_w))$  that is closely related to the original noise distribution  $f_w$ . When we add extra noise  $\eta$  to the observations and the background noise becomes the composite noise  $z = w + \eta$ , the asymptotic variance of the maximum likelihood estimator that is optimal for noise  $z$  is updated as  $1/(nJ(f_z))$ . According to the Fisher information convolution inequality  $J(f_z) \leq \min\{J(f_w), J(f_\eta)\}$  in Refs. [41] (Chapter 16.6) and [52], we can derive that  $1/(nJ(f_z)) \geq 1/(nJ(f_w))$ , which indicates no noise-enhanced effect for the maximum likelihood estimator.

APPENDIX B

ASYMPTOTIC EFFICIENCY OF AN ARRAY OF M-ESTIMATORS

In this Appendix, we present the derivation of the asymptotic efficiency of Eq. (16) for a summing array of M-estimators. The location estimate  $\hat{\theta}_l$  of each M-estimator satisfies  $\sum_{i=1}^n \psi(y_{li} - \hat{\theta}_l) = 0$  for  $l = 1, 2, \dots, L$ . We can perform a Taylor expansion of  $\psi$  as a function of  $\hat{\theta}_l$  around the true value of  $\theta$ , so the result yields [34], [37]

$$\begin{aligned} 0 &= \sum_{i=1}^n \psi(y_{li} - \hat{\theta}_l) \\ &\approx \sum_{i=1}^n \psi(y_{li} - \theta) - (\hat{\theta}_l - \theta) \sum_{i=1}^n \frac{\partial}{\partial \theta} \psi(y_{li} - \theta) \\ &= A_n - (\hat{\theta}_l - \theta) B_n, \end{aligned} \quad (28)$$

where  $A_n = \sum_{i=1}^n \psi(y_{li} - \theta)/n$  and  $B_n = \sum_{i=1}^n (\partial \psi(y_{li} - \theta)/\partial \theta)/n$ . Hence, by Slutsky's lemma and according to the central limit theorem for a sufficiently large  $n$ ,  $B_n$  in probability converges to  $B = E[\partial \psi(y - \theta)/\partial \theta] = E_z[\psi'(z)]$ , and  $A_n$  converges to a Gaussian distribution with zero mean and variance  $E[\psi^2(y - \theta)]/n = E_z[\psi^2(z)]/n$  [34], [37]. Therefore, each estimate  $\hat{\theta}_l$  converges to a Gaussian distribution with mean  $\theta$  and variance  $\text{var}(\hat{\theta}_l) = n^{-1} E_z[\psi^2(z)]/E_z^2[\psi'(z)]$  of Eq. (5). In line with this general proof, the average estimate  $\hat{\theta} = \sum_{l=1}^L \hat{\theta}_l/L$ , also converges to a Gaussian distribution with mean  $\theta$  and variance

$$\begin{aligned} \text{var}_a(\hat{\theta}) &= \text{var} \left( \frac{1}{L} \sum_{l=1}^L \hat{\theta}_l \right) \\ &= \frac{1}{L^2} [L \text{var}(\hat{\theta}_l) + \sum_{\substack{l,p=1 \\ l \neq p}}^L E_z [(\hat{\theta}_l - \theta)(\hat{\theta}_p - \theta)]] \\ &= \frac{1}{L^2} [L \text{var}(\hat{\theta}_l) + L(L-1) E_z [(\hat{\theta}_l - \theta)(\hat{\theta}_p - \theta)]] \\ &= \frac{1}{L} [\text{var}(\hat{\theta}_l) + (L-1) E_w \{E_\eta [(\hat{\theta}_l - \theta)(\hat{\theta}_p - \theta)]\}] \\ &= \frac{1}{L} \text{var}(\hat{\theta}_l) + \frac{L-1}{L} E_w \{E_\eta [(\hat{\theta}_l - \theta)(\hat{\theta}_p - \theta)]\}. \end{aligned} \quad (29)$$

Since for a given  $w_i$ ,  $w_i + \eta_{li}$  are i.i.d. random variables for  $l = 1, 2, \dots, L$  and the Taylor expansion of  $\psi$  in Eq. (28), so

we can obtain

$$\mathbb{E}_w \{ \mathbb{E}_\eta [(\hat{\theta}_l - \theta)(\hat{\theta}_p - \theta)] \} = \frac{\mathbb{E}_w \{ \mathbb{E}_\eta^2 [\psi(w + \eta)] \}}{n \mathbb{E}_z^2 [\psi'(z)]}. \quad (30)$$

Substituting Eqs. (5) and (30) into Eq. (29), the asymptotic variance of  $\hat{\theta}$  can be calculated as

$$\text{var}_a(\hat{\theta}) = \frac{1}{n} \frac{\mathbb{E}_z [\psi^2(z)] + (L-1) \mathbb{E}_w \{ \mathbb{E}_\eta^2 [\psi(w + \eta)] \}}{L \mathbb{E}_z^2 [\psi'(z)]}. \quad (31)$$

Then, the asymptotic efficiency of Eq. (16) is obtained.

