



Cite this article: Xu L, Duan F, Gao X, Abbott D, McDonnell MD. 2017 Adaptive recursive algorithm for optimal weighted suprathreshold stochastic resonance. *R. Soc. open sci.* **4**: 160889.
<http://dx.doi.org/10.1098/rsos.160889>

Received: 8 November 2016

Accepted: 14 August 2017

Subject Category:

Engineering

Subject Areas:

mathematical physics/biophysics

Keywords:

suprathreshold stochastic resonance, adaptive signal processing, Kalman–least mean square, recursive algorithm

Author for correspondence:

Fabing Duan

e-mail: fabing.duan@gmail.com

Adaptive recursive algorithm for optimal weighted suprathreshold stochastic resonance

Liyan Xu¹, Fabing Duan¹, Xiao Gao², Derek Abbott³
and Mark D. McDonnell^{2,3}

¹Institute of Complexity Science, Qingdao University, Qingdao 266071, People's Republic of China

²Computational and Theoretical Neuroscience Laboratory, Institute for Telecommunications Research, School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, South Australia 5095, Australia

³Centre for Biomedical Engineering (CBME) and School of Electrical & Electronic Engineering, The University of Adelaide, Adelaide, South Australia 5005, Australia

LX, 0000-0003-4615-9183; DA, 0000-0002-0945-2674

Suprathreshold stochastic resonance (SSR) is a distinct form of stochastic resonance, which occurs in multilevel parallel threshold arrays with no requirements on signal strength. In the generic SSR model, an optimal weighted decoding scheme shows its superiority in minimizing the mean square error (MSE). In this study, we extend the proposed optimal weighted decoding scheme to more general input characteristics by combining a Kalman filter and a least mean square (LMS) recursive algorithm, wherein the weighted coefficients can be adaptively adjusted so as to minimize the MSE without complete knowledge of input statistics. We demonstrate that the optimal weighted decoding scheme based on the Kalman–LMS recursive algorithm is able to robustly decode the outputs from the system in which SSR is observed, even for complex situations where the signal and noise vary over time.

1. Introduction

Stochastic resonance in multi-threshold systems was initially investigated in [1], where the input signal is subthreshold. More interestingly, the concept of suprathreshold stochastic resonance (SSR) was also introduced in multi-threshold systems [2–4]. Note that SSR is an important variation of stochastic resonance [5, 6], where the output is counterintuitively enhanced by noise,

operating with signals of arbitrary magnitude, not restricted to weak or subthreshold signals. Since its introduction, SSR has received considerable attention in diverse areas concerned with the transmission of signals, and has also been considered in the design of cochlear implants [7], analogue-to-digital converter circuits [8–10], nonlinear detectors [11,12], digital accelerometers [13] and stochastic quantizers [14]. In the seminal works of Stocks and co-workers [2–4], the nonlinearity in each element of the model is assumed to be a binary quantizer. Thus, such threshold systems can be described as stochastic signal quantizers that have been analysed in terms of lossy source coding and quantization theory [14–16]. Following the works in [14–16], we investigated the decoding scheme of quantized signal, named as optimal weighted decoding, by using weighted coefficients [17,18]. Our results show that under certain conditions the performance of the optimally weighted quantizer response is superior to that of the original unweighted arrays [17,18].

However, the previous studies [17,18] have been undertaken assuming that the input characteristics are statistically stationary, and the decoding schemes are based on *a priori* knowledge of the inputs. Specifically, the noise is modelled as white Gaussian distributed [17,18]. However, in practical applications, the statistical characteristics of the input signals are generally unknown or are often varied with time. Furthermore, noise in real systems is coloured, and the idealization of white noise is never exactly realized [5,19]. These constraints severely limit the decoding operation based on performance enhancement of real systems. A highly successful solution to this more difficult problem is found in adaptive filtering, which is a powerful approach with a wide variety of engineering applications [20–24]. Adaptive filtering has the ability to adjust system parameters automatically with no *a priori* knowledge of inputs, and allows processing the case wherein the properties of inputs are unknown, non-stationary or time variable [25–27]. Interestingly, a toy model has been established that illustrates a process of optimization that works without *a priori* knowledge of the input statistical distribution—so we do have a precedent to show mathematical tractability of this class of problem [28].

Specifically, the Kalman filter and the least mean square (LMS) algorithm are two of the most popular adaptive estimation methods in adaptive signal processing, with the former as a realization of the optimal Bayesian estimator and the latter as a recursive solution to the optimal Wiener filtering problem [27]. Recently, Mandic *et al.* subtly developed a joint perspective on these two algorithms, and proposed the Kalman–LMS recursive algorithm that permits the implementation of Kalman filters without any notion of Bayesian statistics [29].

The purpose of this paper is to extend optimal-weighted decoding to more general input characteristics by using the Kalman–LMS recursive algorithm, wherein the weighting coefficients can be adaptively adjusted based on real-time measurements of the input signals. For the typical SSR model of threshold arrays, we find this Kalman–LMS recursive algorithm can deal with not only the simple situation of stationary signals, but also more complicated cases of non-stationary signals and coloured noise. The decoding performance of the mean square error (MSE) distortion obtained by the Kalman–LMS recursive algorithm illustrates interesting progress in optimal weighted SSR that may be of benefit in adaptive signal processing algorithms applied to nonlinear noisy systems.

