



## Noise benefits in combined nonlinear Bayesian estimators

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

### ► To cite this version:

Fabing Duan, Yan Pan, François Chapeau-Blondeau, Derek Abbott. Noise benefits in combined nonlinear Bayesian estimators. IEEE Transactions on Signal Processing, 2019, 67 (17), pp.4611-4623. 10.1109/TSP.2019.2931203 . hal-02278486

**HAL Id: hal-02278486**

**<https://hal.science/hal-02278486>**

Submitted on 21 Jun 2021

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

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

# Noise Benefits in Combined Nonlinear Bayesian Estimators

Fabing Duan<sup>✉</sup>, Yan Pan<sup>✉</sup>, François Chapeau-Blondeau<sup>✉</sup>, and Derek Abbott<sup>✉</sup>, *Fellow, IEEE*

**Abstract**—This paper investigates the benefits of intentionally adding noise to a Bayesian estimator, which comprises a linear combination of a number of individual Bayesian estimators that are perturbed by mutually independent noise sources and multiplied by a set of adjustable weighting coefficients. We prove that the Bayes risk for the mean square error (MSE) criterion is minimized when the same optimum weighting coefficients are assigned to the identical estimators in the combiner. This property leads to a simplified analysis of the noise benefit to the MSE of the combined Bayesian estimator even when the number of individual estimators tends to infinity. It is shown that, for a sufficiently large number of individual estimators, the MSE of the designed Bayesian estimator approaches a plateau for a wide range of added noise levels. This robust feature facilitates the incorporation of the added noise into the design of Bayesian estimators without tuning the noise level. For an easily implementable Bayesian estimator composed of quantizers, the benefit of the symmetric scale-family noise is demonstrated, and the optimal noise probability density function is approximated by solving a constrained nonlinear optimization problem. We further extend this potential Bayesian estimator to the nonlinear filter design. Finally, examples of the noise benefits in random parameter estimation and nonlinear filtering support the theoretical analyses.

**Index Terms**—Noise benefit, Bayesian estimator, linear combination, nonlinear filtering, stochastic resonance.

## I. INTRODUCTION

IT IS well known that the closed form description of an optimal Bayesian estimator is difficult to achieve in general [1]. For instance, the implementation of a minimum mean square error (MMSE) estimator requires the solution to the mean of the posterior probability density function (PDF) of the observation via difficult integrals [1]. Thus, in practice, it is reasonable to seek some suboptimal but feasible nonlinear Bayesian

estimators with tractable forms to estimate random parameters or variables [1]–[3].

Recently, the noise benefit in nonlinear estimators [4]–[18] and detectors [19]–[33] has attracted great attentions of researchers in the field of signal processing, because the accuracy of an estimator and the detectability of a detector can be enhanced by design via intentionally adding noise. Sufficient or necessary conditions have been derived for the existence of the optimal added noise PDF [6]–[8], [16], [23]–[26], and the explicit or approximate forms of optimal added noise PDFs [7]–[13], [15], [16] have also been of great interest. Among these investigations, it was found that a parallel array of estimators can benefit from mutually independent added noise components in comparison to a single estimator [4], [5], [7]–[13]. From the parameter estimation standpoint, Uhlich [7] proposed a novel noise-enhanced estimator by averaging estimates from the same observation added by artificial noise components, and discussed its superiority over the original estimator and the noise-modified estimator derived by Chen *et al.* [6]. Based on the sum of outputs of quantizer arrays, the linear Wiener decoding scheme [9] and the linear MMSE estimation [8] of the random inputs were extensively investigated within the framework of suprathreshold stochastic resonance [34]. We also used the least-square regression algorithm to numerically study the noise benefit in a quantizer array with optimal weights in comparison with the unweighted array [35].

In this paper, we design a linear combination illustrated in Fig. 1 as a potential noise-enhanced Bayesian estimator  $\hat{\theta}_{LC}$ , which consists of two modules: Module 1 exploits the benefit of noise by adding mutually independent noise components  $\eta_i$  into each estimator  $\hat{\theta}_i$ , and module 2 outputs the linear MMSE estimation based on a set of estimates  $\{\hat{\theta}_i\}$  from module 1 multiplied by optimally tuned weighting coefficients  $w_i$  for  $i = 1, 2, \dots, m$  and a bias weighting  $w_0$ . Then, the collective responses of all individual estimators yield the combined Bayesian estimator output  $\hat{\theta}_{LC}$ , as shown in Fig. 1. We first prove that, with the Bayes risk for the MSE criterion and at a given added noise level, the MSE of  $\hat{\theta}_{LC}$  gradually reduces as the number  $m$  of individual estimators increases, leading to the minimum MSE achieved by the combined estimator  $\hat{\theta}_{LC}$  in the limit of an infinite number. It is also proved that any two identical estimators in a combiner require the same optimum weighting coefficients. This characteristic simplifies theoretical analyses of the convergence of the MSE of  $\hat{\theta}_{LC}$  with no requirement for matrix inversions, and makes it possible to recognize the noise benefit of the added noise in the limiting case of an infinite number of individual

F. Duan is with the Institute of Complexity Science, Qingdao University, Qingdao 266071, China (e-mail: fabing.duan@gmail.com).

Y. Pan is with the College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: yanpn.pan@gmail.com).

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

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

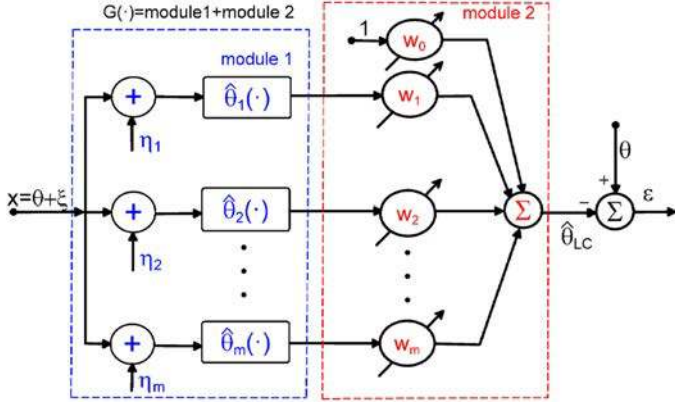


Fig. 1. Block diagram representation of a linear combination of estimators performing as a noise-enhanced Bayesian estimator. In module 1, mutually independent noise components  $\eta_i$  are intentionally added into individual estimators  $\hat{\theta}_i$ , and then each  $\hat{\theta}_i$  is multiplied by optimally tuned weighting coefficient  $w_i$  in module 2, resulting in the combined Bayesian estimator output  $\hat{\theta}_{LC}$ .

estimators. It is interesting to note that the noise-enhanced estimator  $\hat{\theta}_{NE}$  of Ref. [7] is just a special case of the combined estimator  $\hat{\theta}_{LC}$  with same fixed weighting coefficients, while the linear MMSE estimation [8], [9] corresponds to the estimator  $\hat{\theta}_{LC}$  with a combination of homogeneous quantizers. It is illustratively shown that, for a large enough number of individual estimators, the MSE of  $\hat{\theta}_{LC}$  can be reduced to a minimum for an optimal added noise level, i.e. the stochastic resonance effect [34], [36]. When the response of an original estimator  $\hat{\theta}$  has an upper bound, it is demonstrated that, for a very large number of individual estimators, the MSE of the combined estimator  $\hat{\theta}_{LC}$  approaches a plateau for a wide range of large added noise levels. If this plateau consists of ‘local’ minima of the MSE that are tolerable levels for practical applications, the noise-enhanced estimator  $\hat{\theta}_{LC}$  can be operated over a wide range of added noise levels. This robust feature suggests that, as the background noise varies, the designed estimator might be improved by the added noise without tuning levels in practice. For a special combined Bayesian estimator  $\hat{\theta}_{LC}$  with quantizers in number  $m > 1$ , we also prove that  $\hat{\theta}_{LC}$  can always benefit from any type of the zero-mean symmetric scale-family added noise, because the MSE initially decreases by adding a small amount of noise. Moreover, using constrained nonlinear optimization methods, the optimal noise PDF is also approximately solved and the obtained MSE is effectively reduced in the considered cases. The designed Bayesian estimator can be also extended to nonlinear filtering with the multiple samples of observations, and some examples demonstrate the applicability of noise benefits in the proposed Bayesian estimators. The obtained novel results indicate that the added noise can be incorporated into the design of nonlinear Bayesian estimation and nonlinear filtering as a potential technique to enhance the accuracy of estimators [4]–[18].

## II. PARAMETER MODEL AND BAYESIAN ESTIMATOR

We observe the scalar data model as

$$x = \theta + \xi, \quad (1)$$

where the parameter  $\theta$  is a random variable with the prior PDF  $f_\theta$ , and the background white noise  $\xi$ , independent of  $\theta$ , has the PDF  $f_\xi$ . Then, the observation  $x$  accords with the convolved PDF  $f_x(x) = \int f_\theta(\theta) f_\xi(x - \theta) d\theta$ . As in Fig. 1, the same observation  $x$  perturbed by the independent noise component  $\eta_i$  is operated by each estimator  $\hat{\theta}_i(x + \eta_i)$  for  $i = 1, 2, \dots, m$ , and here  $m$  denotes the total number of individual estimators in the combination. Letting the estimate vector  $\hat{\theta} = [\hat{\theta}_1(x + \eta_1), \hat{\theta}_2(x + \eta_2), \dots, \hat{\theta}_m(x + \eta_m)]^\top$  and the weight vector  $\mathbf{w} = [w_1, w_2, \dots, w_m]^\top$ , we design a new Bayesian estimator  $\hat{\theta}_{LC}(x) = w_0 + \mathbf{w}^\top \hat{\theta}$  as an unbiased estimation of the parameter  $\theta$ . Here,  $w_0$  is the bias weight. Then, the unbiased condition of expectations  $E_{x,\eta}(\hat{\theta}_{LC}) = E_\theta(\theta)$  yields the bias weight  $w_0 = E_\theta(\theta) - \mathbf{w}^\top E_{x,\eta}(\hat{\theta})$ , and then the designed estimator  $\hat{\theta}_{LC}$  can be written as

$$\hat{\theta}_{LC}(x) = E_\theta(\theta) + \mathbf{w}^\top [\hat{\theta} - E_{x,\eta}(\hat{\theta})], \quad (2)$$

where  $E_{x,\eta}(\cdot)$  denotes the expectation with respect to the joint PDF of variables  $x$  and  $\eta$  and  $E_\theta(\cdot)$  denotes the expectation with respect to the PDF of variable  $\theta$ .

Define  $\varepsilon = \theta - \hat{\theta}_{LC}$  as the error of the estimator for a particular sample of  $x$ , then the Bayes risk of  $\hat{\theta}_{LC}$  for the MSE criterion is given by

$$\mathcal{R}_m = E_{x,\eta}(\varepsilon^2) = \text{var}(\theta) - 2\mathbf{w}^\top \mathbf{p} + \mathbf{w}^\top \mathbf{C} \mathbf{w}, \quad (3)$$

where the subscript  $m$  is used to denote the number of estimators  $\hat{\theta}_i$  and the variance of  $\theta$  is  $\text{var}(\theta) = E_\theta(\theta^2) - E_\theta^2(\theta)$ . The centralized cross-correlation vector between the parameter  $\theta$  and the estimate vector  $\hat{\theta}$  is  $\mathbf{p} = E_{x,\eta}[(\theta - E_\theta(\theta))(\hat{\theta} - E_{x,\eta}(\hat{\theta}))]$  with its elements calculated as

$$\begin{aligned} p_i &= E_{x,\eta}[(\theta - E_\theta(\theta))(\hat{\theta}_i - E_{x,\eta}(\hat{\theta}_i))] \\ &= E_x\{\theta E_\eta[\hat{\theta}_i(x + \eta)]\} - E_\theta(\theta) E_x\{E_\eta[\hat{\theta}_i(x + \eta)]\}. \end{aligned} \quad (4)$$

The covariance matrix of the estimate vector  $\hat{\theta}$  is a symmetric positive definite matrix  $\mathbf{C} = E_{x,\eta}[(\hat{\theta} - E_{x,\eta}(\hat{\theta}))(\hat{\theta} - E_{x,\eta}(\hat{\theta}))^\top]$  with its diagonal elements

$$\begin{aligned} C_{ii} &= E_{x,\eta}[(\hat{\theta}_i - E_{x,\eta}(\hat{\theta}_i))^2] \\ &= E_x\{E_\eta[\hat{\theta}_i^2(x + \eta)]\} - E_x^2\{E_\eta[\hat{\theta}_i(x + \eta)]\} \end{aligned} \quad (5)$$

and the non-diagonal elements

$$\begin{aligned} C_{ij} &= E_{x,\eta}[(\hat{\theta}_i - E_{x,\eta}(\hat{\theta}_i))(\hat{\theta}_j - E_{x,\eta}(\hat{\theta}_j))] \\ &= E_x\{E_\eta[\hat{\theta}_i(x + \eta)]E_\eta[\hat{\theta}_j(x + \eta)]\} \\ &\quad - E_x\{E_\eta[\hat{\theta}_i(x + \eta)]\}E_x\{E_\eta[\hat{\theta}_j(x + \eta)]\} \end{aligned} \quad (6)$$

for  $i, j = 1, 2, \dots, m$  ( $i \neq j$ ). Note that the positive definite property of  $\mathbf{C}$  requires the individual estimator in the combination not being such degenerate cases as  $\hat{\theta}_i = \kappa$  for an arbitrary constant  $\kappa$ .