#### APPENDIX C

##### OPTIMIZATION OF THE ASYMPTOTIC EFFICIENCY OF A SINGLE M-ESTIMATOR

For a single M-estimator ( $L = 1$ ), the asymptotic efficiency of  $\text{Eff}_a(\hat{\theta})$  of Eq. (16) becomes

$$\text{Eff}_a(\hat{\theta}) = \frac{1}{J(f_w)} \frac{\mathbb{E}_z^2 [\psi'(z)]}{\mathbb{E}_z [\psi^2(z)]}. \quad (32)$$

Thus, it is interesting to find an optimal noise probability density  $f_\eta^{\text{opt}}$  to maximize the asymptotic efficiency of  $\text{Eff}_a(\hat{\theta})$  in Eq. (32) under the constraint of unbiasedness. Since the Fisher information  $J(f_w) > 0$ , then this optimization problem can be expressed as

$$\max_{f_\eta} \frac{\mathbb{E}_z^2 [\psi'(z)]}{\mathbb{E}_z [\psi^2(z)]}, \quad (33)$$

with the constraint of  $\mathbb{E}_z [\psi(z)] = 0$ .

Consider a real-value vector function  $f(\mathbf{Z}) = Z_1^2/Z_2$  with  $\mathbf{Z} = [Z_1, Z_2]^T$  and  $Z_2 > 0$ . It is seen that  $f(\mathbf{Z})$  is convex, because its Hessian matrix  $\nabla^2 f(\mathbf{Z}) = 2[Z_2, -Z_1]^T [Z_2, -Z_1]/Z_2^3$  is positive semidefinite [43]. Then, Jensen's inequality  $\mathbb{E}_\eta^2(Z_1)/\mathbb{E}_\eta(Z_2^2) \leq \mathbb{E}_\eta(Z_1^2/Z_2)$  is valid in the definition domain  $\{(Z_1, Z_2) \in \mathbb{R}^2 | Z_2 > 0\}$  [17], [41], [43], [44]. Therefore, defining  $Z_1 = \mathbb{E}_w[\psi'(w + \eta)]$  and  $Z_2 = \mathbb{E}_w[\psi^2(w + \eta)]$ , we have

$$\frac{\mathbb{E}_z^2 [\psi'(z)]}{\mathbb{E}_z [\psi^2(z)]} = \frac{\mathbb{E}_\eta^2 \{ \mathbb{E}_w [\psi'(w + \eta)] \}}{\mathbb{E}_\eta \{ \mathbb{E}_w [\psi^2(w + \eta)] \}} = \frac{\mathbb{E}_\eta^2(Z_1)}{\mathbb{E}_\eta(Z_2)} \leq \mathbb{E}_\eta(Z_1^2/Z_2) \quad (34a)$$

$$= \mathbb{E}_\eta[g(\eta)] \quad (34b)$$

$$\leq \max_\eta g(\eta), \quad (34c)$$

where  $g(\eta) = Z_1^2/Z_2$ ,  $Z_1^2$  and  $Z_2$  are even functions for the odd function  $\psi$  and the even probability density  $f_w$ . It is noted that  $\mathbb{E}_\eta[g(\eta)]$  cannot be larger than the maximum value of  $g(\eta)$  in Eq. (34c) and the optimal added noise has a density function of one or multiple mass points (depending on the number of points of  $a^{\text{opt}} = \arg \max_\eta g(\eta)$ ). Furthermore, since  $g(\eta)$  is an even function, then it must have an extremum value  $g(0)$  at the origin. If  $g(0)$  is the absolute maximum over the real space  $\eta \in \mathbb{R}$ , then the optimal solution of Eq. (33) is  $f_\eta^{\text{opt}}(\eta) = \delta(\eta)$ . This indicates that there is no benefit in adding noise. Meanwhile, if  $g(\eta)$  has two absolute maximum values of  $g(a^{\text{opt}}) = g(-a^{\text{opt}}) > g(0)$  at  $\pm a^{\text{opt}}$  ( $a^{\text{opt}} \neq 0$ ), then the optimal dichotomous noise

density exists and can be expressed by Eq. (25) for obtaining the maximum of  $g(a^{\text{opt}})$ . Substituting Eq. (25) with the optimal noise level  $a^{\text{opt}}$  into Eq. (34a) and Eq. (34c), both equalities hold and the maximum value of Eq. (33) can be achieved. This result accords with the determination of the optimal noise proven by Chen *et al.* [9]–[11] based on Carathéodory's theorem [53].

