Physica A 534 (2019) 120835

Contents lists available at ScienceDirect

Physica A

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

Benefits of noise in M-estimators: Optimal noise level and probability density

Yan Pan [a,c], Fabing Duan [a,*], Liyan Xu [b], François Chapeau-Blondeau [c]

[a] Institute of Complexity Science, Qingdao University, Qingdao 266071, China
[b] School of Electronic Information, Qingdao University, Qingdao 266071, China
[c] Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), Université d'Angers, 62 avenue Notre Dame du Lac, 49000 Angers, France

HIGHLIGHTS

- Noise as a designable variable of the M-estimator due to its benefits.
- Maximizing the asymptotic efficiency of the M-estimator by optimizing the added noise PDF.
- Non-convex optimization of the optimal added noise PDF based on the Parzen-window density estimation technique.
- Approximate solutions of the optimal added noise PDF showing their feasibility with greatly improved asymptotic efficiency.

ARTICLE INFO

Article history:
Received 9 October 2018
Received in revised form 20 February 2019
Available online 18 October 2019

Keywords:
Noise benefit
Asymptotic efficiency
M-estimator
Optimal noise probability density
Parzen-window estimation

ABSTRACT

For the robust estimation of a location parameter, we consider a parallel array of maximum likelihood type estimators (M-estimators). We investigate the possibility of added noise as a design variable of the M-estimators, and characterize a nonzero optimal amount of added noise maximizing the efficiency for estimation. The added noise shows its benefits to the asymptotic efficiency of the M-estimator when the noise level and the noise probability density are optimally tuned. The optimal noise level can be theoretically derived by maximizing the asymptotic efficiency as the probability density of added noise is given. Based on the Parzen-window density estimation technique, we approximate the infinite-dimensional non-convex optimization of the optimal probability density of added noise as a simpler optimization problem with respect to a finite-dimensional vector under certain constraints. This approximate solution for the optimal probability density of added noise shows its feasibility for various M-estimators with an arbitrary array size, which is also validated by simulation results.

1. Introduction

Noise benefits are manifest in nonlinear signal processing systems if a suitable amount of noise (characterized by the optimal noise level [1–5] or the optimal noise type [6–10]) is added into these systems. This constructive action of noise has essentially been reported in physical systems as stochastic resonance coined by Benzi et al. [11], and now has crossed disciplinary boundaries [2,12–26]. Since the term 'stochastic resonance' has a very specific definition for the frequency of nonlinear system output in accordance with that of the weak periodic input signal, many researchers describe stochastic resonance as one example of the potential benefits of noise. In order to exploit noise benefits in a

* Corresponding author.
E-mail address: fabing.duan@gmail.com (F. Duan).

https://doi.org/10.1016/j.physa.2019.04.071
0378-4371/

single system [3,7,25–29] or an parallel array of nonlinearities [5,10,30–32], some researchers tune the noise level for a given noise type [33–39] or find the optimal noise probability density function (PDF) [7,8,10,25,26,36–38,40–44] to optimize the system performance. Among these studies, Chen et al. [7,25,26] first proved that the optimal dichotomous noise can maximize the performances of suboptimal detectors and estimators based on Carathéodory theorem (convex hull). Later, the optimality and effectiveness of the dichotomous noise was also demonstrated in various signal detection and estimation problems [8,36–38,41–43]. Furthermore, for the M-ary hypothesis-testing problem, Bayram et al. proved that the optimal noise PDF includes at most M mass points in the restricted Bayesian framework [45] and in the minimax framework [46]. For transmitting a stochastic signal in a quantizer array with a large array size, McDonnell et al. [10] found that an optimal noise PDF, depending on the input signal distribution in terms of Fisher information, can achieve the maximum channel capacity. Meanwhile, Patel and Kosko [37] proved that the uniform quantizer noise provides the fastest initial decrease in the mean-squared error for the linear estimator of the input signal. In addition, using the Gateaux differential of functionals, Zhai et al. [44] also demonstrated that the optimal uniform noise in an array of binary quantizers leads to the minimal distortion between the input signal and the output decoding.

Recently, in the robust estimation field, we proved that the optimal noise PDF for improving the asymptotic efficiency of an array of maximum likelihood type estimators (M-estimators) is the solution of the weighted minimum $\mathcal{L}^2$-norm with regularization constraints [38]. However, no general or approximate achievable PDF of the optimal added noise was derived. In this letter, based on the results of our work [38], we further investigate the optimal added noise PDF for improving the asymptotic efficiency of a parallel array of M-estimators. It is noted that finding the optimal added noise for a parallel array of M-estimators is actually a non-convex infinite-dimensional optimization problem, and analytic solutions of the optimal noise PDF are quite difficult to obtain in general. Using the Parzen–window density estimation technique [47], we approximate the optimal noise PDF as a solution of an optimization problem with respect to a finite-dimensional non-negative vector under certain constraints. Then, for this simpler optimization problem, we use the interior-point approach [48,49] to find an approximate, but feasible, solution of the optimal noise PDF for enhancing the asymptotic efficiency of M-estimators. It is emphasized that the optimal added noise PDF has non-trivial complicated shapes and varies non-trivially with the array size, the score function of M-estimators and the background noise PDF. Compared with some common added noise PDFs and the case without added noise but with optimal estimator parameters, the obtained approximate form of optimal added noise does provide more benefits to the asymptotic efficiency of the M-estimator. Thus, we argue that the added noise is a useful option to be exploited and designed as a generalized estimator 'parameter'. The obtained results are also validated by numerical simulation.

2. Array of M-estimators and asymptotic efficiency

Consider a location model of observations [50,51]

$$x_n = \theta + w_n, \quad n = 1, 2, \ldots, N \tag{1}$$

where θ is the unknown location parameter and w_n are independent and identically distributed (i.i.d.) white noise components (errors). Here, the background noise w has a symmetric scale-family PDF $f_w(w) = f_{\overline{w}}(w/\sigma_w)/\sigma_w$ and the scale parameter σ_w. Note that σ_w also measures the level of the background noise, because $w = \sigma_w \overline{w}$ and $\overline{w}$ is distributed as the standardized noise PDF $f_{\overline{w}}$ with unity noise level σ_w [10,36–38,51]. With a loss function ρ satisfying certain regularity conditions [50,51], the M-estimator $\hat{\theta}$ for estimating the location parameter θ is defined as

$$\hat{\theta} = \arg\min_{\theta} \sum_{n=1}^{N} \rho(x_n - \theta). \tag{2}$$

If ρ is differentiable, differentiating Eq. (2) with respect to θ, the M-estimator $\hat{\theta}$ also satisfies $\sum_{n=1}^{N} \psi(x_n - \hat{\theta}) = 0$, where the score function $\psi = d\rho/d\theta$ [50,51]. The maximum likelihood estimator is a special case of M-estimators when $\psi = \psi_{\mathrm{M}} = -f_w'/f_w$ ($f_w' = df_w/dx$) and $\rho = -\log f_w$. Here the score function ψ is assumed to be odd and with a bounded derivative $\psi' = d\psi/dx$ [51]. Under regularity conditions, Huber showed [50,52] that a Fisher-consistent M-estimator $\hat{\theta}$ satisfying $\mathrm{E}_w[\psi(x-\theta)] = 0$ is consistent and tends to a Gaussian distribution with mean θ and variance $V(\psi, f_w)/N$, where $V(\psi, f_w) = \mathrm{E}_w[\psi^2(x)]/\mathrm{E}_w^2[\psi'(x)]$ and $\mathrm{E}_w(\cdot) = \int \cdot f_w(x)dx$ stands for the expectation with respect to f_w. Moreover, the optimal estimator, i.e. maximum likelihood estimator, has the minimum variance $[NJ(f_w)]^{-1}$ with the Fisher information $J(f_w) = \mathrm{E}_w[(f_w')^2/f_w^2]$ of the noise PDF f_w [50,51,53]. Due to the unknown PDF f_w of the background noise w, the optimal estimator is unachievable in practice. Then, some easily implemented M-estimators, for instance, Huber or bisquare estimators, are frequently designed with adjustable estimator parameters for the location estimation [50,51]. Therefore, in order to measure how near the variance $V(\psi, f_w)/N$ of the M-estimator $\hat{\theta}$ is to the minimum variance of the optimal estimator, the asymptotic efficiency of $\hat{\theta}$ is defined as [50,51,53]

$$\mathrm{Eff}(\hat{\theta}) = \frac{[NJ(f_w)]^{-1}}{V(\psi, f_w)/N} = \frac{1}{J(f_w)} \frac{\mathrm{E}_w^2[\psi'(x)]}{\mathrm{E}_w[\psi^2(x)]}. \tag{3}$$

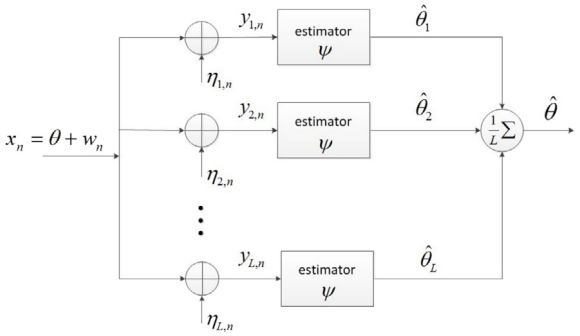

Fig. 1. Block diagram representation of the array of M-estimators with added noises.