Setting the gradient  $\partial \mathcal{R}_m / \partial \mathbf{w} = -2\mathbf{p} + 2\mathbf{C} \mathbf{w}$  to zero, we have the optimum weight vector

$$\mathbf{w}^o = \mathbf{C}^{-1} \mathbf{p}. \quad (7)$$

Substituting  $\mathbf{w}^o$  of Eq. (7) into Eqs. (2) and (3), we have the final form of the designed estimator as

$$\hat{\theta}_{\text{LC}}(x) = E_\theta(\theta) + \mathbf{p}^\top \mathbf{C}^{-1}(\hat{\theta} - E_{x,\eta}(\hat{\theta})), \quad (8)$$

and the optimized MSE with respect to the weight vector is

$$\mathcal{R}_m^o = \text{var}(\theta) - \mathbf{w}^{o\top} \mathbf{C} \mathbf{w}^o = \text{var}(\theta) - \mathbf{p}^\top \mathbf{C}^{-1} \mathbf{p}. \quad (9)$$

*Theorem 1:* When the covariance matrix  $\mathbf{C}$  is positive-definite, the MSE  $\mathcal{R}_m^o$  in Eq. (9) is a monotonically decreasing function of the number  $m$ , i.e.

$$\mathcal{R}_m^o < \mathcal{R}_{m-1}^o, \quad (10)$$

for a given circumstance of the background noise  $\xi$  and the added noise  $\eta$ .

Proof of *Theorem 1* is presented in Appendix A. This theorem implies that, for a given circumstance of the background noise  $\xi$  and the added noise  $\eta$ , the minimum MSE  $\mathcal{R}_\infty^o$  is achieved by the Bayesian estimator  $\hat{\theta}_{\text{LC}}$  in the limiting case of  $m \rightarrow \infty$ . Moreover, the individual estimator  $\hat{\theta}_i$  can differ from each other.

*Theorem 2:* If two estimators  $\hat{\theta}_i = \hat{\theta}_j$  ( $i \neq j$ ) in a combiner are identical, then optimum weighting coefficients  $w_i^o = w_j^o$ .

Proof of *Theorem 2* is given in Appendix B. This theorem suggests that, when a combiner has  $L$  groups and each group has  $m_l$  identical estimators  $\hat{\theta}_{l,i} = \hat{\theta}_l$  for  $l = 1, 2, \dots, L$  and  $i = 1, 2, \dots, m_l$ , then each group has  $m_l$  identical weighting coefficients  $w_{l,i} = w_l$ . Updating the weight vector as  $\mathbf{w} = [w_1, w_2, \dots, w_L]^\top$  and rewriting the estimate vector as  $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_L]$  with  $\hat{\theta}_l = \sum_{i=1}^{m_l} \hat{\theta}_l(x + \eta_i)$ , we then obtain the combined Bayesian estimator of Eq. (8) and its MSE of Eq. (9).

*Corollary 1:* For a combiner composed of  $m$  identical original estimators  $\hat{\theta}_i = \hat{\theta}$ , the Bayesian estimator  $\hat{\theta}_{\text{LC}}$  in Eq. (8) becomes

$$\hat{\theta}_{\text{LC}}(x) = E_\theta(\theta) + w^o \sum_{i=1}^m (\hat{\theta}(x + \eta_i) - E_x\{E_\eta[\hat{\theta}(x + \eta)]\}), \quad (11)$$

where the optimum weighting coefficient is

$$w^o = p_i / [C_{ii} + (m-1)C_{ij}]. \quad (12)$$

The corresponding MSE  $\mathcal{R}_m^o$  of Eq. (9) is now simplified as

$$\mathcal{R}_m^o = \text{var}(\theta) - mp_i^2 / [C_{ii} + (m-1)C_{ij}] \quad (13)$$

with the limit

$$\mathcal{R}_\infty^o = \lim_{m \rightarrow \infty} \mathcal{R}_m^o = \text{var}(\theta) - p_i^2 / C_{ij}. \quad (14)$$

Proof of *Corollary 1* is presented in Appendix C, where  $p_i$ ,  $C_{ii}$  and  $C_{ij}$ , as special cases of Eqs. (4)–(6), are given in Eqs. (37)–(39). *Corollary 1* indicates that, for a combiner of  $m$  identical original estimators, the combined estimator of Eq. (8) with optimally weighting before summation and the estimator of Eq. (11) with optimally weighting after summation can achieve the same MSE  $\mathcal{R}_m^o$  of Eq. (13). Moreover, compared with the general expression of  $\mathcal{R}_m^o$  in Eq. (9), Eq. (13) results in considerable simplification of the computation of the MSE  $\mathcal{R}_m^o$  with no requirement for matrix inversions in the considered case of  $m$  identical estimators  $\hat{\theta}_i = \hat{\theta}$ .

*Corollary 2:* For a combination of identical estimators  $\hat{\theta}_i = \hat{\theta}$  in the limit case of  $m \rightarrow \infty$ , the designed Bayesian estimator  $\hat{\theta}_{\text{LC}}$  in Eq. (8) evolves into

$$\hat{\theta}_{\text{LC}}(x) = E_\theta(\theta) + w^o (\hat{\theta}_{\text{NE}}(x) - E_x[\hat{\theta}_{\text{NE}}(x)]), \quad (15)$$

where  $\hat{\theta}_{\text{NE}}(x)$  is the noise-enhanced estimator

$$\hat{\theta}_{\text{NE}}(x) = E_\eta[\hat{\theta}(x + \eta)] \quad (16)$$

defined by Uhlich [7] and the optimum weight is given by

$$w^o = \frac{E_x[(\theta - E_\theta(\theta))(\hat{\theta}_{\text{NE}}(x) - E_x(\hat{\theta}_{\text{NE}}))]}{E_x[(\hat{\theta}_{\text{NE}}(x) - E_x(\hat{\theta}_{\text{NE}}))^2]}. \quad (17)$$

The designed estimator  $\hat{\theta}_{\text{LC}}$  of Eq. (15) achieves the MSE  $\mathcal{R}_\infty^o$  of Eq. (14), and is never worse than the noise-enhanced estimator  $\hat{\theta}_{\text{NE}}$  of Eq. (16).

*Corollary 3:* The MSE  $\mathcal{R}_\infty^o$  of the estimator  $\hat{\theta}_{\text{LC}}$  in Eq. (15) satisfies the inequality

$$\mathcal{R}_\infty^o \geq \mathcal{R}_{\text{ms}} = E_x[(\theta - \hat{\theta}_{\text{mmse}})^2], \quad (18)$$

where  $\mathcal{R}_{\text{ms}}$  is achieved by the MMSE estimator [1]

$$\hat{\theta}_{\text{mmse}}(x) = E_{\theta|x}(\theta|x) = \int \theta f_{\theta|x}(\theta|x) d\theta \quad (19)$$

with the conditional posterior PDF  $f_{\theta|x}(\theta|x) = f_\theta(\theta)f_\xi(x - \theta) / \int f_\theta(\theta)f_\xi(x - \theta)d\theta$ . The equality of Eq. (18) occurs when

$$\hat{\theta}_{\text{mmse}}(x) - E_\theta(\theta) = \kappa[\hat{\theta}_{\text{NE}}(x) - E_x(\hat{\theta}_{\text{NE}})],$$

where  $\kappa$  is an arbitrary constant.

Proofs of *Corollaries 2* and *3* are presented in Appendices D and E, respectively, which are also illustrated in the following examples.

*Example 1:* Consider an uniformly distributed parameter  $\theta$  buried in the Gaussian white noise  $\xi$  [1], [7]. The prior PDF of  $\theta$  is  $f_\theta(x) = 1/a$  for  $0 \leq x \leq a$  and otherwise zero, and  $\xi$  has the PDF  $f_\xi(x) = \exp(-x^2/2\sigma_\xi^2) / \sqrt{2\pi\sigma_\xi^2}$  with zero-mean and variance  $\sigma_\xi^2$ . Then, the maximum a posteriori (MAP) estimator is given by [1], [7]

$$\hat{\theta}_{\text{map}}(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x \leq a, \\ a, & x > a, \end{cases} \quad (20)$$

and the MMSE estimator of Eq. (19) becomes [1], [7]

$$\hat{\theta}_{\text{mmse}}(x) = x + \sigma_\xi \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{x^2}{2\sigma_\xi^2}} - e^{-\frac{(x-a)^2}{2\sigma_\xi^2}}}{\text{erf}\left(\frac{x}{\sqrt{2}\sigma_\xi}\right) - \text{erf}\left(\frac{x-a}{\sqrt{2}\sigma_\xi}\right)}. \quad (21)$$

For the interval bound  $a = 2$  and the background noise level  $\sigma_\xi = 1$ , it is seen in Fig. 2(a) that the MSE of the MAP estimator  $\hat{\theta}_{\text{map}}$  of Eq. (20) is 0.4832 (magenta dashed line), much higher than the MMSE 0.2492 (red solid line) achieved by the MMSE estimator  $\hat{\theta}_{\text{mmse}}$  of Eq. (21). Substituting this suboptimal estimator  $\hat{\theta}_{\text{map}}$  of Eq. (20) into Eq. (11) and Eq. (16), we can construct the combined estimator  $\hat{\theta}_{\text{LC}}$  and the noise-enhanced estimator  $\hat{\theta}_{\text{NE}}$ . The added noise  $\eta$  is chosen as Gaussian white noise with PDF  $f_\eta(x) = \exp(-x^2/2\sigma_\eta^2) / \sqrt{2\pi\sigma_\eta^2}$  and the standard



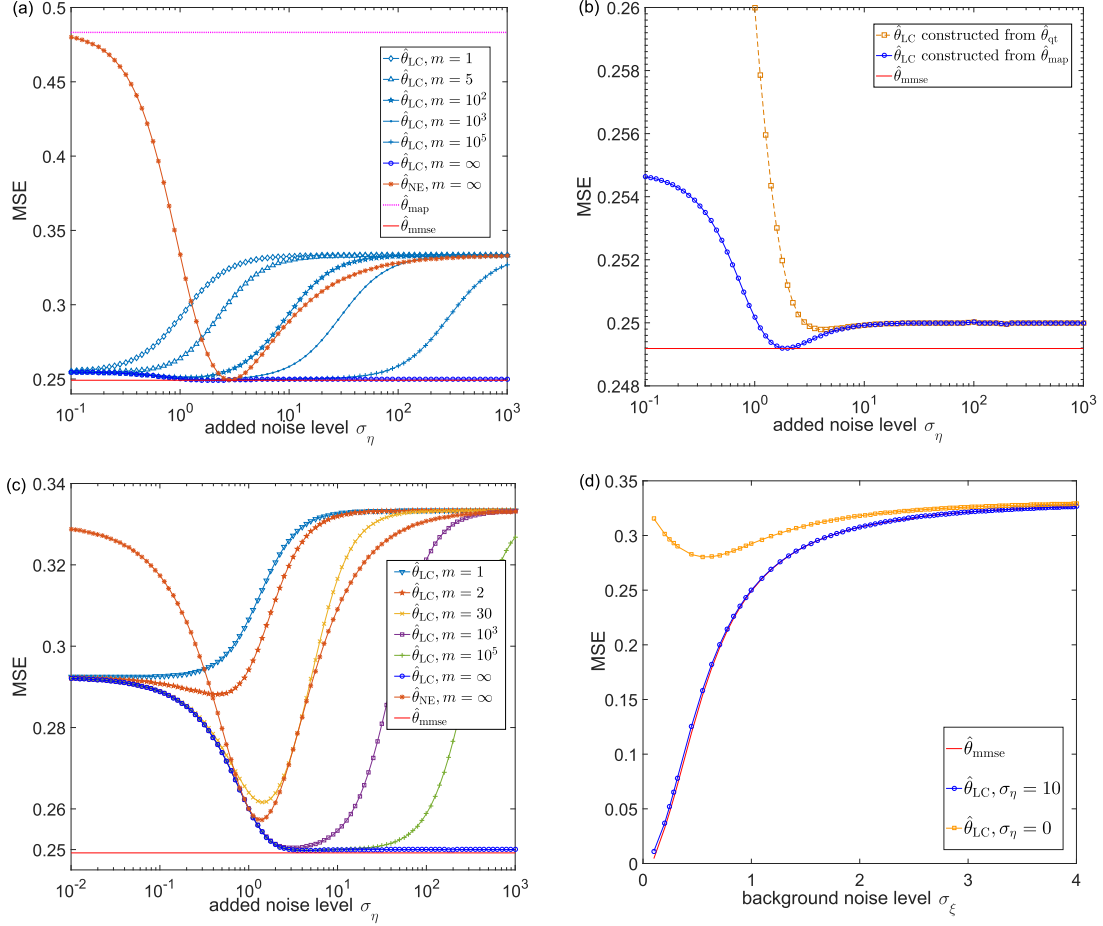


Fig. 2. (a) MSEs of the designed estimator  $\hat{\theta}_{LC}$  and the noise-enhanced estimator  $\hat{\theta}_{NE}$  [7] constructed from the suboptimal MAP estimator  $\hat{\theta}_{map}$  in Eq. (20) versus the added noise level  $\sigma_\eta$ . For comparison, the MSEs of the MAP estimator  $\hat{\theta}_{map}$  in Eq. (20) and the MMSE estimator  $\hat{\theta}_{mmse}$  in Eq. (21) are also plotted. Here, the background noise level  $\sigma_\xi = 1$  and the interval bound  $a = 2$  for the parameter  $\theta$ . Specially, for the limiting case  $m \rightarrow \infty$ , MSEs of  $\hat{\theta}_{LC}$  are also plotted in (b) over a suitable range. (c) MSEs of  $\hat{\theta}_{LC}$  and  $\hat{\theta}_{NE}$  constructed from the quantizer  $\hat{\theta}_{qt}$  of Eq. (22) with the threshold  $\gamma = 0$  versus  $\sigma_\eta$ . (d) For a combiner of identical  $\hat{\theta}_{qt}$ , the MSEs of  $\hat{\theta}_{LC}$  and the MMSE estimator  $\hat{\theta}_{mmse}$  versus the background noise level  $\sigma_\xi$  at a fixed added noise level.