#### APPENDIX D

##### ASYMPTOTIC EFFICIENCY VERSUS THE M-ESTIMATOR NUMBER

For the convex function  $x^2$  and by the Jensen inequality, we obtain

$$\mathbb{E}_\eta[\psi^2(w + \eta)] \geq \mathbb{E}_\eta^2[\psi(w + \eta)], \quad (35)$$

for any fixed variable  $w$  [41]. Then, we have

$$\begin{aligned} \mathbb{E}_z[\psi^2(z)] &= \mathbb{E}_w \{ \mathbb{E}_\eta[\psi^2(w + \eta)] \} \\ &\geq \mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \}. \end{aligned} \quad (36)$$

Immediately, we find

$$\begin{aligned} &\frac{1}{L+1} \mathbb{E}_z[\psi^2(z)] + \frac{L}{L+1} \mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \} \\ &\leq \frac{1}{L} \mathbb{E}_z[\psi^2(z)] + \frac{L-1}{L} \mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \}. \end{aligned} \quad (37)$$

From Eq. (16) and Eq. (37), we derive

$$\begin{aligned} &\frac{1}{J(f_w)} \frac{(L+1) \mathbb{E}_z^2[\psi'(z)]}{\mathbb{E}_z[\psi^2(z)] + L \mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \}} \\ &\geq \frac{1}{J(f_w)} \frac{L \mathbb{E}_z^2[\psi'(z)]}{\mathbb{E}_z[\psi^2(z)] + (L-1) \mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \}}, \end{aligned}$$

which indicates

$$\text{Eff}_a(\hat{\theta}, L+1) \geq \text{Eff}_a(\hat{\theta}, L). \quad (38)$$

Thus, for the given M-estimator function  $\psi$  and fixed noise components  $w$  and  $\eta$ , the asymptotic efficiency  $\text{Eff}_a(\hat{\theta}, L)$  in Eq. (16) is a monotonically increasing function of the M-estimator number  $L$ .

#### APPENDIX E

##### OPTIMIZATION OF THE ASYMPTOTIC EFFICIENCY FOR AN INFINITE NUMBER OF M-ESTIMATORS

By using the Cauchy-Schwarz inequality, we find that the asymptotic efficiency  $\text{Eff}_\infty(\hat{\theta})$  in Eq. (17) satisfies

$$\begin{aligned} \text{Eff}_\infty(\hat{\theta}) &= \frac{1}{J(f_w)} \frac{\mathbb{E}_w^2 \{ d\mathbb{E}_\eta[\psi(w + \eta)]/dw \}}{\mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \}} \\ &= \frac{1}{J(f_w)} \frac{\mathbb{E}_w^2 \{ \mathbb{E}_\eta[\psi(w + \eta)] \psi_M(w) \}}{\mathbb{E}_w \{ \mathbb{E}_\eta^2[\psi(w + \eta)] \}} \\ &\leq \frac{1}{J(f_w)} \mathbb{E}_w [\psi_M^2(w)] = 1, \end{aligned} \quad (39)$$

where the equality occurs as

$$\mathbb{E}_\eta[\psi(w + \eta)] = C \psi_M(w), \quad (40)$$

with an arbitrary constant  $C$ .

Thereby, for the given function  $\psi$  and the background noise distribution  $f_w$ , the optimal additive noise probability density  $f_\eta^{\text{opt}}$  is the solution of Eq. (40). With the assumptions of symmetrical distribution function  $f_w$  and the odd M-estimator function  $\psi$ , the added noise probability density function is also even, i.e.,  $f_\eta(-\eta) = f_\eta(\eta)$ , due to the Fisher-consistent M-estimator satisfying  $E_z[\psi(z)] = 0$  [37]. Then, Eq. (40) can be rewritten as

$$\begin{aligned} E_\eta[\psi(w + \eta)] &= \int_{-\infty}^{+\infty} \psi(w + \eta) f_\eta(-\eta) d\eta \\ &= \int_{+\infty}^{-\infty} \psi(w - u) f_\eta(u) d(-u) \\ &= \psi(w) * f_\eta(w) \\ &= C\psi_M(w), \end{aligned} \quad (41)$$

which indicates that the optimal noise probability density  $f_\eta^{\text{opt}}$  is the deconvolution of  $C\psi_M(x)$  and  $\psi(x)$ . If the Fourier transformations of  $\psi(x)$  and  $\psi_M(x)$  both exist [54], [55], we have

$$f_\eta^{\text{opt}}(x) = C \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}[\psi_M(x)]}{\mathcal{F}[\psi(x)]} \right\}, \quad (42)$$

where  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  denote the Fourier transform and its inverse transform, respectively. Here, we emphasize that, due to the probability density  $f_\eta(\eta) \geq 0$ , the optimal density  $f_\eta^{\text{opt}}$  exists only if the inverse of the Fourier transform in Eq. (42) exists and is a nonnegative function. The constant  $C$  can be calculated by the normalization condition  $\int f_\eta(\eta) d\eta = 1$ . In addition, both  $\psi_M(x)$  and  $\psi(x)$  are odd functions. According to the convolution property [55],  $f_\eta^{\text{opt}}$  is an even function when it exists. For example, consider a kind of heavy-tailed noise with the hyperbolic secant distribution  $f_w(x) = \text{sech}(\pi x/2)$  [42] and the median estimator  $\psi(x) = \text{sign}(x)$ . Then, from Eq. (42), the optimal probability density becomes  $f_\eta^{\text{opt}}(x) = \pi e^{\pi x} / [4(1 + e^{\pi x})^2]$ , which indicates logistic noise as a feasible optimal noise.