2. Model and method

2.1. Model

We consider an adaptively weighted summing array of N noisy nonlinear elements, as shown in figure 1. All elements receive the same input signal x_k representing the k th element in the time series, which is assumed to be a deterministic signal or a stationary stochastic process [29]. Each element of the array is endowed with the same input–output characteristic, modelled by the static (memoryless) function g . The i -th nonlinear element is subject to independent and identically distributed (i.i.d.) additive noise component $\eta_{i,k}$ with standard deviation σ_η , which is independent of the signal x_k . Accordingly, each element produces the output signal $y_{i,k} = g[x_k + \eta_{i,k}]$. The output signal $y_{i,k}$ is multiplied by the weighting coefficient $w_{i,k}$ ($w_{i,k} \in \mathbb{R}$). $w_{i,k}$ are adjusted by the Kalman–LMS recursive algorithm aiming to minimize the performance metric of MSE. All weighted outputs are summed to give the overall output of array $\hat{y}_k = \sum_{i=1}^N w_{i,k} y_{i,k}$.

The Kalman–LMS recursive algorithm combines Kalman filtering with LMS-type algorithms to control both the direction and magnitude of adaptation steps along the shortest path so as to achieve the global minimum of MSE [29]. Our purpose is to adaptively adjust the weights by this algorithm to make the decoding output $\hat{y}_k$ approximate the input x_k , i.e. minimizing the MSE distortion.

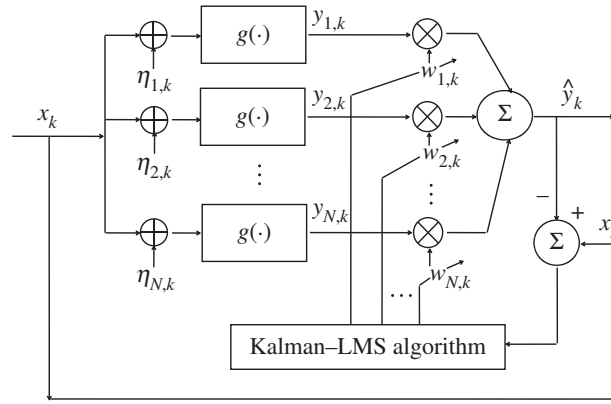


Figure 1. Adaptively weighted summing array of N noisy nonlinear elements, $g(\cdot)$. Each element operates on a common signal x_k subject to additive noise $\eta_{i,k}$ at time k . The output of each individual element $y_{i,k}$ is multiplied by the weighting coefficient $w_{i,k}$, and the overall output $\hat{y}_k = \sum_{i=1}^N w_{i,k} y_{i,k}$.

2.2. Method

To begin, we introduce a column vector $\mathbf{y}_k = [y_{1,k}, y_{2,k}, \dots, y_{N,k}]^T$ to represent the N output signals of nonlinear elements for each random input x_k . We denote a vector of weights as $\mathbf{w}_k = [w_{1,k}, w_{2,k}, \dots, w_{N,k}]^T$ and the optimal weights as $\mathbf{w}_k^0$ that corresponds to the minimum of MSE.

2.2.1. Performance evaluation criteria

Ideally, we wish to achieve

$$x_k = \mathbf{y}_k^T \mathbf{w}_k^0, \quad (2.1)$$

where the aim is to estimate the optimal weight vector $\mathbf{w}_k^0$ for minimizing the MSE distortion. It can be fixed, i.e. $\mathbf{w}_k^0 = \mathbf{w}^0$, or time varying as in equation (2.1) [29]. For the stationary input signal x_k and noise η_k , equation (2.1) can be written as:

$$x_k = \mathbf{y}_k^T \mathbf{w}^0. \quad (2.2)$$

When the inverse matrix $(\mathbf{y}_k \mathbf{y}_k^T)^{-1}$ exists, the Wiener optimal weight vector $\mathbf{w}^0$ is given by $\mathbf{w}^0 = (\mathbf{y}_k \mathbf{y}_k^T)^{-1} \mathbf{y}_k x_k$ [16,21]. For the stationary input x_k , the memoryless nonlinearity g and the stationary noise η_k , we assume the optimal weight vector $\mathbf{w}_k^0$ converges to the Wiener solution of $\mathbf{w}^0$. We desire to estimate the optimal weight vector $\mathbf{w}^0$ recursively, based on the existing weight vector $\mathbf{w}_{k-1}$, the input signal x_k and output signals of nonlinear elements $\mathbf{y}_k$, i.e. $\hat{\mathbf{w}}^0 = \mathbf{w}_k = f(\mathbf{w}_{k-1}, x_k, \mathbf{y}_k)$. Then the decoding output is $\hat{y}_k = \mathbf{y}_k^T \mathbf{w}_{k-1}$. The error between the input x_k , and the decoding output $\hat{y}_k$ is given by

$$e_k = x_k - \hat{y}_k = x_k - \mathbf{y}_k^T \mathbf{w}_{k-1}. \quad (2.3)$$

Similarly, the weight error vector $\tilde{\mathbf{w}}_k$ between the weight vector estimate $\mathbf{w}_k$ and the optimal weight vector $\mathbf{w}^0$ is written by [29]

$$\tilde{\mathbf{w}}_k = \mathbf{w}^0 - \mathbf{w}_k \quad (2.4)$$

and its contribution to the error e_k is given by

$$e_k = \mathbf{y}_k^T \tilde{\mathbf{w}}_{k-1}. \quad (2.5)$$

The decoding performance metric of MSE represents the power of the output error e_k , and is expressed as

$$\text{MSE} = E[e_k^2], \quad (2.6)$$

where the statistical expectation $E(\cdot)$ introduced in equation (2.6) is defined in terms of the joint probability of the output $\mathbf{y}_k$ and the weight error vector $\tilde{\mathbf{w}}_k$. The theoretical calculation of MSE is difficult, and we will numerically obtain the MSE distortion for a sufficiently large observation time in the following experiments.