We here purposefully add mutually independent noise components $\eta_{l,n}$ into an array of M-estimators, as shown in Fig. 1, and consider the added noise as a design variable of the estimator $\hat{\theta}_l$. Then, a new M-estimator averaging L estimators $\hat{\theta}_l$ is given by

$$\hat{\theta} = \frac{1}{L}\sum_{l=1}^{L}\hat{\theta}_l. \tag{4}$$

As illustrated in Fig. 1, each M-estimator $\hat{\theta}_l$ is subject to the same observations x_n added by the noise components $\eta_{l,n}$. Therefore, the input of each M-estimator $\hat{\theta}_l$ is $y_{l,n} = x_n + \eta_{l,n} = \theta + w_n + \eta_{l,n} = \theta + z_{l,n}$, where noise components $\eta_{l,n}$ are with the common PDF f_η and the composite noise components $z_{l,n} = w_n + \eta_{l,n}$ have a convolved PDF $f_z(z) = \int f_w(z-u)f_\eta(u)du$.

In our previous work [38], using the first-order Taylor expansion of ψ for each M-estimator $\hat{\theta}_l$ around the true value of θ and according to the central limit theorem for a sufficiently large observation size N, we proved that the proposed M-estimator $\hat{\theta}$ of Eq. (4) converges to a Gaussian distribution with mean θ and the asymptotic variance $V(\psi, f_w, f_\eta)/N = \left\{\mathrm{E}_z[\psi^2(z)] + \mathrm{E}_w\{\mathrm{E}_\eta^2[\psi(w+\eta)]\}(L-1)\right\}/(NL\mathrm{E}_z^2[\psi'(z)])$ [38]. Then, from Eq. (3), the asymptotic efficiency of the proposed M-estimator $\hat{\theta}$ in Eq. (4) can be expressed as

$$\mathrm{Eff}_a(\hat{\theta}) = \frac{1}{J(f_w)}\frac{\mathrm{E}_z^2[\psi'(z)]}{\frac{1}{L}\mathrm{E}_z[\psi^2(z)] + \frac{L-1}{L}\mathrm{E}_w\left\{\mathrm{E}_\eta^2[\psi(w+\eta)]\right\}}, \tag{5}$$

with expectations $\mathrm{E}_z(\cdot) = \int \cdot f_z(x)dx$ and $\mathrm{E}_\eta(\cdot) = \int \cdot f_\eta(x)dx$. It is interesting to note that, for a given noisy environment of existing background noise w_n and added noise $\eta_{l,n}$, the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ in Eq. (5) is a monotonically increasing function of the array size L [38]. For a very large size L, the asymptotic efficiency of Eq. (5) approaches

$$\mathrm{Eff}_a^{\infty}(\hat{\theta}) = \lim_{L\to\infty}\mathrm{Eff}_a(\hat{\theta}) \approx \frac{\mathrm{E}_z^2[\psi'(z)]}{J(f_w)\mathrm{E}_w\left\{\mathrm{E}_\eta^2[\psi(w+\eta)]\right\}} = \frac{\mathrm{E}_w^2\{\mathrm{E}_\eta[\psi'(w+\eta)]\}}{J(f_w)\mathrm{E}_w\left\{\mathrm{E}_\eta^2[\psi(w+\eta)]\right\}} \tag{6}$$

with $\lim_{L\to\infty}\mathrm{E}_z[\psi^2(z)]/L = 0$ and $\mathrm{E}_z[\psi^2(z)] < \infty$. In practice, we can parallel as many M-estimator $\hat{\theta}_l$ as possible to approach the limit value of $\mathrm{Eff}_a^{\infty}(\hat{\theta})$ given by Eq. (6) [38].

3. Optimal noise level and PDF

We emphasize that the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ of $\hat{\theta}$ in Eq. (5) is a nonlinear functional of the added noise PDF f_η that can act as a design variable of the M-estimator. In the following section, we will optimize the added noise level σ_η and the added noise PDF f_η to maximize the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ of Eq. (5).

3.1. Optimal noise level σ_η

For a given array size L, noise PDFs f_w and f_η, the level σ_η of the added noise is the tunable parameter. The noise benefit to the asymptotic efficiency can be derived by the condition $\partial\mathrm{Eff}_a(\hat{\theta})/\partial\sigma_\eta > 0$, which indicates the occurrence of the noise-enhanced asymptotic efficiency effect in M-estimators. When the inequality of $\partial\mathrm{Eff}_a(\hat{\theta})/\partial\sigma_\eta > 0$ holds for $\sigma_\eta \geq 0$, then σ_η can be gradually raised above zero up to the level $\sigma_\eta^{\mathrm{opt}} > 0$ satisfying the condition

$$\left.\frac{\partial\mathrm{Eff}_a(\hat{\theta})}{\partial\sigma_\eta}\right|_{\sigma_\eta=\sigma_\eta^{\mathrm{opt}}} = 0 \tag{7}$$

and in this circumstance σ_η^{opt} is an optimal noise level achieving a maximum of the asymptotic efficiency. Only when the solution to Eq. (7) is unique can we expect to have a global maximum. In practice, the solution of Eq. (7) we show in Figs. 2(b) and (c), although we cannot rigorously prove it corresponds to a global maximum due to the complicated form of the derivative of the asymptotic efficiency in Eq (7), offers a useful and meaningful improvement obtained by adding noise. Especially it substantiates the claim that enhancement of the efficiency $\text{Eff}_a(\hat{\theta})$ can be obtained by injecting a nonzero level of added noise. For instance, for a sufficiently large size L and from Eq. (6), the condition of Eq. (7) can be rewritten as

$$\frac{\text{E}_w\{\partial \text{E}_\eta[\psi'(w+\eta)]/\partial\sigma_\eta\}}{\text{E}_w\{\text{E}_\eta[\psi'(w+\eta)]\}} = \frac{\text{E}_w\{\text{E}_\eta[\psi(w+\eta)]\partial \text{E}_\eta[\psi(w+\eta)]/\partial\sigma_\eta\}}{\text{E}_w\{\text{E}_\eta^2[\psi(w+\eta)]\}}, \tag{8}$$

which yields an optimal noise level σ_η^{opt} if it exists.

For instance, consider the Huber estimator [52]

$$\psi(x) = \begin{cases} x, & |x| \le \gamma, \\ \gamma\ \text{sgn}(x), & |x| > \gamma, \end{cases} \tag{9}$$

where $\gamma \ge 0$ is the estimator parameter. When $\gamma \to 0$, the Huber function of Eq. (9) becomes $\psi(x) = \text{sgn}(x)$ associated with the median estimator, and as $\gamma \to \infty$ the estimator yields the sample mean of the observation data. Specially, consider the background noise with Cauchy PDF $f_w(w) = \pi^{-1}(1+w^2)^{-1}$, Student PDF $f_w(w) = (3\pi)^{-1/2}\Gamma^{-1}(3/2)(1+w^2/3)^{-2}$ and Laplacian PDF $f_w(w) = \exp(-|w|)/2$, and these noise PDFs all have the unity level of σ_w. The added Gaussian noise PDF is $f_\eta(\eta) = \exp(-\eta^2/2\sigma_\eta^2)/(\sqrt{2\pi}\sigma_\eta)$ with the noise level σ_η. Here, the standard deviation σ_w of the background noise is set as the unit of the signal amplitude, to which all other signal amplitudes occurring in the problem will be referred. We also provide in Appendix C, a mathematical proof that the asymptotic efficiencies of Eq. (5) for the Huber estimator and the bisquare estimator are functions of the dimensionless variables γ/σ_w and σ_η/σ_w.

It is shown in Fig. 2(a) that, as the level σ_η of the added noise increases and for a fixed estimator parameter $\gamma/\sigma_w = 0.1$ in Eq. (9), the asymptotic efficiency $\text{Eff}_a^{\infty}(\hat{\theta})$ of Eq. (6) shows the phenomenon of noise enhancement for Cauchy noise and Student noise. For the situation of Laplacian noise and the given parameter $\gamma/\sigma_w = 0.1$, the addition of noise degrades the asymptotic efficiency, as shown in Fig. 2(a). An alternative criterion, being capable of assessing noise benefit effects, is to take as a reference the performance of the suboptimal array with no added noise instead of the best performance achievable by the optimal estimator in the asymptotic efficiency we used, which can be written as the ratio $G(\psi, f_w, f_\eta) = V(\psi, f_w)/V(\psi, f_w, f_\eta)$. From the definition of the asymptotic efficiency $\text{Eff}_a(\hat{\theta}) = 1/(NJ(f_w))/(V(\psi, f_w, f_\eta)/N)$, the criterion $G = (V(\psi, f_w)J(f_w))\text{Eff}_a(\hat{\theta})$ can be directly deduced from the asymptotic efficiency, where $V(\psi, f_w)J(f_w)$ is a constant that does not vary with the added noise. So G will experience the same evolution with the added noise as the asymptotic efficiency. As in Fig. 2(a), when there is a resonant evolution of $\text{Eff}_a(\hat{\theta})$ at a maximum for an optimal level σ_η^{opt} of the added noise, there will also be a resonant evolution of G culminating at a maximum for the same optimal level σ_η^{opt} of the added noise. Using the condition of the optimal noise level σ_η^{opt} of Eq. (7), we plot the optimal noise level σ_η^{opt} as a function of the estimator parameter γ in Fig. 2(b). Here, the numerical algorithm of σ_η^{opt} in Eq. (7) is a combination of bisection, secant, and inverse quadratic interpolation methods [54,55] within the positive real number field.

It is interesting to note that, as the estimator parameter γ varies, a non-zero optimal noise level exists for Student noise ($0 \le \gamma/\sigma_w \le 4$), Laplacian noise ($\gamma/\sigma_w > 0.2$) and Cauchy noise ($0 \le \gamma/\sigma_w \le 0.4$ and $\gamma/\sigma_w > 1.95$), which closely depends upon the PDF f_w of the background noise, the PDF f_η of the added noise and the array size L. Fig. 2(c) shows the optimal noise level σ_η^{opt} versus the array size L. It is seen in Fig. 2(c) that, only as the array size L is larger than certain values (e.g. $L > 10$ for Cauchy background noise), an optimal noise level σ_η^{opt} exists. Here the Huber estimator parameter $\gamma/\sigma_w = 4$. In Fig. 2(c), we also see that the optimal noise level σ_η^{opt} increases as the array size L grows. As the array size L is sufficiently large, the value of the optimal noise level σ_η^{opt} tends to a constant for the three added noise types.

