

Sequential noise-boosted M-estimation for robust parameter estimation under impulsive noise

Li Zhang(张莉), Yan Pan(潘燕), Fabing Duan(段法兵), François Chapeau-Blondeau, and Derek Abbott
Citation: Chin. Phys. B, 2026, 35 (**3**): 030204. DOI: 10.1088/1674-1056/adfefe
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Probabilistic distribution and stochastic *P*-bifurcation of a nonlinear energy-regenerative suspension system with time-delayed feedback control

Zhao-Bin Zeng(曾昭彬), Ya-Hui Sun(孙亚辉), and Yang Liu(刘洋)
Chin. Phys. B, 2026, 35 (**1**): 010201. DOI: 10.1088/1674-1056/ade669

Adaptive multi-stable stochastic resonance assisted by neural network and physical supervision

Xucan Li(李栩灿), Deming Nie(聂德明), Ming Xu(徐明), and Kai Zhang(张凯)
Chin. Phys. B, 2025, 34 (**5**): 050203. DOI: 10.1088/1674-1056/adbd28

Dynamic properties of rumor propagation model induced by Lévy noise on social networks

Ying Jing(景颖), Youguo Wang(王友国), Qiqing Zhai(翟其清), and Xianli Sun(孙先莉)
Chin. Phys. B, 2024, 33 (**9**): 090203. DOI: 10.1088/1674-1056/ad58c5

Performance enhancement of a viscoelastic bistable energy harvester using time-delayed feedback control

Mei-Ling Huang(黄美玲), Yong-Ge Yang(杨勇歌), and Yang Liu(刘洋)
Chin. Phys. B, 2024, 33 (**6**): 060203. DOI: 10.1088/1674-1056/ad3b89

Consensus for second-order multi-agent systems with position sampled data

Rusheng Wang(王如生), Lixin Gao(高利新), Wenhai Chen(陈文海), Dameng Dai(戴大蒙)
Chin. Phys. B, 2016, 25 (**10**): 100202. DOI: 10.1088/1674-1056/25/10/100202

Sequential noise-boosted M-estimation for robust parameter estimation under impulsive noise

Li Zhang(张莉)[1], Yan Pan(潘燕)[1,†], Fabing Duan(段法兵)[2], François Chapeau-Blondeau[3], and Derek Abbott[4]

[1] *College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao* 266590, *China*
[2] *Institute of Complexity Science, Qingdao University, Qingdao* 266071, *China*
[3] *Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS),Université d'Angers*, 62 *Avenue Notre Dame du Lac*, 49000 *Angers, France*
[4] *Centre for Biomedical Engineering (CBME) and School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA* 5005, *Australia*

(Received 16 July 2025; revised manuscript received 17 August 2025; accepted manuscript online 26 August 2025)

We propose a sequential noise-boosted M-estimation algorithm for estimating system parameters in environments characterized by impulsive (heavy-tailed) noise. This algorithm extends the conventional M-estimation framework by strategically injecting artificial noise into the observations, thereby facilitating the estimation procedure and ensuring convergence to the desired estimator. A fundamental criterion theorem is established to determine the conditions under which injecting scale-family noise enhances the efficacy of the M-estimator in heavy-tailed background noise. For cases where noise injection is beneficial, it is rigorously proved that the sequential noise-boosted M-estimation algorithm converges with probability one. Experimental results demonstrate that the proposed algorithm outperforms traditional M-estimation methods, both under a given injected noise intensity and when the noise injection is adaptively optimized via Bayesian optimization. Furthermore, it is observed that the proposed algorithm can asymptotically achieve the performance of the maximum likelihood estimator (MLE) for system parameter estimation.

Keywords: sequential estimation, noise-boosted m-estimation, convergence analysis, stochastic resonance

PACS: 02.50.–r, 02.50.Fz, 05.40.–a, 05.40.Ca

DOI: 10.1088/1674-1056/adfefe **CSTR:** 32038.14.CPB.adfefe

1. Introduction

In statistical physics and signal processing, optimal estimators or detectors are derived based on strong assumptions, such as the data following a specific probability distribution (e.g., a Gaussian distribution) or the model being accurately specified.[1–5] Assuming that the data follow a specific distribution, the maximum likelihood estimator (MLE) is an efficient method for parameter estimation,[5] as it is asymptotically unbiased and achieves the Cramér–Rao lower bound (CRLB) for large sample sizes. However, these idealized assumptions may not always hold in real-world scenarios due to the presence of outliers, heavy-tailed noise, or model misspecification, which can significantly degrade the performance of traditional estimators.[1,6] To address these challenges, Huber[6] proposed a foundational framework for robust estimation methods by introducing a class of maximum-likelihood-type estimators known as M-estimators. The purpose of these M-estimators is to reduce the effects of outliers or atypical observations while maintaining sufficient estimation accuracy under a variety of conditions.[1,3]

Nevertheless, in many contemporary applications, such as real-time signal processing,[7] sensor networks,[8] and streaming data analysis,[9] data are not acquired in large batches at once but instead arrive sequentially over time. In these dynamic settings,[7–9] existing methods may not be directly applicable, as they generally assume access to the entire dataset. This elicits the need for sequential estimation techniques that update parameter estimates incrementally as new data arrive. While sequential estimation methods such as recursive least squares (RLS),[7] the Kalman filter,[10] and stochastic gradient descent[11] are widely used in the identification of physical systems, they lack sufficient robustness to outliers or distributional shifts in the data stream. Extending from weighted least squares, a class of robust sequential algorithms has been developed, including a robust recursive identification algorithm[12] via M-estimation and schemes[13] using Taylor expansion with matrix inversion approximation. Building on adaptive filtering principles, RLS-based recursive least M-estimation algorithms[14,15] improve robustness by replacing the mean square error (MSE) criterion with M-estimation cost functions. From the perspective of Bayesian estimation, a sequential M-estimation algorithm has been developed by solving a maximum a posteriori problem,[16] and robust Bayesian filtering methods have demonstrated superior performance under heavy-tailed distributions.[17] In addition, genetic algorithms have been incorporated[18] to enhance robustness against heavy-tailed noise and outlier contamination. In these studies,[13–18] there exists a trade-off between robust-

[†]Corresponding author. E-mail: panyan87@sdust.edu.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

ness and efficiency. Moreover, a fundamental limitation of these existing approaches[13–17] is their insufficient adaptability in adjusting the parameters of M-estimators in response to changing data, making it difficult to achieve an optimal balance between robustness and efficiency in robust sequential estimation.

A promising approach, known as stochastic resonance,[19,20] offers a potential solution to this trade-off problem in robust sequential estimation. Stochastic resonance refers to a phenomenon in which noise can enhance the performance of nonlinear systems under certain conditions rather than degrade it.[21–27] By introducing noise, estimation efficiency can be effectively improved in signal estimation problems.[28–33] It has been proven that the injection of additive noise into the threshold array system can effectively reduce the MSE of both optimal and suboptimal linear estimators.[28] Within the Bayesian framework, a noise-boosted estimator obtained by averaging estimates from noisy observations outperforms the conventional maximum a posteriori estimator.[29] Furthermore, the Bayesian risk can be significantly reduced by optimally combining injected noise with weighting coefficients.[30] For robust location estimation, it has been shown that injecting noise can improve efficiency and reduce the bias of robust estimators.[31] In the iterative expectation-maximization algorithm, it has been demonstrated[34] that noise injection can effectively reduce the average convergence time, with its effectiveness depending on the mode of noise injection.[35]

In this paper, we explore the mechanism of stochastic resonance as a noise-boosting principle to enhance the performance of the robust sequential M-estimator in dynamic and uncertain environments. Specifically, we propose a sequential noise-boosted M-estimation algorithm that extends the traditional sequential M-estimation framework by incorporating controlled noise into the score function of M-estimators. A theoretical criterion, rigorously derived using the first-order Taylor series expansion, provides a method for evaluating whether the addition of a small scale-family noise component can effectively improve the performance of the proposed M-estimation algorithm. Under conditions where the addition of noise is beneficial, the convergence of the proposed sequential noise-boosted M-estimation algorithm is rigorously established. Experimental results further validate these theoretical findings and demonstrate the superior performance of the sequential noise-boosted M-estimation algorithm over traditional M-estimators, both under a fixed intensity of added noise and with adaptively optimized noise injection. Using the Bayesian optimization method, the intensity or probability density of added noise can be adaptively tuned to enhance the estimation accuracy of the proposed algorithm in various noisy environments. Moreover, the performance of the proposed sequential noise-boosted M-estimation algorithm asymptotically approaches that of the MLE under certain conditions, demonstrating its robustness and efficiency in the presence of impulsive background noise.

