

Applying the Optimal Injected Noise to Signal Estimation via Adaptive Stochastic Resonance

Yan Pan (✉ panyan87@sdust.edu.cn)

Shandong University of Science and Technology <https://orcid.org/0000-0001-6476-5457>

Liyan Xu

Qingdao University

Fabing Duan

Qingdao University

Francois Chapeau-Blondeau

University of Angers: Universite d'Angers

Research Article

Keywords: Adaptive stochastic resonance, Signal estimation, Mean square error, Optimization space, Learning capacity

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1253141/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Applying the optimal injected noise to signal estimation via adaptive stochastic resonance

Yan Pan^{1*}, Liyan Xu², Fabing Duan³ and François Chapeau-Blondeau⁴

^{1*} College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China.

² School of Electronic Information, Qingdao University, Qingdao 266071, China.

³ Institute of Complexity Science, Qingdao University, Qingdao 266071, China.

⁴ Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), Université d'Angers, 62 avenue Notre Dame du Lac, 49000 Angers, France.

*Corresponding author(s). E-mail(s): panyan87@sdust.edu.cn;

Abstract

In signal estimation, an optimal estimator is frequently unachievable, because its closed form may not be analytically tractable or is too complex to implement. Alternatively, one can turn to suboptimal yet easily implementable estimators for practical signal estimation tasks. In this paper, a general noise-boosted estimator is designed and the adaptive stochastic resonance method is implemented to simultaneously exploit the beneficial role of the injected noise as well as the learning ability of the estimator parameter. Aiming to effectively improve the estimation performance, we use the kernel function method to find an approximate solution for the probability density function (PDF) of the optimal injected noise. During this process, the noise PDF and the estimator parameter establish a finite dimensional non-convex optimization space for maximizing the estimation performance, which is adaptively searched by the sequential quadratic programming algorithm at each iteration. Two representative signal estimation problems are explored, consisting in estimating a random signal from low-resolution observations, or a deterministic parameter in a heavy-tailed noisy environment. The obtained results demonstrate that this adaptive stochastic resonance method can improve the performance of the suboptimal estimators and bring it very close to that of the optimal estimator.

Keywords: Adaptive stochastic resonance, Signal estimation, Mean square error, Optimization space, Learning capacity

1 Introduction

It is a well-established fact nowadays that a suitable amount of noise can result in an improved performance of certain nonlinear systems, and such favorable use of injected noise is the so-called stochastic resonance phenomenon [1–3]. In recent years, the phenomenon of stochastic resonance has been the subject of considerable interest in various research fields, such as physics [4–7], biomedicine [8], signal processing [9, 10] and neural networks [11–13]. However, a complicated question is to characterize the optimal level, or even better the optimal probability density function (PDF), of the noise to be purposefully injected in a given nonlinear process in order to maximize the performance. Most of the time the noise level alone is optimized, analytically or numerically, for a fixed given noise PDF [14–18]. The task of optimizing the noise PDF is more ambitious but a priori can lead to better performances of processors [10, 19–21]. However, analytic solutions of the optimal noise PDF are quite difficult to obtain in general. It can be displayed numerically or by adaptive learning.

In order to adaptively search a non-zero optimal noise level, Mitaim and Kosko proposed the adaptive stochastic resonance method by performing the stochastic gradient ascent algorithm on the signal-to-noise ratio of dynamical systems [2] and the mutual-information of neuron systems [22]. The random noise gradient can tune both the system parameters and the noise level in any adaptive process, and can thus improve the learning capability [2, 22]. Inspired by the parameter-tuning stochastic resonance effect [23], the adaptive stochastic resonance can also be achieved by searching optimal parameters of dynamical systems and processing steps based on different optimization algorithm [24–32], such as the particle swarm optimization [24], the genetic algorithm [25], and the artificial fish swarm algorithm [31, 32]. Most recently, we proposed a noise-boosted backpropagation learning method in feedforward threshold neural networks [12, 33], wherein the injected noise level as well as network weights are adaptively learned by the stochastic gradient descent learning rule.

The stochastic gradient descent algorithm can learn to find the stochastic resonance effects in some simple dynamical systems, but might fail to exploit the beneficial role of noise for complex nonlinear systems [2, 18]. Especially, the adaptive learning process of the optimal noise PDF for some non-convex optimization problems is not sufficiently discussed [18]. In this paper, we design a distributed parameter estimation strategy based on the adaptive stochastic resonance method, wherein the optimal injected noise PDF, along with the

estimator parameters, is “intelligently” updated and searched by the sequential quadratic programming (SQP) optimization algorithm [34, 35]. We here consider two illustrative signal estimation problems, consisting in estimating a random signal from low-resolution observations, or a deterministic parameter in a heavy-tailed noisy environment. It will be shown that, via the adaptive stochastic resonance method, the mean square error (MSE) of the designed estimator can be significantly reduced. The learning scheme converges to an approximate optimal injected noise PDF and to the optimized estimator parameters. It is interesting to note that, using the optimal noise PDF, the designed estimator is almost identical to the optimal estimator and therefore, we can improve the practically implemented estimator to have the same estimation accuracy as the optimal one does. These obtained results enlarge the applicability of adaptive stochastic resonance for signal estimation, and can be further investigated for other complex signal processing tasks.

2 Model and main results

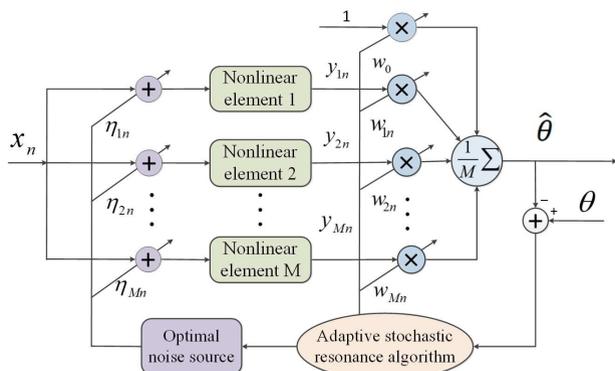


Fig. 1: Block diagram of the noise-boosted estimator $\hat{\theta}$.