3.2. Optimal noise density f_η

The condition of Eq. (7) for the noise benefit is based on the design variable of the added noise level σ_η, but the PDFs f_w, f_η, and the array size L are given. In practice, the background noise is often unavoidable (the noise PDF f_w is fixed), while the added noise PDF f_η can be artificially designed. Naturally, in the following parts, an interesting question is how to optimally design the PDF f_η (including the noise level σ_η) of the added noise to improve the asymptotic efficiency $\text{Eff}_a(\hat{\theta})$ in Eq. (5), which can be viewed as an optimization problem

$$f_\eta^{\text{opt}} = \max_{f_\eta} \frac{\text{E}_z^2[\psi'(z)]}{\frac{1}{L}\text{E}_z[\psi^2(z)] + \frac{L-1}{L}\text{E}_w\left\{\text{E}_\eta^2[\psi(w+\eta)]\right\}} \tag{10}$$

with constraints of the symmetric function $f_\eta(-x) = f_\eta(x)$ for ensuring the Fisher consistency of the estimator, i.e. $\text{E}_z[\psi(z)] = 0$ [38], and regularity conditions of $f_\eta(x) \ge 0$ ($x \in \mathbb{R}$) and $\int f_\eta(x)dx = 1$. It is noted in Eq. (5) that the Fisher information $J(f_w)$ of the noise density f_w is also a fixed quantity, which is independent of the optimal noise PDF f_η^{opt} in Eq. (10).

According to the definition of convex functionals [56], the asymptotic efficiency of Eq. (5) with respect to the added noise PDF f_η is a non-convex functional, and then the optimization problem of Eq. (10) is non-convex. Thus, the

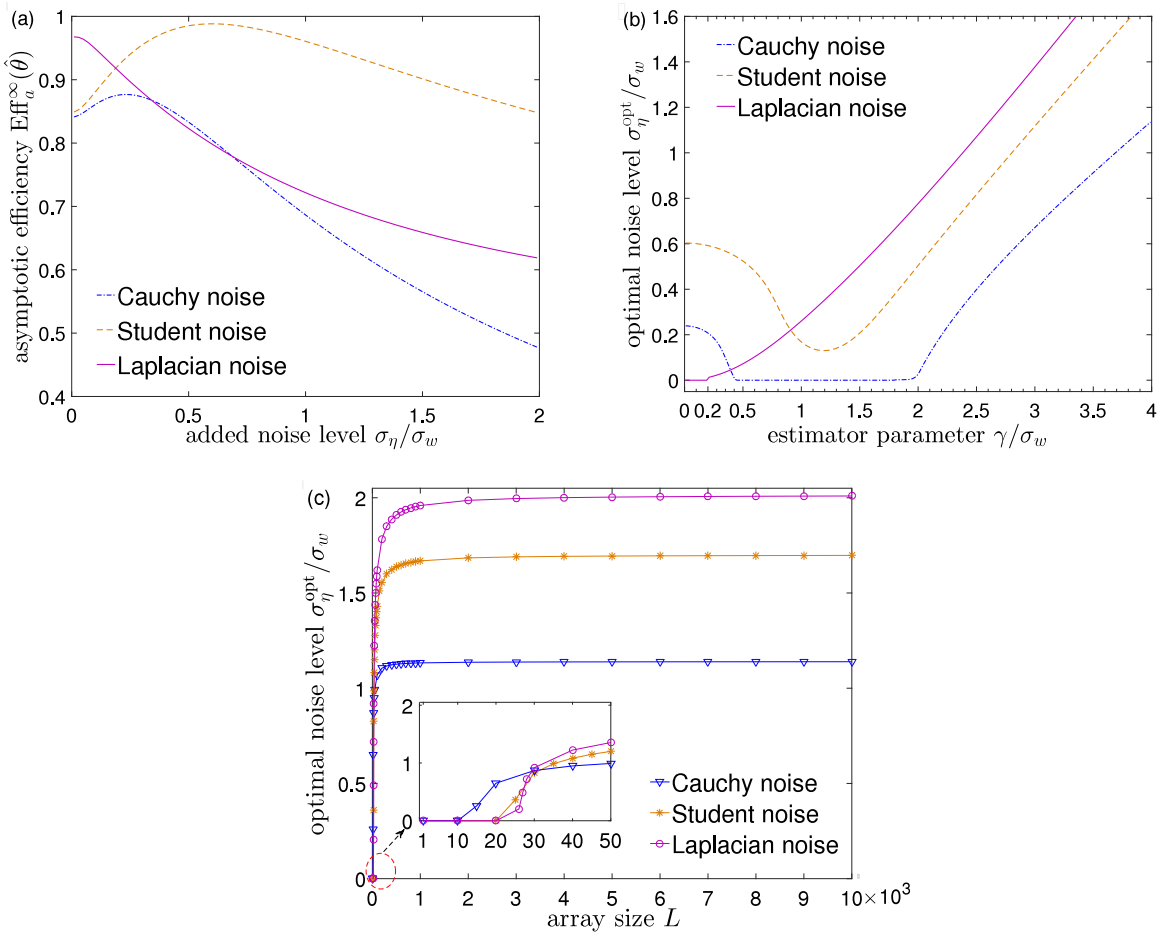

Fig. 2. (a) Asymptotic efficiency $\mathrm{Eff}_a^{\infty}(\hat{\theta})$ of Eq. (6) of the Huber estimator in Eq. (9) with $\gamma/\sigma_w = 0.1$ as a function of the level σ_η of the added noise; (b) The optimal level $\sigma_\eta^{\mathrm{opt}}$ of the added noise versus the Huber estimator parameter γ. (c) The optimal level $\sigma_\eta^{\mathrm{opt}}$ of the added noise versus the array size L for the estimator parameter $\gamma/\sigma_w = 4$.

optimization problem of Eq. (10) over the space of all possible PDFs of f_η has high computational complexity, and is difficult to solve theoretically. Here, we employ an approximation technique of the Parzen–window density estimation [57] to approximate the PDF f_η^{opt} of the optimal added noise [47] as

$$\tilde{f}_\eta^{\mathrm{opt}}(\eta) = \sum_{k=1}^{K} \nu_k r_k(\eta - u_k), \tag{11}$$

where the normalization coefficients $\nu_k \geq 0$ and $\sum_{k=1}^{K} \nu_k = 1$, and window functions $r_k(\cdot)$ satisfy $r_k(u) \geq 0$ and $\int r_k(u)du = 1$ for $k = 1, 2, \ldots, K$ [47]. As the number K of window functions increases, the estimation form $\tilde{f}_\eta^{\mathrm{opt}}(\eta)$ of Eq. (11) gradually converges to $f_\eta^{\mathrm{opt}}(\eta)$ in Eq. (10) under certain conditions [47]. We further solve the approximate PDF $\tilde{f}_\eta^{\mathrm{opt}}$ in a finite symmetric interval $[-a, a]$ ($a > 0$), and the interval bound a needs to be iteratively chosen. For $0 < \Delta a \ll 1$, if the corresponding calculation values of the asymptotic efficiency satisfy $|\mathrm{Eff}_a(\hat{\theta}, a' + \Delta a) - \mathrm{Eff}_a(\hat{\theta}, a')| < \varepsilon$ for a sufficiently small positive number ε, then we obtain the interval bound $a = a'$.

We divide the interval $[-a, a]$ into K sub-intervals $\Lambda_i = [a_{k-1}, a_k]$ with equal width $\Delta u = 2a/K$, where $a_k = a_0 + k\Delta u$ and $a_0 = -a$ and the midpoint of each sub-interval is $u_k = (a_{k-1} + a_k)/2$ for $k = 1, 2, \ldots, K$. Using this discretization method, we prove in Appendix A that the solution of the optimization problem of Eq. (10) becomes the simpler optimization problem of Eq. (A.5) with respect to a finite-dimensional vector $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_K]^\top$. The optimization problem of Eq. (A.5) is still non-convex in general, and some optimization techniques, such as particle-swarm optimization [45,58–60], penalty function approaches [61,62], the sequential quadratic programming algorithm [62,63], can be employed. Here, we use a kind of penalty function, i.e. the interior-point technique [48,49,64], to solve the optimization problem of Eq. (A.5), whereby a barrier function is added into the objective function in order to replace

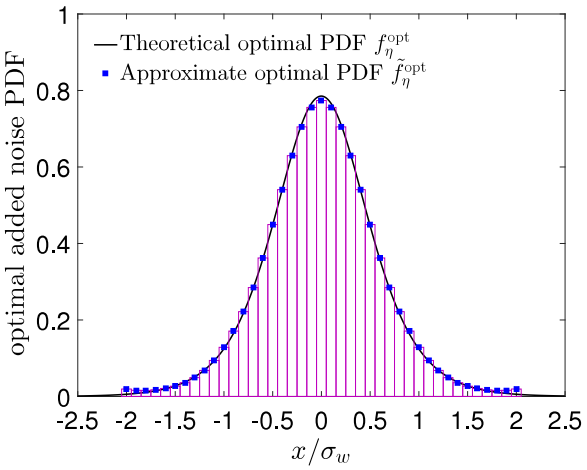

Fig. 3. Approximate form $\tilde{f}_\eta^{\rm opt}$ (□) of the optimal added noise PDF in Eq. (11) and the theoretical optimal noise PDF $f_\eta^{\rm opt}$ (solid line) in Eq. (10) for $L\to\infty$, the hyperbolic secant background noise and the median estimator.

the inequality constraints. Then, we can transfer the problem of Eq. (A.5) to a sequence of barrier subproblems of the form of Eq. (A.7), yielding the approximate form $\tilde{f}_\eta^{\rm opt}$ in Eq. (11) of the optimal noise PDF.