2. Model and algorithm

2.1. M-estimation of observation model

We here consider the scenario in which observations are measured sequentially, following a linear model described as

$$x_n = \boldsymbol{h}_n^{\top}\boldsymbol{\theta} + \omega_n, \ \ n = 1, 2, \ldots, N, \tag{1}$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_p]^{\top}$ is the unknown p-dimensional regression parameter vector, and x_n and $\boldsymbol{h}_n$ represent the observed value and the corresponding system vector for the n-th observation, respectively. Here, ω_n is generally assumed to be independent and identically distributed (i.i.d.) zero-mean measurement noise with the probability density function (PDF) $f_{\omega}(x)$.

In application fields such as communication systems,[36–38] image processing,[39–41] and radar and sonar,[42,43] abnormal data often arise during the measurement process, resulting in a heavy-tailed probability distribution of the observed data. We here focus on symmetric heavy-tailed distributions. In the presence of such heavy-tailed background noise, M-estimation, a well-known robust strategy,[1–3] is introduced to minimize a loss function ρ with respect to the error $x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta}$, which can be defined as

$$\hat{\boldsymbol{\theta}}_M = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \rho(x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta}), \tag{2}$$

or as the solution to the implicit equation

$$\sum_{n=1}^{N} \psi(x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta})\boldsymbol{h}_n = 0, \tag{3}$$

where $\rho(x)$ is a non-negative even function and $\psi(x) = \mathrm{d}\rho(x)/\mathrm{d}x$ is called the score function. The assumption that $\rho(x)$ is even in Eq. (2) lies in the fact that it penalizes positive and negative departures equally, and therefore the score function $\psi(x)$ is naturally odd. When $\rho(x) = x^2$ or equivalently when $\psi(x) = x$, the solution in Eq. (3) simplifies to the least squares estimation algorithm.[3]

2.2. Sequential noise-boosted M-estimation algorithm

In many cases, traditional M-estimation algorithms exhibit limited adaptability to the data distribution,[44] often assuming a fixed score function that may not adequately capture the true underlying characteristics of the data. Moreover, the adjustment of M-estimator parameters is often constrained,

which prevents the M-estimation algorithm from asymptotically achieving the CRLB,[45] especially when the data distribution deviates from the Gaussian assumption.[46]

To overcome this limitation, we here propose a noise-boosted score function $\tilde{\psi}(x)$ defined as the expectation of $\psi(x+\eta)$ with respect to an artificially added random variable η, i.e.,

$$\tilde{\psi}(x) = \mathrm{E}_{\eta}[\psi(x+\eta)] = \psi(x) * f_{\eta}(x). \tag{4}$$

Here, the operator $E_{\eta}[\cdot] = \int_{-\infty}^{\infty} \cdot f_{\eta}(x)\,\mathrm{d}x$ represents the expectation with respect to the PDF $f_{\eta}(x)$ of the added noise η. By introducing a symmetric PDF $f_{\eta}(x)$ that is even, the proposed score function $\tilde{\psi}(x)$ in Eq. (4) can be equivalently expressed as the convolution of the original odd score function $\psi(x)$ and the even PDF $f_{\eta}(x)$. This guarantees that $\tilde{\psi}(x)$ preserves the odd symmetry of the original score function. Crucially, this odd symmetry allows the noise-boosted score function $\tilde{\psi}(x)$ in Eq. (4) to decouple the effects of positive and negative noise perturbations. As a result, $\tilde{\psi}(x)$ provides a tractable framework to analyze how the added noise influences the estimation accuracy and robustness of the M-estimator.

Based on the noise-boosted score function $\tilde{\psi}(x)$ defined in Eq. (4), the M-estimate in Eq. (3) can be rewritten as

$$\sum_{n=1}^{N} \tilde{\psi}(x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta})\boldsymbol{h}_n = \sum_{n=1}^{N} \frac{\tilde{\psi}(x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta})}{x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta}}(x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta})\boldsymbol{h}_n = 0. \tag{5}$$

Define the factor

$$\lambda_n = \frac{\tilde{\psi}(x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta})}{x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta}}, \tag{6}$$

and assume that $\lambda_n = 1$ when $x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta} = 0$. Then, Eq. (5) can be expressed in the simplified form

$$\sum_{n=1}^{N} \lambda_n (x_n - \boldsymbol{h}_n^{\top}\boldsymbol{\theta})\boldsymbol{h}_n = 0. \tag{7}$$

It is noted that the sequential solution to the problem in Eq. (7) is approximately reduced to solving a weighted nonlinear least squares estimation problem.[7] By using the matrix inversion lemma,[47] a new sequential noise-boosted M-estimation $\hat{\boldsymbol{\theta}}[n]$ of the regression parameter vector $\boldsymbol{\theta}$ is computed recursively as

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \frac{\lambda_n \boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}{1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n} e_n, \tag{8}$$

and the covariance matrix $\boldsymbol{\Sigma}[n]$ of $\hat{\boldsymbol{\theta}}[n]$ can be updated as

$$\boldsymbol{\Sigma}[n] = \boldsymbol{\Sigma}[n-1] - \frac{\lambda_n \boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]}{1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}, \tag{9}$$

where the prediction error $e_n = x_n - \boldsymbol{h}_n^{\top}\hat{\boldsymbol{\theta}}[n-1]$, and λ_n in Eq. (6) is updated as $\lambda_n = \tilde{\psi}(x_n - \boldsymbol{h}_n^{\top}\hat{\boldsymbol{\theta}}[n-1])/(x_n - \boldsymbol{h}_n^{\top}\hat{\boldsymbol{\theta}}[n-1])$ using the previous estimate $\hat{\boldsymbol{\theta}}[n-1]$. Let the adaptive gain matrix $\boldsymbol{K}[n]$ be defined as

$$\boldsymbol{K}[n] = \frac{\lambda_n \boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}{1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}, \tag{10}$$

then the estimates $\hat{\boldsymbol{\theta}}[n]$ and the covariance matrix $\boldsymbol{\Sigma}[n]$ at observation n can be updated as

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \boldsymbol{K}[n]e_n, \tag{11}$$

$$\boldsymbol{\Sigma}[n] = \left(\boldsymbol{I} - \boldsymbol{K}[n]\boldsymbol{h}_n^{\top}\right)\boldsymbol{\Sigma}[n-1]. \tag{12}$$

To ensure numerical stability, the initial values are set to $\hat{\boldsymbol{\theta}}[0] = 0$ and $\boldsymbol{\Sigma}[0] = kI$ where k is sufficiently larger than zero, for example $k = 10$.[48]

2.3. Discriminant and convergence analysis of sequential noise-boosted M-estimation

After incorporating added noise into the score function, a fundamental question arises: whether noise benefits parameter estimation and how this effect can be efficiently determined. Here, we establish a criterion theorem for the proposed sequential noise-boosted M-estimation algorithm. This theorem provides a discriminant to determine the beneficial impact of noise on enhancing the performance of M-estimation under the condition of sufficiently small noise intensity.