Consider a general noise-boosted estimator, as shown in Fig. 1, in which M identical nonlinear elements are boosted by the injection of mutually independent noise components η_{mn} at time n for $m = 1, 2, \dots, M$. The injected noise components η_{mn} are independent and identically distributed random variables with a common PDF f_η . The paralleled elements of the designed estimator in Fig. 1 contain a tunable parameter γ . Let $\boldsymbol{\eta}_n = [\eta_{1n}, \eta_{2n}, \dots, \eta_{Mn}]^\top$ denote the injected noise vector, $\mathbf{y}_n = [y_{1n}, y_{2n}, \dots, y_{Mn}]^\top$ be the output vector of the paralleled elements and $\mathbf{w}_n = [w_{1n}, w_{2n}, \dots, w_{Mn}]^\top$ represent the weight vector. The noise-boosted estimator can be calculated as $\hat{\theta} = w_0 + \mathbf{w}_n^\top \mathbf{y}_n$. For estimating the signal parameter θ , the adaptation process is oriented toward

4 Applying the optimal injected noise to signal estimation

minimizing the MSE

$$\mathcal{E} = \mathbb{E}_{x,\eta}[(\theta - \hat{\theta})^2] = \mathbb{E}_{x,\eta}[(\theta - w_0 - \mathbf{w}_n^\top \mathbf{y}_n)^2], \quad (1)$$

of the noise-booster estimator, where $\mathbb{E}_{x,\eta}(\cdot)$ denotes the expectation with respect to the joint PDF $f_{x,\eta}$ of variables x and η .

It is noted that, for a given noisy environment, the MSE of Eq. (1) is closely related to the PDF f_η of the injected noise, the parameter γ of the nonlinear elements and the weights w_0 and \mathbf{w}_n . In practice, by differentiating Eq. (1) with respect to the weights $[\partial\mathcal{E}/\partial w_0, \partial\mathcal{E}/\partial \mathbf{w}_n^\top]^\top$ and set the derivative equal to zero, the optimum solution of $[w_0, \mathbf{w}_n^\top]^\top$ is easily obtained as the Wiener weight vector [36]. In order to minimize the MSE \mathcal{E} of Eq. (1), we here maintain our focus on the role of the injected noise and the tunable estimator parameter. Since the injected noise type is not necessarily Gaussian, then our main objective is finding an optimal injected noise PDF

$$f_\eta^{\text{opt}}(\eta) = \min_{f_\eta(\eta), \gamma} \mathcal{E}(f_\eta, \gamma), \quad (2)$$

to minimize the MSE \mathcal{E} of Eq. (1) with respect to an arbitrary PDF f_η and the adjustable estimator parameter γ .

The optimization problem of the optimal noise PDF $f_\eta^{\text{opt}}(\eta)$ in Eq. (2) is usually analytically intractable, and we here search an approximate optimal noise PDF described by

$$\tilde{f}_\eta^{\text{opt}}(\eta) = \sum_{k=1}^K \nu_k \varrho_k(\eta) \quad (3)$$

with K kernel functions $\varrho_k(\eta)$. Note that $\varrho_k(\eta) \geq 0$ is also a probability density function and satisfies the normalization condition of $\int \varrho_k(\eta) d\eta = 1$ for $k = 1, 2, \dots, K$. As the number K increases, the approximate form $\tilde{f}_\eta^{\text{opt}}$ of the optimal noise PDF converges to the analytical one f_η^{opt} if it exists [37]. The normalization coefficients satisfy $\nu_k \geq 0$ and $\sum_{k=1}^K \nu_k = 1$. For instance, a popular and commonly used kernel function is the Gaussian kernel function [21, 38] $\varrho_k(\eta) = \varphi((\eta - \mu_k)/\sigma_k)/\sigma_k$ with the location parameter μ_k and scale parameter $\sigma_k \geq 0$, and the standard Gaussian function $\varphi(x) = \exp(-x^2/2)/(2\pi)^{1/2}$. In this way, for a given K , the infinite-dimensional optimization problem of Eq. (2) reduces to a constrained finite-dimensional optimization

$$\begin{aligned} & \min_{\Theta} \mathcal{E}(\Theta) \\ & \text{s.t.} \quad \sum_{k=1}^K \nu_k = 1, \nu_k \geq 0, \sigma_k \geq 0 \end{aligned} \quad (4)$$

with respect to the parameter vector $\Theta = [\gamma, \boldsymbol{\nu}, \boldsymbol{\mu}, \boldsymbol{\sigma}]^\top$, where the coefficient vector $\boldsymbol{\nu} = [\nu_1, \nu_2, \dots, \nu_K]^\top$, the position vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_K]^\top$ and the scale vector $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_K]^\top$.

Since the MSE $\mathcal{E}(\Theta)$ in Eq.(4) with respect to parameter vector Θ is frequently non-convex, then the parameter vector Θ can be adaptively searched numerically by the sequential quadratic programming method [34, 35]. At each major iteration of the renewal process of Θ , an approximate positive definite Hessian matrix of the Lagrangian function is computed via a quasi-Newton difference updating method [35, 37] and then the quadratic programming subproblem yields a solution for a line search procedure [35, 37]. A detailed description of the noise-boosted learning method is presented in Algorithm 1 for adaptively searching the optimal noise indicated by the (local) optimal solution of f_η^{opt} .

