Signal estimation and filtering from quantized observations via adaptive stochastic resonance

Fei Li and Fabing Duan*
Institute of Complexity Science, Qingdao University, Qingdao 266071, People's Republic of China

François Chapeau-Blondeau†
Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS), Université d'Angers, 62 avenue Notre Dame du Lac, 49000 Angers, France

Derek Abbott‡
Centre for Biomedical Engineering (CBME) and School of Electrical & Electronic Engineering, University of Adelaide, Adelaide, South Australia 5005, Australia

(Received 20 February 2021; accepted 22 April 2021; published 5 May 2021)

Using a gradient-based algorithm, we investigate signal estimation and filtering in a large-scale summing network of single-bit quantizers. Besides adjusting weights, the proposed learning algorithm also adaptively updates the level of added noise components that are intentionally injected into quantizers. Experimental results show that minimization of the mean-squared error requires a nonzero optimal level of the added noise. The process adaptively achieves in this way a form of stochastic resonance or noise-aided signal processing. This adaptive optimization method of the level of added noise extends the application of adaptive stochastic resonance to some complex nonlinear signal processing tasks.

DOI: 10.1103/PhysRevE.103.052108

I. INTRODUCTION

An optimal amount of existing background noise or intentionally added noise sometimes is beneficial to physical, biological, and engineering systems [1–9]. Such noise benefits, originally named stochastic resonance (SR) [10], offer a possible explanation of the functional role of neuronal noise in nervous systems [11–13] and provide a means of improving information processing capabilities of nonlinear systems [6,9,14–20].

From the practical point of view, a natural way of utilizing the SR effect is adding a suitable amount of noise to a given nonlinear system, and then operating it in an optimally noisy environment. However this is not always feasible. Adding more noise to a given nonlinear system is not useful when the input signal is already corrupted by too much existing noise [4,21]. In other words, a given nonlinear system is operated at a larger noise level than the "optimum" one that corresponds to the stochastic resonance peak, and adding more noise will only degrade the system performance. Moreover, it is noted that an important feature of certain nonlinear systems is the optimal frequency range where the input signal can be transferred [22–26]. When the system operates out of this optimal frequency range, adding noise has a beneficial effect for improving the signal transfer; otherwise, adding noise only degrades the system response [24–27]. Another limitation is that adding noise or reducing noise in the system might not be a useful effort, because the system itself has a good adaptability to a given noisy environment; for instance, the activity-deprived neuron actively lowers its action potential threshold to improve its sensitivity to weak stimuli [2–4,8,11]. A number of issues are known to limit the practical exploitation of SR in a *single* nonlinear system [3,28]. Nevertheless, studies of SR in a summing network of excitable neuron models or threshold devices demonstrate that adding more noise may be of no use for a single system, but can significantly enhance the response of a large-scale summing network [3,13,28]. Moreover, even for conditions wherein a single neuron or device has the capability of a local self-optimization, the addition of noise is actually necessary for maximizing the information flow of the large-scale network [3,4,13,28].

Therefore, adding a suitable amount of noise to a large-scale summing network is gaining the attention of researchers. Since the addition of noise can be artificially controlled, finding the optimal level of added noise or the optimal noise type becomes an interesting question [6,14,16–20,29–39]. Beyond the analytical studies of optimal noise [6,14,16–20,29–40], a gradient-based learning law that continuously finds the optimal noise level has been proposed to adaptively learn the SR effect, which is called adaptive SR [6,17] and is also applied to image enhancement [14,17,20] and nonlinear signal processing [40,41].

In this paper, we investigate the constructive role of added noise in networks of single-bit quantizers for signal parameter estimation and nonlinear filtering by using the approach of adaptive SR, as illustrated in Fig. 1. We first consider a summing network of M quantizers that all receive the same input

*fabing.duan@gmail.com
†f.chapeau@univ-angers.fr
‡derek.abbott@adelaide.edu.au

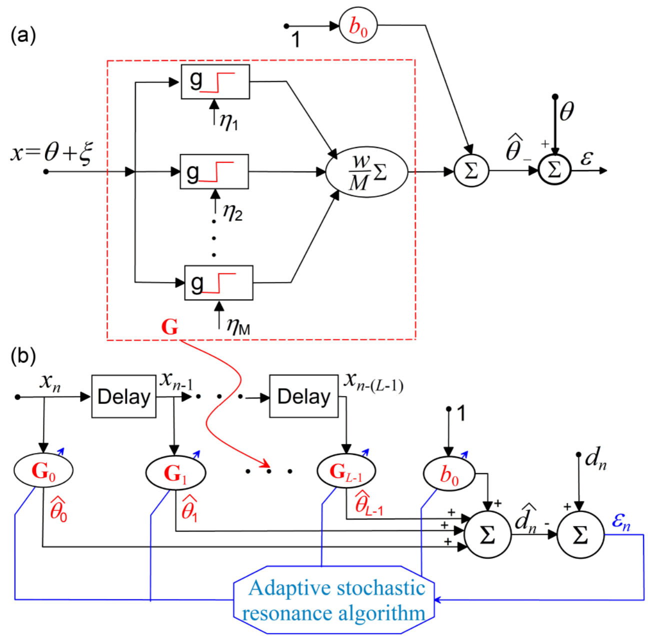

FIG. 1. Block diagram representations of (a) the summing threshold network model G for the noise-enhanced estimator $\hat{\theta}$ and (b) the adaptive noise-enhanced transversal filter $\hat{d}_n$.

x, but each quantizer is subjected to the added noise η_m for $m = 1, 2, \ldots, M$. For a large-scale network with a sufficiently large M, the input-output characteristic of this summing network can be approximated as a differentiable function that allows us to use the gradient-based learning rule to adaptively adjust the quantizer threshold, the network weights, and the level of added noise. We further combine a number of such summing networks into a transversal filter and apply this gradient-based algorithm to tracking the desired signal based on quantized observations. Experimental results show that this learning scheme can adaptively reach the SR peak that corresponds to a global or local minimum mean-squared error (MSE), which is achieved at a nonzero optimal noise level of added noise. These interesting results also suggest the potential applications of the adaptive SR to some complex signal processing problems in practical situations.