Theorem 1 Assume that the added noise η belongs to a scale-family distribution, i.e., $f_{\eta}(x;\sigma_{\eta}) = f_{\eta_0}(x/\sigma_{\eta})/\sigma_{\eta}$ and $E_{\omega}[\tilde{\psi}(x/\sigma_{\omega})] = 0$. Here, $E_{\omega}(\cdot)$ denotes the expectation with respect to the PDF $f_{\omega}(x)$ of the background noise, and $f_{\eta_0}(\cdot)$ denotes the PDF of the random variable $\eta_0 = \eta/\sigma_{\eta}$, which has zero mean and unit scale. Then, for a sufficiently small noise intensity σ_{η}, the condition for improvement by added noise

$$E_{\omega}[\psi(\omega)\psi''(\omega)]E_{\omega}[\psi'(\omega)] - E_{\omega}[\psi'''(\omega)]E_{\omega}[\psi^2(\omega)] < 0, \tag{13}$$

indicates that when the condition of Eq. (13) is satisfied, the covariance of the estimation error in Eq. (A1) decreases as the added noise begins to increase above zero. Here, ψ', ψ'', and ψ''' represent the first, second, and third derivatives of ψ, respectively.

Proof See Appendix A.

For example, consider a heavy-tailed Cauchy background noise with PDF[3,49]

$$f_{\omega}(x) = \frac{1}{\pi\sigma_{\omega}\left[1+\left(\frac{x}{\sigma_{\omega}}\right)^2\right]}, \tag{14}$$

where σ_{ω} is the noise intensity. Furthermore, consider the monotone Huber score function,[6] defined as

$$\psi(x) = \begin{cases} x, & |x| \le \gamma, \\ \gamma\,\mathrm{sgn}(x), & |x| > \gamma, \end{cases} \tag{15}$$

and the redescending bisquare function[1,3]

$$\psi(x)=\begin{cases} x\left[1-\left(\frac{x}{\gamma}\right)^2\right]^2, & |x|\le\gamma,\\ 0, & |x|>\gamma, \end{cases} \tag{16}$$

where γ is an adjustable threshold parameter. For various noise intensities σ_ω of the Cauchy noise in Eq. (14), Figs. 1(a) and 1(b) illustrate the regions of the discriminant function $\Phi(\omega)$ in Eq. (13) versus the threshold parameter γ of the Huber function in Eq. (15) and the bisquare function in Eq. (16), respectively. Based on Theorem 1 and the discriminant inequality of Eq. (13), the negative region marked by "–" of the discriminant function $\Phi(\omega)$ clearly manifests the benefit of added noise across substantial ranges of the parameter pair (γ,σ_ω), as illustrated in Figs. 1(a) and 1(b).

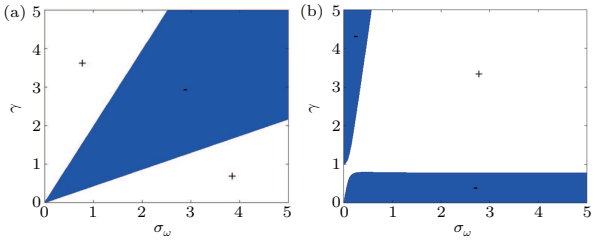


Fig. 1. Regions of negative ("–", blue) and positive ("+", white) values of $\Phi(\omega)$ in Eq. (13), where the added noise is beneficial in the former. These regions are plotted as functions of the noise intensity σ_ω of the Cauchy distribution in Eq. (14) and the threshold parameter γ of (a) the Huber function in Eq. (15) and (b) the bisquare function in Eq. (16).

The condition of Eq. (13) in Theorem 1 is derived in the limit of small added scale-family noise and determines whether, as the noise intensity increases from zero, the covariance of the estimation error in Eq. (A1) begins to decrease. Now we leave the small-noise limit and consider the evolution of the influence factor υ in Eq. (A2) with respect to the added noise intensity σ_η, in order to display the non-monotonic behavior with a minimum at a non-zero optimal noise intensity. For the added Gaussian noise with PDF $f_\eta(x)=f_{\eta_0}(x/\sigma_\eta)/\sigma_\eta$ and $f_{\eta_0}(x)=\exp(-x^2/2)/\sqrt{2\pi}$, the value of the factor υ attains its minimum at the optimal $\sigma_\eta=2.82$, which validates the beneficial effects of the added noise and the potential existence of an optimal noise intensity for the sequential M-estimation algorithm. Here, the threshold parameter of the Huber score function in Eq. (15) is $\gamma=8$, and the background Cauchy noise intensity is $\sigma_\omega=1$.

Although the addition of noise has been shown to enhance M-estimation performance in certain regimes, it is essential to demonstrate that the sequential noise-boosted M-estimation algorithm maintains asymptotic unbiasedness and convergence properties. Theorem 2 proves the convergence of the sequential noise-boosted M-estimation algorithm under certain conditions.

Theorem 2 Given that $\tilde{\psi}(x)$ is an odd function and Riemann integrable over any finite interval, the Fisher-consistent constraint condition $E[\tilde{\psi}(x/\sigma_\omega)]=0$ holds.[3] For the linear observation model in Eq. (1), the sequential noise-boosted M-estimate $\hat{\boldsymbol{\theta}}[n]$ given by Eqs. (10)–(12) converges to its true value $\boldsymbol{\theta}$ with probability one, i.e.,

$$\lim_{n\to\infty}\hat{\boldsymbol{\theta}}[n]=\boldsymbol{\theta}, \tag{17}$$

for a sufficiently large observation number n.

Proof See Appendix B.

To evaluate the performance of the proposed sequential noise-boosted M-estimation algorithm given in Eqs. (10)–(12), we employ the MSE of the M-estimator at time n, defined as

$$\mathrm{MSE}[n]=E_\omega\left[||\hat{\boldsymbol{\theta}}[n]-\boldsymbol{\theta}||_2^2\right]. \tag{18}$$

It is known that the MSE of M-estimation is generally a non-convex functional of the noise intensity or the probability density of the added noise,[33] and therefore it is difficult to adaptively optimize the added noise to minimize the MSE. To address this issue, we employ Bayesian optimization,[50] which uses a probabilistic surrogate model, e.g., a Gaussian process with a prior distribution, to approximate the objective function $\mathrm{MSE}[n]$ in Eq. (18). By iteratively updating the surrogate model via Bayesian posterior inference that combines prior assumptions and observed data, the acquisition function determines the posterior mean value as the next evaluation point.

Here, for an a priori parametric class of scale-family noise distributions, the intensity σ_η of the added noise and the threshold parameter γ of the score function are combined to form the parameter vector $\boldsymbol{\zeta}=[\sigma_\eta,\gamma]^\top$ to be optimized. Then, a Gaussian process surrogate model[51] can be employed to model the objective function $l(\boldsymbol{\zeta})=\mathrm{MSE}(\boldsymbol{\zeta})$, assuming a multivariate Gaussian prior over its function values. For such situations, we utilize the tree-structured Parzen estimator approach to optimize the parameter vector $\boldsymbol{\zeta}=[\sigma_\eta,\gamma]^\top$ by the acquisition function of the expected improvement

$$\alpha(\boldsymbol{\zeta})=E_{l(\boldsymbol{\zeta})|\mathcal{D}}\left[\max\left(0,l(\boldsymbol{\zeta})-l^*\right)\right], \tag{19}$$

where l^* is the current best value, and the expectation $E_{l(\boldsymbol{\zeta})|\mathcal{D}}$ is computed over the posterior distribution of $l(\boldsymbol{\zeta})$ conditioned on the observed data $\mathcal{D}$. The next sample point $\boldsymbol{\zeta}$ is then selected to maximize the acquisition function. The sequential noise-boosted M-estimation algorithm for optimizing the noise intensity σ_η by Bayesian optimization is shown in Algorithm 1 in detail.

Algorithm 1 Sequential noise-boosted M-estimation algorithm.