Algorithm 1 Adaptive Noise-boosted Learning Algorithm

Consider the minimization of MSE of Eq. (4), a quadratic programming(QP) subprogram for updating the parameter vector Θ_t at the t -th iteration is developed

$$\begin{aligned} \min_{\mathbf{q}} \quad & \frac{1}{2} \mathbf{q}^\top \mathbf{H}_t \mathbf{q} + \nabla \mathcal{E}_t(\Theta_t) \mathbf{q} & (5) \\ \text{s.t.} \quad & \nabla \mathbf{c}(\Theta_t) \mathbf{q} + \mathbf{c}(\Theta_t) = 0 \\ & \nabla \mathbf{h}(\Theta_t) \mathbf{q} + \mathbf{h}(\Theta_t) \geq 0 \end{aligned}$$

where the equality constraint $\mathbf{c}(\Theta_t) = \sum_{k=1}^K \nu_k - 1$ and inequality constraints $\mathbf{h}(\Theta_t) = [\boldsymbol{\nu}, \boldsymbol{\sigma}]^\top$. Here, ∇ is the gradient operator and \mathbf{H}_t denotes the Hessian matrix of the Lagrangian function $\mathcal{L}(\Theta_t, \lambda_t, \zeta_t) = \mathcal{E}(\Theta) - \lambda_t^\top \mathbf{c}(\Theta_t) - \zeta_t^\top \mathbf{h}(\Theta_t)$ for Lagrangian coefficient vectors λ_t^\top and ζ_t^\top .

Start: Choose initial point Θ_0 , and Hessian matrix H_0 ;

set $t \leftarrow 0$

repeat

 Evaluate \mathcal{E}_t and the gradient $\nabla \mathcal{E}_t$;

 Solve Eq. (5) to obtain the SQP step \mathbf{q}_t ;

 Calculate the step length β_t by the line search method;

 Set $\Theta_{t+1} \leftarrow \Theta_t + \beta_t^\top \mathbf{q}_t$;

if convergence test satisfied **then**

 Stop with approximate solution Θ_t ;

else

 Update \mathbf{H}_t using a quasi-Newton formula to obtain \mathbf{H}_{t+1} ;

end if

 Set $t \leftarrow t + 1$;

until convergence test satisfied.

In the following parts, we will utilize the proposed adaptive noise-boosted Algorithm 1 to illustratively accomplish the Bayesian estimation of a random parameter and the robust estimation of a determinate signal. The adaptive learning capacity of the injected noise will be clearly demonstrated.

2.1 Random parameter estimation via adaptive stochastic resonance

First, we consider the problem of estimating a random parameter θ from the signal model $x = \theta + \xi$. As a motivating example, the random parameter θ independent of background noise ξ , has a uniform *prior* PDF $f_\theta(x) = 1/a$ for $x \in [0, a]$ and zero otherwise, and the background noise ξ has the generalized Gaussian PDF

$$f_\xi(x) = \frac{c_1}{\sigma_\xi} \exp\left(-c_2 \left|\frac{x}{\sigma_\xi}\right|^\alpha\right), \quad (6)$$

where $c_1 = \frac{\alpha}{2} \Gamma^{\frac{1}{2}}(3/\alpha) / \Gamma^{\frac{3}{2}}(1/\alpha)$, $c_2 = [\Gamma(3/\alpha) / \Gamma(1/\alpha)]^{\frac{\alpha}{2}}$, the shape parameter α and the scale parameter σ_ξ are both positive [39, 40]. As illustrated in Fig. 1, M identical quantizers have the transfer function $g(x) = 1$ for $x \geq \gamma$ and zero otherwise. This kind of low-resolution sensors is widely employed in distributed estimation problems, whereby the analog observation is compressed into a low-precision single-bit information with low occupied bandwidth.

Since the estimator $\hat{\theta} = w_0 + \mathbf{w}^\top \mathbf{y}_n$ is designed to be unbiased, then the bias weight $w_0 = \mathbb{E}_\theta(\theta) - w \mathbb{E}_x\{\mathbb{E}_\eta[g(x+\eta)]\}$. It is proved that, for a sufficiently large number M of identical transfer functions, the optimum weight coefficients for the minimization of the MSE can be derived as Wiener solution [21, 38]. The tedious manipulation of the Wiener weight w is not included here for simplicity, and we derive the corresponding MSE as [21, 38]

$$\mathcal{E} = \text{var}(\theta) - \frac{\{\mathbb{E}_{\theta,x}[\theta \bar{g}(x)] - \mathbb{E}_\theta(\theta) \mathbb{E}_x[\bar{g}(x)]\}^2}{\mathbb{E}_x[\bar{g}^2(x)] - \mathbb{E}_x^2[\bar{g}(x)]}, \quad (7)$$

where $\mathbb{E}_\theta(\theta)$ and $\text{var}(\theta)$ denote the mean and the variance of the random parameter θ , respectively, and the noise-smoothed transfer function $\bar{g}(x) = \mathbb{E}_\eta[g(x+\eta)]$.

In order to learn the optimal PDF of the injected noise that minimizes the MSE \mathcal{E} of Eq. (7), we employ the Gaussian kernel functions $\varrho_k(\eta) = \varphi((\eta - \mu_k)/\sigma_k)$ in the implementation of the proposed learning Algorithm 1. For comparison, the minimum MSE (MMSE) \mathcal{E}_{ms} achieved by the MMSE estimator $\hat{\theta}_{\text{ms}} = \mathbb{E}(\theta|\mathbf{x}) = \int \theta f(\theta|\mathbf{x}) d\theta$ is also calculated, where the conditional posterior PDF $f(\theta|\mathbf{x}) = f_\theta(\theta) f_\xi(\mathbf{x} - \theta) / \int f_\theta(\theta) f_\xi(\mathbf{x} - \theta) d\theta$ [41].