deviation  $\sigma_\eta$ . As in Fig. 2(a), with the benefits of added noise, both  $\hat{\theta}_{LC}$  and  $\hat{\theta}_{NE}$  are better than the MAP estimator  $\hat{\theta}_{map}$ . At an optimal noise level  $\sigma_\eta^{opt} = 2.8184$ , the noise-enhanced estimator  $\hat{\theta}_{NE}$  achieves its minimum MSE value of 0.2494. After the optimal added noise level  $\sigma_\eta > \sigma_\eta^{opt}$ , the MSE of  $\hat{\theta}_{NE}$  goes up again, and finally reaches a stable value of 0.33 for very large added noise levels (e.g.  $\sigma_\eta > 10^2$ ). This can be viewed as a kind of stochastic resonance effect measuring by the MSE that descends to its lowest point at an optimal non-zero noise level. It is also interesting to note in Fig. 2(a) that, as the added noise level  $\sigma_\eta$  increases and the number  $m \geq 5$ , the MSE  $\mathcal{R}_m^o$  of  $\hat{\theta}_{LC}$  calculated by Eq. (13) can be further reduced by the added noise. In order to observe the behavior of  $\mathcal{R}_\infty^o$  in Eq. (14) in detail, Fig. 2(b) redraws the MSE  $\mathcal{R}_\infty^o$  over a suitable range, and clearly shows that the MSE  $\mathcal{R}_\infty^o$  in Eq. (14) achieves a minimum value of 0.2493 at an optimal noise level  $\sigma_\eta^{opt} = 1.995$ . When  $\sigma_\eta > \sigma_\eta^{opt}$ ,  $\mathcal{R}_\infty^o$  also rises and gradually evolves into a stable value of 0.2500 for large noise levels (e.g.  $\sigma_\eta > 10$ ). Therefore, compared with the estimator  $\hat{\theta}_{NE}$  (\*), the designed estimator  $\hat{\theta}_{LC}$  (o) has the similar noise-enhanced effect, but exhibits a quite stable improvement for a wide range of added noise levels, as

shown in Figs. 2(a) and (b). The reason is that, as the added noise level  $\sigma_\eta$  increases, the optimum weighting coefficient  $w^o$  in Eq. (15) can adaptively tune itself for the variety of added noise, and an effectively reduced MSE  $\mathcal{R}_\infty^o$  is achieved by the incorporation of added noise  $\eta_i$  and the weights  $w^o$  in  $\hat{\theta}_{LC}$ . It is shown in Figs. 2(a) and (b) that, even without the help of added noise  $\eta$  (when  $\sigma_\eta = 0$ ), the MSE  $\mathcal{R}_m^o$  of  $\hat{\theta}_{LC}$  can be optimized by the weight vector  $\mathbf{w}^o$  as 0.2550 much lower than the MSE 0.4832 achieved by  $\hat{\theta}_{NE}$  at  $\sigma_\eta = 0$ . This superiority of  $\hat{\theta}_{LC}$  over  $\hat{\theta}_{NE}$  is also theoretically proved by Eq. (43) in Appendix D.

The benefits of added noise might take effect in other suboptimal estimators. For instance, consider a binary quantizer [9], [34]

$$\hat{\theta}_{qt}(x) = \begin{cases} 1, & x \geq \gamma, \\ 0, & x < \gamma, \end{cases} \quad (22)$$

with the threshold  $\gamma$ . Similarly, based on this suboptimal estimator  $\hat{\theta}_{qt}$ , the combined estimator  $\hat{\theta}_{LC}$  and the noise-enhanced estimator  $\hat{\theta}_{NE}$  can be also obtained. Upon increasing the added noise level  $\sigma_\eta$  and the number  $m$ , the MSEs  $\mathcal{R}_m^o$  of  $\hat{\theta}_{LC}$  are

plotted in Fig. 2(c) as a function of  $\sigma_\eta$ . It is seen that the noise benefit to the MSE  $\mathcal{R}_\infty^o$  of  $\hat{\theta}_{LC}$  is more effective, reducing the initial MSE  $\mathcal{R}_\infty^o = 0.2925$  at  $\sigma_\eta = 0$  to the minimum of 0.2498 at the optimal added noise level  $\sigma_\eta^{\text{opt}} = 4.4668$ , as illustrated in Figs. 2(b) and (c).

Naturally, the optimal added noise level  $\sigma_\eta^{\text{opt}}$  is expected to be obtained for achieving the minimum MSE of the Bayesian estimators  $\hat{\theta}_{NE}$  and  $\hat{\theta}_{LC}$ . The solution of  $\sigma_\eta^{\text{opt}}$  by setting the derivative  $\partial R_m^o / \partial \sigma_\eta = 0$  may however identify several extrema. As shown in Figs. 2(a)–(c), the equation of  $\partial R_m^o / \partial \sigma_\eta = 0$  will yield multiple solutions of  $\sigma_\eta$ . In addition, the second-order derivative  $\partial^2 R_m^o / \partial \sigma_\eta^2$  needs to be derived for further determining the optimal level  $\sigma_\eta^{\text{opt}}$ . However, it is also seen in Figs. 2(a)–(c) that, for a sufficiently large number  $m$  (e.g.  $m \geq 10^4$ ), the MSE  $R_m^o$  of the combined estimator  $\hat{\theta}_{LC}$  reaches a plateau with stationary values of 0.25 for a wide range of added noise levels (e.g.  $\sigma_\eta > 1.995$  in Fig. 2(a) or  $\sigma_\eta > 4.4668$  in Fig. 2(c)). This robust feature is very attractive and forms *Corollary 4*.

*Corollary 4:* For the combined estimator  $\hat{\theta}_{LC}$  in Eq. (11), if the original estimator  $\hat{\theta}(x)|_{x \rightarrow \infty} = \Theta < \infty$  and the added noise has a symmetric scale-family PDF  $f_\eta(\eta, \sigma_\eta) = (1/\sigma_\eta) f_{\tilde{\eta}}(\eta/\sigma_\eta)$ , then the MSE  $R_\infty^o$  of Eq. (14) has a plateau of the local extremum for large added noise levels  $\sigma_\eta$ . Here,  $f_{\tilde{\eta}}$  is the PDF of the standardized noise variable  $\tilde{\eta}$  with zero mean and unit variance.

Proof of *Corollary 4* is given in Appendix F. *Corollary 4* only tells us the MSE  $R_\infty^o$  of  $\hat{\theta}_{LC}$  reaches a stable value for a large added noise level  $\sigma_\eta$ , but does not indicate how low this stable value is. Due to the optimum weight  $w^o$  and for a large noise level, it is seen in Figs. 2(a)–(c) that this stable value of 0.2500 is rather near to the MMSE 0.2492 given by  $\hat{\theta}_{\text{mmse}}$ . This interesting characteristic of  $\hat{\theta}_{LC}$  suggests that, instead of finding the optimal noise level  $\sigma_\eta^{\text{opt}}$ , a fixed large noise level  $\sigma_\eta$  can be used to improve the MSE of the combined estimator  $\hat{\theta}_{LC}$  with a tolerable accuracy. In practice, the limiting case of  $m \rightarrow \infty$  is inaccessible and can only be approached. Then, we here use a combiner of quantizers to construct the estimator  $\hat{\theta}_{LC}$  with a finite number  $m = 10^5$ , and plot its MSE  $\mathcal{R}_m^o$  as a function of the background noise level  $\sigma_\xi$  in Fig. 2(d). It is clearly seen in Fig. 2(d) that, with the help of added noise (when  $\sigma_\eta = 10$ ), the combined estimator  $\hat{\theta}_{LC}$  achieves a rather comparable performance in comparison to the MMSE estimator  $\hat{\theta}_{\text{mmse}}$  in the considered range of  $\sigma_\xi \in [0.1, 4]$ .

In Fig. 2, we only consider the added noise  $\eta$  with Gaussian distribution. Since the noise components  $\eta_i$  are artificially added to estimators  $\hat{\theta}_i$ , then the characteristic of added noise  $\eta$  can be selected purposefully for improving the performance of the designed estimator  $\hat{\theta}_{LC}$ . Thus, an important question is whether the addition of any type of noise to the estimators  $\hat{\theta}_i$  is always beneficial to the decrease of MSE. When the initial rate of the MSE is negative, i.e.

$$\lim_{\sigma_\eta \rightarrow 0} \frac{dR_m^o}{d\sigma_\eta} < 0, \quad (23)$$

the answer is affirmative in certain circumstances, as stated in *Corollary 5*.

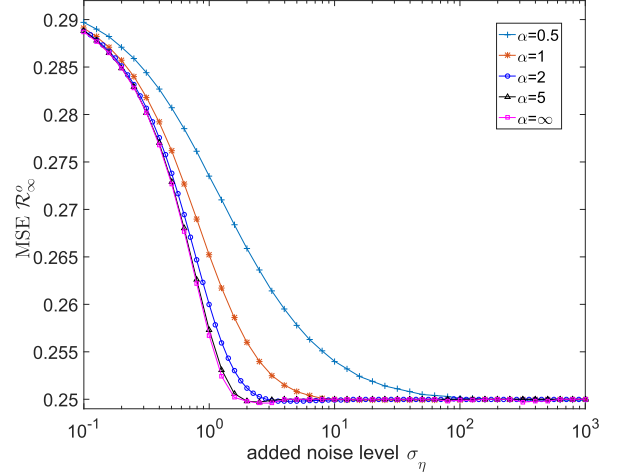


Fig. 3. MSEs  $\mathcal{R}_\infty^o$  of the combined estimator  $\hat{\theta}_{LC}$  as a function of the added noise level  $\sigma_\eta$ . Here,  $\hat{\theta}_{LC}$  is constructed from the quantizer  $\hat{\theta}_{qt}$  of Eq. (22) with the threshold  $\gamma = 0$ . The added noise  $\eta$  has the generalized Gaussian PDF with exponents  $\alpha = 0.5, 1, 2$  and  $\infty$ . The other parameters are the same as in Fig. 2

*Corollary 5:* For the designed estimator  $\hat{\theta}_{LC}$  composed of identical estimators  $\hat{\theta}_i = \hat{\theta}_{qt}$  and the symmetric scale-family added noise  $\eta$ ,  $\hat{\theta}_{LC}$  can benefit from the added noise when the number  $m > 1$  and  $f_x(\gamma) \neq 0$ .

Proof of *Corollary 5* is presented in Appendix G, and an illustrative example is given as follows.

*Example 2:* Consider the symmetric scale-family added noise  $\eta$  with generalized Gaussian PDF  $f_\eta(x) = (c_1/\sigma_\eta) \exp(-c_2|x/\sigma_\eta|^\alpha)$ . Here,  $c_1 = \frac{\alpha}{2} \Gamma(\frac{1}{\alpha}) / \Gamma(\frac{3}{\alpha})$ ,  $c_2 = [\Gamma(\frac{3}{\alpha}) / \Gamma(\frac{1}{\alpha})]^\alpha$ ,  $\Gamma(x)$  is the gamma function and the exponent  $\alpha > 0$  [1], [12]. It is shown in Fig. 3 that the noise-enhanced effects of the MSE also occur clearly for different noise types of  $\alpha = 0.5, 1$  (Laplacian noise), 2 (Gaussian noise) and  $\infty$  (uniform noise). The larger the exponent  $\alpha$  is, the faster the initial rate of the MSE  $\mathcal{R}_\infty^o$  declines, which also accords to the results in Ref. [8]. The corresponding MSE achieved by different noise types also reaches a minimum at an optimal added noise level, and finally tends to the stable value of 0.25 for larger added noise levels, as *Corollary 4* indicated.

However, *Corollary 5* is only applicable to the prescribed symmetric scale-family added noise  $\eta$ . An important problem is whether the benefit of the added noise (not restricted to the symmetric scale-family) exists or not for a combiner of arbitrary suboptimal estimators  $\hat{\theta}_i$ . The resolution of this problem is in general difficult [7]–[10], [12]–[16], [37], even for a combiner of arbitrary identical estimators with the simplified expression of the MSE  $\mathcal{R}_m^o$  in Eq. (13). In this case, since the variance  $\text{var}(\theta)$  is given, finding the optimal added noise PDF to minimize the MSE  $\mathcal{R}_m^o$  of  $\hat{\theta}_{LC}$  with a finite number  $m$  is the optimization problem

$$\begin{aligned} f_\eta^{\text{opt}}(\eta) &= \arg \min_{f_\eta} \mathcal{R}_m^o \\ &= \arg \max_{f_\eta} \frac{m p_i^2}{C_{ii} + (m-1)C_{ij}}, \\ \text{s.t. } f_\eta(\eta) &\geq 0, \int f_\eta(\eta) d\eta = 1, \end{aligned} \quad (24)$$

which is a constrained minimization of the nonlinear functional  $\mathcal{R}_m^o(f_\eta)$  for the given observation data  $x$ , the estimators  $\hat{\theta}_i$  and the number  $m$ . This optimization problem is in general theoretically intractable, even for a single estimator  $\hat{\theta}_i$  with  $m = 1$ . This is due to the fact that the expectation  $E_x\{E_\eta[\hat{\theta}_i(x + \eta)]\}$  has no constraint due to the adaptive bias weight  $w_0$ , and the terms  $E_x^2\{E_\eta[\hat{\theta}_i(x + \eta)]\}$  in Eq. (5) (Eq. (38)) and  $E_x\{E_\eta^2[\hat{\theta}_i(x + \eta)]\}$  in Eq. (6) (Eq. (39)) are nonlinear functional of  $f_\eta$ . Thus, the theoretical determination of optimal added noise PDF  $f_\eta^{\text{opt}}$  that minimizes  $\mathcal{R}_m^o$  can not refer to the convex optimization method proposed by Chen *et al.* [6], [16], [23], [24]. The necessary and sufficient conditions determining the solution of this *non-convex* optimization problem in Eq. (24) are that the first variation of  $\mathcal{R}_m^o$  vanishes and the second variation of  $\mathcal{R}_m^o$  is nonnegative [7], [38]. As Uhlich [7] pointed out, these theoretical conditions are rather difficult to handle in general, and such approximation methods as Parzen widows density estimation [7] and particle-swarm optimization [37] can be employed.