Input:
$\hat{\boldsymbol{\theta}}^{(m)}[0]$: Initial estimate (size $p\times 1$);
$\boldsymbol{\Sigma}^{(m)}[0]$: Initial covariance matrix (size $p\times p$);
$\mathcal{Z}$: Search space for the parameter vector $\boldsymbol{\zeta}=[\sigma_\eta,\gamma]^\top$;
S: Maximum number of trials;
$\mathcal{D}_0 = \{(\boldsymbol{\zeta}_1,l_1),(\boldsymbol{\zeta}_2,l_2),\cdots,(\boldsymbol{\zeta}_S,l_S)\}$, $l_s = \mathrm{MSE}(\boldsymbol{\zeta}_s)$ $(s=1,2,\ldots,S)$;
T: Total number of Monte Carlo runs;
N: Number of samples.

Output: $\boldsymbol{\zeta}^{\mathrm{opt}}$, $\mathrm{MSE}(\hat{\boldsymbol{\theta}};\boldsymbol{\zeta}^{\mathrm{opt}})$.

for iteration $s=1\to S$ **do**
  for Monte Carlo trial $t=1\to T$ **do**
    for sample $n=1\to N$ **do**
      $e_n^{(t)} = x_n^{(t)} - \boldsymbol{h}_n^\top \hat{\boldsymbol{\theta}}^{(t)}[n-1]$
      $\tilde{\psi}(e_n^{(t)}) = \psi(e_n^{(t)}) * f_\eta(e_n^{(t)})$
      $\lambda_n^{(t)} = \tilde{\psi}(e_n^{(t)})/e_n^{(t)}$
      $\boldsymbol{K}^{(t)}[n] \leftarrow \frac{\lambda_n^{(t)}\boldsymbol{\Sigma}^{(t)}[n-1]\boldsymbol{h}_n}{1+\lambda_n^{(t)}\boldsymbol{h}_n^\top\boldsymbol{\Sigma}^{(t)}[n-1]\boldsymbol{h}_n}$
      $\hat{\boldsymbol{\theta}}^{(t)}[n] \leftarrow \hat{\boldsymbol{\theta}}^{(t)}[n-1] + \boldsymbol{K}^{(t)}[n]e_n^{(t)}$
      $\boldsymbol{\Sigma}^{(t)}[n] \leftarrow \left(\boldsymbol{I}-\boldsymbol{K}^{(t)}[n]\boldsymbol{h}_n^\top\right)\boldsymbol{\Sigma}^{(t)}[n-1]$
    $\hat{\boldsymbol{\theta}}^{(t)} = \hat{\boldsymbol{\theta}}^{(t)}[N]$
  $l(\boldsymbol{\zeta}^{(s)}) = \mathrm{MSE}(\boldsymbol{\zeta}^{(s)}) = \frac{1}{T}\sum_{t=1}^{T}||\hat{\boldsymbol{\theta}}^{(t)}-\boldsymbol{\theta}||^2$
  $\boldsymbol{\zeta}^{(s+1)} = \arg\max \alpha(\boldsymbol{\zeta})$
$\boldsymbol{\zeta}^{\mathrm{opt}} \leftarrow \arg\min \mathrm{MSE}(\boldsymbol{\zeta}^{(s)})$

3. Experimental results

For the linear vector model of Eq. (1), we consider a two-dimensional regression parameter vector $\boldsymbol{\theta}=[\theta_1,\theta_2]^\top=[1,3]^\top$ for the sequential M-estimation problem. The system vector $\boldsymbol{h}_n$ is generated as Gaussian random variables without loss of generality. In the following experiments, the performance of the proposed sequential noise-boosted M-estimation algorithm is measured by MSE$[n]$, and each point of MSE$[n]$ is obtained by averaging over 500 independent trials.

3.1. Comparison results of the sequential noise-boosted M-estimation and traditional sequential M-estimation

In order to investigate the noise effects on the performance of the proposed sequential noise-boosted M-estimation algorithm, we here consider two forms of the added noise η: one with Gaussian PDF, and the other with its PDF represented as a linear combination of Delta kernel functions

$$\tilde{f}_\eta^{\mathrm{opt}}(x) = \frac{1}{2}\sum_{k=1}^{K} v_k[\delta(x+\mu_k)+\delta(x-\mu_k)], \tag{20}$$

where $v_k\geq 0$, $\sum_{k=1}^{K}v_k=1$, and $\delta(\cdot)$ denotes the Delta function. When K is sufficiently large, optimizing the coefficients v_k can approximate the unknown PDF $f_\eta^{\mathrm{opt}}(x)$ of the optimal added noise.

Figure 2 presents a comparison of the convergence curves of the MSE for the sequential noise-boosted M-estimation algorithm with the RLS algorithm,[7] the traditional sequential M-estimation algorithm,[12,48] and the recursive constrained generalized maximum correntropy (RCGMC) algorithm.[52] The intensity σ_ω of the background Cauchy noise is set to unity. In the traditional sequential M-estimation algorithm, the threshold parameter of the redescending bisquare score function in Eq. (16) is fixed as $\gamma=2$. In RCGMC, we consider a single constraint, and set the correntropy kernel parameter to 1.5 and the regularization parameter to 2. To evaluate the performance of the sequential noise-boosted M-estimation algorithm in the presence of added Gaussian noise η, we consider two distinct cases: the parameter vector $[\sigma_\eta,\gamma]^\top=[2,2]^\top$ is fixed, and the parameter vector is determined as $[\sigma_\eta^{\mathrm{opt}},\gamma^{\mathrm{opt}}]^\top=[2.6762,8.76]^\top$ via Bayesian optimization described in Algorithm 1. Furthermore, the approximate optimal noise PDF $\tilde{f}_\eta^{\mathrm{opt}}(x)$ defined in Eq. (20) with $K=3$ Delta kernel functions can be obtained with parameters $v_1=0.3512$, $v_2=0.1723$, $v_3=0.4765$, $\mu_1=1.1609$, $\mu_2=3.0505$, $\mu_3=0.6393$, and $\gamma=6.9282$ via Bayesian optimization.

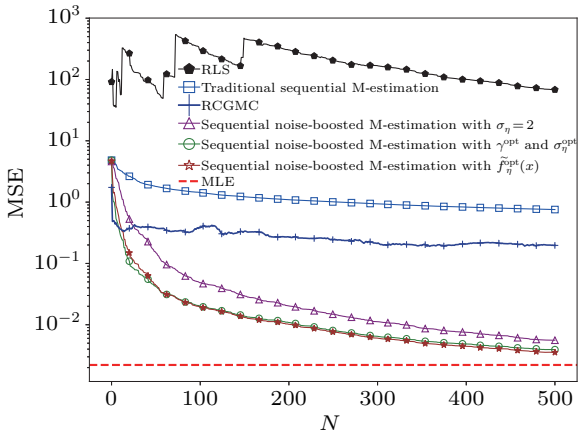


Fig. 2. Convergence curves of the MSE for the RLS algorithm, the traditional sequential M-estimation algorithm, the RCGMC algorithm, and the sequential noise-boosted M-estimation algorithm. For the sequential noise-boosted M-estimation algorithm, three cases are considered: a fixed threshold $\gamma=2$ with Gaussian noise intensity $\sigma_\eta=2$; the approximate optimal noise PDF $\tilde{f}_\eta^{\mathrm{opt}}(x)$ defined in Eq. (20); and the optimized parameters $\gamma^{\mathrm{opt}}=8.76$ and $\sigma_\eta^{\mathrm{opt}}=2.6762$. The approximate optimal noise PDF $\tilde{f}_\eta^{\mathrm{opt}}(x)$ with $K=3$ Delta kernel functions is determined by parameters $v_1=0.3512$, $v_2=0.1723$, $v_3=0.4765$, $\mu_1=1.1609$, $\mu_2=3.0505$, $\mu_3=0.6393$, and $\gamma=6.9282$ via Bayesian optimization.