For instance, consider the interval bound $a = 2$ of the uniform PDF of the random parameter θ and the Laplacian background noise having PDF of Eq. (6) with $\alpha = 1$ and $\sigma_\xi = 1$. Choosing $K = 4$ Gaussian kernel functions, by using the proposed adaptive noise-boosted Algorithm 1, the four kernels finally converge to one Gaussian kernel. It is shown in Figs. 2 (a)-(c) that the learning curves of the MSE \mathcal{E} of Eq. (7), the threshold parameter γ of quantizers and the level σ_η of injected noise are adaptively updated. After 30 iterations, the quantizer threshold γ and the noise level σ_η converge to 1 and 0.5584, respectively, as shown in Figs. 2 (b) and (c). It is also illustrated in

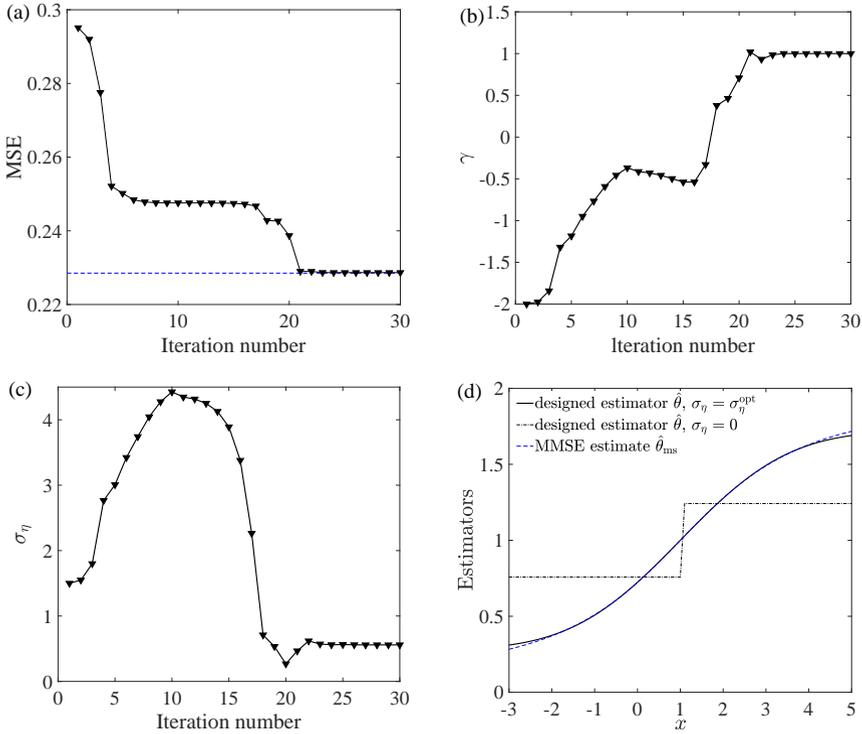


Fig. 2: Learning curves of (a) MSE, (b) the threshold γ of quantizers, (c) the noise level σ_η of the injected noise. (d) Comparison of the transfer functions of the MMSE estimator $\hat{\theta}_{\text{ms}}$, the designed estimator $\hat{\theta}$ without injected noise and $\hat{\theta}$ with the optimal noise level $\sigma_\eta^{\text{opt}} = 0.5584$. Here, the Laplacian background noise level is $\sigma_\xi = 1$, and the injected noise PDF is only approximated by one Gaussian kernel function in Eq. (3).

Fig. 2(a) that, even with only one Gaussian kernel function, the corresponding MSE \mathcal{E} of Eq. (7) can be optimized to 0.2287 at the optimal noise level 0.5584, which almost approaches the minimum MSE 0.2285 (blue dotted line) achieved by the MMSE estimator $\hat{\theta}_{\text{ms}}$. The comparisons of the transfer functions of the MMSE estimator $\hat{\theta}_{\text{ms}}$ and the designed estimator $\hat{\theta}$ at the optimal $\sigma_\eta^{\text{opt}} = 0.5584$ demonstrate that adding noise is advantageous for this example, as shown in Fig. 2(d). It is seen that, with the learned noise level $\sigma_\eta = 0.5584$ and the threshold $\gamma = 1$, the transfer function of the noise-boosted estimator $\hat{\theta}$, far better than the quantizer without the injected noise, approximates to that of the MMSE estimator $\hat{\theta}_{\text{ms}}$ in the concerned domain interval of $[-2, 2]$.

The MSE surface of the noise-boosted estimator $\hat{\theta}$ versus the noise level σ_η and the threshold parameter γ is also illustratively plotted in Fig. 3(a). It is seen that the non-convex MSE of the designed estimator $\hat{\theta}$ is clearly higher

8 Applying the optimal injected noise to signal estimation

as the noise level $\sigma_\eta \rightarrow 0$, which indicates the optimal noise level will be non-zero. Therefore, the injection of noise is necessary for reducing the MSE of the designed estimator $\hat{\theta}$. For a randomly selected initial vector Θ , the learning curve of the MSE in Fig. 2(a) is also shown in Fig. 3(a) with its corresponding projection given in the contour map of Fig. 3(b). The convergence of the MSE demonstrates the aforementioned analyses and the feasibility of the proposed adaptive noise-boosted Algorithm 1.

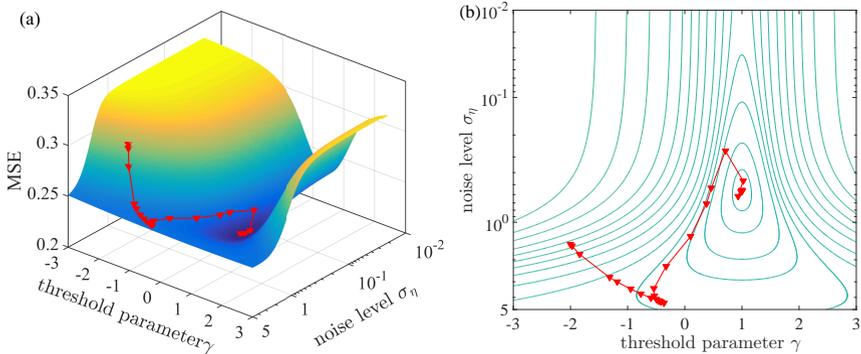


Fig. 3: (a) Surface and (b) the corresponding contour of the MSE \mathcal{E} as a function of the injected noise level σ_η and the threshold parameter γ . The learning curve of the MSE versus the injected noise level σ_η and the threshold parameter γ is also plotted. Other parameters are the same as in Fig. 2.