In order to validate the above-mentioned approach of Eq. (11), the hyperbolic secant background noise PDF $f_w(x) = \mathrm{sech}(\pi x/2)/2$ with the unity level of σ_w and the median estimator $\psi(x) = \mathrm{sign}(x)$ are considered. The hyperbolic secant distribution is an interesting model as it is close to the Gaussian distribution at small amplitude, while it tends to an exponential distribution at large amplitude [65]. It is for instance applied in the context of financial return data [66,67]. Under these circumstances, it is theoretically proven by the Fourier transform and its inverse transform [38] that the optimal Logistic noise PDF, as a solution of the optimization of Eq. (10) for $L\to\infty$, has the form $f_\eta^{\rm opt}(x) = \pi\,\mathrm{sech}^2(\pi x/2)/4$, as shown in Fig. 3 (solid line). The median estimator can benefit from the addition of the Logistic noise components $\eta_{l,n}$ and the asymptotic efficiency in Eq. (6) can be improved to reach the upper bound of unity, i.e. $\mathrm{Eff}_a^\infty(\hat{\theta}) = 1$ [38]. The approximate solution $\tilde{f}_\eta^{\rm opt}$ (□) of Eq. (11) is also plotted in Fig. 3 in the interval $[-a/\sigma_w, a/\sigma_w] = [-2, 2]$, where a rectangular window function $r_k(\eta) = 1/\Delta u$ for $|\eta| \le \Delta u/2$ and zero otherwise, and the step $\Delta u/\sigma_w = 0.1$. It is clearly visible in Fig. 3 that the approximate form $\tilde{f}_\eta^{\rm opt}(\eta)$ of the optimal added noise (□) is consistent with the theoretical one $f_\eta^{\rm opt}$ (solid line) in the considered solution interval. Moreover, substituting the approximate form $\tilde{f}_\eta^{\rm opt}$ of Eq. (11) into Eq. (6), we obtain the corresponding asymptotic efficiency $\mathrm{Eff}_a^\infty(\hat{\theta}) = 0.995$, which is very close to unity. This result shows the feasibility of the approximate form $\tilde{f}_\eta^{\rm opt}$ in Eq. (11). As seen in Fig. 3, the tails of the approximate solution of $\tilde{f}_\eta^{\rm opt}$ go up a little bit. Because of the normalization of the finitely-supported probability density $\tilde{f}_\eta^{\rm opt}$ approximating the theoretical density $f_\eta^{\rm opt}$ with infinite support, the very small probability of the theoretical optimal solution $f_\eta^{\rm opt}$, neglected outside the interval of $[-a/\sigma_w, a/\sigma_w]$ in the tails, shows up around the two limits of the interval supporting the finite-dimensional solution $\tilde{f}_\eta^{\rm opt}$.

Next, we consider the background Cauchy noise and the redescending bisquare M-estimator with the score function

$$\psi(x) = x\left[1 - (x/\gamma)^2\right]^2 \tag{12}$$

for $|x| \le \gamma$ ($\gamma > 0$) and otherwise zero. Without the added noise components $\eta_{l,n}$, the optimal estimator parameter is $\gamma/\sigma_w = 3.3$ and the corresponding maximum asymptotic efficiency $\mathrm{Eff}(\hat{\theta}) = 0.902$ (see Fig. 5(b)). For different array sizes L, Fig. 4 plots the approximate forms of $\tilde{f}_\eta^{\rm opt}$ in Eq. (11). It is seen in Fig. 4(a) that, for the array size $L = 1$, the approximate optimal noise PDF $\tilde{f}_\eta^{\rm opt}(u) = 1/\Delta u$ for $|u| < \Delta u/2$, which means that the addition of noise is not beneficial to the asymptotic efficiency of the M-estimator. For $L = 100$, 1000 and 3000, the corresponding approximate optimal noise PDFs are shown in Figs. 4(b), (c) and (d), respectively. Also of interest, with the obtained approximate optimal noise PDFs and from Eq. (A.5), the corresponding asymptotic efficiencies $\mathrm{Eff}_a(\hat{\theta})$ of Eq. (5) are calculated as 0.934 ($L = 100$), 0.981 ($L = 1000$) and 0.990 ($L = 3000$), which are all higher than the value of 0.902 obtained without added noise. It is seen in Fig. 4 that the approximate PDFs $\tilde{f}_\eta^{\rm opt}(u)$ of the optimal noise have non-trivial complicated shapes and vary non-trivially with the array size L. We also calculate the optimal noise PDFs in other configurations, for instance, the Huber estimator under Cauchy or Logistic background noise (not shown here), which also present complicated shapes and vary with the array size, the background noise PDF and the score function of the M-estimator. Consequently, it is difficult to analytically characterize the optimal noise PDF in general, whence the approximate from of the optimal noise PDF is very practical.

In Fig. 5(a), the asymptotic efficiency, as the array size L increases, benefits more and more from the addition of optimal noise components $\eta_{l,n}$ with PDF $\tilde{f}_\eta^{\rm opt}$. In practice, for an arbitrary estimator parameter γ and a given array size $L = 1000$,

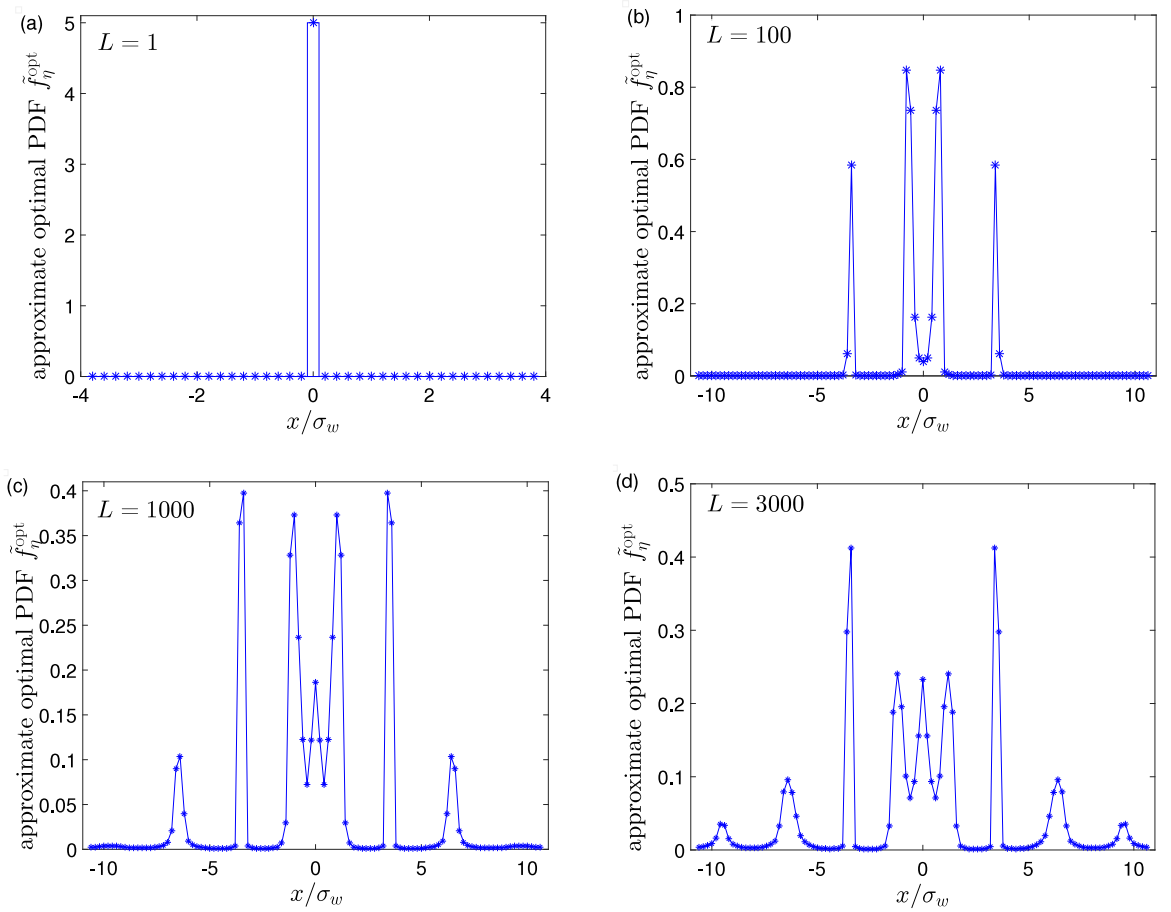

Fig. 4. Approximate forms $\tilde{f}_\eta^{\rm opt}$ in Eq. (11) of the optimal noise PDF for different array sizes (a) $L = 1$, (b) $L = 100$, (c) $L = 1000$ and (d) $L = 3000$. Here, the background noise is Cauchy distributed and the redescending bisquare M-estimator of Eq. (12) has the estimator parameter $\gamma/\sigma_w = 3.3$. The interval bound value $a/\sigma_w = 10.7$ and the step $\Delta u/\sigma_w = 0.2$ in Eq. (A.5).