For low computational cost, we here employ an approximate form of added noise PDF as

$$\tilde{f}_\eta^{\text{opt}}(\eta) = \sum_{k=1}^K \nu_k \varrho_k(\eta, \mu_k, \sigma_k), \quad (25)$$

where the normalization coefficients  $\nu_k \geq 0$ ,  $\sum_{k=1}^K \nu_k = 1$ , and the Gaussian window function  $\varrho_k(\eta, \mu_k, \sigma_k) = \exp[-(\eta - \mu_k)^2 / 2\sigma_k^2] / \sqrt{2\pi\sigma_k^2}$  with means  $\mu_k$  and standard deviations  $\sigma_k \geq 0$  [37], [39]. As the number  $K$  of  $\varrho_k$  increases, the estimation form  $\tilde{f}_\eta^{\text{opt}}(\eta)$  of Eq. (25) gradually converges to  $f_\eta^{\text{opt}}(\eta)$  in Eq. (24) under certain conditions [7], [37], [39].

*Example 3:* Reconsider the parameter estimation given in *Example 1*, and choose the estimator  $\hat{\theta}_i = \hat{\theta}_{\text{qt}}$  of Eq. (22) with the threshold  $\gamma = 0$ . The number of  $\varrho_k$  is chosen as  $K = 4$ , and the constrained nonlinear optimization algorithm of sequential quadratic programming [40] is employed to find the approximation PDF of the optimal added noise. Note that, for randomly selected initial values of coefficients  $0 \leq c_k \leq 1$ , means  $-\infty < \mu_k < \infty$  and standard deviation  $\sigma_k \geq 0$ , both the sequential quadratic programming method and the particle-swarm optimization method are carried out for finding the minimum MSE  $\mathcal{R}_m^o(\tilde{f}_\eta^{\text{opt}})$  and the corresponding optimal PDF  $\tilde{f}_\eta^{\text{opt}}$  of added noise  $\eta$ . At each major iteration of the sequential quadratic programming method, a positive definite quasi-Newton approximation of the Hessian of the objective function is calculated using the BFGS method [40], [41]. For different estimator numbers  $m = 1, 2, 10^4$  and  $\infty$ , we randomly choose 20 groups of initial coefficients  $c_k$ , means  $\mu_k$  and standard deviation  $\sigma_k$ , the minimum MSEs  $\mathcal{R}_m^o(\tilde{f}_\eta^{\text{opt}})$  with the corresponding vector parameters  $\boldsymbol{\nu} = [\nu_1, \nu_2, \dots, \nu_K]^\top$ ,  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_K]^\top$  and  $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_K]^\top$  are recorded in Table I. It is interesting to note that, for a single estimator with  $m = 1$ , the optimal added noise PDF is  $\tilde{f}_\eta^{\text{opt}}(\eta) = \delta(\eta + 1.0091)$  with the Dirac delta function  $\delta(\eta)$ . As the estimator number  $m = 10^4$ , the approximate optimal added noise PDF  $\tilde{f}_\eta^{\text{opt}}(\eta)$  have two nonzero normalization coefficients and is composed of two Gaussian window functions, yielding a rather small MSE of 0.2493. Interestingly, for the limiting case  $m \rightarrow \infty$ , the optimal added noise PDF obtained in Table I, as shown in Fig. 4 (red

TABLE I  
OPTIMAL ADDED NOISE VIA THE SEQUENTIAL QUADRATIC PROGRAMMING ALGORITHM

$m$	$\boldsymbol{\nu}$	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	MSE $\mathcal{R}_m^o(\tilde{f}_\eta^{\text{opt}})$
1	0	1.1728	0.3567	0.2748
	1	-1.0091	0	
	0	0.5220	0.8433	
	0	-0.7426	0.0838	
2	0	-17.9468	0.1656	0.2718
	0	8.4038	0.6020	
	0	-8.8803	0.2630	
	1	-1.0000	0.2584	
$10^4$	0.8423	-0.7481	1.9117	0.2493
	0.1577	-2.1813	1.6678	
	0	3.2796	0.5884	
	0	3.1288	0.6654	
$\infty$	0.6303	-1.6186	2.8524	0.2492
	0.0340	1.5593	0.8829	
	0.3329	-0.8833	1.3872	
	0.0027	1.8127	0.1960	

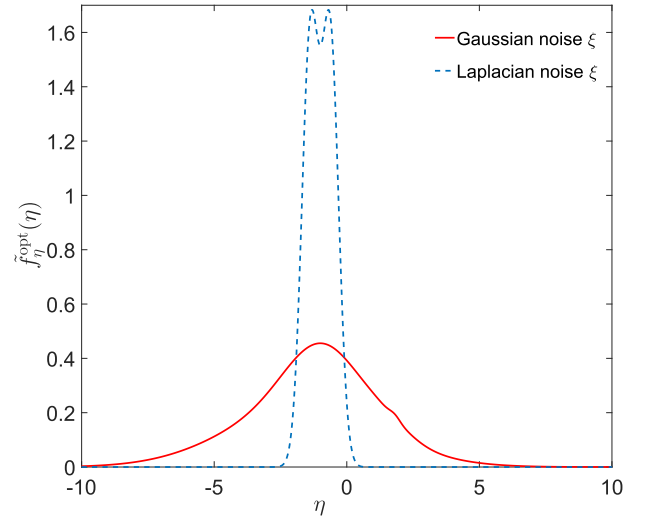


Fig. 4. Approximate PDFs  $\tilde{f}_\eta^{\text{opt}}(\eta)$  of the optimal added noise for the estimator  $\hat{\theta}_i = \hat{\theta}_{\text{qt}}$  of Eq. (22) with the threshold  $\gamma = 0$  and the number  $m = \infty$ . Here,  $\tilde{f}_\eta^{\text{opt}}(\eta)$  is solved for the Gaussian (red solid line) and the Laplacian (blue dashed line) background noise  $\xi$ . The other parameters are the same as in Fig. 2.

solid line), presents a minimum MSE  $\mathcal{R}_\infty^o(\tilde{f}_\eta^{\text{opt}}) = 0.2492$ , which is just the MMSE achieved by the optimal estimator of Eq. (19). We note that the solution  $\tilde{f}_\eta^{\text{opt}}(\eta)$  of the optimal added noise PDF is not unique, because the particle-swarm optimization method also presents another solution as  $\tilde{f}_\eta^{\text{opt}}(\eta) = \exp[-(\eta - \mu_\eta)^2 / 2\sigma_\eta^2] / \sqrt{2\pi\sigma_\eta^2}$  with  $\mu_\eta = -1$  and  $\sigma_\eta = 2.004$ . Using this solution, the corresponding MSE also achieves the MMSE 0.2492. Of course, for various estimators  $\hat{\theta}_i$ , different prior PDF  $f_\theta$  and background noise PDF  $f_\xi$ , both the sequential quadratic programming method and the particle-swarm optimization method need to be applied anew.

*Example 4:* In Examples 1–3, the background noise  $\xi$  is assumed to be Gaussian distributed. However, the background noise  $\xi$  in the real data model of Eq. (1) can in some situations come with non-Gaussian distributions, and its type results in the variety of MMSE Bayesian estimators  $\hat{\theta}_{\text{mmse}}$  and the corresponding MMSE values. It is noted that the approximate PDF

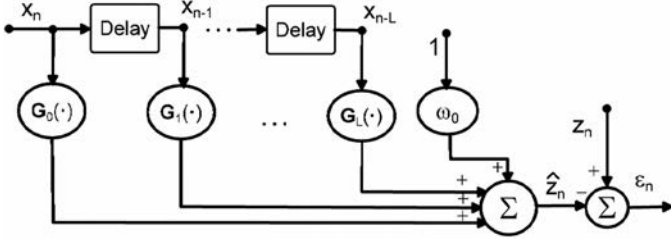


Fig. 5. Diagram of the noise-enhanced FIR filter. The combiner  $G_\ell$ , representing module 1 plus module 2 in Fig. 1, processes the input sample  $x_{n-\ell}$ . All outputs of combiners  $G_\ell$  and the bias weight  $\omega_0$  applied at the filter output  $\hat{z}_n$ .

form in Eq. (25) of the optimal added noise and the constrained nonlinear optimization algorithms remain valid. For a uniformly distributed parameter  $\theta$  buried in a given background noise  $\xi$  with its PDF  $f_\xi$  and cumulative distribution function  $F_\xi$ , the MMSE Bayesian estimator of Eq. (19) becomes

$$\hat{\theta}_{\text{mmse}}(x) = E_{\theta|x}[\theta|x] = x + \frac{\int_x^{x-a} u f_\xi(u) du}{F_\xi(x) - F_\xi(x-a)}. \quad (26)$$

For instance, consider the uniform distributed parameter  $\theta$  with its PDF  $f_\theta(u) = 1/2$  ( $u \in [0, 2]$ ) corrupted by the Laplace noise  $\xi$  with its PDF  $f_\xi(\xi) = \exp(-\sqrt{2}|x|/\sigma_\xi)/(\sqrt{2}\sigma_\xi)$  and standard deviation  $\sigma_\xi = \sqrt{2}$ , the MMSE achieved by the estimator of Eq. (26) is 0.2679. For the limiting case of  $m \rightarrow \infty$ , we use the constrained nonlinear optimization algorithms as in Example 3 with the number  $K = 4$  to optimize the designed Bayesian estimator  $\hat{\theta}_{\text{LC}}$  constructed by quantizer with threshold  $\gamma = 0$ . Then, the approximate form of the optimal added noise PDF is  $\tilde{f}_\eta^{\text{opt}}(\eta) = [e^{-(x-m_1)^2/(2\sigma_\eta^2)} + e^{-(x-m_2)^2/(2\sigma_\eta^2)}]/(2\sqrt{2\pi\sigma_\eta^2})$  with two nonzero coefficients  $\nu_1 = \nu_2 = 0.5$ , means  $m_1 = -1.3784$ ,  $m_2 = -0.6217$  and standard deviations  $\sigma_1 = \sigma_2 = \sigma_\eta = 0.3195$  (see Fig. 4 in blue dashed line), which is effective for obtaining the corresponding MSE 0.2679, the same as the MMSE achieved by  $\hat{\theta}_{\text{mmse}}(x)$  in Eq. (26). Therefore, a natural question is that, for any type of background noise  $\xi$ , can we always find an approximation form  $\tilde{f}_\eta^{\text{opt}}$  of optimal noise PDF for the designed estimator  $\hat{\theta}_{\text{LC}}$  to achieve the MMSE? We have shown this is feasible for the cases of great practical relevance of a Gaussian background noise and a non-Gaussian (Laplacian) background noise. For other cases, we here leave these interesting open questions of the solution of Eq. (24) for further study. We can extend the design of the Bayesian estimator to the finite impulse response (FIR) filter as follows.

### III. NOISE-ENHANCED FIR FILTERS

We can extend the noise-benefit combiner  $G$ , i.e. module 1 plus module 2 marked in Fig. 1, to estimate a desired signal  $z_n$  by the current sample  $x_n$  and the past samples  $x_{n-1}, x_{n-2}, \dots, x_{n-L}$ , as illustrated in Fig. 5. Here, we assume that the straight-forward observation  $x_n$  can not be directly obtained and goes through the distortion endowed by the nonlinearity  $\hat{\theta}$  in the combiner, representing for example neuronal models [42], nonlinear estimators [1]–[3] or sensors [5], [9], [34]. In this situation, for  $L + 1$  samples of  $x_n$ ,  $L + 1$  combiners  $G_\ell$  are

equipped and each combiner  $G_\ell$  contains  $m$  original estimators  $\hat{\vartheta}_{\ell i}$ ,  $m$  added noise components  $\eta_{\ell i}$  and the corresponding adjustable weighting coefficients  $\omega_{\ell i}$  for  $\ell = 0, 1, \dots, L$  and  $i = 1, 2, \dots, m$ . Then, collecting all outputs of combiners  $G_\ell$  and adding the bias weight  $\omega_0$ , we have the designed noise-enhanced finite impulse response (FIR) filter

$$\hat{z}_n = \omega_0 + \sum_{\ell=0}^L \sum_{i=1}^m \omega_{\ell i} \hat{\vartheta}_{\ell i}(x_{n-\ell} + \eta_{\ell i}). \quad (27)$$

Here, the number  $L + 1$  of combiners  $G_\ell$  is the order of the designed FIR filter of Eq. (27) and the error signal is defined as  $\varepsilon_n = z_n - \hat{z}_n$ .