Furthermore, for a wide range of non-Gaussian environments, Fig. 4(a) illustrates the minimum MSE achieved by the MMSE estimator $\hat{\theta}_{\text{ms}}$ (blue dotted line) as the benchmark. The MSEs of the designed estimator $\hat{\theta}$ without injected noise (\circ) and with optimal noise (∇) versus the background noise type described by the shape parameter α of Eq. (6) are also plotted. It is clearly seen in Fig. 4(a) that, with the assistance of the optimal injected noise, the MSE of the designed estimator achieves a rather similar performance in comparison to the MMSE estimator $\hat{\theta}_{\text{ms}}$ for a wide range of background noise types. In addition, as shown in Fig. 4(b), the MSEs of three concerned estimators versus the bound parameter a of the uniformly distributed θ are compared. Upon increasing a from 0 to 8, the curve of the MSE of the designed estimator with the optimal injected noise (∇) almost always coincides with the minimum MSE achieved by the MMSE estimator.

2.2 Robust estimation via adaptive stochastic resonance

The proposed adaptive noise-boosted Algorithm 1 can also be applied to robust estimation of a determinate signal θ . It is known that many areas of engineering, for instance, mobile communication channels, result in outliers or atypical

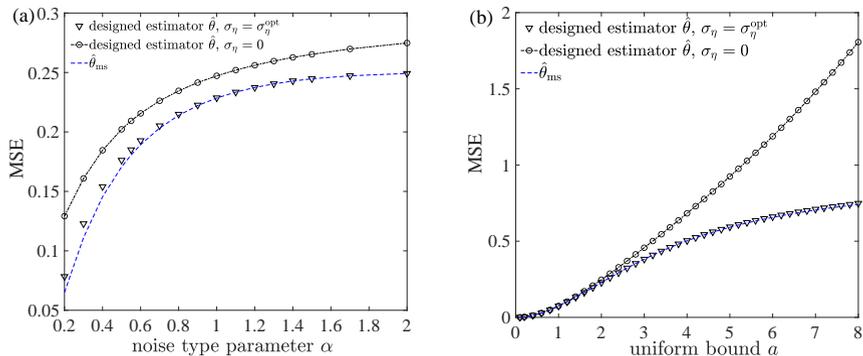


Fig. 4: Curves of MSE of the MMSE estimator $\hat{\theta}_{\text{ms}}$ (dash-dotted line) and the designed estimator $\hat{\theta}$ without (\circ) or with optimal injected noise (∇) versus (a) the shape parameter α of the background generalized Gaussian noise in Eq. (6) and (b) the bound a of the prior uniform PDF f_θ in the presence of Laplacian background noise ($\alpha = 1$). Here, the scale parameter of the background noise in Eq. (6) is $\sigma_\xi = 1$.

observations that do not obey the original assumptions. Thus, the need for robust estimation techniques that are insensitive to outliers is great. Consider a signal model $x_n = \theta + \xi_n$ denoting a determinate signal θ corrupted by heavy-tailed background noise observations ξ_n for $n = 1, 2, \dots, N$. Here, the Cauchy noise ξ_n , as an example of the heavy-tailed background noise, has a common PDF $f_\xi(x) = \sigma_\xi / [\pi(\sigma_\xi^2 + x^2)]$ with zero mean and noise level σ_ξ . The paralleled elements in Fig. 1 are chosen as the bisquare maximum-likelihood type estimator (M-estimator) [42, 43], which has the score function $\psi(x) = x[1 - (x/\gamma)^2]^2$ for $|x| \leq \gamma$ and zero otherwise. Here, the parameter γ is adjustable. Of course, other heavy-tailed noise types and M-estimators can be employed for verifying the practicability of the proposed adaptive noise-boosted Algorithm 1 for improving the performance of the M-estimator.

For this considered robust estimation of the determinate signal θ buried in the heavy-tailed background noise, the Fisher-consistent condition $E_{\xi, \eta}[\psi(\xi + \eta)] = 0$ needs to be satisfied. Therefore, we here select the bias $w_0 = 0$ and the weights $w_{nm} = 1$. Then, for a sufficiently large number M of elements, we proved in [44] that the designed estimator $\hat{\theta}$ in Fig. 1 converges to the true signal θ . With a sufficiently large number M of the injected noise components η_{nm} , the score function $\psi(\xi + \eta)$ can be boosted as $\phi(\xi) = E_\eta[\psi(\xi + \eta)]$ with its derivative $\phi'(\xi) = dE_\eta[\psi(\xi + \eta)]/d\xi$. Then, by taking the first-order Taylor expansion of ϕ around the true signal θ , the MSE or variance of the designed

estimator $\hat{\theta}$ can be calculated as [44]

$$\mathcal{E} = \frac{1}{N} \frac{\mathbb{E}_{\xi}[\phi^2(\xi)]}{\mathbb{E}_{\xi}^2[\phi'(\xi)]} = \frac{1}{N} \frac{\mathbb{E}_{\xi}[\phi^2(\xi)]}{\mathbb{E}_{\xi}^2 \left[\phi(\xi) \left(-\frac{f'_{\xi}(\xi)}{f_{\xi}(\xi)} \right) \right]} \quad (8)$$

$$\geq \frac{1}{N \mathbb{E}_{\xi} \left[\left(-\frac{f'_{\xi}(\xi)}{f_{\xi}(\xi)} \right)^2 \right]} = \frac{1}{NJ(f_{\xi})}, \quad (9)$$

where the inequality in Eq. (9) holds according to the Cauchy-Schwarz inequality. The Cramér-Rao lower bound $(NJ(f_{\xi}))^{-1}$ in Eq. (9) is achieved by $\phi(\xi) = Cf'_{\xi}(\xi)/f_{\xi}(\xi)$ for an arbitrary constant C and $f'_{\xi}(x) = df_{\xi}(x)/dx$, in which $\psi_{\text{ML}}(\xi) = f'_{\xi}(\xi)/f_{\xi}(\xi)$ denotes the score function of the maximum-likelihood estimator (MLE) [42, 43]. Here, $J(f_{\xi})$ is the Fisher information of the noise PDF f_{ξ} .