For the memoryless function g , we assume the output y_k at the recursion step k is not related to the weight error vector $\tilde{\mathbf{w}}_{k-1}$ at the time step $k-1$. Thus, the instantaneous time-varying MSE is given in [29]

$$\zeta_k = E[(y_k^\top \tilde{\mathbf{w}}_{k-1})^2] = \mathbf{y}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k, \quad (2.7)$$

where $\mathbf{P}_{k-1} = E[\tilde{\mathbf{w}}_{k-1} \tilde{\mathbf{w}}_{k-1}^\top]$ is the symmetric and positive semi-definite weight error covariance matrix, and the statistical expectation $E(\cdot)$ is calculated in terms of the probability density of $\tilde{\mathbf{w}}_{k-1}$. In [29], another performance evaluating metric—the mean square deviation J_k —is introduced, which represents the power of the weight error vector $\tilde{\mathbf{w}}_k$ and is given by

$$J_k = E[\tilde{\mathbf{w}}_k^\top \tilde{\mathbf{w}}_k] = \text{tr}(\mathbf{P}_k). \quad (2.8)$$

The mean square deviation J_k is related to the instantaneous MSE in equation (2.7) through the weight error covariance matrix $\mathbf{P}_k$. Therefore, minimizing J_k is equivalent to minimizing the instantaneous MSE [29]. Based on equation (2.8), we will deduce the weight vector estimate $\mathbf{w}_k$ at each recursion step k .

2.2.2. Optimal learning gain

The LMS algorithm uses the stochastic gradient descent and uses a recursive estimation of the optimal weight vector, $\mathbf{w}^0$ in equation (2.2), in the form $\mathbf{w}_k = \mathbf{w}_{k-1} - \mu_k \nabla_{\mathbf{w}} E[e_k^2]$, where μ_k is a step size and $\nabla_{\mathbf{w}}$ is a gradient vector [29]. Based on the instantaneous estimate $E[e_k^2] \approx e_k^2$, the LMS solution is given in [21]

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mu_k \mathbf{y}_k e_k. \quad (2.9)$$

Note that the second term in equation (2.9) the weight update, $\mu_k \mathbf{y}_k e_k$, has the same direction as the vector $\mathbf{y}_k$. It turns out that the gradient descent performs locally optimal steps but has no means to follow the globally optimal shortest path to the solution $\mathbf{w}^0$. Therefore, it is necessary to control both the direction and magnitude of the adaptive steps μ_k to follow the shortest, optimal path to $\mathbf{w}^0$ [29]. In this way, Mandic *et al.* introduce a positive definite learning gain matrix $\mathbf{G}_k$, in the context of Kalman filters, to replace the scalar step size μ_k so as to control both the direction and magnitude of the gradient descent [29]. Thus, the weight update recursion in equation (2.9) generalizes to

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mathbf{G}_k \mathbf{y}_k e_k = \mathbf{w}_{k-1} + \mathbf{g}_k e_k, \quad (2.10)$$

where the gain vector $\mathbf{g}_k = \mathbf{G}_k \mathbf{y}_k$ [29].

Subtracting $\mathbf{w}^0$ from both sides of equation (2.10) and replacing the error with equation (2.5), we can rewrite equation (2.10) in terms of the weight error vector as [29]

$$\tilde{\mathbf{w}}_k = \tilde{\mathbf{w}}_{k-1} - \mathbf{g}_k \mathbf{y}_k^\top \tilde{\mathbf{w}}_{k-1}. \quad (2.11)$$

Utilizing equation (2.11), we can obtain the recursion for the weight error covariance matrix $\mathbf{P}_k$ as [29]

$$\mathbf{P}_k = E[\tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^\top] = \mathbf{P}_{k-1} - (\mathbf{P}_{k-1} \mathbf{y}_k \mathbf{g}_k^\top + \mathbf{g}_k \mathbf{y}_k^\top \mathbf{P}_{k-1}) + \mathbf{g}_k \mathbf{g}_k^\top \mathbf{y}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k. \quad (2.12)$$

Substituting equation (2.12) into equation (2.8), the mean square deviation is obtained as [29]

$$J_k = J_{k-1} - 2\mathbf{g}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k + \|\mathbf{g}_k\|^2 \mathbf{y}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k. \quad (2.13)$$

This derivation of equation (2.13) uses the facts that $\text{tr}(\mathbf{P}_{k-1} \mathbf{y}_k \mathbf{g}_k^\top) = \text{tr}(\mathbf{g}_k \mathbf{y}_k^\top \mathbf{P}_{k-1}) = \mathbf{g}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k$ and $\text{tr}(\mathbf{g}_k \mathbf{g}_k^\top) = \mathbf{g}_k^\top \mathbf{g}_k = \|\mathbf{g}_k\|^2$.

Based on equation (2.13), the optimal learning gain vector $\mathbf{g}_k$ can be obtained by differentiating J_k with respect to $\mathbf{g}_k$, setting to zero and solving for $\mathbf{g}_k$, to give [29]

$$\mathbf{g}_k = \frac{\mathbf{P}_{k-1} \mathbf{y}_k}{\mathbf{y}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k}. \quad (2.14)$$

This optimal gain vector is precisely the Kalman gain [22]. Substituting equation (2.14) into equation (2.12), the update for $\mathbf{P}_k$ is then obtained

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{g}_k \mathbf{y}_k^\top \mathbf{P}_{k-1}. \quad (2.15)$$

2.2.3. Kalman–LMS recursive algorithm

From equations (2.10), (2.14) and (2.15), Kalman–LMS recursive algorithm of estimating the optimal weights $\mathbf{w}^0$ is outlined as:

At each instant $k > 0$, based on measurements $(x_k, \mathbf{y}_k)$

- (1) compute the optimal learning gain (Kalman gain):

$$\mathbf{g}_k = \frac{\mathbf{P}_{k-1} \mathbf{y}_k}{(\mathbf{y}_k^\top \mathbf{P}_{k-1} \mathbf{y}_k + \delta)},$$

where the constant $\delta > 0$ is an initial disturbance for preventing from recursion stop. Without loss of generality, the symmetric and positive semi-definite matrix $\mathbf{P}_0$ is chosen as unit matrix.