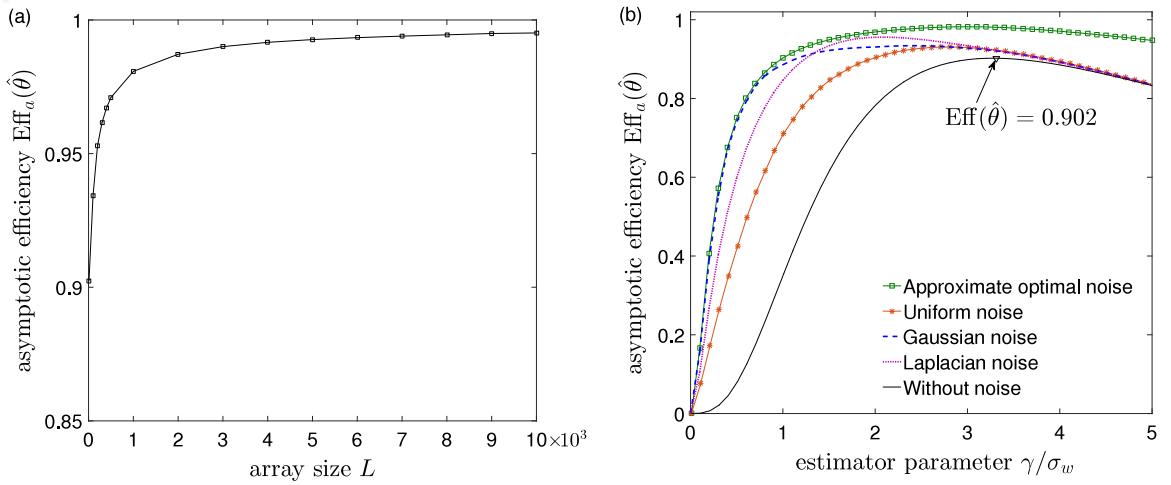

Fig. 5. Asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ of Eq. (5) obtained by the approximate optimal noise PDF $\tilde{f}_\eta^{\rm opt}$ of Eq. (11) versus (a) the array size L for the parameter $\gamma/\sigma_w = 3.3$ of the bisquare estimator in Eq. (12); (b) the parameter γ of the bisquare M-estimator in Eq. (12) for the approximate optimal noise (□ plus solid line), uniform noise (∗ plus solid line), Gaussian noise (dashed line), Laplacian noise (dotted line) and the case of no added noise (solid line). Here, the array size $L = 1000$.

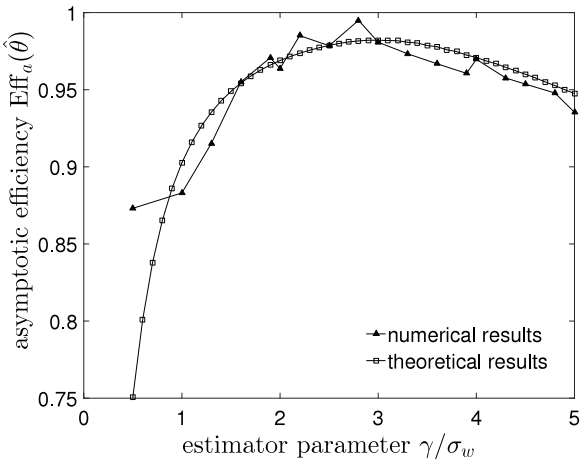

Fig. 6. Numerical (▲) and theoretical (□) results of the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ of Eq. (5) obtained by the approximate form $\tilde{f}_\eta^{\mathrm{opt}}$ of the optimal nose PDF as a function of the parameter γ of the bisquare estimator in Eq. (12). Here, the simulation results of the asymptotic efficiency (▲) are computed by 10^4 Monte Carlo trials, the observation size $N = 6000$ and the tolerance parameter $\zeta/\sigma_w = 10^{-7}$. The other parameters are the same as in Fig. 5(b).

we always find an approximate PDF $\tilde{f}_\eta^{\mathrm{opt}}$ of the optimal noise that yields an improved asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ (□ plus solid line), as shown in Fig. 5(b). For comparison, we also plot the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ for the addition of uniform noise (∗ plus solid line), Gaussian noise (dashed line), Laplacian noise (dotted line) by optimally tuning their noise levels. Moreover, the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ in the case of no added noise (solid line) is also illustrated in Fig. 5(b). The comparison results show that the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ (□ plus solid line) obtained by the approximate optimal noise PDF $\tilde{f}_\eta^{\mathrm{opt}}$ is actually better.

In Appendix B, the procedure of the numerical simulation for the asymptotic efficiency $\mathrm{Eff}_a(\hat{\theta})$ of the estimator $\hat{\theta}$ in Eq. (4) by the iterative reweighting method is briefly introduced. The simulation results are presented in Fig. 6, and the optimal noise components $\eta_{l,n}$ are generated according to the approximate solution $\tilde{f}_\eta^{\mathrm{opt}}$ of the optimal noise. It is visible in Fig. 6 that the numerical results of $\mathrm{Eff}_a(\hat{\theta})$ are almost well consistent with the theoretical values, which validates the practicability of the approximate noise PDF $\tilde{f}_\eta^{\mathrm{opt}}$ for enhancing the asymptotic efficiency of the M-estimator. We also note that the small deviation between numerical and theoretical results might be caused by the limited simulation numbers of the random samples and the difference between the sample distribution and the theoretical distribution. Due to some of the approximations made in the derivations of Eq. (11), there might be also a deviation between the approximate and the true PDF of the optimal noise. Therefore, a more accurate theoretical model needs to be established to reduce the deviation in future studies.

4. Conclusion and discussion

In this paper, we investigate the noise benefit to the asymptotic efficiency of M-estimators arrayed in parallel and treat the added noise as a design variable of M-estimators. It is shown that the asymptotic efficiency of an array of M-estimators can be optimized by not only adjusting the added noise level, but also the added noise PDF. With the theoretical expression of the asymptotic efficiency, the optimal added noise level can be derived as the noise type is given. For the non-convex optimization of the added noise PDF, we use the Parzen-window density estimation approach to approximate the optimal added noise PDF. Then, the infinite-dimensional non-convex optimization of the optimal noise PDF becomes a simpler optimization problem with respect to a finite-dimensional vector under certain constrained conditions. Using this approximate approach, we characterize the optimal noise PDF for various M-estimators, background noise types, and different array sizes. The obtained optimal noise PDFs varies non-trivially with the array size, the background noise distribution and the selected estimator. This result indicates that the optimal noise PDF has a complicated shape that is difficult to describe theoretically, and the proposed approximate approach is very meaningful for practical signal estimation problems. Compared with some common noise distributions and the M-estimator with optimal parameter (without added noise), the optimal noise indeed provides a greatly improved asymptotic efficiency of M-estimators. The numerical simulation results also support the above-mentioned theoretical analysis.

For the non-convex optimization problem of Eq. (A.5), various optimization approaches, such as particle-swarm optimization [68–70], feasible point pursuit [71], genetic algorithms and differential evolution [72], could also be tested for finding the optimal solution of added noise PDFs. In this paper, we only focus on the parallel array of identical estimators in the field of robust estimation. Now, the addition of noise in nonlinear signal processor is recognized as a possibly useful

option to contribute to signal processing techniques. Noise, as well as the processor parameters [5,7,8,10,25,26,30–32,34–44] can be explored as a new design variable for the optimization of the performances of processors. Thus, the considered approximate approach that finds the optimal noise PDF deserves to be further investigated in the weak signal detection and transmission. The Parzen-window density estimation approach applies to the cases of the optimal noise probability density f_η^{opt} being mainly contained in the considered finite interval, and the accuracy of the approximate form $\tilde{f}_\eta^{\text{opt}}$ is quite high. Beyond these cases, some feasible density estimation approaches need to be developed. The results we show are specific of a given background noise and a given score function for the estimator, but the approach we developed can be applied to any other choices for these aspects. We propose an interesting hypothesis that, if the considered estimator is not the maximum likelihood estimator for the location parameter in the background noise, an optimal added noise certainly exists and is of great benefit to the asymptotic efficiency of an array of suboptimal estimators. The generality of this hypothesis deserves to be tested and verified.

Acknowledgment

This work is sponsored by the National Science Foundation of China under the Grant No. 61573202.

Appendix A. Approximate form of the optimal noise PDF

In an interval $[-a, a]$ and for a fixed equal step Δu, the approximate form of noise PDF $\tilde{f}_\eta^{\text{opt}}$ in Eq. (11) can be utilized for the simplification of the expectations in Eq. (10). Then, we have

$$\begin{aligned}\mathrm{E}_\eta[\psi(w+\eta)] &\approx \int_{-a}^{a}\psi(w+\eta)\tilde{f}_\eta^{\text{opt}}(\eta)d\eta = \int_{-a}^{a}\psi(w+\eta)\sum_{i=1}^{K}\nu_i r_i(\eta-u_i)d\eta \\ &= \sum_{i=1}^{K}\nu_i q_i(w) = \boldsymbol{\nu}^\top \boldsymbol{q}(w),\end{aligned} \tag{A.1}$$

where vectors $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_K]^\top$ and $\boldsymbol{q}(w) = [q_1(w), q_2(w), \ldots, q_K(w)]^\top$ with its element $q_i(w) = \int_{\Lambda_i}\psi(w+\eta)r_i(\eta-u_i)d\eta$ and $\Lambda_i = [a_{i-1}, a_i]$. The expectation $\mathrm{E}_w\{\mathrm{E}_\eta^2[\psi(w+\eta)]\}$ in Eq. (10) can be simplified as

$$\mathrm{E}_w\{\mathrm{E}_\eta^2[\psi(w+\eta)]\} = \mathrm{E}_w[\boldsymbol{\nu}^\top\boldsymbol{q}(w)\boldsymbol{q}^\top(w)\boldsymbol{\nu}] = \boldsymbol{\nu}^\top\mathrm{E}_w[\boldsymbol{q}(w)\boldsymbol{q}^\top(w)]\boldsymbol{\nu} = \boldsymbol{\nu}^\top\boldsymbol{\varphi}\boldsymbol{\nu}, \tag{A.2}$$

where the matrix $\boldsymbol{\varphi}$ has its element $\varphi_{i,j} = \mathrm{E}_w[q_i(w)q_j(w)]$ for $1 \le i, j \le K$. The expectation $\mathrm{E}_z^2[\psi'(z)]$ in Eq. (10) can be rewritten as