II. SIGNAL PARAMETER ESTIMATION VIA ADAPTIVE SR

We first consider a summing network model G as shown in Fig. 1(a) with a common scalar observation

$$x = \theta + \xi, \tag{1}$$

where the parameter θ is an unknown random variable with the *prior* probability density function (PDF) $f_\theta(\theta)$, and the zero-mean background noise ξ, independent of the θ, has its PDF $f_\xi(\xi)$. We here add a number of artificial noise components η_m $(m = 1, 2, \ldots, M)$ into the common observation x and send these mixed inputs $x + \eta_m$, respectively, to M single-bit quantizers described by

$$g(x+\eta_m) = \begin{cases} 1, & x+\eta_m \geqslant \gamma, \\ 0, & x+\eta_m < \gamma, \end{cases} \tag{2}$$

with the response threshold γ. The noise components η_m are mutually independent and accord to a common PDF $f_\eta(\eta)$. Then, multiplying the average value of the summing network by the weight w and adding a variable bias b_0, we design a noise-enhanced estimator

$$\hat{\theta}(x) = b_0 + \frac{w}{M}\sum_{m=1}^{M} g(x+\eta_m) \tag{3}$$

to estimate the parameter θ, as illustrated in Fig. 1(a). Moreover, the estimator $\hat{\theta}$ in Eq. (3) is designed to be unbiased, i.e., $\mathrm{E}_{x,\eta}(\hat{\theta}) = b_0 + w\mathrm{E}_{x,\eta}[g(x+\eta)] = \mathrm{E}_\theta(\theta)$, because of $\mathrm{E}_{x,\eta}[g(x+\eta_m)] = \mathrm{E}_{x,\eta}[g(x+\eta)]$ for $m = 1, 2, \ldots, M$. Here, the mean $\mathrm{E}_\theta(\theta)$ of θ is known, the operator $\mathrm{E}_\theta(\cdot)$ denotes the expectation with respect to the PDF f_θ, and the operator $\mathrm{E}_{x,\eta}(\cdot)$ represents the expectation with respect to the joint PDF $f_{x,\eta}(x,\eta)$ of random variables x and η. We then find the bias

$$b_0 = \mathrm{E}_\theta(\theta) - w\mathrm{E}_{x,\eta}[g(x+\eta)]. \tag{4}$$

Substituting Eq. (4) into $\hat{\theta}$ of Eq. (3), we have

$$\hat{\theta} = \mathrm{E}_\theta(\theta) + \frac{w}{M}\sum_{m=1}^{M} \widetilde{g}(x+\eta_m), \tag{5}$$

where $\widetilde{g}(x+\eta_m) = g(x+\eta_m) - \mathrm{E}_{x,\eta}[g(x+\eta)]$.

Defining the error signal $\varepsilon = \theta - \hat{\theta}$, the MSE of the designed estimator $\hat{\theta}$ in Eq. (5) is given by

$$\begin{aligned}
\mathcal{R} &= \mathrm{E}_{x,\eta}[(\theta-\hat{\theta})^2] \\
&= \mathrm{E}_{x,\eta}\left[\left(\widetilde{\theta} - \frac{w}{M}\sum_{m=1}^{M}\widetilde{g}(x+\eta_m)\right)^2\right] \\
&= \mathrm{E}_\theta(\widetilde{\theta}^2) - 2w\mathrm{E}_{x,\eta}[\widetilde{\theta}\widetilde{g}(x+\eta)] \\
&\quad + \frac{w^2}{M^2}\Bigg\{\sum_{m=1}^{M}\mathrm{E}_{x,\eta}[\widetilde{g}^2(x+\eta_m)] \\
&\quad + \mathrm{E}_x\left(\sum_{m=1}^{M}\mathrm{E}_\eta[\widetilde{g}(x+\eta_m)]\sum_{n=1}^{M}\mathrm{E}_\eta[\widetilde{g}(x+\eta_n)]\right)\Bigg\}(m \neq n) \\
&= \mathrm{E}_\theta(\widetilde{\theta}^2) - 2w\mathrm{E}_{x,\eta}[\widetilde{\theta}\widetilde{g}(x+\eta)] + w^2\mathrm{E}_x\{\mathrm{E}_\eta^2[\widetilde{g}(x+\eta)]\} \\
&\quad + \frac{w^2}{M}\left(\mathrm{E}_{x,\eta}[\widetilde{g}^2(x+\eta)] - \mathrm{E}_x\{\mathrm{E}_\eta^2[\widetilde{g}(x+\eta)]\}\right),
\end{aligned} \tag{6}$$

where $\widetilde{\theta} = \theta - \mathrm{E}_\theta(\theta)$, $\widetilde{g}(x+\eta) = g(x+\eta) - \mathrm{E}_{x,\eta}[g(x+\eta)]$ and $\mathrm{E}_\eta(\cdot)$ is the expectation with respect to the PDF $f_\eta(\eta)$.

Using the Jensen inequality and for the given observation x, we obtain the inequality $\mathrm{E}_\eta[\widetilde{g}^2(x+\eta)] \geqslant \mathrm{E}_\eta^2[\widetilde{g}(x+\eta)]$ based on the convex function $f(x) = x^2$. Then, $\mathrm{E}_{x,\eta}[\widetilde{g}^2(x+\eta)] \geqslant \mathrm{E}_x\{\mathrm{E}_\eta^2[\widetilde{g}(x+\eta)]\}$. Therefore, it is indicated in Eq. (6) that, for a given observation x and the added noise components η_m, the MSE $\mathcal{R}$ is a monotonically decreasing function of the number M. We here are interested in the limit of the MSE

$$\begin{aligned}
\mathcal{R}_\infty &= \lim_{M\to\infty}\mathcal{R} \\
&= \mathrm{E}_\theta(\widetilde{\theta}^2) - 2w\mathrm{E}_{x,\eta}[\widetilde{\theta}\widetilde{g}(x+\eta)] + w^2\mathrm{E}_x\{\mathrm{E}_\eta^2[\widetilde{g}(x+\eta)]\} \\
&= \mathrm{E}_x\{(\widetilde{\theta} - w\mathrm{E}_\eta[\widetilde{g}(x+\eta)])^2\}
\end{aligned} \tag{7}$$

for a sufficiently large number M. Interestingly, as $M \to \infty$, the average in Eq. (5) asymptotically converges to

$$\lim_{M\to\infty} \frac{1}{M}\sum_{m=1}^{M} \widetilde{g}(x+\eta_m) = \mathrm{E}_\eta[\widetilde{g}(x+\eta)] \tag{8}$$

and the designed estimator $\hat{\theta}$ of Eq. (5) reduces to

$$\hat{\theta}_\infty = \lim_{M\to\infty} \hat{\theta} = \mathrm{E}_\theta(\theta) + w\mathrm{E}_\eta[\widetilde{g}(x+\eta)]. \tag{9}$$

Substituting Eq. (9) into Eq. (6), the MSE of estimator $\hat{\theta}_\infty$ is just the limit $\mathcal{R}_\infty$ given by Eq. (7).