Furthermore, in each combiner  $G_\ell$ , Theorem 2 also holds. Thus, consider the individual estimators  $\hat{\vartheta}_{\ell i} = \hat{\vartheta}_\ell$  in each  $G_\ell$  are identical, we have  $m$  identical optimum weight coefficients  $\omega_{\ell i} = w_\ell$ . In this case, we can label these weight coefficients  $w_\ell$  as the vector  $\mathbf{w} = [w_0, w_1, \dots, w_L]^\top$  and rewrite the estimate vector as  $\hat{\boldsymbol{\theta}} = [\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_L]^\top$  with  $\hat{\theta}_\ell = \sum_{i=1}^m \hat{\vartheta}_\ell(x + \eta_i)$ . Using the unbiased condition  $E_{\hat{z}}(\hat{z}_n) = E_z(z_n)$ , the designed filter in Eq. (27) becomes

$$\hat{z}_n = E_z(z_n) + \mathbf{w}^\top (\hat{\boldsymbol{\theta}} - E_{x,\eta}(\hat{\boldsymbol{\theta}})). \quad (28)$$

Under this circumstance, the MSE  $\mathcal{J}_{m,L+1}$  of the FIR filter in Eq. (28) is

$$\mathcal{J}_{m,L+1} = E[\varepsilon_n^2] = \text{var}(z_n) - 2\mathbf{w}^\top \mathbf{p} + \mathbf{w}^\top \mathbf{C} \mathbf{w}. \quad (29)$$

The covariance matrix  $\mathbf{C} = E_{x,\eta}\{[\hat{\boldsymbol{\theta}} - E_{x,\eta}(\hat{\boldsymbol{\theta}})][\hat{\boldsymbol{\theta}} - E_{x,\eta}(\hat{\boldsymbol{\theta}})]^\top\}$  has diagonal elements

$$\begin{aligned} C_{\ell+1,\ell+1} &= E_{x_{n-\ell}}[E_\eta(\hat{\theta}_\ell^2)] - E_{x_{n-\ell}}^2[E_\eta(\hat{\theta}_\ell)] \\ &= E_{x_{n-\ell}}\{mE_\eta[\hat{\vartheta}_\ell^2(x_{n-\ell} + \eta)] + m(m-1) \\ &\quad \times E_\eta^2[\hat{\vartheta}_\ell(x_{n-\ell} + \eta)]\} \\ &\quad - m^2 E_{x_{n-\ell}}^2\{E_\eta[\hat{\vartheta}_\ell(x_{n-\ell} + \eta)]\} \end{aligned}$$

and the non-diagonal elements ( $\ell \neq \kappa$ )

$$\begin{aligned} C_{\ell+1,\kappa+1} &= E_{x_{n-\ell}, x_{n-\kappa}}[E_\eta(\hat{\theta}_\ell)E_\eta(\hat{\theta}_\kappa)] - E_{x_{n-\ell}}[E_\eta(\hat{\theta}_\ell)]E_{x_{n-\kappa}}[E_\eta(\hat{\theta}_\kappa)] \\ &= m^2 E_{x_{n-\ell}, x_{n-\kappa}}\{E_\eta[\hat{\vartheta}_\ell(x_{n-\ell} + \eta)]E_\eta[\hat{\vartheta}_\kappa(x_{n-\kappa} + \eta)]\} \\ &\quad - m^2 E_{x_{n-\ell}}\{E_\eta[\hat{\vartheta}_\ell(x_{n-\ell} + \eta)]\}E_{x_{n-\kappa}}\{E_\eta[\hat{\vartheta}_\kappa(x_{n-\kappa} + \eta)]\}, \end{aligned}$$

where  $E_{x_{n-\ell}, x_{n-\kappa}}(\cdot) = \int \cdot f_x(x_{n-\ell}, x_{n-\kappa}) dx_{n-\ell} dx_{n-\kappa}$  is with respect to the second-order joint PDF  $f_x(x_{n-\ell}, x_{n-\kappa})$ . The centralized cross-correlation vector  $\mathbf{p} = E_{x,\eta}\{[z_n - E_z(z_n)][\hat{\boldsymbol{\theta}} - E_{x,\eta}(\hat{\boldsymbol{\theta}})]\}$  has elements

$$\begin{aligned} p_{\ell+1} &= mE_{z_n, x_{n-\ell}}\{z_n E_\eta[\hat{\vartheta}_\ell(x_{n-\ell} + \eta)]\} - mE_z(z_n) \\ &\quad \times E_{x_{n-\ell}}\{E_\eta[\hat{\vartheta}_\ell(x_{n-\ell} + \eta)]\}. \end{aligned}$$

Then, we obtain the optimum weight vector  $\mathbf{w}^o = \mathbf{C}^{-1}\mathbf{p}$  and the optimized MSE  $\mathcal{J}_{m,L+1}^o$  of the considered FIR filter in Eq. (28) is given by

$$\mathcal{J}_{m,L+1}^o = E(\varepsilon_n^2) = \text{var}(z_n) - \mathbf{p}^\top \mathbf{C}^{-1} \mathbf{p}. \quad (30)$$



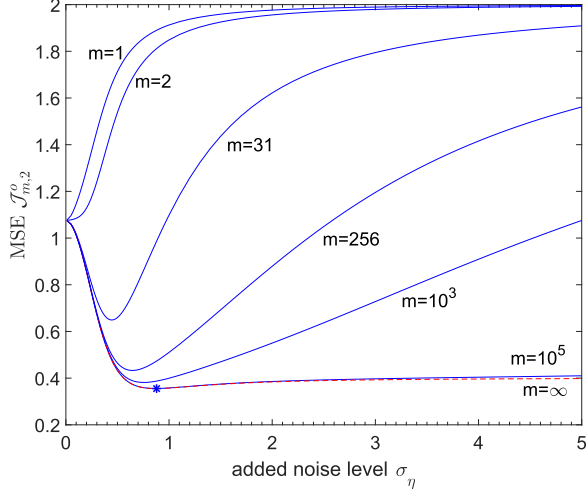


Fig. 6. Plots of the MSE  $\mathcal{J}_{m,2}^o$  of the designed filter in Eq. (28) as a function of the added noise level  $\sigma_\eta$  for different numbers  $m = 1, 2, 31, 256, 10^3, 10^5$  and  $\infty$ . Here, the saturation estimators of Eq. (31) are with the identical parameters  $\beta = 5$ , and the external Gaussian noise level  $\sigma_\xi = 0.1$ .

Notice the multiplier  $m$  in each element  $p_{\ell+1}$  and the term  $\mathbf{p}^\top \mathbf{C}^{-1} \mathbf{p}$  in Eq. (30), we find limits of elements as

$$\begin{aligned} C_{\ell+1,\ell+1}^\infty &= \lim_{m \rightarrow \infty} C_{\ell+1,\ell+1}/m^2 \\ &= \mathbb{E}_{x_{n-\ell}} \{ \mathbb{E}_\eta^2 [\hat{\vartheta}(x_{n-\ell} + \eta)] \} - \mathbb{E}_{x_{n-\ell}}^2 \{ \mathbb{E}_\eta [\hat{\vartheta}(x_{n-\ell} + \eta)] \}, \\ C_{\ell+1,\kappa+1}^\infty &= \lim_{m \rightarrow \infty} C_{\ell+1,\kappa+1}/m^2 \\ &= \mathbb{E}_{x_{n-\ell}, x_{n-\kappa}} \{ \mathbb{E}_\eta [\hat{\vartheta}_\ell(x_{n-\ell} + \eta)] \mathbb{E}_\eta [\hat{\vartheta}_\kappa(x_{n-\kappa} + \eta)] \} \\ &\quad - \mathbb{E}_{x_{n-\ell}} \{ \mathbb{E}_\eta [\hat{\vartheta}_\ell(x_{n-\ell} + \eta)] \} \mathbb{E}_{x_{n-\kappa}} \{ \mathbb{E}_\eta [\hat{\vartheta}_\kappa(x_{n-\kappa} + \eta)] \}, \\ p_{\ell+1}^\infty &= p_{\ell+1}/m \\ &= \mathbb{E}_{z_n, x_{n-\ell}} \{ z_n \mathbb{E}_\eta [\hat{\vartheta}_\ell(x_{n-\ell} + \eta)] \} - \mathbb{E}_z(z_n) \\ &\quad \times \mathbb{E}_{x_{n-\ell}} \{ \mathbb{E}_\eta [\hat{\vartheta}_\ell(x_{n-\ell} + \eta)] \}. \end{aligned}$$

Substituting these limits into Eq. (30), we can also compute the limit value of  $\mathcal{J}_{\infty,L+1}^o$  as  $m \rightarrow \infty$ .

*Example 5:* 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  $N = 16$  ( $N > 2$ ) samples per period and the external white-noise process  $\xi_n$  is Gaussian distributed. The desired signal is also assumed to be the sampled sinusoid  $z_n = 2 \cos(2\pi n/N)$  at the same frequency [43]. Due to the periodicity of the input and desired signals, the expectations must be computed by averaging over one period, i.e the operator  $\sum_{n=1}^N (\cdot)/N$ . Then, the variance of the desired signal is  $\text{var}(z_n) = 2$ . When the external Gaussian noise  $\xi_n$  has a fixed noise level  $\sigma_\xi = 0.1$  and the added noise components  $\eta_{\ell i}$  are also Gaussian distributed, we plot the MSE  $\mathcal{J}_{m,2}^o$  of the filter in Eq. (28) as a function of the added noise level  $\sigma_\eta$  for combiners with  $m$  identical saturation estimators

$$\hat{\vartheta}_\ell(x) = \tanh(\beta x), \quad (31)$$

where the slope parameter  $\beta > 0$ . It is clearly seen in Fig. 6 that the MSE  $\mathcal{J}_{m,2}^o$  of the designed filter in Eq. (28) benefits from the increase of the added noise level  $\sigma_\eta$  as the number  $m$  is large enough. For instance, as  $m = 10^5$  and at an optimal added

TABLE II  
MSES OF FIR FILTERS

Filter order $L+1$	MSE of designed filter at optimal noise levels	MSE of designed filter at a fixed level $\sigma_\eta = 5$	MSE of Wiener filter of Ref. [43]
2	0.3558	0.3980	0.4011
3	0.1037	0.1143	0.1113
4	0.0401	0.0437	0.0425
5	0.0191	0.0212	0.0206
6	0.0128	0.0140	0.0129
7	0.0073	0.0107	0.0104
8	0.0069	0.0102	0.0100
9	0.0069	0.0102	0.0100

noise level  $\sigma_\eta = 0.890$ , a minimum value of the MSE is achieved as  $\mathcal{J}_{m,2}^o = 0.3558$  (\* marked in Fig. 6). In addition, Ref. [43] (pp. 103, Fig. 6.3 in Sect. 6) also considered this example of the two-sample FIR filter, namely Wiener filter, yielding the MSE value of 0.4011. Notice that the weights of Wiener filter are directly deduced by the covariance matrix of  $x_n$  and the cross-correlation vector between  $z_n$  and samples of  $x_n$  [43]. Obviously, with the benefit of added noise and for a sufficient large order  $m = 10^5$ , the obtained result in Fig. 6 is better. In fact, the sinusoid  $s_n = A \sin(2\pi n/N)$  can be viewed as a variable with PDF  $f_s(u) = 1/(\pi\sqrt{A^2 - u^2})$  over the interval  $u \in [-A, A]$ , and the estimate of the desired signal from the observation data is also a Bayesian estimation problem. The corresponding MMSE filter is not the linear Wiener filtering. The designed filter of Eq. (28) might achieve a better MSE than the linear Wiener filter, as shown in Fig. 6.

Upon increasing the order  $L+1$  and at the corresponding optimal added noise levels, we list the minimum values of the MSE  $\mathcal{J}_{m,L+1}^o$  in Table II for the designed FIR filter with the number  $m = 10^5$ . As the order  $L+1$  increases, *Theorem 1* also holds and the minimum MSE  $\mathcal{J}_{m,L+1}^o$  of Eq. (28) gradually decreases from 0.3558 at order  $L+1 = 2$  to 0.0069 at order  $L+1 = 9$ . Compared with the results of Wiener filter, the added noise  $\eta$ , as a potential designable variable, clearly manifests its benefits with a performance that is always superior in Table II for the nonlinear FIR filter. In addition, it is also seen in Fig. 6 that, for a large enough level of added noise, the MSE  $\mathcal{J}_{m,2}^o$  keeps a stable value for a sufficiently large number  $m \rightarrow \infty$ . This robust feature of the noise-enhanced filter of Eq. (28) also suggests no need of tuning the added noise level. Then, in Table II, we choose the number  $m = 10^5$  and list the MSE  $\mathcal{J}_{m,L+1}^o$  of the noise-enhanced filter in Eq. (28) at a fixed added noise level  $\sigma_\eta = 5$ , yielding a comparative results of the MSE in Table II. Specially, for  $L+1 = 5$ , Fig. 7(a) shows a realization of the input  $x_n$ , and the saturation estimator  $\vartheta_\ell(x_n) = \tanh(5x_n)$  is illustrated in Fig. 7(b) at the added noise level  $\sigma_\eta = 0$ . When the added noise level  $\sigma_\eta = 5$  and  $m = 10^5$ , we obtain the optimum weight vector  $\mathbf{w}^o = \mathbf{C}^{-1} \mathbf{p} = 10^{-5} \times [3.1082, 0.3873, -2.3912, -4.80578, -6.4904]^\top$  and plot the output  $\hat{z}_n$  (o) of the designed filter of Eq. (28) in Fig. 7(c). Note that  $\hat{z}_n$  starts from  $n = 5$ , since it is a function of  $x_n$  and the past four samples  $x_{n-\ell}$ . For comparison, the desired signal  $z_n = 2 \cos(2\pi n/N)$  (red solid line) is also given in Fig. 7(c). In this numerical realization of  $\hat{z}_n$ , the corresponding MSE 0.0261 agrees well with (is slightly larger than) the MSE  $\mathcal{J}_{m,5}^o = 0.0212$  at the order  $L+1 = 5$  shown in Table II.

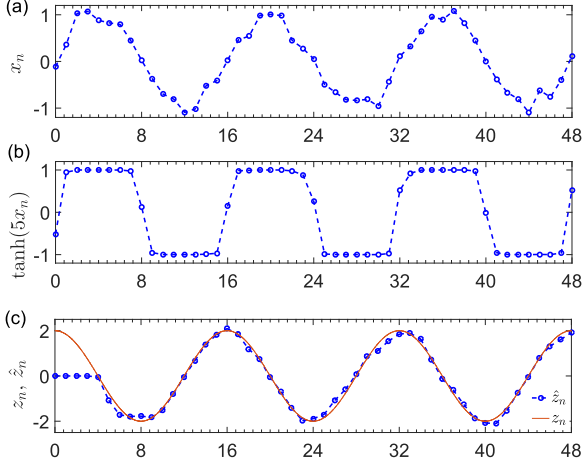


Fig. 7. (a) A realization of the input  $x_n$ ; (b) The saturation estimator  $\vartheta_\ell(x_n) = \tanh(5x_n)$  at the added noise level  $\sigma_\eta = 0$ ; (c) Outputs of the designed filter  $\hat{z}_n$  of Eq. (28) at the added noise level  $\sigma_\eta = 5$  and for  $m = 10^5$ , and the desired signal  $z_n = 2 \cos(2\pi n/N)$ . The other parameter are the same as in Fig. 6.