$$\begin{aligned}\mathrm{E}_z^2[\psi'(z)] &= \mathrm{E}_w^2\{\mathrm{E}_\eta[\psi'(w+\eta)]\} = \mathrm{E}_w^2\{\mathrm{E}_\eta[\psi(w+\eta)]\psi_{\mathrm{M}}(w)\} \\ &\approx \mathrm{E}_w^2[\boldsymbol{\nu}^\top\boldsymbol{q}(w)\psi_{\mathrm{M}}(w)] = \boldsymbol{\nu}^\top\mathrm{E}_w[\boldsymbol{q}(w)\psi_{\mathrm{M}}(w)]\mathrm{E}_w[\boldsymbol{q}(w)\psi_{\mathrm{M}}(w)]^\top\boldsymbol{\nu} \\ &= \boldsymbol{\nu}^\top\overline{\boldsymbol{\varphi}}\boldsymbol{\nu},\end{aligned} \tag{A.3}$$

where the matrix $\overline{\boldsymbol{\varphi}}$ is with the element $\overline{\varphi}_{i,j} = \mathrm{E}_w[q_i(w)\psi_{\mathrm{M}}(w)]\mathrm{E}_w[q_j(w)\psi_{\mathrm{M}}(w)]$. Similarly, $\mathrm{E}_z[\psi^2(z)]$ can be approximated as

$$\begin{aligned}\mathrm{E}_z[\psi^2(z)] &= \mathrm{E}_\eta\{\mathrm{E}_w[\psi^2(w+\eta)]\} \approx \int_{-a}^{a}\sum_{i=1}^{K}\nu_i r_i(\eta-u_i)\mathrm{E}_w[\psi^2(w+\eta)]d\eta \\ &= \sum_{i=1}^{K}\nu_i s_i = \boldsymbol{s}^\top\boldsymbol{\nu}\end{aligned} \tag{A.4}$$

with $\boldsymbol{s} = [s_1, \ldots, s_K]^\top$ and its ith element $s_i = \int_{\Lambda_i}\mathrm{E}_w[\psi^2(w+\eta)]r_i(\eta-u_i)d\eta$. Substituting Eqs. (A.2)–(A.4) into Eq. (10), the infinite-dimensional non-convex optimization of Eq. (10) becomes a simpler optimization problem

$$\begin{aligned}&\max_{\boldsymbol{\nu}} \ \frac{\boldsymbol{\nu}^\top\overline{\boldsymbol{\varphi}}\boldsymbol{\nu}}{\frac{1}{L}\boldsymbol{s}^\top\boldsymbol{\nu} + \frac{L-1}{L}\boldsymbol{\nu}^\top\boldsymbol{\varphi}\boldsymbol{\nu}}, \\ &\text{s.t. } \nu_i \ge 0\ ,\ i = 1, 2, \ldots, K, \\ &\quad \sum_{i=1}^{K}\nu_i = 1,\ \boldsymbol{\nu} = \widetilde{\boldsymbol{\nu}},\end{aligned} \tag{A.5}$$

where $\widetilde{\boldsymbol{\nu}} = [\nu_K, \nu_{K-1}, \ldots, \nu_1]^\top$ because of the symmetry of the added noise PDF f_η.

Due to the positive terms $\boldsymbol{\nu}^\top\overline{\boldsymbol{\varphi}}\boldsymbol{\nu}$, $\boldsymbol{s}^\top\boldsymbol{\nu}$ and $\boldsymbol{\nu}^\top\boldsymbol{\varphi}\boldsymbol{\nu}$, the maximization problem of Eq. (A.5) can be converted into the equivalent minimization of the reciprocal of its objective function, i.e.

$$\min_{\boldsymbol{\nu}} B(\boldsymbol{\nu}) = \frac{1}{L}\frac{\boldsymbol{s}^\top\boldsymbol{\nu}}{\boldsymbol{\nu}^\top\overline{\boldsymbol{\varphi}}\boldsymbol{\nu}} + \frac{L-1}{L}\frac{\boldsymbol{\nu}^\top\boldsymbol{\varphi}\boldsymbol{\nu}}{\boldsymbol{\nu}^\top\overline{\boldsymbol{\varphi}}\boldsymbol{\nu}}, \tag{A.6}$$

with the same constraints in Eq. (A.5). Here, we use the interior-point approach [48,49,64] to solve the optimization problem of Eq. (A.6) as

$$\begin{aligned} &\min_{\boldsymbol{\nu},\lambda} g(\boldsymbol{\nu},\lambda) = B(\boldsymbol{\nu}) - \lambda \sum_{i=1}^{K} \log(\nu_i) \\ &\text{s.t.} \sum_{i=1}^{K} \nu_i = 1, \ \ \boldsymbol{\nu} = \widetilde{\boldsymbol{\nu}}, \end{aligned} \tag{A.7}$$

where $-\lambda \sum_{i=1}^{K} \log(\nu_i)$ is the barrier function and the barrier parameter λ is a small positive scalar. When λ decreases to zero [48,49,62,64], the minimum of Eq. (A.7) converges to the solution of the problem of Eq. (A.6). To solve the constrained minimization problem of (A.7), many methods such as Newton algorithm [73], Davidon–Fletcher–Powell algorithm (DFP) [74,75] and Broyden–Fletcher–Goldfarb–Shanno (BFGS) method [75] can be used. We can also make use of the professional software of Matlab or Lingo. Here, we use the Matlab function of 'fmincon' to solve this minimization constrained problem of Eq. (A.6) and set the option parameter of 'Algorithm' as 'interior-point' [48,49,62,64]. Interestingly, for the limit case of $L \to \infty$ and the terms $1/L \to 0$ in $B(\boldsymbol{\nu})$, the optimization problem of (A.5) becomes

$$\max_{\boldsymbol{\nu}} \frac{\boldsymbol{\nu}^{\top} \overline{\boldsymbol{\varphi}} \boldsymbol{\nu}}{\boldsymbol{\nu}^{\top} \boldsymbol{\varphi} \boldsymbol{\nu}}, \tag{A.8}$$

which can be regarded as the maximization of the generalized Rayleigh quotient with the same constraints as in Eq. (A.5).

Appendix B. Numerical simulation of the asymptotic efficiency with the benefit of the optimal noise

The estimator $\hat{\theta}$ in Eq. (2) can be computed as a weighted mean [51]

$$\hat{\theta} = \frac{\sum_{n=1}^{N} W_n \, x_n}{\sum_{n=1}^{N} W_n}, \tag{B.1}$$

with $W_n = W(x_n - \hat{\theta})$ and

$$W(x) = \begin{cases} \psi(x)/x, & x \neq 0, \\ \psi'(x), & x = 0. \end{cases} \tag{B.2}$$

The weighted average expression of Eq. (B.1) suggests an iterative procedure for computing the M-estimator $\hat{\theta}$ in Eq. (2) by using the updated samples generated by the optimal noise distribution. From Eq. (11) and Eq. (A.7), the approximate form of the optimal noise PDF $\tilde{f}_{\eta}^{\text{opt}}$ is solved. Then the random noise components are generated by the acceptance–rejection method [76] according to $\tilde{f}_{\eta}^{\text{opt}}$, yielding L groups of mutually independent noise samples $\eta_{l,n}$ with the length of N ($1 \le l \le L$ and $n = 1, 2, \ldots, N$).

Then, by using the iterative reweighting method of Eq. (B.1), compute the M-estimators $\hat{\theta}_l$ ($1 \le l \le L$) with L groups of noisy observations, respectively. For each estimator $\hat{\theta}_l$, the steps of the iterative reweighting method start from an initial estimator $\hat{\theta}_l(0)$ as the median of the updated observations $y_l = (y_{l,1}, y_{l,2}, \ldots, y_{l,N})$. According to Eq. (B.1), compute the kth iteration value of $\hat{\theta}_l(k) = \sum_{n=1}^{N} W_{k-1,ln} \, y_{l,n} / \sum_{n=1}^{N} W_{k-1,ln}$ with $W_{k,ln} = W(y_{l,n} - \hat{\theta}_l(k))$. The iteration procedure continues until $|\hat{\theta}_l(k) - \hat{\theta}_l(k-1)| < \zeta$, where ζ is a small positive tolerance parameter. Finally, the proposed estimator $\hat{\theta}$ is computed by averaging L estimators $\hat{\theta}_l$ as $\sum_{l=1}^{L} \hat{\theta}_l / L$. The numerical results of the M-estimator $\hat{\theta}$ are achieved by 10^4 Monte Carlo trials, and then the sample mean and variance of the M-estimator $\hat{\theta}$ can be numerically calculated. Substituting the sample variance into Eq. (3), the numerical value of the asymptotic efficiency $\text{Eff}_a(\hat{\theta})$ is obtained.

Appendix C. Asymptotic efficiency of M-estimators for the scale family of noise PDFs

We consider the scale-family PDFs $f_w(w) = f_{\overline{w}}(w/\sigma_w)/\sigma_w$ of the background noise and $f_\eta(\eta) = f_{\overline{\eta}}(\eta/\sigma_\eta)/\sigma_\eta$ of the added noise, where $f_{\overline{w}}$ and $f_{\overline{\eta}}$ are the standardized PDFs with unity scale. The noise cumulative distribution function (CDF) of the added noise satisfies $F_\eta(\eta) = F_{\overline{\eta}}(\eta/\sigma_\eta)$ and with $F_{\overline{\eta}}$ being the standardized CDF of $f_{\overline{\eta}}$. $J(f_w) = J(f_{\overline{w}})/\sigma_w^2$ and $J(f_{\overline{w}})$ is the Fisher information of $f_{\overline{w}}$ [53]. We can derive that the asymptotic efficiencies $\text{Eff}_a(\hat{\theta})$ of the Huber estimator in Eq. (9) and the bisquare estimator in Eq. (12) are a function of the dimensionless variables γ/σ_w and σ_η/σ_w as follows.