- (2) update the weight vector estimate:

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mathbf{g}_k (x_k - \mathbf{y}_k^\top \mathbf{w}_{k-1}).$$

- (3) update the weight error covariance matrix:

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{g}_k \mathbf{y}_k^\top \mathbf{P}_{k-1}.$$

The above operating process exhibits that Kalman–LMS recursive algorithm iteratively updates the weights after each sample, which requires little or no *a priori* knowledge of the signal or noise characteristics. A proof of convergence of the mean square deviation for stationary inputs is provided in appendix A. In the meanwhile, the fastest convergence time corresponding to the array of N static nonlinear elements in figure 1 is also analysed. Our analysis shows that the fastest convergence time is about N times sampling time Δt . It means that, for a given sampling time Δt , the larger the parallel array size N is, the longer the convergent time is. This argument will be shown in the following experiments.

3. Results

It is interesting to note that the above-mentioned decoding scheme using the Kalman–LMS recursive algorithm can be applied to an array composed of arbitrary nonlinear elements. Here, we consider the static function g of figure 1 as Heaviside function that is a typical threshold element [2,3]. The individual output $y_{i,k}$ is given by the response function

$$y_{i,k} = \begin{cases} 1 & x_k + \eta_{i,k} > \theta_i, \\ 0 & \text{otherwise,} \end{cases} \quad (3.1)$$

where θ_i ($i = 1, 2, \dots, N$) is the threshold level for each Heaviside function g .

In the following, we will explore two cases of input characteristics, i.e. stationary and non-stationary, to examine the MSE distortion performance of optimal weighted decoding scheme based on the Kalman–LMS recursive algorithm. In addition, for the noise components $\eta_{i,k}$ in figure 1, white Gaussian and coloured noises are considered, respectively.

3.1. Gaussian noise with identical thresholds

We first consider the case where all threshold levels in equation (3.1) are identical, i.e. $\theta_i = \theta$, and the noises $\eta_{i,k}$ are Gaussian distributed.

It is well known that the stationary assumption of inputs is ideal and is inadequate for dealing with situations in which non-stationarity of the signal and or noise is intrinsic to the problem. The ability to adapt in a non-stationary environment is an important function of an adaptive algorithm. Specifically in a non-stationary environment it can offer a tracking capability for the time-varying input signal, provided that the variations are sufficiently slow [24].

We now consider the case where the input x_k is non-stationary stochastic signal. For instance, x_k is Gaussian distributed, but the standard deviation $\sigma_x(t)$ is time varying. Here, the signal standard deviation is chosen as $\sigma_x(t) = \sqrt{2} \sin(2\pi ft)$ with the modulation frequency f , and the sampling time $\Delta t = 10^{-3}$ s. Of course, we can have other forms of $\sigma_x(t)$ provided that it is slow time-varying. In appendix A, we investigate the tracking capability of the proposed Kalman–LMS algorithm, and an intrinsic time scale of $N\Delta t$ is approximately provided for analysing the temporal dynamics of the adaptive process. For different time variations of the non-stationary input and the array size $N = 63$, a performance comparison

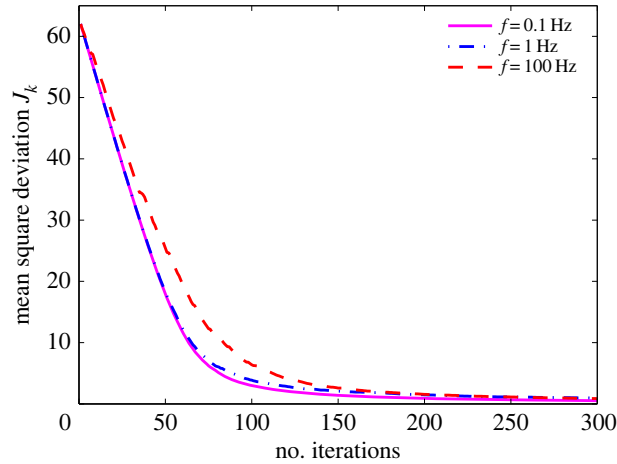


Figure 2. Comparison of mean square deviation for non-stationary inputs with various time variations. The non-stationary input is Gaussian distributed with standard deviation $\sigma_x(t) = \sin(2\pi ft)$ ($f = 0.1$ Hz, 1 Hz and 100 Hz). Here, the array size $N = 63$, the noise standard deviation $\sigma_\eta = 1$ and sampling time $\Delta t = 10^{-3}$ s.

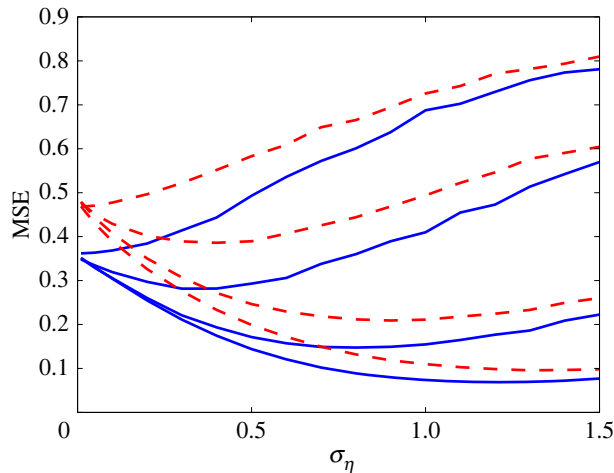


Figure 3. MSE distortion versus σ_η for Gaussian noise with identical thresholds. The dashed red lines represent the MSE distortion for non-stationary input that is Gaussian distributed with standard deviation $\sigma_x(t) = \sqrt{2} \sin(2\pi t)$. The solid blue lines correspond to the MSE distortion for stationary input that is Gaussian distributed with $\sigma_x = 1$. For the two cases of input characteristics, from top to bottom the array sizes are $N = 1, 3, 15$ and 63 .