From Eq. (7), it is seen that the MSE $\mathcal{R}_\infty$ is a function of the weight w, the added noise PDF $f_\eta(\eta)$, and the response threshold γ. We here consider a given noise PDF $f_\eta(\eta)$ with the tunable level σ_η of the added noise and simplify $g(x+\eta)$ as g in the following deductions. Then, the partial derivative of $\mathcal{R}_\infty$ with respect to the weight w can be written as

$$\frac{\partial \mathcal{R}_\infty}{\partial w} = -2\mathrm{E}_{x,\eta}(\widetilde{\theta}\widetilde{g}) + 2w\mathrm{E}_x[\mathrm{E}_\eta^2(\widetilde{g})]. \tag{10}$$

The partial derivative of $\mathcal{R}_\infty$ with respect to the level σ_η of the added noise is given by

$$\begin{aligned}\frac{\partial \mathcal{R}_\infty}{\partial \sigma_\eta} &= -2w\frac{\partial \mathrm{E}_{x,\eta}(\widetilde{\theta}\widetilde{g})}{\partial \sigma_\eta} + w^2\frac{\partial \mathrm{E}_x[\mathrm{E}_\eta^2(\widetilde{g})]}{\partial \sigma_\eta}\\ &= -2w\left(\mathrm{E}_x\left[\theta\frac{\partial \mathrm{E}_\eta(g)}{\partial \sigma_\eta}\right] - \mathrm{E}_\theta(\theta)\mathrm{E}_x\left[\frac{\partial \mathrm{E}_\eta(g)}{\partial \sigma_\eta}\right]\right)\\ &\quad +2w^2\left(\mathrm{E}_x\left[\mathrm{E}_\eta(g)\frac{\partial \mathrm{E}_\eta(g)}{\partial \sigma_\eta}\right] - \mathrm{E}_{x,\eta}(g)\mathrm{E}_x\left[\frac{\partial \mathrm{E}_\eta(g)}{\partial \sigma_\eta}\right]\right)\end{aligned} \tag{11}$$

with

$$\frac{\partial \mathrm{E}_\eta(g)}{\partial \sigma_\eta} = \int_{\gamma-x}^{+\infty} \frac{\partial f_\eta(\eta)}{\partial \sigma_\eta} d\eta = \frac{\gamma - x}{\sqrt{2\pi}\sigma_\eta^2}\exp\left(-\frac{(\gamma-x)^2}{2\sigma_\eta^2}\right). \tag{12}$$

Substituting γ for σ_η in Eq. (11) and noting

$$\frac{\partial \mathrm{E}_\eta(g)}{\partial \gamma} = \frac{\partial}{\partial \gamma}\int_{\gamma-x}^{+\infty} f_\eta(\eta) d\eta = \frac{-1}{\sqrt{2\pi}\sigma_\eta}\exp\left(-\frac{(\gamma-x)^2}{2\sigma_\eta^2}\right), \tag{13}$$

we can also obtain the partial derivatives of $\partial\mathcal{R}_\infty/\partial\gamma$. Then, the learning rule of the weight w, the noise level σ_η, and the quantizer threshold γ can be expressed as

$$\Theta(k) = \Theta(k-1) - \mu\frac{\partial \mathcal{R}_\infty}{\partial \Theta}\bigg|_{\Theta=\Theta(k-1)}, \tag{14}$$

where the network parameter $\Theta \in \{w, \sigma_\eta, \gamma\}$, $\Theta(0)$ denotes the initial values, and the learning rate $\mu > 0$ for the iteration times $k \geqslant 1$.

For instance, consider a uniformly distributed parameter θ buried in the Gaussian white noise ξ [37,42]. The *prior* PDF of θ is $f_\theta(x) = 1/a$ for $0 \leqslant x \leqslant a$ ($a = 2$) and otherwise zero, and ξ has the PDF $f_\xi(x) = \exp(-x^2/2\sigma_\xi^2)/\sqrt{2\pi\sigma_\xi^2}$ with zero mean and variance $\sigma_\xi^2 = 1$. When the threshold of quantizers

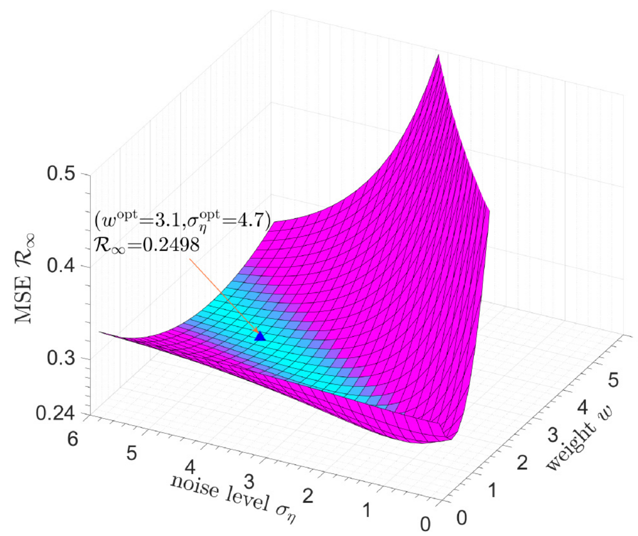

FIG. 2. MSE $\mathcal{R}_\infty$ in Eq. (7) as a function of the weight w and the noise level σ_η for a given quantizer threshold $\gamma = 0$. The minimum $\mathcal{R}_\infty = 0.2498$ (▲) at the optimal weight $w^{\mathrm{opt}} = 3.1$ and the optimal level $\sigma_\eta^{\mathrm{opt}} = 4.7$ of the added Gaussian noise is marked.