It is noted that the optimal noise indicated in Eq. (42) corresponding to the upper bound of unity is only a special case of

$$f_\eta^{\text{opt}}(x) = \arg \max_{f_\eta} \text{Eff}_\infty(\hat{\theta}). \quad (43)$$

An achievable optimal noise density that achieves the maximum of the asymptotic efficiency (not more than unity) of Eq. (17) among all feasible densities is practically significant. Define the inner product of two functions  $f$  and  $g$  as  $\langle f, g \rangle = E_w[f(w)g(w)]$ , and the  $\mathcal{L}^2$ -norm of a function  $g$  as  $\|g\| = \langle g, g \rangle^{1/2}$ . Here, we define  $\phi(w) = E_\eta[\psi(w + \eta)]$  and its normalized function can be expressed as  $\bar{\phi}(w) = \phi(w) / \|\phi(w)\|$ . Correspondingly, the maximum likelihood estimator  $\psi_M(w)$  can be also normalized as  $\bar{\psi}_M(w) = \psi_M(w) / \|\psi_M(w)\|$  with its norm  $\|\psi_M(w)\| = \sqrt{J(f_w)}$ . From Eqs. (17) and (41), we have

$$\text{Eff}_\infty(\hat{\theta}) = \frac{\langle E_\eta[\psi(w + \eta)], \psi_M(w) \rangle^2}{\|E_\eta[\psi(w + \eta)]\|^2 \|\psi_M(w)\|^2} = \langle \bar{\phi}(w), \bar{\psi}_M(w) \rangle^2.$$

Note that  $\|\bar{\phi}(w) - \bar{\psi}_M(w)\|^2 = 2 - 2\langle \bar{\phi}(w), \bar{\psi}_M(w) \rangle$ . Thus, the optimal noise density indicated in Eq. (43) is just the

minimum  $\mathcal{L}^2$ -norm solution of  $\|\bar{\phi}(w) - \bar{\psi}_M(w)\|^2$ , expressed as

$$f_\eta^{\text{opt}}(x) = \arg \min_{f_\eta} \|\bar{\phi}(w) - \bar{\psi}_M(w)\|^2, \quad (44)$$

with the constrains of  $\int f_\eta(x) dx = 1$  and  $f_\eta(x) \geq 0$  for  $x \in \mathbb{R}$ . A further point is to consider the set of all possible noise densities, and then in this set there is necessarily at least one that realizes the maximum efficiency, which is equivalent to solving the constrained minimization of Eq. (44). How to calculate such an achievable optimal density of the additive noise remains an interesting open question for future study.

## APPENDIX F

### BENEFIT CONDITION OF ADDING UNIFORM NOISE

Based on Eq. (19), the derivative of  $\text{Eff}_\infty(\hat{\theta}, d)$  with respect to the uniform noise level  $d$  is given by

$$\frac{\partial \text{Eff}_\infty(\hat{\theta}, d)}{\partial d} = \frac{1}{J(f_w)} \frac{A}{E_w^2 \{ [\rho(w+d) - \rho(w-d)]^2 \}}, \quad (45)$$

with  $A = 2E_w[\psi(w+d) - \psi(w-d)]E_w[\psi'(w+d) + \psi'(w-d)]E_w\{[\rho(w+d) - \rho(w-d)]^2\} - 2E_w^2[\psi(w+d) - \psi(w-d)]E_w\{[\rho(w+d) - \rho(w-d)][\psi(w+d) + \psi(w-d)]\}$ . The optimal noise level  $d^{\text{opt}}$  is solved by  $\partial \text{Eff}_\infty(\hat{\theta}, d) / \partial d = 0$ . Since  $J(f_w) > 0$  and  $E_w^2\{[\rho(w+d) - \rho(w-d)]^2\} > 0$ , then the optimal noise level  $d^{\text{opt}}$  is the solution of  $A = 0$ . We further consider the small noise level  $d \rightarrow 0$  ( $d > 0$ ). In this situation, the derivative of  $\text{Eff}_\infty(\hat{\theta}, d)$  in Eq. (45) can be approximated with the second-order Taylor expansion as

$$\frac{\partial \text{Eff}_\infty}{\partial d} \approx \frac{d}{J(f_w)} \frac{E_w^2[\psi'(w)]}{E_w[\psi^2(w)]} \left\{ \frac{E_w[\psi'''(w)]}{E_w[\psi'(w)]} - \frac{E_w[\psi(w)\psi''(w)]}{E_w[\psi^2(w)]} \right\}.$$