It is seen in Fig. 2 that the proposed sequential noise-boosted M-estimation algorithm, either enhanced by the added Gaussian noise or by the approximate optimal noise defined in Eq. (20), can significantly outperform the RLS algorithm, the traditional sequential M-estimation algorithm, and the RCGMC algorithm. The MSE of the proposed sequential noise-boosted M-estimation algorithm with the optimal threshold parameter $\gamma^{\mathrm{opt}}=8.76$ and Gaussian noise intensity

$\sigma_\eta^{\rm opt} = 2.6762$ decreases to MSE$[n] = 0.0039$ at $n = 500$. Similarly, the MSE of the proposed sequential noise-boosted M-estimation algorithm optimized by the approximate optimal noise decreases to 0.0035. Both of these MSE values are comparable to the MSE$[n] = 0.0022$ obtained by the MLE.

3.2. Noise-boosted M-estimation and stochastic resonance

As demonstrated in Subsection 3.1, the injection of noise can effectively reduce the MSE of the sequential M-estimation algorithm compared to the RLS algorithm and the traditional M-estimation algorithm. This enhancement in parameter estimation[32,53,54] can also be interpreted as a stochastic resonance phenomenon,[19,20,26] which reveals the constructive role of noise in enhancing estimator performance by adjusting the intensity of specific types of noise or seeking the optimal noise.

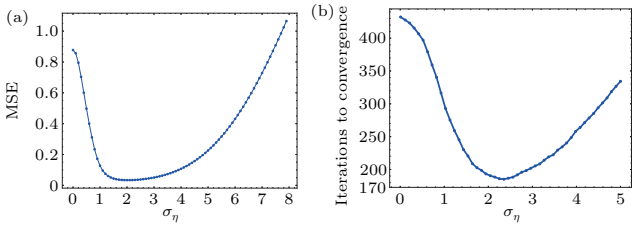


Fig. 3. (a) The MSE and (b) the number of iterations required for convergence versus the injected noise intensity σ_η. The threshold parameter of the bisquare function in Eq. (16) is set to $\gamma = 2$, and the background noise ω_n follows a standard Cauchy distribution with intensity $\sigma_\omega = 1$.

For instance, Fig. 3(a) illustrates the relationship between the MSE of the proposed sequential noise-boosted M-estimation algorithm and the Gaussian noise intensity σ_η. It is observed that, with the threshold parameter γ fixed at 2, the MSE initially takes the value 0.8775 at $\sigma_\eta = 0$, decreases to a minimum of 0.03 at the optimal noise intensity $\sigma_\eta = 2$, and then increases for $\sigma_\eta > 2$. This phenomenon indicates the existence of a nonzero optimal noise intensity that minimizes the MSE of the proposed sequential noise-boosted algorithm, which can be interpreted within the framework of stochastic resonance.

Furthermore, we numerically demonstrate that noise injection can improve the convergence speed of the sequential M-estimation algorithm. The convergence speed[34] is evaluated as the minimum number of iterations required for the MSE to fall below a predefined threshold over multiple Monte Carlo runs. Fewer iterations to convergence indicate faster reduction of the MSE. In our simulations, the threshold is set to 10^{-2}. For the sequential Gaussian noise-boosted M-estimation algorithm, the average number of iterations to convergence exhibits a non-monotonic dependence on the injected noise intensity σ_η, as depicted in Fig. 3(b). The number of iterations decreases as the noise intensity increases, reaching a minimum value of 185 at an optimal nonzero intensity $\sigma_\eta = 2.3522$. Three types of sequential noise-boosted algorithms are compared. For the Gaussian noise-boosted algorithm with fixed parameters $[\sigma_\eta, \gamma]^\top = [2,2]^\top$, the average number of iterations to convergence is 185. When the noise intensity σ_η and threshold parameter γ are simultaneously optimized, the average number of iterations decreases to 91. Furthermore, using the approximate optimal noise with PDF defined in Eq. (20), the average number of iterations is further reduced to 57. These results highlight the importance of identifying the optimal noise PDF to accelerate convergence.

To illustrate the stochastic resonance phenomenon in M-estimation more intuitively, Fig. 4 depicts the adaptive optimization trajectory of the threshold parameter γ and noise intensity σ_η obtained via Bayesian optimization over the search space $\sigma_\eta \in [0,5]$ and $\gamma \in [0,10]$. As shown, the MSE is relatively high when the added noise intensity is near zero. However, as the noise intensity increases and the threshold parameter adapts, the optimal MSE of 0.0039 is achieved at the corresponding optimal values $\sigma_\eta^{\rm opt} = 2.6762$ and $\gamma^{\rm opt} = 8.76$. This result demonstrates an adaptive stochastic resonance mechanism.

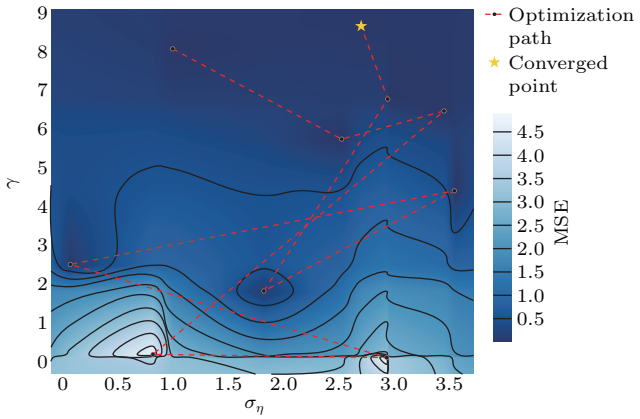


Fig. 4. Adaptive optimization trajectory of the MSE for the noise-boosted M-estimation algorithm as a function of the Gaussian added noise intensity σ_η and the threshold parameter γ. The optimization is performed over 10 Bayesian optimization trials, and the yellow star indicates the final converged optimal coordinate $(\sigma_\eta^{\rm opt}, \gamma^{\rm opt}) = (2.6762, 8.76)$.

Furthermore, to explore the optimal noise PDF approximated by Eq. (20), Bayesian optimization is employed to determine the optimal parameter vector

$$\zeta = [v_1, v_2, v_3, \mu_1, \mu_2, \mu_3, \gamma]^\top$$

for $K = 3$ Delta kernel functions. In this setting, the minimum MSE is achieved as 0.0037 after 10 Bayesian optimization trials, with the optimal parameter vector

$$\zeta^{\rm opt} = [0.3512, 0.1723, 0.4765, 1.1609, 3.0505, 0.6393, 6.9282]^\top,$$

as shown in Fig. 2. The corresponding PDF of the optimal noise is illustrated in Fig. 5. The existence of a non-zero optimal noise further demonstrates that noise injection can

enhance the performance of M-estimators, which can be interpreted as a manifestation of the stochastic resonance phenomenon.

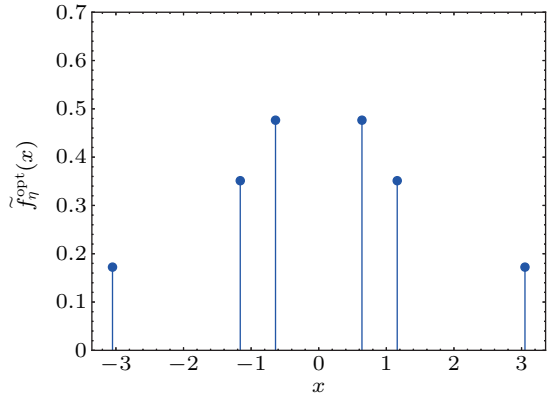


Fig. 5. Approximation of the optimal noise PDF using Eq. (20) with $K = 3$ Delta kernel functions. The parameters are given by $[v_1, v_2, v_3, \mu_1, \mu_2, \mu_3]^\top = [0.3512, 0.1723, 0.4765, 1.1609, 3.0505, 0.6393]^\top$.

3.3. Comparison of the sequential noise-boosted M-estimation algorithm and MLE

To further evaluate the robustness of the proposed noise-boosted M-estimation under varying background noise intensities, we define the input signal-to-noise ratio (SNR) as

$$\mathrm{SNR_{in}} = 10\log_{10}\left(\frac{\theta_1^2+\theta_2^2}{\sigma_\omega^2}\right), \tag{21}$$

where θ_1 and θ_2 denote the deterministic signal components, and σ_ω is the intensity of the background noise.