For a given form of the score function ψ with a tunable parameter γ , the parameter γ can be optimized to obtain a minimum of the MSE of the M-estimators. We proved in [44] that the addition of extra noise, along with the adjustment of the estimator parameter γ , can bring a further reduction of the MSE of the M-estimators. Therefore, we here explore the optimal noise PDF f_{η}^{opt} to minimize the MSE of the designed estimator $\hat{\theta}$ in Eq. (8) with the constraints of the Fisher-consistency. In order to satisfy the Fisher-consistent condition $\mathbb{E}_{\xi, \eta}[\psi(\xi + \eta)] = 0$, we here adopt in Eq. (3) a symmetric double-Gaussian kernel function $\varrho_k(\eta) = [\varphi((\eta + \mu_k)/\sigma_k) + \varphi((\eta - \mu_k)/\sigma_k)]/(2\sigma_k)$. Moreover, to simplify the calculation of the MSE of the noise-booster estimator $\hat{\theta}$ in Fig. 1, set the scale parameter $\sigma_k \rightarrow 0$, and then the kernel function reduces to $\varrho_k(\eta) = [\delta(\eta + \mu_k) + \delta(\eta - \mu_k)]/2$, where $\delta(\cdot)$ denotes the Delta function.

For a given number $K = 4$ of kernel functions, the adaptive process of the parameter $\Theta = [\gamma, \boldsymbol{\nu}, \boldsymbol{\mu}]^{\top}$ is realized by the proposed adaptive noise-booster Algorithm 1 starting from the initial parameter vector Θ_0 ($\gamma = 2.5$, $\boldsymbol{\nu} = [0.25, 0.25, 0.25, 0.25]^{\top}$, $\boldsymbol{\mu} = [0.6, 1, 3, 6]^{\top}$). The adaptive learning curves of the variance \mathcal{E} in Eq. (8) (for simplicity, the sample length $N = 1$), the estimator parameter γ and the coefficients ν_k versus the iteration times are plotted in Figs. 5 (a)-(c), respectively. It is seen in Fig. 5 that, as the M-estimator parameter γ and the noise parameters adaptively learn, the variance \mathcal{E} of the designed estimator $\hat{\theta}$ can be reduced to the optimum value of 2.008, which is very close to the Cramér-Rao lower bound 2 (blue dotted line in Fig. 5(a)) obtained by the MLE. In Fig. 5(d), the normalized score functions are illustrated for $\bar{\psi}_{\text{ML}}(x) = \psi_{\text{ML}}(x)/\|\psi_{\text{ML}}(x)\|$ of the MLE, $\bar{\phi}(x) = \phi(x)/\|\phi(x)\|$ of the designed estimator $\hat{\theta}$ with the optimal noise and $\bar{\psi}(x) = \psi(x)/\|\psi(x)\|$ of the bisquare M-estimator. Here, $\|g(x)\| = [\int g^2(x)dx]^{1/2}$ denotes the \mathcal{L}^2 -norm of a function $g(x)$. As shown in Fig. 5(d), the score function of the designed estimator $\hat{\theta}$ with the optimal noise is found very similar to that of the MLE.

The boosting role of the injection of noise is clearly manifested again in the M-estimator for robust estimation.

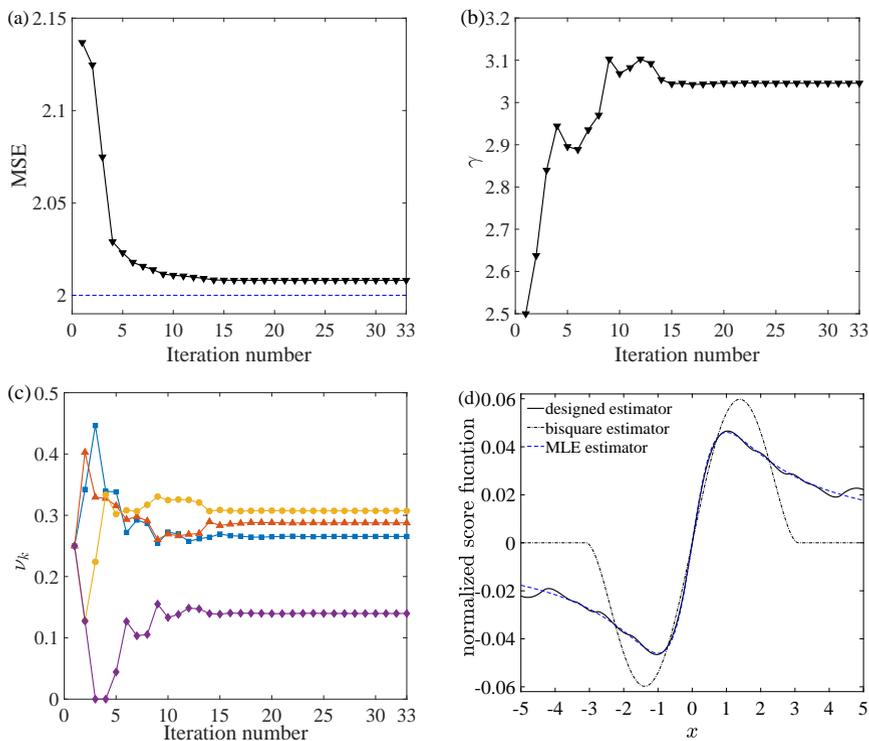


Fig. 5: Learning curves of (a) MSE \mathcal{E} of Eq. (8), (b) the parameter γ of the bisquare estimator, (c) the 4 coefficients ν_k versus the iteration number. (d) The score functions of MLE, the bisquare M-estimator without injected noise and the designed estimator $\hat{\theta}$ with the approximate optimal noise of Eq. (3). Here, the scale parameter of the background Cauchy noise is $\sigma_\xi = 1$, and the kernel function number is $K = 4$.