in Eq. (2) is fixed as $\gamma = 0$ and the added Gaussian noise is with its PDF $f_\eta(\eta) = \exp(-\eta^2/2\sigma_\eta^2)/\sqrt{2\pi\sigma_\eta^2}$, the MSE $\mathcal{R}_\infty$ can be illustratively plotted in Fig. 2 as a function of the weight w and the level σ_η of the added Gaussian noise. It is seen in Fig. 2 that the minimum MSE $\mathcal{R}_\infty = 0.2498$ is achieved at an optimal nonzero noise level $\sigma_\eta^{\mathrm{opt}} = 4.7$ and the optimal weight $w^{\mathrm{opt}} = 3.1$. However, $\mathcal{R}_\infty = 0.2498$ is still larger than the minimum MSE 0.2492 obtained by the minimum MSE estimator [37,42]

$$\begin{aligned}\hat{\theta}_{\mathrm{mmse}}(x) &= \mathrm{E}_{\theta|x}(\theta|x) = \int \theta f_{\theta|x}(\theta|x)d\theta\\ &= x + \sigma_\xi\sqrt{\frac{2}{\pi}}\frac{e^{-\frac{x^2}{2\sigma_\xi^2}} - e^{-\frac{(x-a)^2}{2\sigma_\xi^2}}}{\mathrm{erf}\left(\frac{x}{\sqrt{2}\sigma_\xi}\right) - \mathrm{erf}\left(\frac{x-a}{\sqrt{2}\sigma_\xi}\right)}\end{aligned} \tag{15}$$

with the conditional posterior PDF $f_{\theta|x}(\theta|x) = f_\theta(\theta)f_\xi(x-\theta)/\int f_\theta(\theta)f_\xi(x-\theta)d\theta$.

However, the optimized MSE $\mathcal{R}_\infty = 0.2498$ in Fig. 2 is manually searched by griding the weight w and the noise level σ_η for a fixed quantizer threshold $\gamma = 0$. Next, we show that the learning rule in Eq. (14) can adaptively find a smaller MSE $\mathcal{R}_\infty$ with the corresponding optimal noise level $\sigma_\eta^{\mathrm{opt}}$, weight w^{opt}, and threshold γ^{opt}. Using the learning rule in Eq. (14), it is shown in Fig. 3(a) that, after 69 iterations, the MSE $\mathcal{R}_\infty$ of the designed filter $\hat{\theta}_\infty$ in Eq. (9) reaches the minimum value 0.2492 achieved by the minimum MSE estimator of Eq. (15). Interestingly, the converged noise level $\sigma_\eta^{\mathrm{opt}} = 1.806$ as shown in Fig. 3(b) clearly manifests that the benefit of added noise exists for the considered parameter estimation problem. The converged optimal weight $w^{\mathrm{opt}} = 1.362$ and threshold $\gamma^{\mathrm{opt}} = 0.976$ are also obtained in 69 iterations, as shown in Figs. 3(c) and 3(d), respectively. Compared with the method of finding an optimal noise PDF in Refs. [19,37], the feasibility and efficacy of this adaptive SR algorithm for signal estimation is demonstrated. We also numerically demonstrate

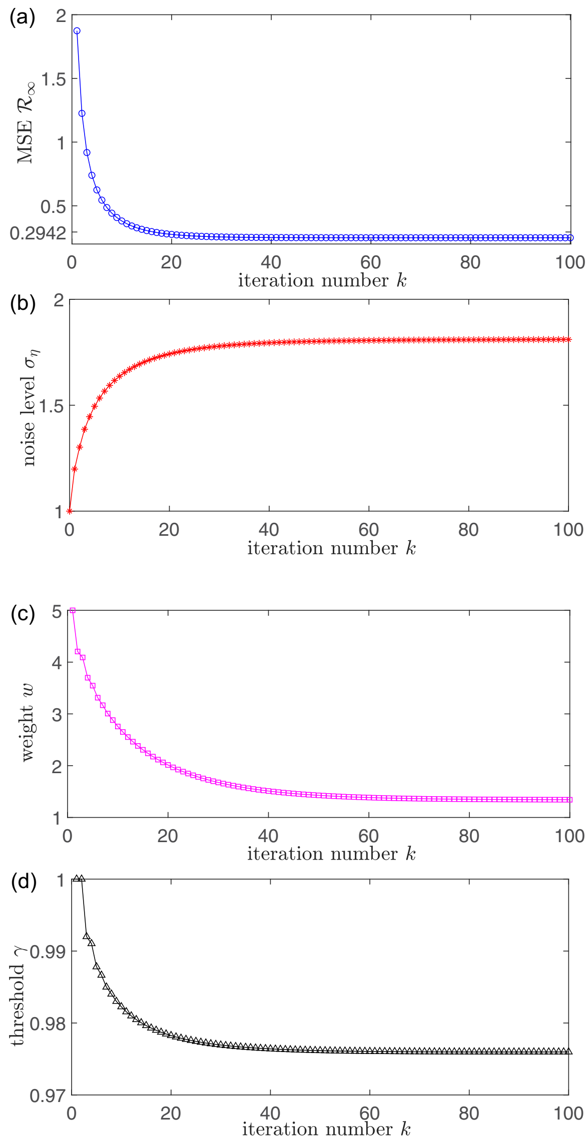

FIG. 3. Learning curves of (a) MSE $\mathcal{R}_\infty$ in Eq. (7), (b) the level σ_η of the added noise, (c) the weight w, and (d) the quantizer threshold γ for estimating an uniformly distributed parameter θ buried in the Gaussian white noise ξ. The learning rate μ in Eq. (14) is chosen as 1 for the weight w, 0.1 for the noise level σ_η, and 0.01 for the threshold γ.

the noise-enhanced parameter estimation with the converged parameters σ_η^{opt}, w^{opt}, and threshold γ^{opt} for 10^6 estimation experiments for 10^6 values of the parameter θ randomly drawn according to the PDF $f_\theta(\theta)$. In the numerical realization, the finite number $M = 10^4$ of the summing network is employed. The numerical MSE of the designed estimator in Eq. (3) is 0.2494, which is very close to the minimum one 0.2492.