is shown in figure 2. It apparently shows that, when the modulation frequency $f = 0.1$ Hz and 1 Hz, the mean square deviation J_k can track non-stationary changes of the environment well, since such time variations of standard deviation $\sigma_x(t)$, compared with the intrinsic time scale of $N\Delta t$, occur slowly enough.

Unless specifically mentioned, all results in the following parts are obtained for the example case where the input signal for non-stationary characteristics is Gaussian distributed with standard deviation $\sigma_x(t) = \sqrt{2} \sin(2\pi t)$, and the input signal for stationary characteristics is Gaussian distributed with $\sigma_x = 1$.

The MSE distortions for the non-stationary stochastic input are plotted in figure 3 (dashed red lines). Figure 3 clearly illustrates that the optimal value of σ_η for minimizing the MSE distortion is nonzero for $N > 1$, and thus SSR occurs. As N increases, the MSE distortion at the optimal σ_η decreases, while the optimal value of σ_η will gradually increase, even to a value larger than unity for $N = 63$.

For comparison, the solid blue lines correspond to the MSE distortion curves for stationary Gaussian input signal with standard deviation $\sigma_x = 1$. It is noted that these two inputs have been chosen to give the same average power. For a sufficient time duration, their average powers are all unity. It is clear in

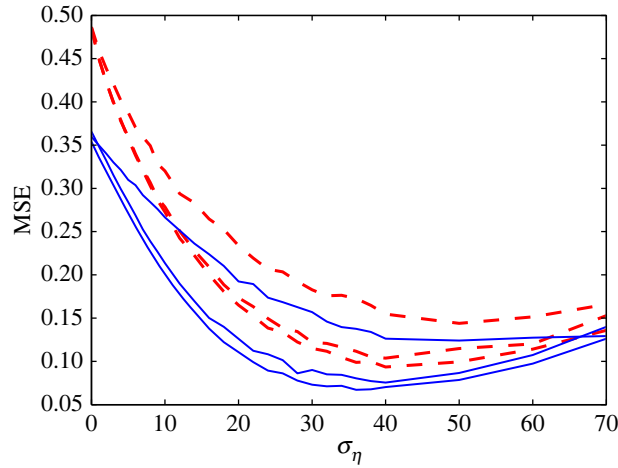


Figure 4. MSE distortion versus standard deviation σ_η for coloured noise with different values of correlation time τ . The dashed red lines represent the non-stationary input, which is Gaussian distributed with time-varying standard deviation $\sigma_x(t) = \sqrt{2} \sin(2\pi t)$. The solid blue lines correspond to the stationary input, which is Gaussian distributed with $\sigma_x = 1$. For the two cases of input characteristics, from bottom to top $\tau = 0.1$ s, 1 s and 10 s. Here, the array size $N = 63$.

figure 3 that the MSE value is far larger for the non-stationary case than stationary case. In addition, for two cases of the input characteristics, very similar qualitative behaviours are all seen. Thus, this also validates that the SSR effect is quite general, not restricted to the stationary signals considered in [17,18].

3.2. Coloured noise with identical thresholds

In the above studies, noise is assumed to be white Gaussian distributed. In the physical world, however, such an idealization is never exactly realized. The effects of various noise correlation times on stochastic resonance have been previously investigated [5,19,30–37]. Owing to the practical importance of coloured noise, we next evaluate the optimal weighted decoding performance under the coloured noise circumstances.

We consider the model driven by an additive exponentially correlated Gaussian noise, i.e. Ornstein–Uhlenbeck noise (OU noise). The archetypal source for OU noise is given in [19]

$$\dot{\xi}(t) = -\frac{1}{\tau}\xi + \frac{\sqrt{D}}{\tau}\xi_w(t), \quad (3.2)$$

where $\xi_w(t)$ denotes Gaussian white noise with autocorrelation $\langle \xi_w(t)\xi_w(s) \rangle = 2\delta(t-s)$ and D is the noise intensity. The stationary autocorrelation of $\xi(t)$ with correlation time τ is then represented by [19]

$$\langle \xi(t)\xi(s) \rangle = \left(\frac{D}{\tau}\right) \exp\left(-\frac{|t-s|}{\tau}\right). \quad (3.3)$$

When correlation time $\tau \rightarrow 0$, equation (3.3) reproduces the white-noise source frequently used in stochastic resonance studies.

Figure 4 shows the MSE distortion against the standard deviation σ_η of the OU noise with different correlation time $\tau = 0.1$ s, 1 s and 10 s. Here, the array size is $N = 63$. The dashed red lines represent the MSE distortion performance for non-stationary inputs, while the solid blue lines correspond to the decoding performance for stationary inputs. As the correlation time τ increases, it is seen in figure 4 that the minimum MSE value at the optimal σ_η increases. The characteristic behaviour for stochastic resonance is in good agreement with results in [5,19,30–32]. The reason is that, since the variance σ_η^2 of the OU noise is D/τ in equation (3.3), the noise strength D increases proportionally with the increase of τ for a given σ_η^2 . However, we assume the output $\mathbf{y}_k$ at the recursion step k is not related to the weight error vector $\tilde{\mathbf{w}}_{k-1}$ at the time step $k-1$ in equation (2.7). This restriction will hinder the application of this adaptive algorithm to the case of coloured noise. Despite this, this adaptive algorithm still plays, and presents the corresponding MSE distortions in figure 4. It is seen that, as the correlation time τ increases, the corresponding MSE distortions increase for both the non-stationary (dashed lines) and stationary (solid lines) inputs.