3 Conclusion

In this paper, a noise-boosted learning algorithm is implemented for the exploration of noise benefits in estimating deterministic and random signal parameters. The injected noise, as well as the estimator parameter, are adaptively learned for improving the performance of the designed estimator. It is shown that, in the learning process, the performance of the designed estimator can be improved to come very close to that of the optimal estimator (e.g. Fig. 5(d) and Fig. 2(d)). For the random signal estimation with *prior*

knowledge, the MSE of the designed estimator based on low-resolution observations can adaptively converge to the minimum MSE achieved by the MMSE estimator using the full analog observations. For the problem of estimating a deterministic signal buried in heavy-tailed background noise, the MSE of the designed estimator almost reaches the Cramér-Rao lower bound by the injection of the optimal noise herein. The beneficial effect of the injected noise is adaptively implemented by the proposed noise-boosted algorithm in signal estimation. The effectiveness of this adaptive learning process deserves to be intensively studied in various other complex signal processing tasks.

Several open questions still need to be addressed in future studies. The proposed adaptive learning algorithm is based on the Jacobian and Hessian matrices of the MSE of the designed estimator, and the theoretical expression of the MSE needs to be known. When the theoretical expression of the MSE is too complicated to obtain, this learning algorithm is inapplicable. In this situation, an open and interesting question is how to adaptively learn the optimal noise to be injected into the observed data. In addition, in many practical signal processing tasks, the observed data are acquired sequentially as time progresses. Then, the sequential implementation of adaptively choosing and injecting the optimal noise as time progresses will be of great interest.

Acknowledgments. This work is sponsored by the National Science Foundation of China under the Grant No. 62001271.

Data availability statement. The datasets analysed during the current study are available from the corresponding author on reasonable request.

Declarations

- **Funding:** This study was funded by National Science Foundation of China (Grant Number 62001271).
- **Conflict of interest:** The authors declare that they have no conflict of interest.
- **Ethics approval:** Not applicable.
- **Consent to participate:** Not applicable.
- **Consent for publication:** Not applicable.
- **Availability of data and materials:** The datasets analysed during the current study are available from the corresponding author on reasonable request.
- **Code availability:** The code generated or used during the study is available from the corresponding author on reasonable request.
- **Authors' contributions:** Material preparation, formal analysis, methodology were performed by Yan Pan. The first draft of the manuscript was written by Yan Pan and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

- [1] Benzi, R., Sutera, A., Vulpiani, A.: The mechanism of stochastic resonance. *Journal of Physics A: Mathematical and General* **14**, 453–457 (1981)
- [2] Mitaim, S., Kosko, B.: Adaptive stochastic resonance. *Proceedings of the IEEE* **86**, 2152–2183 (1998)
- [3] Stocks, N.G.: Suprathreshold stochastic resonance in multilevel threshold systems. *Physical Review Letters* **84**, 2310–2313 (2000)
- [4] Lu, S., He, Q., Wang, J.: A review of stochastic resonance in rotating machine fault detection. *Mechanical Systems and Signal Processing* **116**, 230–260 (2018)
- [5] He, M., Xu, W., Sun, Z.: Dynamical complexity and stochastic resonance in a bistable system with time delay. *Nonlinear Dynamics* **79**, 1787–1795 (2015)
- [6] Yang, J.H., Sanjun, M., Liu, H.G., Zhu, H.: Noise-induced resonance at the subharmonic frequency in bistable systems. *Nonlinear Dynamics* **87**, 1721–1730 (2017)
- [7] Xiao, L., Bajric, R., Zhao, J., Tang, J., Zhang, X.: An adaptive vibrational resonance method based on cascaded varying stable-state nonlinear systems and its application in rotating machine fault detection. *Nonlinear Dynamics* **103**, 715–739 (2021)
- [8] Collins, J.J., Priplata, A.A., Gravelle, D.C., Niemi, J., Harry, J., Lipsitz, L.A.: Noise-enhanced human sensorimotor function. *IEEE Engineering in Medicine and Biology Magazine* **22**, 76–83 (2003)
- [9] Chen, H., Varshney, L.R., Varshney, P.K.: Noise-enhanced information systems. *Proceedings of the IEEE* **102**, 1607–1621 (2014)
- [10] Uhlich, S.: Bayes risk reduction of estimators using artificial observation noise. *IEEE Transactions on Signal Processing* **63**, 5535–5545 (2015)
- [11] Audhkhasi, K., Osoba, O., Kosko, B.: Noise-enhanced convolutional neural networks. *Neural Networks* **78**, 15–23 (2016)
- [12] Duan, L., Duan, F., Chapeau-Blondeau, F., Abbott, D.: Noise-boosted backpropagation learning of feedforward threshold neural networks for function approximation. *IEEE Transactions on Instrumentation and Measurement* **70**, 1010612 (2021)
- [13] Yang, X.L., Li, N., Sun, Z.K.: Extended analysis of stochastic resonance