III. NONLINEAR FILTERING VIA ADAPTIVE SR

The summing network model G in Fig. 1(a) and the proposed adaptive SR learning rule in Eq. (14) can be further extended to the nonlinear filtering problem based on quantized observations. An L-order transversal filter is implemented with the network model G_ℓ and unit delay elements for $\ell = 0, 1, \dots, L-1$, as shown in Fig. 1(b). The input vector $\boldsymbol{x} = [x_n, x_{n-1}, \dots, x_{n-(L-1)}]^\top$ is the data sequence for estimating a desired signal d_n by the current and past samples $x_{n-\ell}$. Here, the subscripts are used as time indexes. This nonlinear filter might be meaningful in the context of the widespread use of low-power and low-complexity sensors (e.g., quantizers) in practical engineering systems to face the demanding requirements of cost constraints and bandwidth limitations [19,43–46].

We assume that the input $\boldsymbol{x}$ and the added noise components $\eta_{\ell m}$ in each network model G_ℓ are statistically stationary, and set each network model G_ℓ with its weight w_ℓ and output $\hat{\theta}_\ell$ for $\ell = 0, 1, \dots, L-1$. Then, we have the weight vector $\boldsymbol{w} = [w_0, w_1, \dots, w_{L-1}]^\top$ and the output vector $\hat{\boldsymbol{\theta}} = [\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_{L-1}]^\top$. Here, we still consider that each threshold network G_ℓ incorporates a sufficiently large number M of one-bit quantizers, and thus its output can be also approximated as $\hat{\theta}_\ell = \mathrm{E}_\eta[g(x_{n-\ell} + \eta)]$ as indicated in Eq. (8). Then, the designed filter output at the time index n can be expressed as

$$\hat{d}_n = b_0 + \boldsymbol{w}^\top \hat{\boldsymbol{\theta}}. \tag{16}$$

Since the expected value $\mathrm{E}_{x,\eta}(\hat{d}_n)$ of Eq. (16) equals the expectation $\mathrm{E}_d(d_n)$ of the desired signal d_n, then we set the bias $b_0 = \mathrm{E}_d(d_n) - \boldsymbol{w}^\top \mathrm{E}_{x,\eta}(\hat{\boldsymbol{\theta}})$ and rewrite the filter output $\hat{d}_n$ in Eq. (16) as

$$\hat{d}_n = \mathrm{E}_d(d_n) + \boldsymbol{w}^\top [\hat{\boldsymbol{\theta}} - \mathrm{E}_{x,\eta}(\hat{\boldsymbol{\theta}})]. \tag{17}$$

Define the error signal $\varepsilon = d_n - \hat{d}_n$, the MSE of the designed filter in Eq. (17) is given by

$$\begin{aligned} \mathcal{J} &= \mathrm{E}_{x,\eta}[(d_n - \hat{d}_n)^2] \\ &= \mathrm{var}(d_n) - 2\boldsymbol{w}^\top \boldsymbol{p} + \boldsymbol{w}^\top \boldsymbol{C} \boldsymbol{w}, \end{aligned} \tag{18}$$

where the variance $\mathrm{var}(d_n) = \mathrm{E}_d(d_n^2) - \mathrm{E}_d^2(d_n)$, and the $L \times 1$ cross-correlation vector $\boldsymbol{p} = \mathrm{E}_{x,\eta}\{[d_n - \mathrm{E}_d(d_n)][\hat{\boldsymbol{\theta}} - \mathrm{E}_{x,\eta}(\hat{\boldsymbol{\theta}})]\}$ has elements

$$\begin{aligned} [\boldsymbol{p}]_{\ell+1} &= \mathrm{E}_{x_{n-\ell},d_n}\{d_n \mathrm{E}_\eta[g(x_{n-\ell} + \eta)]\} - \mathrm{E}_d(d_n) \\ &\quad \times \mathrm{E}_{x_{n-\ell}}\{\mathrm{E}_\eta[g(x_{n-\ell} + \eta)]\}. \end{aligned} \tag{19}$$

The $L \times L$ covariance matrix $\boldsymbol{C} = \mathrm{E}_{x,\eta}\{[\hat{\boldsymbol{\theta}} - \mathrm{E}_{x,\eta}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\theta}} - \mathrm{E}_{x,\eta}(\hat{\boldsymbol{\theta}})]^\top\}$ has the L diagonal elements

$$\begin{aligned} [\boldsymbol{C}]_{\ell+1,\ell+1} &= \mathrm{E}_{x_{n-\ell}}\{\mathrm{E}_\eta^2[g(x_{n-\ell} + \eta)]\} \\ &\quad - \mathrm{E}_{x_{n-\ell}}^2\{\mathrm{E}_\eta[g(x_{n-\ell} + \eta)]\}, \end{aligned} \tag{20}$$

and the $L(L-1)$ nondiagonal elements ($\ell \neq \kappa$)

$$\begin{aligned} &[\boldsymbol{C}]_{\ell+1,\kappa+1} \\ &= \mathrm{E}_{x_{n-\ell},x_{n-\kappa}}\{\mathrm{E}_\eta[g(x_{n-\ell} + \eta)]\mathrm{E}_\eta[g(x_{n-\kappa} + \eta)]\} \\ &\quad - \mathrm{E}_{x_{n-\ell}}\{\mathrm{E}_\eta[g(x_{n-\ell} + \eta)]\}\mathrm{E}_{x_{n-\kappa}}\{\mathrm{E}_\eta[g(x_{n-\kappa} + \eta)]\}. \end{aligned} \tag{21}$$

To develop the adaptive filtering algorithm, we calculate gradients

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}} = -2\boldsymbol{p} + 2\boldsymbol{C}\boldsymbol{w}, \tag{22}$$

$$\frac{\partial \mathcal{J}}{\partial \sigma_\eta} = -2\boldsymbol{w}^\top \frac{\partial \boldsymbol{p}}{\partial \sigma_\eta} + \boldsymbol{w}^\top \frac{\partial \boldsymbol{C}}{\partial \sigma_\eta} \boldsymbol{w}, \tag{23}$$

$$\frac{\partial \mathcal{J}}{\partial \gamma} = -2\boldsymbol{w}^\top \frac{\partial \boldsymbol{p}}{\partial \gamma} + \boldsymbol{w}^\top \frac{\partial \boldsymbol{C}}{\partial \gamma}\boldsymbol{w}, \tag{24}$$

where the $L \times 1$ gradient vectors

$$\frac{\partial \boldsymbol{p}}{\partial \Theta} = \left[\frac{\partial [\boldsymbol{p}]_1}{\partial \Theta}, \frac{\partial [\boldsymbol{p}]_2}{\partial \Theta}, \ldots, \frac{\partial [\boldsymbol{p}]_L}{\partial \Theta}\right]^\top \tag{25}$$