For Cauchy background noise, Fig. 6(a) shows the resulting MSE as a function of increasing $\mathrm{SNR_{in}}$. We also consider a symmetric α-stable (SαS) noise model,[55,56] whose characteristic function is given by

$$\Psi_\omega(u) = \exp(-\sigma_\omega|u|^\alpha), \tag{22}$$

where $0 < \alpha \leq 2$ is the characteristic exponent controlling the tail thickness, and σ_ω is the noise intensity. When $\alpha = 2$, the SαS noise reduces to Gaussian noise; for $\alpha < 2$, the distribution exhibits heavy tails, with $\alpha = 1$ corresponding to the Cauchy distribution. In this experiment, we set $\alpha = 1.2$, and the resulting MSE curves are presented in Fig. 6(b) as a function of $\mathrm{SNR_{in}}$.

From Figs. 6(a) and 6(b), it can be observed that the sequential noise-boosted M-estimation algorithm, enhanced by both Gaussian and approximate optimal noise from Eq. (20), outperforms the traditional sequential M-estimation algorithm. In particular, at low input SNR ($\mathrm{SNR_{in}} < 0$), the noise-boosted sequential M-estimation with approximate optimal noise approaches the performance of the MLE. As $\mathrm{SNR_{in}}$ increases, the influence of the added optimal noise gradually diminishes, reflecting the reduced intensity of the background noise σ_ω.

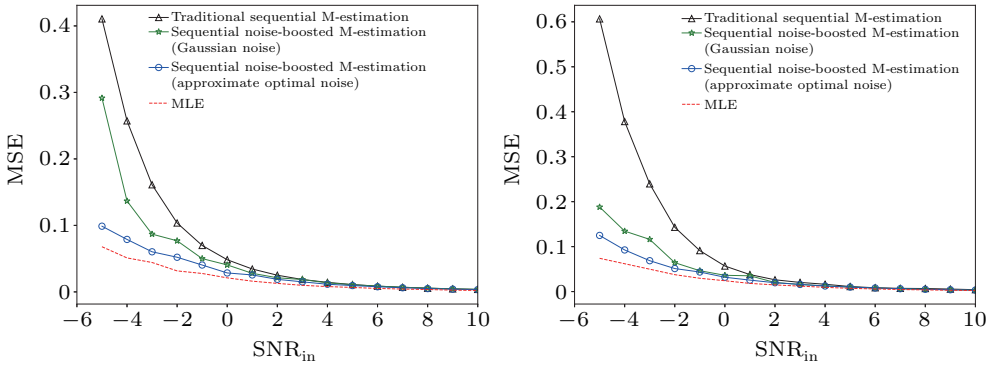


Fig. 6. Performance comparison of the sequential noise-boosted M-estimation algorithm with optimal added Gaussian noise (☆) and optimal added noise distribution in Eq. (20) with $K = 4$ (○), traditional sequential M-estimation algorithm (△), and MLE (dashed line) under (a) Cauchy background noise and (b) SαS noise with $\alpha = 1.2$. Other parameters are the same as in Fig. 4.

4. Conclusion

In this paper, we proposed a sequential noise-boosted M-estimation algorithm to enhance parameter estimation performance in impulsive noise environments. The injection of noise into the observations induces a noise-boosted score function, defined as the convolution of the conventional score function with the probability density function (PDF) of the added noise. For added scale-family noise, we derived a criterion theorem to determine whether noise injection is beneficial at sufficiently small noise intensities. Furthermore, we presented a theoretical proof of the asymptotic convergence of the sequential noise-boosted M-estimation algorithm under the assumption of a symmetric PDF for the added noise.

Experimental results demonstrate that the proposed sequential noise-boosted M-estimation algorithm achieves a mean squared error (MSE) close to that of the maximum likelihood estimator (MLE) when an appropriate amount of noise is added. Its performance surpasses that of the traditional sequential M-estimation and recursive least squares (RLS) algorithms, thereby illustrating the beneficial role of noise in sequential M-estimation. The proposed algorithm is of significant importance for dynamic system state estimation, system

identification, and signal feature extraction in impulsive noise environments, where it achieves enhanced robustness and accuracy. Future research may focus on developing adaptive robust estimators that dynamically adjust noise injection parameters in time-varying impulsive noise conditions to further improve estimation accuracy.

However, several open problems remain. First, when employing Bayesian optimization to search for the optimal noise intensity or noise PDF, the convergence speed of the MSE requires further improvement, and the computational complexity of the optimization process needs to be reduced. Second, in the current approach, the noise intensity at each step is updated using the Bayesian optimization model, rather than being iteratively calculated using information from previous steps as in Eqs. (10)–(12). Exploring single-step iterative noise intensity injection using a stochastic gradient descent method to approximate the MSE gradient represents an interesting direction for future research.

Appendix A: Criterion theorem for determining the beneficial impact of the added scale-family noise

If $E_\omega[\tilde{\psi}(\omega/\sigma_\omega)]=0$ holds and for a large sample size N, the proposed noise-boosted M-estimation method is consistent with respect to the true parameter value, then the covariance matrix of the estimation error $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}$ is given by[3,31]

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})=\sigma_\omega^2\frac{E_\omega\left[\tilde{\psi}^2(\omega,\sigma_\eta)\right]}{E_\omega^2\left[\tilde{\psi}'(\omega,\sigma_\eta)\right]}\left(H^\top H\right)^{-1}. \tag{A1}$$

Given the intensity σ_ω of the background noise and the fixed system matrix H, the covariance matrix in Eq. (A1) is only influenced by the factor

$$\upsilon=\frac{E_\omega\left[\tilde{\psi}^2(\omega,\sigma_\eta)\right]}{E_\omega^2\left[\tilde{\psi}'(\omega,\sigma_\eta)\right]}. \tag{A2}$$

Assume that the order of integration and differentiation can be interchanged, we have $\tilde{\psi}'(\omega,\sigma_\eta)=E_\eta[\psi'(\omega+\eta)]$. To study how noise intensity affects the covariance of the estimation error, we calculate the partial derivative of υ with respect to σ_η as

$$\begin{aligned}\frac{\partial\upsilon}{\partial\sigma_\eta}&=\frac{\partial}{\partial\sigma_\eta}\frac{E_\omega\left[\tilde{\psi}^2(\omega,\sigma_\eta)\right]}{E_\omega^2\left[\tilde{\psi}'(\omega,\sigma_\eta)\right]}\\&=2\upsilon\left\{\frac{E_\omega\left[\tilde{\psi}(\omega,\sigma_\eta)\frac{\partial\tilde{\psi}(\omega,\sigma_\eta)}{\partial\sigma_\eta}\right]}{E_\omega\left[\tilde{\psi}^2(\omega,\sigma_\eta)\right]}\right.\\&\quad\left.-\frac{E_\omega\left[\frac{\partial\tilde{\psi}'(\omega,\sigma_\eta)}{\partial\sigma_\eta}\right]}{E_\omega\left[\tilde{\psi}'(\omega,\sigma_\eta)\right]}\right\}.\end{aligned} \tag{A3}$$

Since $\upsilon>0$, the derivative in Eq. (A3) becomes negative when

$$\begin{aligned}&E_\omega\left[\tilde{\psi}(\omega,\sigma_\eta)\frac{\partial\tilde{\psi}(\omega,\sigma_\eta)}{\partial\sigma_\eta}\right]E_\omega\left[\tilde{\psi}'(\omega,\sigma_\eta)\right]\\&-E_\omega\left[\frac{\partial\tilde{\psi}'(\omega,\sigma_\eta)}{\partial\sigma_\eta}\right]E_\omega\left[\tilde{\psi}^2(\omega,\sigma_\eta)\right]<0,\end{aligned} \tag{A4}$$

which indicates that when the added noise intensity σ_η increases, the covariance matrix of the estimation error in Eq. (A1) decreases.