#### IV. DISCUSSION

Using the Bayes risk for the MSE criterion, this paper mainly investigates the benefits of added noise in the designed Bayesian estimator composed of a combination of individual Bayesian estimators multiplied by the corresponding optimum weighting coefficients. We first demonstrate that the MSE of the combined estimator gradually decreases as the number of individual estimators increases, and then find two equivalent optimum weight coefficients tied with any two identical estimators in the combination. Then, some easily tractable properties of a combination of identical estimators are deduced. It is shown that the MSE of the combined estimator can be optimized not only by the linear weighting coefficients but also by the added noise. Moreover, when the individual estimator has an upper bound and the number of the individual estimators is sufficiently large, we find a robust feature of the combined Bayesian estimator, i.e. the MSE approaching a plateau for a wide range of larger added noise levels. This feature facilitates improvement of the MSE of the designed estimator without tuning the added noise level, and greatly extends the operating range of the designed estimator in practical applications. For the Bayesian estimator composed of identical quantizers, we prove that the injection of a small amount of symmetrical scale-family noise into the observation always improves the MSE of the Bayesian estimator. In order to find the optimal added noise that achieves a “minimum” MSE as close to the MMSE as possible, we adopt the sequential quadratic programming and the particle-swarm optimization method to obtain the approximate form of the optimal noise PDF. The illustrative results show the effectiveness of these constrained nonlinear optimization algorithms. We also extend the combined Bayesian estimator to the nonlinear filtering. In addition, we can extend the results of the scalar parameter model of Eq. (1) to the vector observation model

$$\mathbf{x} = \mathbf{H}\boldsymbol{\vartheta} + \boldsymbol{\xi}, \quad (32)$$

where the observation vector  $\mathbf{x} = [x_1, x_2, \dots, x_N]^\top$ , the unknown vector parameter  $\boldsymbol{\vartheta} = [\vartheta_1, \vartheta_2, \dots, \vartheta_k]^\top$ , the  $N \times k$

observation matrix is  $\mathbf{H}$  and the background noise vector  $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_N]^\top$ . In Appendix H, the Bayesian estimator  $\hat{\boldsymbol{\vartheta}}$  for the parameter vector  $\boldsymbol{\vartheta}$  is derived in Eq. (56). The further discussions of this general Bayesian estimator vector  $\hat{\boldsymbol{\vartheta}}$  in Eq. (56) is significant to the theoretical research and practical applications of noise benefits in the parameter estimation. Finally, two specially important questions that remain in this extended vector model are how to prejudge whether the added noise is beneficial or not, and how to find the exact expression of the optimal added noise PDF that minimizes the MSE of the designed Bayesian estimator formed by a combiner of suboptimal but easily implementable nonlinear estimators.

#### APPENDIX A PROOF OF Theorem 1

Let the subscript  $m$  to denote the dimension of a matrix of  $m \times m$  or a vector of  $m \times 1$ . The positive definite covariance matrix  $\mathbf{C}_m$  has a unique lower-diagonal-upper decomposition  $\mathbf{C}_m = \mathbf{L}_m \mathbf{D}_m \mathbf{L}_m^\top$ , where  $\mathbf{L}_m$  is a unit lower triangular matrix and  $\mathbf{D}_m = \text{diag}(d_1, d_2, \dots, d_m)$  is a diagonal matrix [44]. Since the determinant  $\det(\mathbf{L}_m) = 1$ , then  $\det(\mathbf{C}_m) = \det(\mathbf{D}_m) = \prod_{i=1}^m d_i$  and elements  $d_i = \det \mathbf{C}_i / \det \mathbf{C}_{i-1} > 0$  for  $i = 2, 3, \dots, m$  [44]. Notice the matrix  $\mathbf{C}_m = \mathbf{L}_m \mathbf{D}_m \mathbf{L}_m^\top$  can be also partitioned as

$$\begin{aligned} \mathbf{C}_m &= \begin{bmatrix} \mathbf{C}_{m-1} & \boldsymbol{\gamma}_{m-1} \\ \boldsymbol{\gamma}_{m-1}^\top & C_{mm} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{L}_{m-1} & \mathbf{0} \\ \boldsymbol{\ell}_{m-1}^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{D}_{m-1} & \mathbf{0} \\ \mathbf{0} & d_m \end{bmatrix} \begin{bmatrix} \mathbf{L}_{m-1}^\top & \boldsymbol{\ell}_{m-1} \\ \mathbf{0} & 1 \end{bmatrix}, \quad (33) \end{aligned}$$

where the sub-matrix  $\mathbf{C}_{m-1}$  is just the covariance matrix of the combiner with  $m-1$  individual estimators. This is the optimum nesting property of a matrix (vector) [44]. The diagonal matrix  $\mathbf{D}_m$  in Eq. (33) and the cross-correlation vector  $\mathbf{p}_m = [\mathbf{p}_{m-1}^\top, p_m]^\top$  have the optimum nesting property. From Eq. (33), the row  $\boldsymbol{\ell}_{m-1}$  is uniquely given by  $\mathbf{L}_{m-1} \mathbf{D}_{m-1} \boldsymbol{\ell}_{m-1} = \boldsymbol{\gamma}_{m-1}$ , and then  $\mathbf{L}_m$  also possesses the optimum nesting property. Defining an intermediate vector  $\mathbf{k}_m = \mathbf{L}_m^\top \mathbf{w}_m^o$  and from Eq. (7), we find

$$\mathbf{C}_m \mathbf{w}_m^o = \mathbf{L}_m \mathbf{D}_m \mathbf{L}_m^\top \mathbf{w}_m^o = \mathbf{L}_m \mathbf{D}_m \mathbf{k}_m = \mathbf{p}_m \quad (34)$$

and  $\mathbf{k}_m = [\mathbf{k}_{m-1}^\top, k_m]^\top$  has the optimum nesting property. Therefore, the MSE  $\mathcal{R}_m^o$  of Eq. (9) can be expressed as

$$\begin{aligned} \mathcal{R}_m^o &= \text{var}(\boldsymbol{\theta}) - (\mathbf{w}_m^o)^\top \mathbf{C}_m \mathbf{w}_m^o \\ &= \text{var}(\boldsymbol{\theta}) - \mathbf{k}_m^\top \mathbf{D}_m \mathbf{k}_m \\ &= \text{var}(\boldsymbol{\theta}) - \begin{bmatrix} \mathbf{k}_{m-1} \\ k_m \end{bmatrix}^\top \begin{bmatrix} \mathbf{D}_{m-1} & \mathbf{0} \\ \mathbf{0}^\top & d_m \end{bmatrix} \begin{bmatrix} \mathbf{k}_{m-1} \\ k_m \end{bmatrix} \\ &= \text{var}(\boldsymbol{\theta}) - \mathbf{k}_{m-1}^\top \mathbf{D}_{m-1} \mathbf{k}_{m-1} - d_m k_m^2 \\ &= \mathcal{R}_{m-1}^o - d_m k_m^2, \quad (35) \end{aligned}$$

where  $\mathcal{R}_{m-1}^o = \text{var}(\boldsymbol{\theta}) - \mathbf{k}_{m-1}^\top \mathbf{D}_{m-1} \mathbf{k}_{m-1}$  is just the MSE of the combined estimator  $\hat{\boldsymbol{\theta}}_{LC}$  with the number  $m-1$ . From the positive-definiteness of  $\mathbf{C}_m$ ,  $d_m > 0$  and  $k_m^2 > 0$ . Thus, Eq. (10) holds.

APPENDIX B  
PROOF OF *Theorem 2*

Without loss of generality, we label two identical estimators as  $\hat{\theta}_1 = \hat{\theta}_2$ . Let  $C_{ij}$  denote the element of  $\mathbf{C}$  at the  $i$ th row and the  $j$ th column for  $i, j = 1, 2, \dots, m$ . From Eqs. (5) and (6), it is seen that the covariance matrix  $\mathbf{C}$  has equivalent elements  $C_{11} = C_{22}$ ,  $C_{21} = C_{12}$  and  $C_{1k} = C_{2k}$  for  $k = 3, 4, \dots, m$ . The inverse matrix  $\mathbf{C}^{-1} = \mathbf{A}/\det(\mathbf{C})$ , and the matrix  $\mathbf{A}$  is the cofactor of  $\mathbf{C}$  with its elements  $A_{ij} = (-1)^{i+j}B_{ij}$ . The minor  $B_{ij}$  of  $C_{ij}$  is obtained by deleting the  $i$ th row and the  $j$ th column of  $\mathbf{C}$ . Immediately, we find  $B_{11} = B_{22}$ ,  $B_{21} = B_{12}$  and  $B_{1k} = -B_{2k}$ , yielding  $A_{11} = A_{22}$ ,  $A_{21} = A_{12}$  and  $A_{1k} = A_{2k}$ . Furthermore, for identical estimators  $\hat{\theta}_1 = \hat{\theta}_2$ , the cross correlation vector  $\mathbf{p}$  has equivalent elements  $p_1 = p_2$ . Thus, the optimum weight vector  $\mathbf{w}^o = \mathbf{C}^{-1}\mathbf{p} = \mathbf{A}\mathbf{p}/\det(\mathbf{C})$  has equivalent elements

$$w_1^o = \sum_{j=1}^m \frac{A_{1j}p_j}{\det(\mathbf{C})} = \sum_{j=1}^m \frac{A_{2j}p_j}{\det(\mathbf{C})} = w_2^o. \quad (36)$$

APPENDIX C  
PROOF OF COROLLARY 1

For a combiner of  $m$  identical estimators  $\hat{\theta}_i = \hat{\theta}$  and according to *Theorem 1*, all weight coefficients  $w_i = w$  are equivalent for  $i = 1, 2, \dots, m$ . In this case, the weight vector  $\mathbf{w} = w\mathbf{1}$  for a  $m \times 1$  dimensional vector  $\mathbf{1}$  of all ones and the estimate vector  $\hat{\boldsymbol{\theta}} = [\hat{\theta}(x + \eta_1), \hat{\theta}(x + \eta_2), \dots, \hat{\theta}(x + \eta_m)]^\top$ . Thus, the combined estimator  $\hat{\theta}_{LC}$  in Eq. (2) can be simplified as Eq. (11). In addition, due to identical estimators  $\hat{\theta}_i = \hat{\theta}$ , Eq. (4) becomes

$$p_i = E_x\{\theta E_\eta[\hat{\theta}(x + \eta)]\} - E_\theta(\theta)E_x\{E_\eta[\hat{\theta}(x + \eta)]\}, \quad (37)$$

which indicates the centralized cross-correlation vector  $\mathbf{p}$  has  $m$  equivalent elements  $p_i$ . Similarly, from Eqs. (5) and (6), the covariance matrix  $\mathbf{C}$  has  $m$  equivalent diagonal elements

$$C_{ii} = E_x\{E_\eta[\hat{\theta}^2(x + \eta)]\} - E_x^2\{E_\eta[\hat{\theta}(x + \eta)]\} \quad (38)$$

and  $m(m-1)$  equivalent non-diagonal elements

$$C_{ij} = E_x\{E_\eta^2[\hat{\theta}(x + \eta)]\} - E_x^2\{E_\eta[\hat{\theta}(x + \eta)]\}. \quad (39)$$

Then, the MSE  $\mathcal{R}_m$  of  $\hat{\theta}_{LC}$  in Eq. (11) can be computed as

$$\begin{aligned} \mathcal{R}_m &= E_{x,\eta}[(\theta - \hat{\theta}_{LC})^2] \\ &= \text{var}(\theta) - 2w \sum_{i=1}^m p_i + w^2 \sum_{i=1}^m \sum_{j=1}^m C_{ij} \\ &= \text{var}(\theta) - 2mwp_i + mw^2[C_{ii} + (m-1)C_{ij}]. \end{aligned} \quad (40)$$

Since the covariance matrix  $\mathbf{C}$  is positive definite and  $\mathbf{1}^\top \mathbf{C} \mathbf{1} > 0$ , then  $C_{ii} + (m-1)C_{ij} > 0$ . Setting the derivative  $\partial \mathcal{R}_m / \partial w = 0$ , we obtain the optimum weight  $w^o$  of Eq. (12) achieving the minimum of Eq.(40). Substituting the optimum weight  $w^o$  into Eq. (40), the MSE  $\mathcal{R}_m^o$  can be simplified as Eq. (13). As the number  $m \rightarrow \infty$  and  $C_{ii} < \infty$ , we have the limit  $\mathcal{R}_\infty^o = \lim_{m \rightarrow \infty} \mathcal{R}_m^o$  in Eq. (14).

APPENDIX D  
PROOF OF COROLLARY 2

For the Bayesian estimator in Eq. (15), the corresponding MSE is given by

$$\begin{aligned} \mathcal{R} &= E_x[(\theta - \hat{\theta}_{LC})^2] \\ &= \text{var}(\theta) - 2wE_x[(\theta - E_\theta(\theta))(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))] \\ &\quad + w^2E_x[(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))^2]. \end{aligned} \quad (41)$$

Then, setting the derivative  $\partial \mathcal{R} / \partial w = 0$ , the optimum weight  $w^o$  is solved as Eq. (17). Substituting the optimum weight  $w^o$  of Eq. (17) into Eq. (41), we have the minimum MSE

$$\mathcal{R}^o = \text{var}(\theta) - \frac{E_x^2[(\theta - E_\theta(\theta))(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))]}{E_x[(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))^2]}. \quad (42)$$

Substituting  $\hat{\theta}_{NE}(x) = E_\eta[\hat{\theta}(x + \eta)]$  of Eq. (16) into Eq. (42), we find the numerator  $E_x^2[(\theta - E_\theta(\theta))(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))] = p_i$  of Eq. (37) and the denominator  $E_x[(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))^2] = C_{ij}$  of Eq. (39). Therefore, the MSE  $\mathcal{R}^o$  of Eq. (42) is just the limit  $\mathcal{R}_\infty^o$  of Eq. (14). We also find that  $\mathcal{R}^o$  of Eq. (42) satisfies

$$\begin{aligned} \mathcal{R}^o &= \min_w E_x[(\theta - \hat{\theta}_{LC})^2] \\ &= \min_w E_x\{[\theta - E_\theta(\theta) - w(\hat{\theta}_{NE} - E_x(\hat{\theta}_{NE}))]^2\} \\ &\leq E_x\{[\theta - E_\theta(\theta) - (\hat{\theta}_{NE} - E_x(\hat{\theta}_{NE}))]^2\}_{w=1} \\ &= E_x[(\theta - \hat{\theta}_{NE})^2], \end{aligned} \quad (43)$$

where the unbiased condition  $E_x(\hat{\theta}_{NE}) = E_\theta(\theta)$ . This inequality of Eq. (43) clearly shows that the estimator  $\hat{\theta}_{LC}$  in Eq. (15) presents a better or at least equivalent MSE compared to that of the estimator  $\hat{\theta}_{NE}$  in Eq. (16), and this is obtained thanks to purposeful addition of noise and the adjustable weighting coefficients in the combiner of Fig. 1.