and the $L \times L$ symmetric gradient matrix

$$\frac{\partial \boldsymbol{C}}{\partial \Theta} = \begin{pmatrix} \frac{\partial [\boldsymbol{C}]_{11}}{\partial \Theta} & \frac{\partial [\boldsymbol{C}]_{12}}{\partial \Theta} & \cdots & \frac{\partial [\boldsymbol{C}]_{1L}}{\partial \Theta} \\ & \frac{\partial [\boldsymbol{C}]_{22}}{\partial \Theta} & \cdots & \frac{\partial [\boldsymbol{C}]_{2L}}{\partial \Theta} \\ * & & \ddots & \vdots \\ & & & \frac{\partial [\boldsymbol{C}]_{LL}}{\partial \Theta} \end{pmatrix} \tag{26}$$

for $\Theta \in \{\sigma_\eta, \gamma\}$. Using Eqs. (12) and (13) and the derivation rule of compound function, we can deduce the exact expressions of $\partial[\boldsymbol{p}]_{\ell+1}/\partial\Theta$ and $\partial[\boldsymbol{C}]_{\ell+1,\kappa+1}/\partial\Theta$, and the tedious manipulation is not included here for simplicity. Substituting Eqs. (22)–(24) into the learning rule of Eq. (14), the level σ_η of the added noise, the filter weight vector $\boldsymbol{w}$, and the threshold γ can be adaptively updated.

For example, consider the noisy input $x_n = s_n + \xi_n$, where the input signal $s_n = \sin(2\pi n/N)$ is a sampled sinusoid with $N = 16$ samples per period and the external white-noise process ξ_n is Gaussian distributed. The desired signal is assumed to be the sampled sinusoid $d_n = 2\cos(2\pi n/N)$ at the same frequency [47]. Due to the periodicity of the input and desired signals, the expectations are computed by averaging over one period, i.e., the operator $\sum_{n=1}^{N}(\cdot)/N$. Then, the variance of the desired signal is $\mathrm{var}(d_n) = 2$. When the external Gaussian noise ξ_n has a fixed level $\sigma_\xi = 0.1$ and the added noise components $\eta_{\ell i}$ are also Gaussian distributed, the learning curves of the 5-order filter MSE $\mathcal{J}$ and the noise level σ_η are plotted in Figs. 4(a) and 4(b), respectively. It is seen in Fig. 4 that the MSE $\mathcal{J}$ approaches 0.021 in 10 iterations and the level σ_η of the added noise converges to a stationary nonzero value 1.979 after 16 iterations. It is noted that the root-mean-squared error $\sqrt{J} = 0.145$, which can be appreciated in relation to the maximum amplitude 2 of $d_n = 2\cos(2\pi n/N)$. This result also clearly demonstrates the benefit of adding a suitable amount of noise into the nonlinear filtering on quantized observations.

At each iteration number k, with the iterated weight vector $\boldsymbol{w}(k)$, the level $\sigma_\eta(k)$ of the added noise, and the threshold $\gamma(k)$, we plot a numerical realization of the output $\hat{d}_n$ (circles plus solid line) of the designed filter of Eq. (17) in Fig. 5. For comparison, the desired signal $d_n = 2\cos(2\pi n/N)$ (dashed line) is also plotted. Here, each network model consists of $M = 10^4$ quantizers, and the output $\hat{d}_n$ of the 5-order designed filter starts from $n = 5$. It is noted that the iteration number k is identical to the time n, and the on-line tracking ability of the designed filter in Eq. (17) performs well by comparison with the desired signal d_n, as shown in Fig. 5. After eliminating the first 5 points of the output $\hat{d}_n$, the corresponding numerical MSE is 0.023, which agrees well with the theoretical MSE $\mathcal{J} = 0.021$.

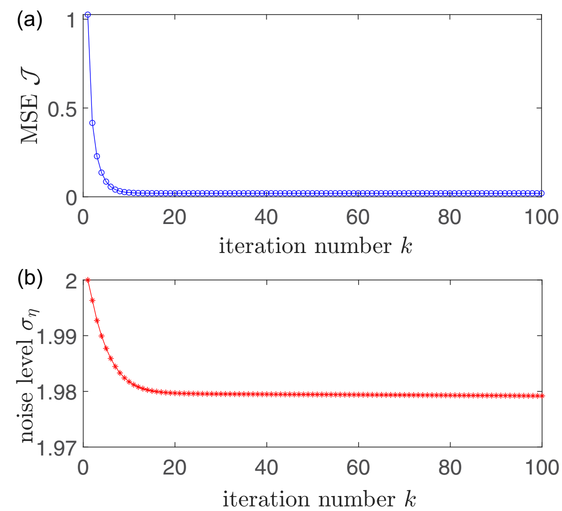

FIG. 4. Learning curves of (a) MSE $\mathcal{J}$ of the 5-order filter of Eq. (17) and (b) the level σ_η of the added noise. The learning rate $\mu = 5$ for updating the weight w, 0.01 for the noise level σ_η, and 0.001 for the threshold γ. Here, the initial noise level $\sigma_\eta(0) = 2$, the initial weight vector $\boldsymbol{w}(0) = [1, -1, -1, -1, -1]^\top$, and the initial threshold $\gamma(0) = 0$.

IV. DISCUSSION AND CONCLUSION

In this paper, an adaptive SR learning algorithm is proposed for estimating unknown parameters or filtering the output signal based on the observations from a summing network of single-bit quantizers. When a large-scale summing network has a sufficiently large number of quantizers, it is found that the network output can be approximately viewed as a differentiable function of the added noise level and network model parameters. Thus, the gradient-based learning rule is applied to the added noise level and adaptively searches the SR peak of the MSE. Results of two examples show the

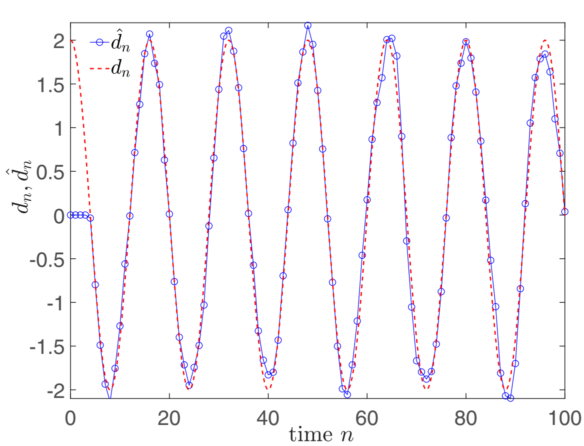

FIG. 5. A realization of the output of the designed filter $\hat{d}_n$ (circles plus solid line) and the desired signal $d_n = 2\cos(2\pi n/N)$ (dashed line). At each iteration number k (i.e., the time n), the output of the designed filter $\hat{d}_n$ is calculated by Eq. (17) with the iterated weight vector $\boldsymbol{w}(k)$, the level $\sigma_\eta(k)$ of the added noise and the threshold $\gamma(k)$, and $M = 10^5$ quantizers in each network model G_ℓ. The other parameter are the same as in Fig. 4.