14 *Applying the optimal injected noise to signal estimation*

- in a modular neuronal network at different scales. *Nonlinear Dynamics* **98**, 1–11 (2019)
- [14] Collins, J.J., Chow, C.C., Imhoff, T.T.: Stochastic resonance without tuning. *Nature* **376**, 236–238 (1995)
- [15] Kay, S.: Can detectability be improved by adding noise? *IEEE Signal Processing Letters* **7**, 8–10 (2000)
- [16] Chapeau-Blondeau, F., Godivier, X.: Theory of stochastic resonance in signal transmission by static nonlinear systems. *Physical Review E* **55**, 1478–1495 (1997)
- [17] Patel, A., Kosko, B.: Optimal mean-square noise benefits in quantizer-array linear estimation. *IEEE Signal Processing Letters* **17**, 1005–1009 (2010)
- [18] Li, F., Duan, F., Chapeau-Blondeau, F., Abbott, D.: Signal estimation and filtering from quantized observations via adaptive stochastic resonance. *Physical Review E* **103**, 052108 (2021)
- [19] Chen, H., Varshney, P.K., Michels, J.H.: Noise enhanced parameter estimation. *IEEE Transactions on Signal Processing* **56**, 5074–5081 (2008)
- [20] Soganci, H., Gezici, S., Arikan, O.: Optimal stochastic parameter design for estimation problems. *IEEE Transactions on Signal Processing* **60**, 4950–4956 (2012)
- [21] Duan, F., Pan, Y., Chapeau-Blondeau, F., Abbott, D.: Noise benefits in combined nonlinear Bayesian estimators. *IEEE Transactions on Signal Processing* **67**, 4611–4623 (2019)
- [22] Mitaim, S., Kosko, B.: Adaptive stochastic resonance in noisy neurons based on mutual information. *IEEE Transactions on Neural Networks* **15**, 1526–1540 (2004)
- [23] Xu, B., Duan, F., Bao, R., Li, J.L.: Stochastic resonance with tuning system parameters: the application of bistable systems in signal processing. *Chaos, Solitons & Fractals* **13**, 633–644 (2002)
- [24] Chen, X.H., Cheng, G., Shan, X.L., Hu, X., Guo, Q., Liu, H.G.: Research of weak fault feature information extraction of planetary gear based on ensemble empirical mode decomposition and adaptive stochastic resonance. *Measurement* **73**, 55–67 (2015)
- [25] Zhou, P., Lu, S., Liu, F., Liu, Y., Li, G., Zhao, J.: Novel synthetic index-based adaptive stochastic resonance method and its application in bearing

- fault diagnosis. *Journal of Sound & Vibration* **391**, 194–210 (2017)
- [26] Li, J., Chen, X., He, Z.: Adaptive stochastic resonance method for impact signal detection based on sliding window. *Mechanical Systems and Signal Processing* **36**, 240–255 (2013)
- [27] Wang, J., Ren, X., Zhang, S.W., Zhang, D.M., Li, H.S.: Adaptive bistable stochastic resonance aided spectrum sensing. *IEEE Transactions on Wireless Communications* **13**, 4014–4024 (2014)
- [28] Lei, Y., Han, D., Lin, J., He, Z.: Planetary gearbox fault diagnosis using an adaptive stochastic resonance method. *Mechanical Systems and Signal Processing* **38**, 113–124 (2013)
- [29] Liu, J., Hu, B., Y.G., W.: Optimum adaptive array stochastic resonance in noisy grayscale image restoration. *Physics Letters A* **383**, 1457–1465 (2019)
- [30] Li, J., Zhang, J., Li, M., Zhang, Y.: A novel adaptive stochastic resonance method based on coupled bistable systems and its application in rolling bearing fault diagnosis. *Mechanical Systems and Signal Processing* **114**, 128–145 (2019)
- [31] Wang, J., He, Q., Kong, F.: Adaptive multiscale noise tuning stochastic resonance for health diagnosis of rolling element bearings. *IEEE Transactions on Instrumentation and Measurement* **64**, 564–577 (2014)
- [32] Liu, X.L., Liu, H.G., Yang, J.H., Litak, G., Cheng, G., Han, S.: Improving the bearing fault diagnosis efficiency by the adaptive stochastic resonance in a new nonlinear system. *Mechanical Systems and Signal Processing* **96**, 58–76 (2017)
- [33] Liu, X.J., Duan, L., Duan, F., Chapeau-Blondeau, F., Abbott, D.: Enhancing threshold neural network via suprathreshold stochastic resonance for pattern classification. *Physics Letters A* **403**, 127387 (2021)
- [34] Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
- [35] Boggs, P.T., Tolle, J.W.: *Sequential quadratic programming*. *Acta Numerica* **4**, 1–51 (1995)
- [36] Widrow, B., Stearns, S.D.: *Adaptive Signal Processing*. Prentice-Hall, Upper Saddle River, New Jersey (1985)
- [37] Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, New York (2006)

- [38] Liu, J., Duan, F., Chapeau-Blondeau, F., Abbott, D.: Distributed Bayesian vector estimation using noise-optimized low-resolution sensor observations. *Digital Signal Processing* **118**, 103224 (2021)
- [39] Kay, S.M.: *Fundamentals of Statistical Signal Processing: Detection Theory* vol. 2. Prentice-Hall, New Jersey (1998)
- [40] Kassam, S.A.: *Signal Detection in Non-Gaussian Noise*. Springer, New York (1988)
- [41] Kay, S.M.: *Fundamentals of Statistical Signal Processing: Estimation Theory* vol. 1. Prentice-Hall PTR, New Jersey (1993)
- [42] Huber, P.J.: *Robust Statistics*. Springer, Berlin Heidelberg (1981)
- [43] Maronna, R., Martin, D., Yohai, V.: *Robust Statistics: Theory and Methods*. Wiley, New York (2006)
- [44] Pan, Y., Duan, F., Chapeau-Blondeau, F., Abbott, D.: Noise enhancement in robust estimation of location. *IEEE Transactions on Signal Processing* **66**, 1953–1966 (2018)