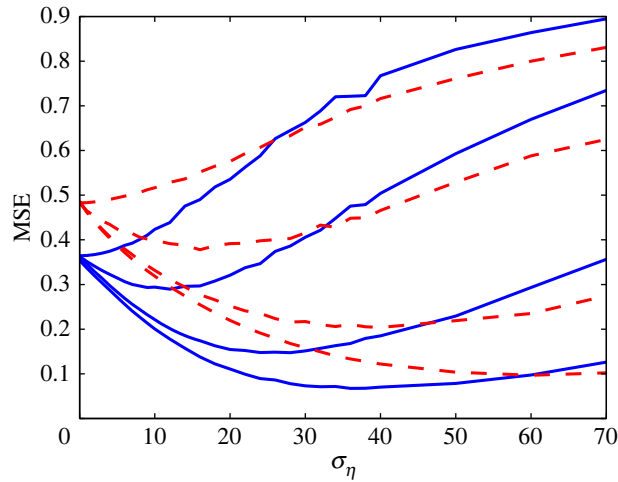


Figure 5. The MSE distortion versus σ_η for coloured noise with the correlation time $\tau = 0.1$ s in the case of identical thresholds. The solid blue lines are for the case of stationary input, and the dashed red lines correspond to the case of non-stationary input. From top to bottom, the array sizes $N = 1, 3, 15$ and 63 .

Under coloured noise conditions, to compare the MSE distortion performance for stationary and non-stationary input situations, figure 5 shows MSE for OU noise with the correlation time $\tau = 0.1$ s and the array sizes $N = 1, 3, 15$ and 63 .

Note from figure 5 that, as N increases, very similar decoding performance appears when compared with figure 3. Specifically, these two figures both indicate that the MSE is far larger for the non-stationary case than in the stationary case. The results in figure 5 show that this Kalman–LMS recursive algorithm can deal with not only the simple situation of stationary signals, but also more complicated cases of non-stationary signal buried in coloured noise.

3.3. Gaussian noise with grouped thresholds

Our recent work [18] showed that the decoding performance for multigroup parameter settings is superior to that of identical parameter settings. Following the approach presented in [18], we next study the decoding performance of the Kalman–LMS recursive algorithm for the case of grouped thresholds.

We divide the set of threshold elements into M ($M \leq N$) groups. Within each group, the threshold element size is equal, i.e. N/M , and the threshold levels are equally spaced and set as $\theta_m = m\sigma_x/(M+1)$ for $m = 1, 2, \dots, M$.

Figure 6a exhibits the MSE distortion for Gaussian noise with various group sizes M in the cases of non-stationary and stationary inputs. Here, the array size $N = 120$, and the group sizes $M = 1, 2, 3, 5, 10$ and 120 . It is apparently illustrated in figure 6a that, for small noise levels, the decoding performance greatly improves as the group size M increases. While, for very large noise levels, all of the MSE values that correspond to different groups tend to the same observation, which equals the MSE distortion of the identical threshold level setting. This fact tells us that, for weak and moderate noise intensities, the multigroup setting scheme can reduce the MSE distortion. Figure 6a also reveals that the grouped threshold setting can be extended to adaptively weighted summing arrays.

3.4. Coloured noise with grouped thresholds

It is interesting to study the performance of weighted decoding for the case of grouped thresholds under the coloured noise circumstances. Similarly, we also illustrate the MSE curves for OU noise with various group sizes M in the cases of non-stationary and stationary inputs in figure 6b. It can be seen from figure 6a,b that, as the group sizes M increase, the grouped decoding performance under the coloured noise circumstances is similar to that under Gaussian noise circumstances.

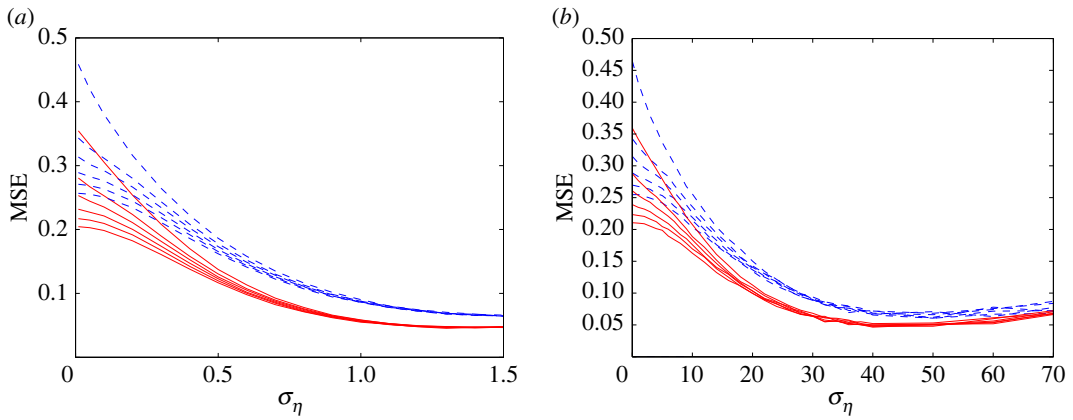


Figure 6. MSE distortion versus the noise level for group sizes $M = 1, 2, 3, 5, 10$ and 120 (from top to bottom), and the array size $N = 120$. Within each group, the threshold element size is equal, i.e. N/M , and the threshold levels are equally spaced and set as $\theta_m = m\sigma_x/(M + 1)$ for $m = 1, 2, \dots, M$ ($M \leq N$). (a) White Gaussian noise. (b) Coloured noise with the correlation time $\tau = 0.1$ s. The solid red lines correspond to the MSE distortion for stationary input, and the dashed blue lines represent the MSE distortion for non-stationary input.