For the Huber estimator in Eq. (9), we have

$$\begin{aligned} \mathrm{E}_z[\psi'(z)] &= \mathrm{E}_w[F_\eta(\gamma - w) - F_\eta(-\gamma - 2)] \\ &= \mathrm{E}_w[F_{\overline{\eta}}((\gamma - w)/\sigma_\eta) - F_{\overline{\eta}}((-\gamma - w)/\sigma_\eta)] \\ &= \int_{-\infty}^{\infty} \left[F_{\overline{\eta}}((\gamma - w)/\sigma_\eta) - F_{\overline{\eta}}((-\gamma - w)/\sigma_\eta)\right] f_{\overline{w}}(w/\sigma_w) d(w/\sigma_w) \end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \left[F_{\overline{\eta}}((\gamma - \overline{w}\sigma_w)/\sigma_\eta) - F_{\overline{\eta}}((-\gamma - \overline{w}\sigma_w)/\sigma_\eta)\right] f_{\overline{w}}(\overline{w}) d\overline{w} \\
&= \mathrm{E}_{\overline{w}}\left[F_{\overline{\eta}}\left(\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)\right) - F_{\overline{\eta}}\left(-\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)\right)\right] \\
&= h_A(\gamma/\sigma_w, \sigma_\eta/\sigma_w). \qquad \text{(C.1)}
\end{aligned}
$$

We also find

$$
\begin{aligned}
\mathrm{E}_w\{\mathrm{E}_\eta^2[\psi(w+\eta)]\} &= \mathrm{E}_w\left[\left(\gamma - \int_{-\gamma-w}^{\gamma-w} F_\eta(\eta) d\eta\right)^2\right] \\
&= \mathrm{E}_w\left[\left(\gamma - \int_{-\gamma-w}^{\gamma-w} F_{\overline{\eta}}(\eta/\sigma_\eta) d\eta\right)^2\right] \\
&= \mathrm{E}_w\left[\left(\gamma - \sigma_\eta \int_{(-\gamma-w)/\sigma_\eta}^{(\gamma-w)/\sigma_\eta} F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right)^2\right] \\
&= \sigma_\eta^2 \mathrm{E}_w\left[\left(\gamma/\sigma_\eta - \int_{(-\gamma-w)/\sigma_\eta}^{(\gamma-w)/\sigma_\eta} F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right)^2\right] \\
&= \sigma_\eta^2 \mathrm{E}_{\overline{w}}\left[\left(\gamma/\sigma_\eta - \int_{-\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)}^{\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)} F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right)^2\right] \\
&= \sigma_\eta^2 h_B(\gamma/\sigma_w, \sigma_\eta/\sigma_w), \qquad \text{(C.2)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{E}_w\{\mathrm{E}_\eta[\psi^2(w+\eta)]\} &= \mathrm{E}_w\left[\gamma^2 - 2\int_{-\gamma-w}^{\gamma-w} (w+\eta) F_\eta(\eta) d\eta\right] \\
&= \mathrm{E}_w\left[\gamma^2 - 2\sigma_\eta \int_{(-\gamma-w)/\sigma_\eta}^{(\gamma-w)/\sigma_\eta} (w + \sigma_\eta\overline{\eta}) F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right] \\
&= \sigma_\eta^2 \mathrm{E}_w\left[(\gamma/\sigma_\eta)^2 - 2\int_{(-\gamma-w)/\sigma_\eta}^{(\gamma-w)/\sigma_\eta} \left(w/\sigma_\eta + \overline{\eta}\right) F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right] \\
&= \sigma_\eta^2 \mathrm{E}_{\overline{w}}\left[(\gamma/\sigma_\eta)^2 - 2\int_{(-\gamma-\overline{w}\sigma_w)/\sigma_\eta}^{(\gamma-\overline{w}\sigma_w)/\sigma_\eta} \left(\overline{w}\sigma_w/\sigma_\eta + \overline{\eta}\right) F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right] \\
&= \sigma_\eta^2 \mathrm{E}_{\overline{w}}\left[(\gamma/\sigma_\eta)^2 - 2\int_{-\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)}^{\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)} \left(\overline{w}/(\sigma_\eta/\sigma_w) + \overline{\eta}\right) F_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right] \\
&= \sigma_\eta^2 h_C(\gamma/\sigma_w, \sigma_\eta/\sigma_w). \qquad \text{(C.3)}
\end{aligned}
$$

Then, by substituting Eqs. (C.1)–(C.3) in Eq. (5), the asymptotic efficiency of the Huber estimator in Eq. (9) can be calculated as

$$
\mathrm{Eff}_a(\gamma/\sigma_w, \sigma_\eta/\sigma_w, L) = \frac{1}{J(f_{\overline{w}})} \frac{1}{(\sigma_\eta/\sigma_w)^2} \frac{h_A^2}{\frac{1}{L}h_C + \frac{L-1}{L}h_B}, \qquad \text{(C.4)}
$$

which can be expressed as a function of γ/σ_w and σ_η/σ_w. Similarly, for the bisquare estimator in Eq. (12), we also derive the asymptotic efficiency as

$$
\mathrm{Eff}_a(\gamma/\sigma_w, \sigma_\eta/\sigma_w, L) = \frac{1}{J(f_{\overline{w}})} \frac{h_A^2}{\frac{1}{L}h_C + \frac{L-1}{L}h_B},
$$

where the functions

$$
h_A = \mathrm{E}_{\overline{w}}\left[\int_{-\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)}^{\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)} \left[1 - \left(\overline{w}/(\gamma/\sigma_w) + \overline{\eta}/(\gamma/\sigma_\eta)\right)^2\right]\left[1 - 5\left(\overline{w}/(\gamma/\sigma_w) + \overline{\eta}/(\gamma/\sigma_\eta)\right)^2\right] f_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right],
$$

$$
h_B = \mathrm{E}_{\overline{w}}\left[\left(\int_{-\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)}^{\gamma/\sigma_\eta - \overline{w}/(\sigma_\eta/\sigma_w)} \left(\overline{w} + (\sigma_\eta/\sigma_w)\overline{\eta}\right)\left[1 - \left(\overline{w}/(\gamma/\sigma_w) + \overline{\eta}/(\gamma/\sigma_\eta)\right)^2\right]^2 f_{\overline{\eta}}(\overline{\eta}) d\overline{\eta}\right)^2\right]
$$

and

$$h_C = \mathrm{E}_{\overline{w}}\Big[\int_{-\gamma/\sigma_\eta-\overline{w}/(\sigma_\eta/\sigma_w)}^{\gamma/\sigma_\eta-\overline{w}/(\sigma_\eta/\sigma_w)}\Big(\overline{w}+(\sigma_\eta/\sigma_w)\overline{\eta}\Big)^2\Big[1-\Big(\overline{w}/(\gamma/\sigma_w)+\overline{\eta}/(\gamma/\sigma_\eta)\Big)^2\Big]^2 f_{\overline{\eta}}(\overline{\eta})d\overline{\eta}\Big]$$

are all functions of γ/σ_w and σ_η/σ_w.