Since  $E_w^2[\psi'(w)]/E_w[\psi^2(w)] > 0$  and the noise level  $d > 0$ , then the condition of a benefit by adding uniform noise is given by

$$D(\psi, f_w) = \frac{E_w[\psi'''(w)]}{E_w[\psi'(w)]} - \frac{E_w[\psi(w)\psi''(w)]}{E_w[\psi^2(w)]} > 0, \quad (46)$$

which indicates when the addition of uniform noise leads to the increase of the asymptotic efficiency  $\text{Eff}_\infty(\hat{\theta}, d)$ . If the second or third derivative of  $\psi$  in Eq. (46) does not exist, we can reduce the derivative order of  $\psi$ , and explore the derivative of  $f_w$  by applying the subsection integral method. For example,  $E_w[\psi'''(w)] = -\int \psi''(w) f'_w(w) dw$  and note  $f_w(\pm\infty) = 0$ . In practice, we can simply calculate the sign of  $D(\psi, f_w)$  in Eq. (46) to determine whether the addition of uniform noise is beneficial or not for a background noise distribution  $f_w$ , as illustrated in Fig. 2.

APPENDIX G  
NUMERICAL SIMULATION

A. Numerical Computation of the Asymptotic Efficiency

We adopt the iterative reweighting method to obtain the numerical estimation  $\hat{\theta}$  of a location parameter [37]. Assuming

$$W(x) = \begin{cases} \psi(x)/x, & x \neq 0, \\ \psi'(x), & x = 0, \end{cases} \quad (47)$$

then Eq. (3) can be rewritten as [37]

$$\sum_{i=1}^n W(x_i - \hat{\theta})(x_i - \hat{\theta}) = 0. \quad (48)$$

From Eq. (48), the estimated parameter can be obtained as a weighted mean [37]

$$\hat{\theta} = \frac{\sum_{i=1}^n W_i x_i}{\sum_{i=1}^n W_i}, \quad (49)$$

with  $W_i = W(x_i - \hat{\theta})$ .

For the method of adding extra noise to an array of M-estimators, we first generate  $L$  groups of updated samples with the addition of  $L$  samples of uniform noise components  $\eta_{li}$  with the level of  $d$ . Then, compute  $L$  values of parallel estimators  $(\hat{\theta}_1, \dots, \hat{\theta}_L)$  by using  $L$  groups of observations respectively. For each group of observations  $y_l = (y_{l1}, y_{l2}, \dots, y_{ln})$ , the  $l$ th estimator  $\hat{\theta}_l$  is obtained by the iterative reweighting method as follows:

1) Set the initial estimator  $\hat{\theta}_l(0)$  as the median of the updated observations  $y_l$ .

2) For  $k = 0, 1, 2, \dots$ , compute the location parameter  $\hat{\theta}_l(k+1)$  at iteration  $k+1$  as

$$\hat{\theta}_l(k+1) = \frac{\sum_{i=1}^n W_{k,li} y_{li}}{\sum_{i=1}^n W_{k,li}}, \quad (50)$$

with the weight function  $W_{k,li} = W[y_{li} - \hat{\theta}_l(k)]$ .

3) Stop when  $|\hat{\theta}_l(k+1) - \hat{\theta}_l(k)| < \zeta$ , where  $\zeta$  is a small tolerance parameter, and the numerical M-estimator  $\hat{\theta}_l(k+1)$  is assumed to be the  $l$ th expected M-estimator  $\hat{\theta}_l$ .

4) Compute the estimator  $\hat{\theta}$  of Eq. (15) by averaging  $L$  estimators  $(\hat{\theta}_1, \dots, \hat{\theta}_L)$ .

Then, we can realize the numerical M-estimator  $\hat{\theta}$  for  $M = 10^4$  Monte Carlo trails by the above iterative reweighting method, and the variance of  $\hat{\theta}$  is evaluated as  $\text{var}(\hat{\theta}) = \sum_{m=1}^M (\hat{\theta}_m - \bar{\hat{\theta}})^2 / (M-1)$  with the mean  $\bar{\hat{\theta}} = \sum_{m=1}^M \hat{\theta}_m / M$ . Substituting the variance  $\text{var}(\hat{\theta})$  into Eq. (7), the numerical asymptotic efficiency  $\text{Eff}_a(\hat{\theta})$  can be obtained, as shown in Fig. 3(c).