Since the PDF of the scale-family noise satisfies $f_\eta(x;\sigma_\eta)=f_{\eta_0}(x/\sigma_\eta)/\sigma_\eta$, with $f_{\eta_0}(x)=\exp(-x^2/2)/\sqrt{2\pi}$, we have $E_\eta[\psi(\omega+\eta)]=E_{\eta_0}[\psi(\omega+\eta_0\sigma_\eta)]$. For a small noise intensity $0<\sigma_\eta\ll 1$, performing a first-order Taylor expansion of $\psi(\omega+\eta_0\sigma_\eta)$ and $\psi'(\omega+\eta_0\sigma_\eta)$ with respect to σ_η, we find

$$\begin{aligned}\tilde{\psi}(\omega,\sigma_\eta)&=E_\eta[\psi(\omega+\eta)]\\&=E_{\eta_0}[\psi(\omega+\eta_0\sigma_\eta)]\\&\approx\psi(\omega),\end{aligned} \tag{A5}$$

$$\frac{\partial\tilde{\psi}(\omega,\sigma_\eta)}{\partial\sigma_\eta}\approx\psi''(\omega)\sigma_\eta, \tag{A6}$$

$$\tilde{\psi}'(\omega,\sigma_\eta)\approx\psi'(\omega), \tag{A7}$$

$$\frac{\partial\tilde{\psi}'(\omega,\sigma_\eta)}{\partial\sigma_\eta}\approx\psi'''(\omega)\sigma_\eta. \tag{A8}$$

Substituting Eqs. (A5)–(A8) into Eq. (A4), the condition for the reduction of the covariance at a small noise intensity $\sigma_\eta\to 0$ is

$$E_\omega[\psi(\omega)\psi''(\omega)]E_\omega[\psi'(\omega)]-E_\omega[\psi'''(\omega)]E_\omega[\psi^2(\omega)]<0. \tag{A9}$$

When the condition of Eq. (A9) is satisfied, the covariance of the estimation error decreases as the added noise rises above zero, revealing the noise-boosted effect in parameter estimation. If the second or third derivative of ψ in Eq. (A9) does not exist, one can alternatively reduce the order of differentiation applied to ψ and evaluate the derivatives of f_ω using integration by parts. For instance

$$E_\omega[\psi'''(\omega)]=-\int_{-\infty}^{\infty}\psi''(\omega)\,\mathrm{d}f_\omega(\omega),$$

under the assumption that $f_\omega(\pm\infty)=0$.

Appendix B: Convergence analysis of the sequential noise-boosted M-estimation

Rewrite the iterative formula of Eq. (8) for parameter estimation as

$$\begin{aligned}\hat{\boldsymbol{\theta}}[n]&=\hat{\boldsymbol{\theta}}[n-1]+\frac{\lambda_n\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}{1+\lambda_n\boldsymbol{h}_n^\top\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}e_n\\&=\hat{\boldsymbol{\theta}}[n-1]+\frac{\frac{\tilde{\psi}(-\varepsilon_n)}{-\varepsilon_n}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}{1+\frac{\tilde{\psi}(-\varepsilon_n)}{-\varepsilon_n}\boldsymbol{h}_n^\top\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}(-\varepsilon_n)\end{aligned}$$

$$= \hat{\boldsymbol{\theta}}[n-1] - \frac{\lambda_n \boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}{1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}\varepsilon_n, \tag{B1}$$

where $e_n = -\varepsilon_n$. Since $\tilde{\psi}(x)$ is an odd function and $\lambda_n = \tilde{\psi}(e_n)/e_n = \tilde{\psi}(\varepsilon_n)/\varepsilon_n$. Subtracting the true value $\boldsymbol{\theta}$ from both sides of Eq. (B1) and let $\widetilde{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n] - \boldsymbol{\theta}$, we obtain

$$\widetilde{\boldsymbol{\theta}}[n] = \widetilde{\boldsymbol{\theta}}[n-1] - \frac{\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n \tilde{\psi}(\varepsilon_n)}{1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n}. \tag{B2}$$

Let $d[n] = 1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n$ and note that $d[n] > 0$, and define $\tilde{\psi}_1(\varepsilon_n) = \tilde{\psi}(\varepsilon_n)/d[n]$.[48] From Eq. (B2), the expectation of the sum of variances at the current time n is

$$E\left[\widetilde{\boldsymbol{\theta}}^{\top}[n]\widetilde{\boldsymbol{\theta}}[n]\right] = E\left[\widetilde{\boldsymbol{\theta}}^{\top}[n-1]\widetilde{\boldsymbol{\theta}}[n-1]\right] + Q[n] - 2q[n], \tag{B3}$$

where

$$\begin{aligned} \mathrm{Q[n]} &= E\left[\frac{\boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}^2[n-1]\boldsymbol{h}_n \tilde{\psi}^2(\varepsilon_n)}{d^2[n]}\right] \\ &= E\left[\boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}^2[n-1]\boldsymbol{h}_n\right] E\left[\tilde{\psi}_1^2(\varepsilon_n)\right], \end{aligned} \tag{B4}$$

$$q[n] = E\left\{E\left[\tilde{\psi}_1(\varepsilon_n)\right]\boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\widetilde{\boldsymbol{\theta}}[n-1]\right\}. \tag{B5}$$

From Eq. (B3), we obtain

$$E\left[\widetilde{\boldsymbol{\theta}}^{\top}[n]\widetilde{\boldsymbol{\theta}}[n]\right] = E\left[\widetilde{\boldsymbol{\theta}}^{\top}[1]\widetilde{\boldsymbol{\theta}}[1]\right] + \sum_{k=2}^{n} Q[k] - 2\sum_{k=2}^{n} q[k] \geq 0, \tag{B6}$$

from which it follows that

$$\sum_{k=2}^{n} q[k] \leq \frac{1}{2}\left\{E\left[\widetilde{\boldsymbol{\theta}}^{\top}[1]\widetilde{\boldsymbol{\theta}}[1]\right] + \sum_{k=2}^{n} Q[k]\right\}. \tag{B7}$$

Let $\sum_{n=1}^{N} |\boldsymbol{h}_n|^2/N < \infty$,[48] and $\boldsymbol{h}_n$ is statistically independent of the prediction error ω_n. We can obtain $\lim_{n\to\infty} ||\boldsymbol{h}_n^2|| < \infty$. Thus, we have

$$\lim_{n\to\infty} \boldsymbol{h}_{n+1}^{\top}\boldsymbol{\Sigma}[n]\boldsymbol{h}_{n+1} < \infty. \tag{B8}$$

Then assume that $\psi(x)$ is bounded, i.e., $|\psi(x)| \leq C$, where C is a constant. Since $0 < d[n] = 1+\lambda_n \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\boldsymbol{h}_n < \infty$, $\tilde{\psi}_1(\varepsilon_n)$ is then bounded.

Finally, let $r(n) = \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}^2[n-1]\boldsymbol{h}_n$, and we can obtain $\sum_{n=1}^{\infty} r(n) = \sum_{n=1}^{\infty} \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}^2[n-1]\boldsymbol{h}_n < \infty$ under the condition of $\sup_{n\in[0,\infty]}\{e_n/\psi[e_n]\} < \infty$.[48] To summarize the above, we can conclude that $\sum_{k=2}^{n} Q[k]$ converges, and hence $\sum_{k=2}^{n} q[k]$ also converges. Therefore, $E\left[\tilde{\boldsymbol{\theta}}^{\top}[n]\tilde{\boldsymbol{\theta}}[n]\right]$ satisfies

$$\begin{aligned} &\lim_{n\to\infty} E\left[\tilde{\boldsymbol{\theta}}^{\top}[n]\widetilde{\boldsymbol{\theta}}[n]\right] = \\ &\lim_{n\to\infty}\left\{E\left[\tilde{\boldsymbol{\theta}}^{\top}[1]\widetilde{\boldsymbol{\theta}}[1]\right] + \sum_{k=2}^{n} Q[k] - 2\sum_{k=2}^{n} q[k]\right\} = b, \end{aligned} \tag{B9}$$

for a constant $b \geq 0$.