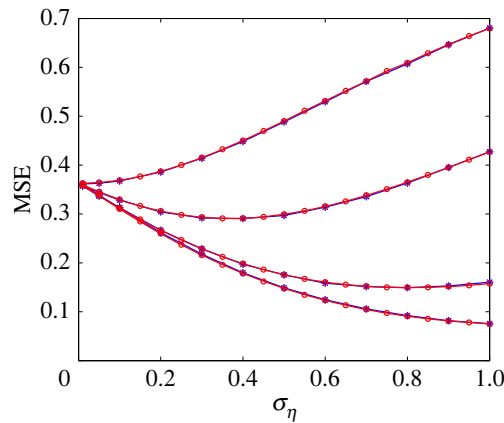


Figure 7. MSE distortion versus the noise level σ_η of Gaussian white noise in the case of identical thresholds. Here, the array sizes $N = 1, 3, 15$ and 63 (from top to bottom), and the stationary input signal is Gaussian distributed. The circled red lines correspond to the MSE distortion for Kalman–LMS recursive algorithm, and the blue stars represent the MSE distortion for optimal weighted decoding presented in [17].

4. Conclusion and discussion

In this paper, we extend the optimal weighted decoding approach to more general input characteristics in the model of a weighted summing array of N noisy nonlinear elements based on the Kalman–LMS recursive algorithm. A proof of convergence of the mean square deviation for stationary inputs is derived. We especially apply the algorithm to a parallel array of threshold elements and investigate the decoding performance for inputs with stationary, non-stationary characteristics under Gaussian noise and coloured noise circumstances. In the case of stationary inputs, after successive iterations of the Kalman–LMS recursive algorithm it converges to the optimum Wiener solution in some statistical sense [24]. Our previous work [17] has shown that, for the case of identical thresholds, the optimal weighted decoding is equivalent to Wiener linear decoding. Therefore, for stationary inputs, the decoding performance exploiting Kalman–LMS recursive algorithm is consistent with that of optimal weighted decoding in the case of identical thresholds. Figure 7 clearly illustrates that the results for these two methods are the same, and MSE distortion curves completely overlap.

Notably, although for the case of stationary inputs, the decoding performance of these two methods is equal under the condition of identical thresholds, the Kalman–LMS recursive algorithm is simple and generally easy to implement since the output weights are updated following each input signal sample,

instead of calculating the output weights using all the training data in one shot. Therefore, compared with the optimal weighted decoding scheme [17], this Kalman–LMS recursive algorithm is a practical method for finding close approximate solutions to equation (2.10) in real time. Moreover, this Kalman–LMS recursive algorithm can adaptively adjust the weighting coefficients so as to minimize the MSE distortion without complete knowledge of input statistics and thus can be applicable to the non-stationary inputs and the coloured noise situations as shown in figures 2 and 3–6 (dashed lines). Additionally, we also apply the Kalman–LMS recursive algorithm to grouped threshold setting for minimizing MSE distortion. The obtained results show that the multigroup setting scheme can be extended to adaptively weighted summing array, and is a significant scheme for many potential applications inspired by SSR mechanism. Although the Kalman–LMS recursive algorithm is applied in the summed array of N noisy nonlinear elements, we expect that our findings may be important for future work on more complex models that will stimulate further studies on SSR.

Beyond the memoryless nonlinearity considered in figure 1, we argue that this Kalman–LMS adaptive algorithm may potentially be extended to other nonlinear systems, for instance, the Hodgkin–Huxley neuron model [5], the reduced FitzHugh–Nagumo neuron model [4,32,34], the auditory model [6,7,16] or biomedical devices [38]. However, the dynamical nonlinear system has its intrinsic time scale that controls the evolution of the system state. The system outputs at adjacent time steps are also correlated with each other to a certain extent, even the input signal or the input noise are independently identically distributed. These factors will lead to larger misalignment between the true and estimated weights. Thus, a more general Kalman–LMS adaptive algorithm for estimating the time-varying weight vector needs to be developed.

Data accessibility. Figures input files: Figshare doi:10.6084/m9.figshare.4214229; Figshare doi:10.6084/m9.figshare.4214232; Figshare doi:10.6084/m9.figshare.4214244; Figshare doi:10.6084/m9.figshare.4214253; Figshare doi:10.6084/m9.figshare.4214262; Figshare doi:10.6084/m9.figshare.4989803.

Authors' contributions. All authors performed the theoretical analyses and participated in the writing of the manuscript. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. F.D. and L.X. are sponsored by the National Natural Science Foundation of China (no. 61573202) and the Science and Technology Development Program of Shandong Province (no. 2014GGX101031).