B. Numerical Simulation of the Maximum Bias

We use  $10^4$  Monte Carlo trails to evaluate the maximum bias with optimal noise and without added noise, respectively. For each trail, the observation size  $n = 5 \times 10^4$  and the fraction of contamination  $\epsilon = 0.1$ . We generate  $n$  observations obeying the Gaussian mixture distribution, then replacing  $0.1n$  of observations with the outliers, resulting in the contaminated

data  $x_i$ . According to the contamination model of Eq. (20), the maximum bias is theoretically evaluated with the extreme outliers obeying the point mass distribution  $q = \delta(x_0)$  at  $x_0 \rightarrow \infty$ . In practice, the outliers are taken as real numbers with sufficiently large magnitude as  $10^6 \gg \mu + 3\tau$ . Then, we generate  $n$  random components  $\eta_i$  composed by two values of  $a^{\text{opt}}$  and  $-a^{\text{opt}}$  with equal probabilities. Then, the updated observations  $y_i = x_i + \eta_i$  are obtained. Next, calculate the median of the samples by the original data  $x_i$  and the updated observations  $y_i$ , respectively. Finally, the maximum bias with optimal noise and without added noise are obtained for  $10^4$  numerical trials, respectively, as shown in Fig. 6.

APPENDIX H  
NOISE-REDUCED MAXIMUM BIAS OF THE MEDIAN ESTIMATOR

For the median estimator and a given proportion  $0 < \epsilon < 1/2$ , we expect that the updated bias  $\tilde{b}_\epsilon$  of Eq. (24) is smaller than the bias  $b_\epsilon$  of Eq. (23), i.e.,  $F_z^{-1}[1/2(1-\epsilon)] < F_w^{-1}[1/2(1-\epsilon)]$ . Here, the median estimator has the breakdown point of  $\epsilon^* = 1/2$  that yields an unbounded estimator  $\hat{\theta} \rightarrow \infty$ , thus requiring  $0 < \epsilon < 1/2$ . Since the probability density  $f_w$  or  $f_z$  is symmetric, the cumulative distribution function is then analyzed for the nonnegative domain of  $x \geq 0$ . It is known that the cumulative distribution function is nondecreasing, then finding a smaller updated bias of  $\tilde{b}_\epsilon < b_\epsilon$  is equivalent to satisfying the condition of  $F_z(b_\epsilon) > F_w(b_\epsilon)$  for  $b_\epsilon = \tilde{b}_\epsilon$ . The cumulative distribution function  $F_z$  of the composite noise  $z$  can be calculated as

$$\begin{aligned} F_z(x) &= \int_{-\infty}^x f_z(z) dz \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f_w(z-\eta) f_\eta(\eta) d\eta dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^x f_w(z-\eta) dz f_\eta(\eta) d\eta \\ &= \int_{-\infty}^{\infty} F_w(x-\eta) f_\eta(\eta) d\eta \\ &= E_\eta[F_w(x-\eta)]. \end{aligned} \quad (51)$$

For a given positive value of  $b_\epsilon \in (U, V)$  for  $0 \leq U < V$  and a function  $F_w(x)$  convex over the interval  $(U, V)$ , then, according to the Jensen inequality [41], we obtain

$$F_z(b_\epsilon) = E_\eta[F_w(b_\epsilon - \eta)] > F_w[E_\eta(b_\epsilon - \eta)] = F_w(b_\epsilon) \quad (52)$$

with the zero mean  $E_\eta[\eta] = 0$ . This indicates that if there exists a positive interval  $(U, V)$  where the cumulative distribution function  $F_w$  is strictly convex, then the maximum bias of the median estimator can be reduced by adding extra noise  $\eta$ .

Under the above convexity condition of  $F_w$  over a certain positive interval  $(U, V)$ , a naturally emerging problem is to find an optimal noise distribution  $f_\eta^{\text{opt}}$ , aiming to minimize the maximum bias of Eq. (24) without changing the consistency of the estimator under  $f_z$ , i.e.,  $E_z[\psi(z)] = 0$  with  $\psi(x) = \text{sgn}(x)$ . It is noted that minimizing the maximum bias is equivalent to maximizing  $F_z(b_\epsilon)$  for a given positive bias  $b_\epsilon \in (U, V)$ . Thus,

the optimal distribution  $f_{\eta}^{\text{opt}}$  is given by

$$\arg \max_{f_{\eta}} F_z(b_{\epsilon}) = \mathbb{E}_{\eta}[F_w(b_{\epsilon} - \eta)], \quad (53)$$

with the constraint  $\mathbb{E}_z[\psi(z)] = \mathbb{E}_{\eta}[2F_w(\eta) - 1] = 0$ . Next, we will demonstrate that the solution of Eq. (53) is just the noise distribution  $f_{\eta}$  of Eq. (25).