Let $s_n = \boldsymbol{h}_n^{\top}\boldsymbol{\Sigma}[n-1]\widetilde{\boldsymbol{\theta}}[n-1]$ and $b_n = E\left[\tilde{\boldsymbol{\theta}}^{\top}[n]\tilde{\boldsymbol{\theta}}[n]\right]$,[48] there exists a sequence $\{k_n\}$ that satisfies

$$s_n \geq k_n b_n, \quad \sum_{n=1}^{\infty} E\left[\tilde{\psi}_1(\varepsilon_n)\right] k_n = \infty. \tag{B10}$$

From the first condition of Eqs. (B10) and (B6), we can obtain

$$\sum_{n=1}^{\infty} E\left[\tilde{\psi}_1(\varepsilon_n)\right] k_n b_n < \sum_{n=1}^{\infty} E\left[\tilde{\psi}_1(\varepsilon_n)\right] s_n < \infty. \tag{B11}$$

From the second condition of Eqs. (B10) and (B11), it follows that for any $\varepsilon > 0$, there must exist infinitely many values of n such that $b_n < \varepsilon$. Since $\lim_{n\to\infty} b_n = b$, we have $b = 0$. Thus, we proved that

$$\lim_{n\to\infty} \tilde{\boldsymbol{\theta}}[n] = 0. \tag{B12}$$

Acknowledgment

This project was supported by the National Natural Science Foundation of China (Grant No. 62001271).

References

[1] Huber P J 1981 *Robust Statistics* (New York: Wiley)
[2] Kassam S A and Poor H V 1985 *Proc. IEEE* **73** 433
[3] Maronna R A, Martin R D and Yohai V J 2006 *Robust Statistics: Theory and Methods* (New York: Wiley)
[4] Zoubir A M, Koivunen V, Chakhchoukh Y and Muma M 2012 *IEEE Signal Process. Mag.* **29** 61
[5] Kay S M 2013 *Fundamentals of Statistical Signal Processing* Vol. 3 (New Jersey: Prentice Hall)
[6] Huber P J 1964 *Ann. Math. Stat.* **35** 73
[7] Haykin S S 2002 *Adaptive Filter Theory* 4th edn. (New Jersey: Prentice Hall)
[8] Diallo O, Rodrigues J J P C and Sene M 2012 *J. Netw. Comput. Appl.* **35** 1013
[9] Hassan A A and Hassan T M 2022 *Eur. J. Inf. Technol. Comput. Sci.* **2** 1
[10] Kalman R E 1960 *J. Basic Eng.* **82** 35
[11] Mathews V J and Xie Z 1993 *IEEE Trans. Signal Process.* **41** 2075
[12] Puthenpura S, Sinha N K and Vidal O P 1986 *Proc. IFAC Symp. Stochastic Control* (Vilnius: IFAC) p.23
[13] Pham D S and Zoubir A M 2004 *IEEE Signal Process. Lett.* **12** 21
[14] Zou Y X, Chan S C and Ng T S 2000 *IEEE Signal Process. Lett.* **7** 324
[15] Chan S C and Zou Y X 2004 *IEEE Trans. Signal Process.* **52** 975
[16] Deng G 2008 *EURASIP J. Adv. Signal Process.* **2008** 1
[17] West M 1981 *J. R. Stat. Soc. B* **43** 157
[18] Mitra S, Mitra A and Kundu D 2011 *Commun. Nonlinear Sci. Numer. Simul.* **16** 2796
[19] Benzi R, Sutera A and Vulpiani A 1981 *J. Phys. A: Math. Gen.* **14** L453
[20] Wiesenfeld K and Moss F 1995 *Nature* **373** 33
[21] Shi P M, Zhang W Y, Han D Y and Li M D 2019 *Chaos Solitons Fract.* **128** 155
[22] Rebolledo-Herrera L F and Espinosa G F V 2016 *Digit. Signal Process.* **52** 55
[23] Li W J, Ren Y H and Duan F B 2022 *Chin. Phys. B* **31** 080503
[24] Xu Z, Fu Y, Mei R, Zhai Y and Kang Y 2025 *Nonlinear Dyn.* **113** 497
[25] Jin Y F, Xu P F, Li Y G, Ma J Z and Xu Y 2023 *Adv. Mech.* **53** 357
[26] Jin Y F, Wang H T and Zhang T T 2024 *Chin. Phys. B* **33** 010501
[27] Zhang Y Q and Lei Y M 2024 *Chin. Phys. B* **33** 038201
[28] Patel A and Kosko B 2010 *IEEE Signal Process. Lett.* **17** 1005
[29] Uhlich S 2015 *IEEE Trans. Signal Process.* **63** 5535
[30] Duan F, Pan Y, Chapeau-Blondeau F and Abbott D 2019 *IEEE Trans. Signal Process.* **67** 4611
[31] Pan Y, Duan F, Chapeau-Blondeau F and Abbott D 2018 *IEEE Trans. Signal Process.* **66** 1953

[32] Chen H, Varshney P K and Michels J H 2008 *IEEE Trans. Signal Process.* **56** 5074
[33] Pan Y, Duan F, Xu L and Chapeau-Blondeau F 2019 *Physica A* **534** 120835
[34] Osoba O, Mitaim S and Kosko B 2013 *Fluct. Noise Lett.* **12** 1350012
[35] Osoba O and Kosko B 2016 *Fluct. Noise Lett.* **15** 1650007
[36] Yaseen M, Canbilen A E and Ikki S 2023 *IEEE Trans. Veh. Technol.* **72** 14330
[37] Ashraf U and Begh G R 2022 *IEEE Trans. Veh. Technol.* **72** 648
[38] Al-Rubaye G A, ALRikabi H T S and Hazim H T 2023 *Herit. Sustain. Dev.* **5** 239
[39] Yuksel M E 2006 *IEEE Trans. Image Process.* **15** 928
[40] Bar L, Brook A, Sochen N and Kiryati N 2007 *IEEE Trans. Image Process.* **16** 1101
[41] Hirakawa K and Parks T W 2006 *IEEE Trans. Image Process.* **15** 2730
[42] Rogers J, Ball J E and Gurbuz A C 2021 *IET Radar Sonar Navig.* **15** 431
[43] Sud S 2017 *Eur. J. Eng. Technol. Res.* **2** 40
[44] Kosorok M R 2008 *Introduction to Empirical Processes and Semiparametric Inference* (New York: Springer)
[45] Newey W K and McFadden D 1994 *Handbook of Econometrics* vol 4 (Amsterdam: Elsevier) pp. 2111–2245
[46] Wasserman L 2013 *All of Statistics: A Concise Course in Statistical Inference* (New York: Springer)
[47] Ljung L and Söderström T 1983 *Theory of Recursive Identification* (Cambridge: MIT Press)
[48] Puthenpura S C 1985 *Robust Identification of Dynamic Systems* (Hamilton: McMaster University)
[49] Kassam S A 1988 *Signal Detection in Non-Gaussian Noise* (New York: Springer)
[50] Balandat M, Karrer B, Jiang D, *et al.* 2020 *Adv. Neural Inf. Process. Syst.* **33** 21524
[51] Schulz E, Speekenbrink M and Krause A 2018 *J. Math. Psychol.* **85** 1
[52] Zhao J, Zhang J A, Li Q, Zhang H B and Wang X Y 2022 *Signal Processing* **199** 108611
[53] Chapeau-Blondeau F and Rousseau D 2004 *IEEE Trans. Signal Process.* **52** 1327
[54] Rousseau D and Chapeau-Blondeau F 2007 *IEEE Trans. Instrum. Meas.* **56** 2658
[55] Nikias C L and Shao M 1995 *Signal Processing with Alpha-Stable Distributions and Applications* (New York: Wiley-Interscience)
[56] Xu Z, Huo J, Kang Y and Wang C 2025 *Cogn. Neurodyn.* **19** 92