APPENDIX E  
PROOF OF COROLLARY 3

From the MSE  $\mathcal{R}_\infty^o$  of Eq. (14), i.e. Eq. (42), achieved by the estimator  $\hat{\theta}_{LC}$  of Eq. (15) and by using the Cauchy-Schwarz inequality, we find the numerator

$$\begin{aligned} &E_x^2[(\theta - E_\theta(\theta))(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))] \\ &= E_x^2[E_{\theta|x}(\theta - E_\theta(\theta))(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))] \\ &= E_x^2[(\hat{\theta}_{mmse} - E_\theta(\theta))(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))] \\ &\leq E_x[(\hat{\theta}_{mmse}(x) - E_\theta(\theta))^2]E_x[(\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE}))^2], \end{aligned} \quad (44)$$

where the equality occurs when  $\hat{\theta}_{mmse}(x) - E_\theta(\theta) = \kappa[\hat{\theta}_{NE}(x) - E_x(\hat{\theta}_{NE})]$ , i.e. Eq. (20), and  $\kappa$  is an arbitrary constant. Sub-

stituting the inequality of Eq. (44) into Eq. (42), we have

$$\begin{aligned}
\mathcal{R}_\infty^o &\geq \text{var}(\theta) - \mathbb{E}_x[(\hat{\theta}_{\text{mmse}}(x) - \mathbb{E}_\theta(\theta))^2] \\
&= \text{var}(\theta) - \mathbb{E}_x(\hat{\theta}_{\text{mmse}}^2) + 2\mathbb{E}_x(\hat{\theta}_{\text{mmse}})\mathbb{E}_\theta(\theta) - \mathbb{E}_\theta^2(\theta) \\
&= \mathbb{E}_\theta(\theta^2) - \mathbb{E}_x(\hat{\theta}_{\text{mmse}}^2) \\
&= \mathbb{E}_x[(\theta - \hat{\theta}_{\text{mmse}})^2] \\
&= \mathcal{R}_{\text{ms}},
\end{aligned}$$

where the unbiased condition  $\mathbb{E}_\theta(\theta) = \mathbb{E}_x(\hat{\theta}_{\text{mmse}})$  and  $\mathbb{E}_x(\theta\hat{\theta}_{\text{mmse}}) = \mathbb{E}_x[\mathbb{E}_{\theta|x}(\theta)\hat{\theta}_{\text{mmse}}] = \mathbb{E}_x(\hat{\theta}_{\text{mmse}}^2)$ . Thus, Eq. (18) holds.

#### APPENDIX F PROOF OF COROLLARY 4

From Eq. (14), we have the derivative

$$\frac{\partial \mathcal{R}_\infty^o}{\partial \sigma_\eta} = \frac{p_i^2 \frac{\partial C_{ij}}{\partial \sigma_\eta} - 2p_i C_{ij} \frac{\partial p_i}{\partial \sigma_\eta}}{C_{ij}^2}. \quad (45)$$

From Eq. (37) and Eq. (39) and noting the interchange of the order of differentiation and integration in Eq. (45), both derivatives  $\partial p_i / \partial \sigma_\eta$  and  $\partial C_{ij} / \partial \sigma_\eta$  contain the term

$$\begin{aligned}
\frac{\partial \mathbb{E}_\eta[\hat{\theta}(x + \eta)]}{\partial \sigma_\eta} &= \int \hat{\theta}(x + \eta) \frac{\partial f_\eta(\eta, \sigma_\eta)}{\partial \sigma_\eta} d\eta \\
&= \int \hat{\theta}(x + \eta) \left( -\frac{f_\eta(\frac{\eta}{\sigma_\eta})}{\sigma_\eta^2} - \frac{\eta f'_\eta(\frac{\eta}{\sigma_\eta})}{\sigma_\eta^3} \right) d\eta \\
&= \int \hat{\theta}(x + \sigma_\eta \tilde{\eta}) \left( -\frac{f_\eta(\tilde{\eta})}{\sigma_\eta} - \frac{\tilde{\eta} f'_\eta(\tilde{\eta})}{\sigma_\eta} \right) d\tilde{\eta}, \quad (46)
\end{aligned}$$

where  $\tilde{\eta} = \eta / \sigma_\eta$  and  $f'_\eta = df_\eta / d\eta$ . Since  $\hat{\theta}(x)|_{x \rightarrow \infty} = \Theta < \infty$ ,  $\int f_\eta(\tilde{\eta}) d\tilde{\eta} = 1$  and the derivative with respect to  $\sigma_\eta$  allowing us to interchange the order of integration and differentiation [8], we find

$$\begin{aligned}
\lim_{\sigma_\eta \rightarrow \infty} \int \hat{\theta}(x + \sigma_\eta \tilde{\eta}) \frac{f_\eta(\tilde{\eta})}{\sigma_\eta} d\tilde{\eta} &= \int \lim_{\sigma_\eta \rightarrow \infty} \hat{\theta}(x + \sigma_\eta \tilde{\eta}) \frac{f_\eta(\tilde{\eta})}{\sigma_\eta} d\tilde{\eta} \\
&= \lim_{\sigma_\eta \rightarrow \infty} \frac{\Theta \int f_\eta(\tilde{\eta}) d\tilde{\eta}}{\sigma_\eta} = 0,
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\sigma_\eta \rightarrow \infty} \int \hat{\theta}(x + \sigma_\eta \tilde{\eta}) \frac{\tilde{\eta} f'_\eta(\tilde{\eta})}{\sigma_\eta} d\tilde{\eta} \\
&= \lim_{\sigma_\eta \rightarrow \infty} \frac{\Theta [\tilde{\eta} f_\eta(\tilde{\eta})|_{-\infty}^{\infty} - \int f_\eta(\tilde{\eta}) d\tilde{\eta}]}{\sigma_\eta} \\
&= \lim_{\sigma_\eta \rightarrow \infty} \Theta \left[ \frac{\eta f_\eta(\eta)|_{-\infty}^{\infty}}{\sigma_\eta^2} - \frac{1}{\sigma_\eta} \right] = 0,
\end{aligned}$$

where  $f_\eta(\pm\infty) = 0$  and  $\lim_{\sigma_\eta \rightarrow \infty, \eta \rightarrow \pm\infty} \eta f_\eta(\eta) / \sigma_\eta^2$  is infinitesimal of higher order. Thus,  $\lim_{\sigma_\eta \rightarrow \infty} \partial \mathbb{E}_\eta[\hat{\theta}(x + \eta)] / \partial \sigma_\eta = 0$ . Substituting this limit into Eq. (45), we have

$$\lim_{\sigma_\eta \rightarrow \infty} \frac{\partial \mathcal{R}_\infty^o}{\partial \sigma_\eta} = 0. \quad (47)$$

Therefore, under the considered conditions in Corollary 4, the MSE  $\mathcal{R}_\infty^o$  of Eq. (14) tends to a (local) extremum asymptotically for large added noise levels.

#### APPENDIX G PROOF OF COROLLARY 5

For a combiner of identical quantizer estimators  $\hat{\theta}_i = \hat{\theta}_{\text{qt}}$  of Eq. (22), the terms  $\mathbb{E}_\eta[\hat{\theta}_{\text{qt}}(x + \eta)] = \mathbb{E}_\eta[\hat{\theta}_{\text{qt}}^2(x + \eta)] = 1 - F_\eta(\gamma - x) = F_\eta(x - \gamma) = F_\eta[(x - \gamma) / \sigma_\eta]$ , where  $F_\eta(x) = F_{\tilde{\eta}}(x / \sigma_\eta)$  is the cumulative distribution function of added noise  $\eta$  and  $F_{\tilde{\eta}}$  corresponds to the cumulative distribution function of standardized noise  $\tilde{\eta} = \eta / \sigma_\eta$ . From Eq. (13), the derivative of  $\mathcal{R}_m^o$  with respect to the added noise level  $\sigma_\eta$  is

$$\frac{\partial \mathcal{R}_m^o}{\partial \sigma_\eta} = \frac{mp_i^2 \left[ \frac{\partial C_{ii}}{\partial \sigma_\eta} + (m-1) \frac{\partial C_{ij}}{\partial \sigma_\eta} \right]}{[C_{ii} + (m-1)C_{ij}]^2} - \frac{2mp_i \frac{\partial p_i}{\partial \sigma_\eta}}{C_{ii} + (m-1)C_{ij}}.$$

Since the symmetrically-distributed noise  $\tilde{\eta}$  has zero mean  $\int f_{\tilde{\eta}}(\tilde{\eta}) \tilde{\eta} d\tilde{\eta} = 0$  and let  $\tilde{\eta} = (x - \gamma) / \sigma_\eta$ , we have

$$\begin{aligned}
&\lim_{\sigma_\eta \rightarrow 0} \frac{\partial \mathbb{E}_x\{\theta \mathbb{E}_\eta[\hat{\theta}_{\text{qt}}(x + \eta)]\}}{\partial \sigma_\eta} \\
&= \lim_{\sigma_\eta \rightarrow 0} \int \theta f_\theta(\theta) \int \frac{\partial F_{\tilde{\eta}}\left(\frac{x-\gamma}{\sigma_\eta}\right)}{\partial \sigma_\eta} f_\xi(x - \theta) dx d\theta \\
&= \int \theta f_\theta(\theta) \int \lim_{\sigma_\eta \rightarrow 0} f_{\tilde{\eta}}\left(\frac{x-\gamma}{\sigma_\eta}\right) \frac{-(x-\gamma)}{\sigma_\eta^2} f_\xi(x - \theta) dx d\theta \\
&= \int \theta f_\theta(\theta) \int \lim_{\sigma_\eta \rightarrow 0} f_{\tilde{\eta}}(\tilde{\eta}) (-\tilde{\eta}) f_\xi(\sigma_\eta \tilde{\eta} + \gamma - \theta) d\tilde{\eta} d\theta \\
&= \int -\theta f_\theta(\theta) f_\xi(\gamma - \theta) d\theta \int f_{\tilde{\eta}}(\tilde{\eta}) \tilde{\eta} d\tilde{\eta} = 0, \quad (48)
\end{aligned}$$

and

$$\lim_{\sigma_\eta \rightarrow 0} \frac{\partial \mathbb{E}_x\{\mathbb{E}_\eta[\hat{\theta}_{\text{qt}}(x + \eta)]\}}{\partial \sigma_\eta} = -f_x(\gamma) \int f_{\tilde{\eta}}(\tilde{\eta}) \tilde{\eta} d\tilde{\eta} = 0. \quad (49)$$

Then, we find

$$\lim_{\sigma_\eta \rightarrow 0} \frac{\partial p_i}{\partial \sigma_\eta} = 0, \quad \lim_{\sigma_\eta \rightarrow 0} \frac{\partial C_{ii}}{\partial \sigma_\eta} = 0. \quad (50)$$

Due to  $F_{\tilde{\eta}}(-\tilde{\eta}) = 1 - F_{\tilde{\eta}}(\tilde{\eta})$  and  $f_{\tilde{\eta}}(-\tilde{\eta}) = f_{\tilde{\eta}}(\tilde{\eta})$ , the derivative

$$\begin{aligned}
&\lim_{\sigma_\eta \rightarrow 0} \frac{\partial \mathbb{E}_x\{\mathbb{E}_\eta^2[\hat{\theta}_{\text{qt}}(x + \eta)]\}}{\partial \sigma_\eta} \\
&= \lim_{\sigma_\eta \rightarrow 0} \int_{-\infty}^{\infty} 2F_{\tilde{\eta}}\left(\frac{x-\gamma}{\sigma_\eta}\right) \frac{\partial F_{\tilde{\eta}}\left(\frac{x-\gamma}{\sigma_\eta}\right)}{\partial \sigma_\eta} f_x(x) dx \\
&= \int_{-\infty}^{\infty} \lim_{\sigma_\eta \rightarrow 0} 2F_{\tilde{\eta}}\left(\frac{x-\gamma}{\sigma_\eta}\right) f_{\tilde{\eta}}\left(\frac{x-\gamma}{\sigma_\eta}\right) \frac{-(x-\gamma)}{\sigma_\eta^2} f_x(x) dx \\
&= 2f_x(\gamma) \left[ \int_{-\infty}^0 F_{\tilde{\eta}}(\tilde{\eta}) f_{\tilde{\eta}}(\tilde{\eta}) (-\tilde{\eta}) d\tilde{\eta} - \int_0^{\infty} F_{\tilde{\eta}}(\tilde{\eta}) f_{\tilde{\eta}}(\tilde{\eta}) \tilde{\eta} d\tilde{\eta} \right] \\
&= 2f_x(\gamma) \int_0^{\infty} [1 - 2F_{\tilde{\eta}}(\tilde{\eta})] f_{\tilde{\eta}}(\tilde{\eta}) \tilde{\eta} d\tilde{\eta}. \quad (51)
\end{aligned}$$



Thus, for the quantizer number  $m > 1$ ,  $f_x(\gamma) \neq 0$  and  $F_{\tilde{\eta}}(\tilde{\eta}) > 1/2$  over the range of  $\tilde{\eta} > 0$ , we find

$$\begin{aligned} \lim_{\sigma_\eta \rightarrow 0} \frac{\partial \mathcal{R}_m^o}{\partial \sigma_\eta} &= \lim_{\sigma_\eta \rightarrow 0} \frac{m(m-1)p_i^2 \frac{\partial C_{ij}}{\partial \sigma_\eta}}{[C_{ii} + (m-1)C_{ij}]^2} \\ &= \int_0^\infty [1 - 2F_{\tilde{\eta}}(\tilde{\eta})] f_{\tilde{\eta}}(\tilde{\eta}) \tilde{\eta} d\tilde{\eta} \\ &\quad \times \frac{2m(m-1)f_x(\gamma)p_i^2}{[C_{ii} + (m-1)C_{ij}]^2} \Big|_{\sigma_\eta=0} < 0, \end{aligned} \quad (52)$$

where  $p_i^2 > 0$  and  $C_{ii} + (m-1)C_{ij} > 0$  for the positive definite covariance matrix  $\mathbf{C}$ .