*Proof:* Consider the set of all pairs of  $(F_w(b_{\epsilon} - \eta), 2F_w(\eta) - 1)$  named  $P = \{(p_1, p_2) | p_1 = F_w(b_{\epsilon} - \eta), p_2 = 2F_w(\eta) - 1\}$ .  $T = \{(t_1, t_2) | t_1 = F_z(b_{\epsilon}), t_2 = \mathbb{E}_z[\psi(z)]\}$  denotes the set of pairs  $(F_z(b_{\epsilon}), \mathbb{E}_z[\psi(z)])$ . Let  $E$  be the convex hull of  $P$  and the dimension  $\text{Dim}(E) \leq 2$ . As discussed in [9]–[11], [19], we have  $T = E$ . From Carathéodory’s theorem [53], any point in  $T$  can be expressed as a combination of at most three elements of  $P$ . Since our aim is to maximize  $F_z(b_{\epsilon})$  with the constraint of  $\mathbb{E}_z[\psi(z)] = 0$ , the optimum point can only occur on the boundary of  $T$ . Therefore, the optimal pair  $(F_z(b_{\epsilon}), \mathbb{E}_z[\psi(z)])$  can be expressed as the convex combination of two points of  $P$ . Based on the constraint such that  $\mathbb{E}_z[\psi(z)] = 0$ , i.e.,  $f_{\eta}(-\eta) = f_{\eta}(\eta)$ , we derive the optimal noise distribution  $f_{\eta}$  of Eq. (25).

#### REFERENCES

- [1] S. Mitaïm and B. Kosko, “Adaptive stochastic resonance,” *Proc. IEEE*, vol. 86, no. 11, pp. 2152–2183, Nov. 1998.
- [2] F. Chapeau-Blondeau and X. Godivier, “Theory of stochastic resonance in signal transmission by static nonlinear systems,” *Phys. Rev. E*, vol. 55, pp. 1478–1495, 1997.
- [3] V. Galdi, V. Pierro, and I. M. Pinto, “Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise,” *Phys. Rev. E*, vol. 57, pp. 6470–6479, 1998.
- [4] S. Kay, “Can detectability be improved by adding noise?” *IEEE Signal Process. Lett.*, vol. 7, no. 1, pp. 8–10, Jan. 2000.
- [5] S. Zozor and P. O. Amblard, “Stochastic resonance in locally optimal detectors,” *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3177–3181, Dec. 2003.
- [6] H. Chen, L. R. Varshney, and P. K. Varshney, “Noise-enhanced information systems,” *Proc. IEEE*, vol. 102, no. 10, pp. 1607–1621, Oct. 2014.
- [7] R. Benzi, A. Sutera, and A. Vulpiani, “The mechanism of stochastic resonance,” *J. Phys. A, Math. Gen.*, vol. 14, pp. L453–L457, 1981.
- [8] S. Zozor and P. O. Amblard, “Noise-aided processing: Revisiting dithering in a  $\Sigma\Delta$  quantizer,” *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3202–3210, Aug. 2005.
- [9] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, “Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3172–3184, Jul. 2007.
- [10] H. Chen and P. K. Varshney, “Theory of the stochastic resonance effect in signal detection—Part II: Variable detectors,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5031–5041, Oct. 2008.
- [11] H. Chen, P. K. Varshney, and J. H. Michels, “Noise enhanced parameter estimation,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5074–5081, Oct. 2008.
- [12] F. Chapeau-Blondeau and D. Rousseau, “Noise-enhanced performance for an optimal Bayesian estimator,” *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1327–1334, May 2004.
- [13] D. Rousseau and F. Chapeau-Blondeau, “Noise-improved Bayesian estimation with arrays of one-bit quantizers,” *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2658–2662, Dec. 2007.
- [14] A. Patel and B. Kosko, “Optimal noise benefits in Neyman–Pearson and inequality-constrained statistical signal detection,” *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1655–1669, May 2009.
- [15] B. Xu, H. Zhang, L. Zeng, and J. Li, “Application of parameter-induced stochastic resonance to target detection in shallow-water reverberation,” *Appl. Phys. Lett.*, vol. 91, 2007, Art. no. 091 908.
- [16] S. Uhlir, “Bayes risk reduction of estimators using artificial observation noise,” *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5535–5545, Oct. 2015.
- [17] G. O. Balkan and S. Gezici, “CRLB based optimal noise enhanced parameter estimation using quantized observations,” *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 477–480, May 2010.
- [18] A. B. Akbay and S. Gezici, “Noise benefits in joint detection and estimation problems,” *Signal Process.*, vol. 118, pp. 235–247, 2016.
- [19] H. Soganci, S. Gezici, and O. Arikan, “Optimal stochastic parameter design for estimation problems,” *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4950–4956, Sep. 2012.
- [20] H. Soganci, S. Gezici, and O. Arikan, “Optimal signal design for multi-parameter estimation problems,” *IEEE Trans. Signal Process.*, vol. 63, pp. 6074–6085, 2015.
- [21] G. Zeitler, G. Kramer, and A. C. Singer, “Bayesian parameter estimation using single-bit dithered quantization,” *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.
- [22] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, “Sequential signal encoding from noisy measurements using quantizers with dynamic bias control,” *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.
- [23] A. Patel and B. Kosko, “Optimal mean-square noise benefits in quantizer-array linear estimation,” *IEEE Signal Process. Lett.*, vol. 17, no. 12, pp. 1005–1009, Dec. 2010.
- [24] F. Chapeau-Blondeau, S. Blanchard, and D. Rousseau, “Fisher information and noise-aided power estimation from one-bit quantizers,” *Digit. Signal Process.*, vol. 18, pp. 434–443, 2008.
- [25] H. Chen and P. K. Varshney, “Performance limit for distributed estimation systems with identical one-bit quantizers,” *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 466–471, Jan. 2010.
- [26] H. Chen and P. K. Varshney, “Nonparametric one-bit quantizers for distributed estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3777–3787, Jul. 2010.
- [27] D. Rousseau, F. Duan, and F. Chapeau-Blondeau, “Suprathreshold stochastic resonance and noise-enhanced fisher information in arrays of threshold devices,” *Phys. Rev. E*, vol. 68, 2003, Art. no. 031 107.
- [28] F. Chapeau-Blondeau, S. Blanchard, and D. Rousseau, “Noise-enhanced fisher information in parallel arrays of sensors with saturation,” *Phys. Rev. E*, vol. 74, 2006, Art. no. 031 102.
- [29] J. Zhu, X. Li, R. S. Blum, and Y. Gu, “Parameter estimation from quantized observations in multiplicative noise environments,” *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4037–4050, Aug. 2015.
- [30] O. Dabeer and A. Karnik, “Signal parameter estimation using 1-bit dithered quantization,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5389–5405, Dec. 2006.
- [31] L. Xu, F. Duan, X. Gao, D. Abbott, and M. D. McDonnell, “Adaptive recursive algorithm for optimal weighted suprathreshold stochastic resonance,” *Roy. Society Open Sci.*, vol. 4, 2017, Art. no. 160 889.
- [32] M. D. McDonnell, “Is electrical noise useful?” *Proc. IEEE*, vol. 99, no. 2, pp. 242–246, Feb. 2011.
- [33] G. P. Harmer, B. R. Davis, and D. Abbott, “A review of stochastic resonance: Circuits and measurement,” *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 299–309, Apr. 2002.
- [34] P. J. Huber, *Robust Statistics*. Berlin, Germany: Springer-Verlag, 1981.
- [35] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. New York, NY, USA: Wiley, 1986.
- [36] S. A. Kassam and H. V. Poor, “Robust techniques for signal processing: A survey,” *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, Mar. 1985.
- [37] R. Maronna, D. Martin, and V. Yohai, *Robust Statistics: Theory and Methods*. New York, NY, USA: Wiley, 2006.
- [38] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, “Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts,” *IEEE Signal Process. Mag.*, vol. 29, no. 4, pp. 61–80, Jul. 2012.
- [39] N. G. Stocks, “Suprathreshold stochastic resonance in multilevel threshold systems,” *Phys. Rev. Lett.*, vol. 84, pp. 2310–2313, 2000.
- [40] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic Resonance To Stochastic Signal Quantization*. New York, NY, USA: Cambridge Univ. Press, 2008.
- [41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.
- [42] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. New York, NY, USA: Springer-Verlag, 1988.
- [43] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [44] M. D. Perlman, “Jensen’s inequality for a convex vector-valued function on an infinite-dimension,” *J. Multivariate Anal.*, vol. 4, pp. 52–65, 1974.