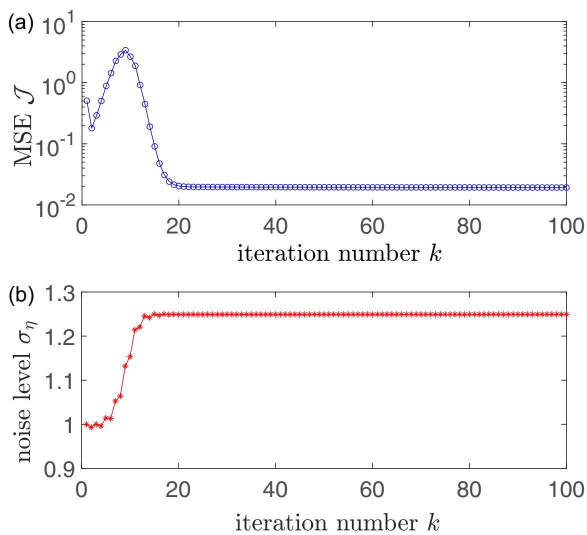

FIG. 6. Learning curves of (a) MSE $\mathcal{J}$ of the 5-order filter of Eq. (17) and (b) the level σ_η of the added noise. Here, the initial level $\sigma_\eta(0) = 1$ of the added noise and the other parameters are the same as in Fig. 4.

practicability of the proposed adaptive SR learning algorithm for signal estimation and filtering problems. The benefits of added noise are also manifested in the nonlinear signal processing. Therefore, we argue that the incorporation of noise in the design of nonlinear signal processors deserves to be further studied.

Several open questions remain. In Fig. 3 and Fig. 4, it is noted that the learning rate μ for the weight (vector) is greatly larger than that of the level of the added noise and the response threshold of quantizers. As indicated by the MSE performance surface in Fig. 2, the reason can be explained by the steep gradient of the MSE with respect to the weight (vector). The larger the learning rate for the weight (vector) is, the faster the convergence of the MSE takes place. However, how to chose suitable learning rates for the weight (vector), the level of the added noise and the threshold of quantizers, and the necessary and sufficient conditions for convergence of the proposed adaptive SR learning algorithm still remain unsolved. We must emphasize that the MSE $\mathcal{J}$ of Eq. (18) is nonconvex, because the inequality

$$\mathcal{J}[\lambda z_1 + (1-\lambda)z_2] \leqslant \lambda\mathcal{J}(z_1) + (1-\lambda)\mathcal{J}(z_2) \qquad (27)$$

does not hold for any two points $z = [\sigma_\eta, \gamma, \boldsymbol{w}^\top]^\top$ in the domain space of definition of the MSE $\mathcal{J}$. For instance, for $\lambda = 0.5$, $z_1 = [0.2964, 0.0394, 0.0339, -2.2067, -0.0130, -1.892, -1.425]^\top$, and $z_2 = [1.0723, 0.0175, 1.386, -5.989, -9.011, -9.394, -2.212]^\top$, Eq. (27) does not hold. This fact implies that the converged values $\mathcal{J} = 0.021$ and $\sigma_\eta = 1.979$ in Fig. 4 are local optimum solutions that depend crucially on the initial values in Eq. (14). For example, we only change the initial noise level $\sigma_\eta(0) = 1.5$ and keep the other parameters the same as in Fig. 4. Then, the learning curves of the MSE $\mathcal{J}$ and the noise level σ_η are plotted in Figs. 6(a) and 6(b), respectively. We find that the level σ_η of the added noise converges to a stationary nonzero value 1.249 after 16 iterations, and the MSE $\mathcal{J}$ can also approach 0.019 within 20 iterations. Although the converged solutions of the noise level σ_η are different, the local convergence of the filter MSE $\mathcal{J}$ shows general usefulness of a nonzero added noise in such large-scale summing networks of single-bit quantizers.

ACKNOWLEDGMENT

This work is supported in part by the Taishan Scholar Project of Shandong Province of China (No. ts20190930).

[1] K. Wiesenfeld and F. Moss, Stochastic resonance and the benefits of noise: From ice ages to crayfish and squids, Nature (London) **373**, 33 (1995).
[2] P. Jung, Stochastic resonance and optimal design of threshold detectors, Phys. Lett. A **207**, 93 (1995).
[3] J. J. Collins, C. C Chow, and T. T Imhoff, Stochastic resonance without tuning, Nature (London) **376**, 236 (1995).
[4] A. R. Bulsara and L. Gammaitoni, Tuning in to noise, Phys. Today **49**(3), 39 (1996).
[5] F. Chapeau-Blondeau and X. Godivier, Theory of stochastic resonance in signal transmission by static nonlinear systems, Phys. Rev. E **55**, 1478 (1997).
[6] S. Mitaim and B. Kosko, Adaptive stochastic resonance, Proc. IEEE **86**, 2152 (1998).
[7] S. M. Kay, Can detectability be improved by adding noise? IEEE Signal Process. Lett. **7**, 8 (2000).
[8] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization* (Cambridge University Press, New York, 2008).
[9] H. Chen, L. R. Varshney, and P. K. Varshney, Noise-enhanced information systems, Proc. IEEE **102**, 1607 (2014).
[10] R. Benzi, A. Sutera, and A. Vulpiani, The mechanism of stochastic resonance, J. Phys. A: Math. Gen. **14**, L453 (1981).
[11] M. D. McDonnell and D. Abbott, What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology, PLoS Comput. Biol. **5**, e1000348 (2009).
[12] G. Wenning and K. Obermayer, Activity Driven Adaptive Stochastic Resonance, Phys. Rev. Lett. **90**, 120602 (2003).
[13] N. G. Stocks, Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance, Phys. Rev. E **63**, 041114 (2001).
[14] T. Yang, Adaptively optimizing stochastic resonance in visual system, Phys. Lett. A **245**, 79 (1998).
[15] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, Sequential signal encoding from noisy measurements using quantizers with dynamic bias control, IEEE Trans. Inf. Theory **47**, 978 (2001).
[16] A. Patel and B. Kosko, Optimal mean-square noise benefits in quantizer-array linear estimation, IEEE Signal Process. Lett. **17**, 1005 (2010).