## APPENDIX H

### DERIVATION OF PARAMETER VECTOR ESTIMATION

For the observation data model in Eq. (32), the unknown  $k \times 1$  parameter vector  $\boldsymbol{\vartheta}$  is with the  $k \times 1$  known mean vector  $\mathbf{E}_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta})$  and the  $k \times k$  covariance matrix  $\mathbf{C}_{\boldsymbol{\vartheta}}$ . Let the  $N \times k$  observation matrix  $\mathbf{H}$  be represented in row form  $\mathbf{H} = [\mathbf{h}_1^\top, \mathbf{h}_2^\top, \dots, \mathbf{h}_N^\top]^\top$  with its  $1 \times k$  row vectors  $\mathbf{h}_n$ , and assume that the  $N \times 1$  noise vector  $\boldsymbol{\xi}$  is with the common distribution  $f_\xi$  for mutually independent samples  $\xi_n$  for  $n = 1, 2, \dots, N$ , we have the scalar observation

$$x_n = \mathbf{h}_n \boldsymbol{\vartheta} + \xi_n, \quad (53)$$

which can be also analyzed by the theory developed in Sec. II with the parameter  $\theta$  replaced by  $\mathbf{h}_n \boldsymbol{\vartheta}$  in Eq. (53). We add  $m$  noise components  $\eta_{in}$  to the observation  $x_n$ , respectively, resulting in  $m \times 1$  estimate vector

$$\hat{\boldsymbol{\vartheta}}_n = [\hat{\vartheta}_{1n}(x_n + \eta_{1n}), \hat{\vartheta}_{2n}(x_n + \eta_{2n}), \dots, \hat{\vartheta}_{mn}(x_n + \eta_{mn})]^\top.$$

Then, for the observation  $x_n$ , the centralized correlation vector between  $\boldsymbol{\vartheta}$  and  $\hat{\boldsymbol{\vartheta}}_n$  is

$$\mathbf{p}_n = \mathbf{E}_{x_n, \eta}[\mathbf{h}_n(\boldsymbol{\vartheta} - \mathbf{E}_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}))(\hat{\boldsymbol{\vartheta}}_n - \mathbf{E}_{x_n, \eta}(\hat{\boldsymbol{\vartheta}}_n))] = \mathbf{G}_n^\top \mathbf{h}_n^\top$$

with the  $k \times m$  cross-covariance matrix  $\mathbf{G}_n = \mathbf{E}_{x_n, \eta}[(\boldsymbol{\vartheta} - \mathbf{E}_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}))(\hat{\boldsymbol{\vartheta}}_n - \mathbf{E}_{x_n, \eta}(\hat{\boldsymbol{\vartheta}}_n))^\top]$ , and the  $m \times m$  covariance matrix  $\mathbf{C}_n = \mathbf{E}_{x_n, \eta}[(\hat{\boldsymbol{\vartheta}}_n - \mathbf{E}_{x_n, \eta}(\hat{\boldsymbol{\vartheta}}_n))(\hat{\boldsymbol{\vartheta}}_n - \mathbf{E}_{x_n, \eta}(\hat{\boldsymbol{\vartheta}}_n))^\top]$ . Thus, using the designed Bayesian estimator of Eq. (8), we obtain

$$\begin{aligned} \hat{\vartheta}_{\text{LC}, n} &= \mathbf{h}_n \hat{\boldsymbol{\vartheta}} \\ &= \mathbf{E}_{\boldsymbol{\vartheta}}(\mathbf{h}_n \boldsymbol{\vartheta}) + \mathbf{p}_n^\top \mathbf{C}_n^{-1} (\hat{\boldsymbol{\vartheta}}_n - \mathbf{E}_{x_n, \eta}(\hat{\boldsymbol{\vartheta}}_n)) \\ &= \mathbf{h}_n \mathbf{E}_{\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) + \mathbf{h}_n \mathbf{G}_n \mathbf{C}_n^{-1} (\hat{\boldsymbol{\vartheta}}_n - \mathbf{E}_{x_n, \eta}(\hat{\boldsymbol{\vartheta}}_n)). \end{aligned} \quad (54)$$

These scalar estimators of Eq. (54) can be combined into a vector estimator as

$$\hat{\boldsymbol{\vartheta}}_{\text{LC}} = \begin{bmatrix} \hat{\vartheta}_{\text{LC}, 1} \\ \hat{\vartheta}_{\text{LC}, 2} \\ \vdots \\ \hat{\vartheta}_{\text{LC}, N} \end{bmatrix} = \mathbf{H} \hat{\boldsymbol{\vartheta}}$$

$$= \mathbf{H} \mathbf{E}(\boldsymbol{\vartheta}) + \begin{bmatrix} \mathbf{h}_1 \mathbf{G}_1 \mathbf{C}_1^{-1} (\hat{\boldsymbol{\vartheta}}_1 - \mathbf{E}_{x_1, \eta}(\hat{\boldsymbol{\vartheta}}_1)) \\ \mathbf{h}_2 \mathbf{G}_2 \mathbf{C}_2^{-1} (\hat{\boldsymbol{\vartheta}}_2 - \mathbf{E}_{x_2, \eta}(\hat{\boldsymbol{\vartheta}}_2)) \\ \vdots \\ \mathbf{h}_N \mathbf{G}_N \mathbf{C}_N^{-1} (\hat{\boldsymbol{\vartheta}}_N - \mathbf{E}_{x_N, \eta}(\hat{\boldsymbol{\vartheta}}_N)) \end{bmatrix}. \quad (55)$$

Assuming that  $\mathbf{H}^\top \mathbf{H}$  is invertible, the estimate vector  $\hat{\boldsymbol{\vartheta}}$  is given by

$$\begin{aligned} \hat{\boldsymbol{\vartheta}} &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \hat{\boldsymbol{\vartheta}}_{\text{LC}} \\ &= \mathbf{E}(\boldsymbol{\vartheta}) + (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \begin{bmatrix} \mathbf{h}_1 \mathbf{G}_1 \mathbf{C}_1^{-1} (\hat{\boldsymbol{\vartheta}}_1 - \mathbf{E}_{x_1, \eta}(\hat{\boldsymbol{\vartheta}}_1)) \\ \mathbf{h}_2 \mathbf{G}_2 \mathbf{C}_2^{-1} (\hat{\boldsymbol{\vartheta}}_2 - \mathbf{E}_{x_2, \eta}(\hat{\boldsymbol{\vartheta}}_2)) \\ \vdots \\ \mathbf{h}_N \mathbf{G}_N \mathbf{C}_N^{-1} (\hat{\boldsymbol{\vartheta}}_N - \mathbf{E}_{x_N, \eta}(\hat{\boldsymbol{\vartheta}}_N)) \end{bmatrix}, \end{aligned} \quad (56)$$

with the MSE of each estimation parameter  $\hat{\vartheta}_i$  as the  $i$ -th diagonal element of covariance matrix  $\mathbf{E}_{x, \eta}[(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta} - \hat{\boldsymbol{\vartheta}})^\top]$  for  $i = 1, 2, \dots, k$ .

## REFERENCES

- [1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [2] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Control*, vol. AC-17, no. 4, pp. 439–448, Aug. 1972.
- [3] H. Z. Fang, N. Tian, Y. B. Wang, M. C. Zhou, and M. A. Haile, "Non-linear Bayesian estimation: From Kalman filtering to a broader horizon," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 2, pp. 401–407, Mar. 2018.
- [4] D. Rousseau and F. Chapeau-Blondeau, "Noise-improved Bayesian estimation with arrays of one-bit quantizers," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2658–2662, Dec. 2007.
- [5] F. Chapeau-Blondeau, S. Blanchard, and D. Rousseau, "Noise-enhanced Fisher information in parallel arrays of sensors with saturation," *Phys. Rev. E*, vol. 74, no. 3, 2006, Art. no. 031 102.
- [6] H. Chen, P. K. Varshney, and J. H. Michels, "Noise enhanced parameter estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5074–5081, Oct. 2008.
- [7] S. Uhlich, "Bayes risk reduction of estimators using artificial observation noise," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5535–5545, Oct. 2015.
- [8] A. Patel and B. Kosko, "Optimal mean-square noise benefits in quantizer-array linear estimation," *IEEE Signal Process. Lett.*, vol. 17, no. 12, pp. 1005–1009, Dec. 2010.
- [9] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization*. New York, NY, USA: Cambridge Univ. Press, 2008.
- [10] G. O. Balkan and S. Gezici, "CRLB based optimal noise enhanced parameter estimation using quantized observations," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 477–480, May 2010.
- [11] L. Xu, F. Duan, X. Gao, D. Abbott, and M. D. McDonnell, "Adaptive recursive algorithm for optimal weighted suprathreshold stochastic resonance," *Roy. Soc. Open Sci.*, vol. 4, no. 9, 2017, Art. no. 160 889.
- [12] Y. Pan, F. Duan, F. Chapeau-Blondeau, and D. Abbott, "Noise enhancement in robust estimation of location," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 1953–1966, Apr. 2018.
- [13] Q. Zhai and Y. Wang, "Noise effect on signal quantization in an array of binary quantizers," *Signal Process.*, vol. 152, no. 11, pp. 265–272, 2018.
- [14] J. Zhu, X. Li, R. S. Blum, and Y. Gu, "Parameter estimation from quantized observations in multiplicative noise environments," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4037–4050, Aug. 2015.
- [15] H. Soganci, S. Gezici, and O. Arikan, "Optimal stochastic parameter design for estimation problems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4950–4956, Sep. 2012.
- [16] H. Chen, L. R. Varshney, and P. K. Varshney, "Noise-enhanced information systems," *Proc. IEEE*, vol. 102, no. 10, pp. 1607–1621, Oct. 2014.

- [17] D. Wang, L. Li, Y. Ji, and Y. Yan, "Model recovery for Hammerstein systems using the auxiliary model based orthogonal matching pursuit method," *Appl. Math. Model.*, vol. 54, no. 2, pp. 537–550, 2018.
- [18] M. Zeng, N. Nam-Phong, O. A. Dobre, and H. V. Poor, "Securing downlink massive MIMO-NOMA networks with artificial noise," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 685–699, Jun. 2019.
- [19] A. Patel and B. Kosko, "Optimal noise benefits in Neyman–Pearson and inequality-constrained statistical signal detection," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1655–1669, May 2009.
- [20] S. Kay, "Can detectability be improved by adding noise?" *IEEE Signal Process. Lett.*, vol. 7, no. 1, pp. 8–10, Jan. 2000.
- [21] S. Zozor and P. O. Amblard, "Stochastic resonance in locally optimal detectors," *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3177–3181, Dec. 2003.
- [22] F. Chapeau-Blondeau and D. Rousseau, "Noise-enhanced performance for an optimal Bayesian estimator," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1327–1334, May 2004.
- [23] H. Chen, P. K. Varshney, S. M. Kay, and J. H. Michels, "Theory of the stochastic resonance effect in signal detection: Part I—Fixed detectors," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3172–3184, Jul. 2007.
- [24] H. Chen and P. K. Varshney, "Theory of the stochastic resonance effect in signal detection—Part II: Variable detectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5031–5041, Oct. 2008.
- [25] A. B. Akbay and S. Gezici, "Noise benefits in joint detection and estimation problems," *Signal Process.*, vol. 118, no. 1, pp. 235–247, 2016.
- [26] H. Soganci, S. Gezici, and O. Arikan, "Optimal signal design for multi-parameter estimation problems," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6074–6085, Nov. 2015.
- [27] G. Zeitler, G. Kramer, and A. C. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.
- [28] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.
- [29] H. Chen and P. K. Varshney, "Performance limit for distributed estimation systems with identical one-bit quantizers," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 466–471, Jan. 2010.
- [30] H. Chen and P. K. Varshney, "Nonparametric one-bit quantizers for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3777–3787, Jul. 2010.
- [31] O. Dabeer and A. Karnik, "Signal parameter estimation using 1-bit dithered quantization," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5389–5405, Dec. 2006.
- [32] M. D. McDonnell, "Is electrical noise useful?" *Proc. IEEE*, vol. 99, no. 2, pp. 242–246, Feb. 2011.
- [33] G. P. Harmer, B. R. Davis, and D. Abbott, "A review of stochastic resonance: Circuits and measurement," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 299–309, Apr. 2002.
- [34] N. G. Stocks, "Suprathreshold stochastic resonance in multilevel threshold systems," *Phys. Rev. Lett.*, vol. 84, no. 11, pp. 2310–2313, 2000.
- [35] L. Xu, T. Vladusich, F. Duan, L. J. Gunn, D. Abbott, and M. D. McDonnell, "Decoding suprathreshold stochastic resonance with optimal weights," *Phys. Lett. A*, vol. 379, no. 38, pp. 2277–2283, 2015.
- [36] R. Benzi, A. Sutera, and A. Vulpiani, "The mechanism of stochastic resonance," *J. Phys. A, Math. General*, vol. 14, no. 11, pp. L453–L457, 1981.
- [37] S. Bayram, S. Gezici, and H. V. Poor, "Noise enhanced hypothesis-testing in the restricted Bayesian framework," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3972–3989, Aug. 2010.
- [38] E. Zeidler, *Nonlinear Functional Analysis and Its Applications, Part 3: Variational Methods and Optimization*. Berlin, Germany: Springer-Verlag, 1984.
- [39] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: Wiley, 2000.
- [40] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 2006.
- [41] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983.
- [42] J. J. Collins, C. C. Chow, and T. T. Imhoff, "Stochastic resonance without tuning," *Nature*, vol. 376, no. 6537, pp. 236–238, 1995.
- [43] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1985.
- [44] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. New York, NY, USA: McGraw-Hill, 2000.