- [45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [46] I. K. Baldry *et al.*, "Quantifying the bimodal color-magnitude distribution of galaxies," *Astrophys. J.*, vol. 600, pp. 681–694, 2004.
- [47] P. Seeman *et al.*, "Bimodal distribution of dopamine receptor densities in brains of schizophrenics," *Science*, vol. 225, pp. 728–731, 1984.
- [48] S. K. Sharma, "Neuronal model with distributed delay: Emergence of unimodal and bimodal ISI distributions," *IEEE Trans. Nanobiosci.*, vol. 12, no. 1, pp. 1–12, Mar. 2013.
- [49] A. Karakoca, U. Erisoglu, and M. Erisoglu, "A comparison of the parameter estimation methods for bimodal mixture Weibull distribution with complete data," *J. Appl. Statist.*, vol. 42, pp. 1472–1489, 2015.
- [50] A. L. M. Vilela, F. G. B. Moreira, and A. J. F. de Souza, "Majority-vote model with a bimodal distribution of noises," *Physica A, Statistical Mech. Appl.*, vol. 391, pp. 6456–6462, 2012.
- [51] H. Lategahn, S. Gross, T. Stehle, and T. Aach, "Texture classification by modeling joint distributions of local patterns with Gaussian mixtures," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1548–1557, Jun. 2010.
- [52] N. M. Blachman, "The convolution inequality for entropy power," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 2, pp. 267–271, Apr. 1965.
- [53] A. W. Roberts and D. E. Varberg, *Convex Functions*. New York, NY, USA: Academic, 1973.
- [54] C. Corduneanu, *Integral Equations and Applications*. New York, NY, USA: Cambridge Univ. Press, 1991.
- [55] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*. Upper Saddle River, NJ, USA: Prentice-Hall, 1997.