References

[1] K. Wiesenfeld, F. Moss, Stochastic resonance and the benefits of noise: from ice ages to crayfish and squids, Nature 373 (6509) (1995) 33–36.
[2] L. Gammaitoni, P. Hänggi, P. Jung, F. Marchesoni, Stochastic resonance, Rev. Modern Phys. 70 (1) (1998) 223–287.
[3] S. Kay, Can detectability be improved by adding noise? IEEE Signal Process. Lett. 7 (2000) 8–10.
[4] F. Chapeau-Blondeau, X. Godivier, Theory of stochastic resonance in signal transmission by static nonlinear systems, Phys. Rev. E 55 (2) (1997) 1478–1495.
[5] J.J. Collins, C.C. Chow, T.T. Imhoff, Stochastic resonance without tuning, Nature 376 (6537) (1995) 236–238.
[6] H. Chen, L.R. Varshney, P.K. Varshney, Noise-enhanced information systems, Proc. IEEE 102 (2014) 1607–1621.
[7] H. Chen, P.K. Varshney, J.H. Michels, Noise enhanced parameter estimation, IEEE Trans. Signal Process. 56 (2008) 5074–5081.
[8] H. Soganci, S. Gezici, O. Arikan, Optimal stochastic parameter design for estimation problems, IEEE Trans. Signal Process. 60 (2012) 4950–4956.
[9] A. Patel, B. Kosko, Noise benefits in quantizer-array correlation detection and watermark decoding, IEEE Trans. Signal Process. 59 (2) (2011) 488–505.
[10] M.D. McDonnell, N.G. Stocks, C.E.M. Pearce, D. Abbott, Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization, Cambridge University Press, New York, 2008.
[11] R. Benzi, A. Sutera, A. Vulpiani, The mechanism of stochastic resonance, J. Phys. A: Math. Gen. 14 (11) (1981) L453–L457.
[12] R.K. Singh, Noise enhanced stability of a metastable state containing coupled Brownian particles, Physica A 473 (2017) 445–450.
[13] I. Gudyma, A. Maksymov, Stochastic resonance in photo-switchable spin-crossover solids, Physica A 477 (2017) 34–41.
[14] L. Zeng, R. Bao, B. Xu, Effects of Lévy noise in aperiodic stochastic resonance, J. Phys. A: Math. Theor. 40 (26) (2007) 7175–7185.
[15] L. Zeng, B. Xu, Effects of asymmetric Lévy noise in parameter-induced aperiodic stochastic resonance, Physica A 389 (22) (2010) 5128–5136.
[16] J. Zhu, X. Lin, R.S. Blum, Y. Gu, Parameter estimation from quantized observations in multiplicative noise environments, IEEE Trans. Signal Process. 63 (2015) 4037–4050.
[17] M.D. McDonnell, Is electrical noise useful? Proc. IEEE 99 (2011) 242–246.
[18] F. Moss, L.M. Ward, W.G. Sannita, Stochastic resonance and sensory information processing: A tutorial and review of application, Clin. Neurophysiol. 115 (2) (2004) 267–281.
[19] R. Löfstedt, S.N. Coppersmith, Quantum stochastic resonance, Phys. Rev. Lett. 72 (13) (1994) 1947.
[20] N. Gillard, E. Belin, F. Chapeau-Blondeau, Qubit state detection and enhancement by quantum thermal noise, Electron. Lett. 54 (1) (2018) 38–39.
[21] F. Chapeau-Blondeau, Optimization of quantum states for signaling across an arbitrary qubit noise channel with minimum-error detection, IEEE Trans. Inform. Theory 61 (8) (2015) 4500–4510.
[22] J.K. Douglass, L. Wilkens, E. Pantazelou, F. Moss, Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance, Nature 365 (6444) (1993) 337.
[23] M.D. McDonnell, L.M. Ward, The benefits of noise in neural systems: Bridging theory and experiment, Nat. Rev. Neurosci. 12 (7) (2009) 415–425.
[24] S. Kay, Noise enhanced detection as a special case of randomization, IEEE Signal Process. Lett. 15 (2008) 709–712.
[25] H. Chen, P.K. Varshney, S.M. Kay, J.H. Michels, Theory of the stochastic resonance effect in signal detection: Part I–Fixed detectors, IEEE Trans. Signal Process. 55 (2007) 3172–3184.
[26] H. Chen, P.K. Varshney, Theory of the stochastic resonance effect in signal detection–Part II: Variable detectors, IEEE Trans. Signal Process. 56 (2008) 5031–5041.
[27] D. Nozaki, Y. Yamamoto, Enhancement of stochastic resonance in a Fitzhugh-Nagumo neuronal model driven by colored noise, Phys. Lett. A 243 (5) (1998) 281–287.
[28] S. Zozor, P.O. Amblard, Stochastic resonance in locally optimal detectors, IEEE Trans. Signal Process. 51 (2003) 3177–3181.
[29] S. Uhlich, Bayes risk reduction of estimators using artificial observation noise, IEEE Trans. Signal Process. 63 (2015) 5535–5545.
[30] N.G. Stocks, Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance, Phys. Rev. E 63 (4) (2001) 041114.
[31] F. Chapeau-Blondeau, S. Blanchard, D. Rousseau, Noise-enhanced Fisher information in parallel arrays of sensors with saturation, Phys. Rev. E 74 (1–10) (2006) 031102.
[32] D. Rousseau, F. Chapeau-Blondeau, Suprathreshold stochastic resonance and signal-to-noise ratio improvement in arrays of comparators, Phys. Lett. A 321 (5–6) (2004) 280–290.
[33] F. Duan, F. Chapeau-Blondeau, D. Abbott, Exploring weak-periodic-signal stochastic resonance in locally optimal processors with a Fisher information metric, Signal Process. 92 (2012) 3049–3055.
[34] F. Duan, F. Chapeau-Blondeau, D. Abbott, Weak signal detection: Condition for noise induced enhancement, Digit. Signal Process. 23 (5) (2013) 1585–1591.
[35] F. Chapeau-Blondeau, D. Rousseau, Noise-aided SNR amplification by parallel arrays of sensors with saturation, Phys. Lett. A 351 (2006) 231–237.
[36] A. Patel, B. Kosko, Error-probability noise benefits in threshold neural signal detection, Neural Netw. 22 (5–6) (2009) 697–706.
[37] A. Patel, B. Kosko, Optimal mean-square noise benefits in quantizer-array linear estimation, IEEE Signal Process. Lett. 17 (2010) 1005–1009.
[38] Y. Pan, F. Duan, F. Chapeau-Blondeau, D. Abbott, Noise enhancement in robust estimation of location, IEEE Trans. Signal Process. 66 (8) (2018) 1953–1966.
[39] V.N. Hari, G.V. Anand, A.B. Premkumar, A.S. Madhukumar, Design and performance analysis of a signal detector based on suprathreshold stochastic resonance, Signal Process. 92 (7) (2012) 1745–1757.
[40] A. Patel, B. Kosko, Optimal noise benefits in Neyman-Pearson and inequality-constrained statistical signal detection, IEEE Trans. Signal Process. 57 (2009) 1655–1669.
[41] G.O. Balkan, S. Gezici, CRLB based optimal noise enhanced parameter estimation using quantized observations, IEEE Signal Process. Lett. 17 (2010) 477–480.
[42] A.B. Akbay, S. Gezici, Noise benefits in joint detection and estimation problems, Signal Process. 118 (2016) 235–247.
[43] T. Yang, S. Liu, H. Liu, M. Tang, X. Tan, X. Zhou, Noise benefits parameter estimation in LMMSE sense, Digit. Signal Process. 73 (2018) 153–163.
[44] Q. Zhai, Y. Wang, Noise effect on signal quantization in an array of binary quantizers, Signal Process. 152 (2018) 265–272.
[45] S. Bayram, S. Gezici, H.V. Poor, Noise enhanced hypothesis-testing in the restricted Bayesian framework, IEEE Trans. Signal Process. 58 (8) (2010) 3972–3989.

[46] S. Bayram, S. Gezici, Noise-enhanced M-ary hypothesis-testing in the minimax framework, in: Proc. International Conference on Signal Processing and Communication Systems, IEEE, Omaha, Nebraska, 2009, pp. 1–6.
[47] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley-Interscience, New York, 2000.
[48] R.H. Byrd, J.C. Gilbert, J. Nocedal, A trust region method based on interior point techniques for nonlinear programming, Math. Program. 89 (1) (2000) 149–185.
[49] R.A. Waltz, J.L. Morales, J. Nocedal, D. Orban, An interior algorithm for nonlinear optimization that combines line search and trust region steps, Math. Program. 107 (3) (2006) 391–408.
[50] P.J. Huber, Robust Statistics, Springer, Berlin, Heidelberg, 1981.
[51] R. Maronna, D. Martin, V. Yohai, Robust Statistics: Theory and Methods, Wiley, New York, 2006.
[52] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Stat. 35 (1) (1964) 73–101.
[53] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, vol. 1, Prentice-Hall PTR, NJ, 1993.
[54] R.P. Brent, Algorithms for Minimization Without Derivatives, Prentice Hall, Englewood Cliffs, NJ, 1973.
[55] G.E. Forsythe, C.B. Moler, M.A. Malcolm, Computer Methods for Mathematical Computations, Prentice Hall, Englewood Cliffs, NJ, 1977.
[56] A.W. Roberts, D.E. Varberg, Convex Functions, Academic Press, New York, 1973.
[57] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.
[58] A.I.F. Vaz, E.M.G.P. Fernandes, Optimization of nonlinear constrained particle swarm, Technol. Econ. Dev. Econ. 12 (1) (2006) 30–36.
[59] K.E. Parsopoulos, M.N. Vrahatis, Particle swarm optimization method for constrained optimization problems, Intell. Technol.–Theory Appl. : New Trends Intell. Technol. 76 (1) (2002) 214–220.
[60] X. Hu, R. Eberhart, Solving constrained nonlinear optimization problems with particle swarm optimization, in: Proceedings of the Sixth World Multiconference on Systemics, Cybernetics and Informatics, vol. 5, Orlando, USA, 2002, pp. 203–206.
[61] W.I. Zangwill, Non-linear programming via penalty functions, Manage. Sci. 13 (5) (1967) 344–358.
[62] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, 1999.
[63] P.T. Boggs, J.W. Tolle, Sequential quadratic programming for large-scale nonlinear optimization, J. Comput. Appl. Math. 124 (1–2) (2000) 123–137.
[64] R.H. Byrd, M.E. Hribar, J. Nocedal, An interior point algorithm for large-scale nonlinear programming, SIAM J. Optim. 9 (4) (1999) 877–900.
[65] P. Ding, Three occurrences of the hyperbolic-secant distribution, Amer. Statist. 68 (1) (2014) 32–35.
[66] M.J. Fischer, Generalized Hyperbolic Secant Distributions: With Applications to Finance, Springer Science & Business Media, Berlin, 2013.
[67] W.L. Harkness, M.L. Harkness, Generalized hyperbolic secant distributions, J. Amer. Statist. Assoc. 63 (321) (1968) 329–337.
[68] K.E. Parsopoulos, M.N. Vrahatis, Particle swarm optimization method for constrained optimization problems, Intell. Technol.–Theory Appl. : New Trends Intell. Technol. 76 (1) (2002) 214–220.
[69] S. Koziel, Z. Michalewicz, Evolutionary algorithms homomorphous mappings and constrained parameter optimization, Evol. Comput. 7 (1) (1999) 19–44.
[70] X. Hu, R. Eberhart, Solving constrained nonlinear optimization problems with particle swarm optimization, in: Proceedings of the sixth world multiconference on systemics, cybernetics and informatics, vol. 5, Orlando, FL, 2002, pp. 203–206.
[71] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, N.D. Sidiropoulos, Feasible point pursuit and successive approximation of non-convex qcqps, IEEE Trans. Signal Process. 22 (2015) 804–808.
[72] K. Price, R.M. Storn, J.A. Lampinen, Differential Evolution: A Practical Approach to Global Optimization, Springer, New York, 2005.
[73] C.T. Kelley, Solving Nonlinear Equations with Newton's Method, SIAM, Philadelphia, 2003.
[74] W.C. Davidon, Variable metric method for minimization, SIAM J. Optim. 1 (1) (1991) 1–17.
[75] R. Fletcher, Practical Methods of Optimization, second ed., John Wiley & Sons, New York, 1987.
[76] A.M. Law, W.D. Kelton, W.D. Kelton, Simulation Modeling and Analysis, vol. 2, McGraw-Hill, New York, 1991.