[17] S. Mitaim and B. Kosko, Adaptive stochastic resonance in noisy neurons based on mutual information, IEEE Trans. Neural Networks **15**, 1526 (2004).

[18] G. O. Balkan and S. Gezici, CRLB based optimal noise enhanced parameter estimation using quantized observations, IEEE Signal Process. Lett. **17**, 477 (2010).

[19] F. Duan, Y. Pan, F. Chapeau-Blondeau, and D. Abbott, Noise benefits in combined nonlinear Bayesian estimators, IEEE Trans. Signal Process. **67**, 4611 (2019).

[20] Y. Fu, Y. Kang, and G. Chen, Stochastic resonance based visual perception using spiking neural networks, Front. Comput. Neurosci. **14**, 24 (2020).

[21] B. Xu, F. Duan, and F. Chapeau-Blondeau, Comparison of aperiodic stochastic resonance in a bistable system realized by adding noise and by tuning system parameters, Phys. Rev. E **69**, 061110 (2004).

[22] M. H. Devoret, J. M. Martinis, D. Esteve, and J. Clarke, Resonant Activation from the Zero-Voltage State of a Current-Biased Josephson Junction, Phys. Rev. Lett. **53**, 1260 (1984).

[23] A. L. Pankratov and M. Salerno, Resonant activation in overdamped systems with noise subjected to strong periodic driving, Phys. Lett. A **273**, 162 (2000).

[24] A. L. Pankratov and M. Salerno, Adiabatic approximation and parametric stochastic resonance in a bistable system with periodically driven barrier, Phys. Rev. E **61**, 1206 (2000).

[25] A. L. Pankratov, Suppression of noise in nonlinear systems subjected to strong periodic driving, Phys. Rev. E **65**, 022101 (2002).

[26] A. A. Yablokov, V. M. Mylnikov, A. L. Pankratov, E. V. Pankratova, and A. V. Gordeeva, Suppression of switching errors in weakly damped Josephson junctions, Chaos, Solitons Fractals **136**, 109817 (2020).

[27] E. V. Pankratova, V. N. Belykh, and E. Mosekilde, Role of the driving frequency in a randomly perturbed Hodgkin-Huxley neuron with suprathreshold forcing, Eur. Phys. J. B **53**, 529 (2006).

[28] N. G. Stocks, Suprathreshold Stochastic Resonance in Multilevel Threshold Systems, Phys. Rev. Lett. **84**, 2310 (2000).

[29] L. Xu, F. Duan, X. Gao, D. Abbott, and M. D. McDonnell, Adaptive recursive algorithm for optimal weighted suprathreshold stochastic resonance, R. Soc. Open Sci. **4**, 160889 (2017).

[30] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors, IEEE Trans. Signal Process. **55**, 3172 (2007).

[31] H. Chen and P. K. Varshney, Theory of the stochastic resonance effect in signal detection: Part II—Variable detectors, IEEE Trans. Signal Process. **56**, 5031 (2008).

[32] H. Chen, P. K. Varshney, and J. H. Michels, Noise enhanced parameter estimation, IEEE Trans. Signal Process. **56**, 5074 (2008).

[33] A. Patel and B. Kosko, Optimal noise benefits in Neyman-Pearson and inequality-constrained statistical signal detection, IEEE Trans. Signal Process. **57**, 1655 (2009).

[34] S. Bayram, S. Gezici, and H. V. Poor, Noise enhanced hypothesis-testing in the restricted Bayesian framework, IEEE Trans. Signal Process. **58**, 3972 (2010).

[35] J. Zhu, X. Li, R. S. Blum, and Y. Gu, Parameter estimation from quantized observations in multiplicative noise environments, IEEE Trans. Signal Process. **63**, 4037 (2015).

[36] H. Soganci, S. Gezici, and O. Arikan, Optimal signal design for multi-parameter estimation problems, IEEE Trans. Signal Process. **63**, 6074 (2015).

[37] S. Uhlich, Bayes risk reduction of estimators using artificial observation noise, IEEE Trans. Signal Process. **63**, 5535 (2015).

[38] A. B. Akbay and S. Gezici, Noise benefits in joint detection and estimation problems, Signal Process. **118**, 235 (2016).

[39] D. Simon, J. Sulam, Y. Romano, Y. Lu, and M. Elad, MMSE approximation for sparse coding algorithms using stochastic resonance, IEEE Trans. Signal Process. **67**, 4597 (2019).

[40] X. Wu, Z. Jiang, D. W. Repperger, and Y. Guo, Enhancement of stochastic resonance using optimization theory, Commun. Inf. Syst. **6**, 1 (2006).

[41] B. Kosko and S. Mitaim, Robust stochastic resonance: Signal detection and adaptation in impulsive noise, Phys. Rev. E **64**, 051110 (2001).

[42] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Vol. 1 (Prentice-Hall, Upper Saddle River, NJ, 1998).

[43] O. Dabeer and A. Karnik, Signal parameter estimation using 1-bit dithered quantization, IEEE Trans. Inf. Theory **52**, 5389 (2006).

[44] A. Sanzi and A. Vosoughi, Distributed vector estimation for power- and bandwidth-constrained wireless sensor networks, IEEE Trans. Signal Process. **64**, 3879 (2016).

[45] G. Zeitler, G. Kramer, and A. C. Singer, Bayesian parameter estimation using single-bit dithered quantization, IEEE Trans. Signal Process. **60**, 2713 (2012).

[46] S. Engelberg, An introduction to dither, IEEE Instrum. Meas. Mag. **15**, 50 (2012).

[47] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Upper Saddle River, NJ, 1985).