Appendix A

First, we analyse the convergence of the proposed adaptive algorithm. Based on equations (2.4) and (2.8), the convergence of $\mathbf{w}_k$ to the optimal weight vector $\mathbf{w}^0$ is equivalent to the convergence of the mean square deviation J_k . From equation (2.14), the mean square deviation J_k in equation (2.13) can be written as

$$\begin{aligned} J_k &= J_{k-1} - \mathbf{g}_k^\top P_{k-1} \mathbf{y}_k \\ &= J_{k-1} - \frac{\mathbf{y}_k^\top P_{k-1}^\top P_{k-1} \mathbf{y}_k}{\mathbf{y}_k^\top P_{k-1} \mathbf{y}_k} \\ &= J_{k-1} - \frac{\|P_{k-1} \mathbf{y}_k\|^2}{\varsigma_k}. \end{aligned} \quad (\text{A } 1)$$

Note that $J_k = E[\|\tilde{\mathbf{w}}_k\|^2] \geq 0$ and $\mathbf{y}_k^\top P_{k-1}^\top P_{k-1} \mathbf{y}_k = \|P_{k-1} \mathbf{y}_k\|^2 \geq 0$, thus we find

$$0 \leq J_k \leq J_{k-1}, \quad (\text{A } 2)$$

for $\varsigma_k \neq 0$ and any recursion k . In practice, we update the learning gain as

$$\mathbf{g}_k = \frac{P_{k-1} \mathbf{y}_k}{\mathbf{y}_k^\top P_{k-1} \mathbf{y}_k + \delta} = \frac{P_{k-1} \mathbf{y}_k}{\varsigma_k + \delta} \quad (\text{A } 3)$$

and the mean square deviation J_k becomes

$$J_k = J_{k-1} - \frac{\|P_{k-1} \mathbf{y}_k\|^2}{\varsigma_k + \delta}, \quad (\text{A } 4)$$

for $\delta > 0$ and $\varsigma_k + \delta > 0$. The inequality of equation (A 2) still holds. Therefore, as $k \rightarrow \infty$, the mean square deviation $J_k \rightarrow 0$ is convergent.

Next, we discuss the convergence time of the adaptive algorithm. In order to obtain the fastest convergence time, we need to maximize the term of $\|P_{k-1}\mathbf{y}_k\|^2/\varsigma_k$ in equation (A 1). Since P_{k-1} is a symmetric positive semi-definite matrix, thus we have the normal form

$$P_{k-1} = Q\Lambda_{k-1}Q^T, \quad (\text{A } 5)$$

with the normalized orthonormal matrix $QQ^T = I$ and the eigenvalue matrix $\Lambda_{k-1} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_N]$. Then, we can rewrite the term

$$\frac{\mathbf{y}_k^T P_{k-1}^T P_{k-1} \mathbf{y}_k}{\mathbf{y}_k^T P_{k-1} \mathbf{y}_k} = \frac{\mathbf{y}_k^T Q \Lambda_{k-1} Q^T Q \Lambda_{k-1} Q^T \mathbf{y}_k}{\mathbf{y}_k^T Q \Lambda_{k-1} Q^T \mathbf{y}_k} = \frac{\mathbf{z}_k^T \Lambda_{k-1}^2 \mathbf{z}_k}{\mathbf{z}_k^T \Lambda_{k-1} \mathbf{z}_k}, \quad (\text{A } 6)$$

with $\mathbf{z}_k = Q^T \mathbf{y}_k$. Using the Rayleigh quotient, we obtain

$$0 \leq \lambda_{\min} \leq \frac{\mathbf{z}_k^T \Lambda_{k-1}^2 \mathbf{z}_k}{\mathbf{z}_k^T \Lambda_{k-1} \mathbf{z}_k} \leq \lambda_{\max} \leq \frac{\text{tr}(\mathbf{y}_k^T P_{k-1}^T P_{k-1} \mathbf{y}_k)}{\varsigma_k + \delta} = \frac{\varsigma_k}{\varsigma_k + \delta} \leq 1, \quad (\text{A } 7)$$

with $\lambda_{\min} = \min\{\lambda_1, \lambda_2, \dots, \lambda_N\}$ and $\lambda_{\max} = \max\{\lambda_1, \lambda_2, \dots, \lambda_N\}$. Thus, the convergence time is related to both the output vector $\mathbf{y}_k$ and the weight error covariance matrix P_{k-1} .

In practice, we choose the unit matrix $P_0 = I$, thus the initial mean square deviation $J_0 = \text{tr}(P_0) = N$. From equation (A 4), we find $J_1 = N - \|\mathbf{y}_1\|^2/(\|\mathbf{y}_1\|^2 + \delta) \approx N - 1$. For each recursion step k and from equation (A 7), we assume $J_k \approx N - k$. Therefore, the fastest convergence time is about N times sampling time Δt , i.e. the recursion step $k = N$. The larger the parallel array size N is, the longer the convergent time is. We must note that this is an ideal assumption of convergence time $N\Delta t$, and the experimental results of figure 2 show that the intrinsic time scale for the adaptive algorithm is larger than $N\Delta t$.

References

- Gammaitoni L. 1995 Stochastic resonance in multi-threshold systems. *Phys. Lett. A* **208**, 315–322. (doi:10.1103/PhysRevLett.84.2310)
- Stocks NG. 2000 Suprathreshold stochastic resonance in multilevel threshold systems. *Phys. Rev. Lett.* **84**, 2310–2313. (doi:10.1103/PhysRevLett.84.2310)
- Stocks NG. 2001 Information transmission in parallel threshold arrays: suprathreshold stochastic resonance. *Phys. Rev. E* **63**, 041114. (doi:10.1103/PhysRevE.63.041114)
- Stocks NG, Mannella R. 2001 Generic noise-enhanced coding in neuronal arrays. *Phys. Rev. E* **64**, 030902. (doi:10.1103/PhysRevE.64.030902)
- Gammaitoni L, Hänggi P, Jung P, Marchesoni F. 1998 Stochastic resonance. *Rev. Mod. Phys.* **70**, 223–287. (doi:10.1103/RevModPhys.70.223)
- McDonnell MD, Abbott D. 2009 What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Comput. Biol.* **5**, e1000348. (doi:10.1371/journal.pcbi.1000348)
- Stocks NG, Allingham D, Morse RP. 2002 The application of suprathreshold stochastic resonance to cochlear implant coding. *Fluct. Noise Lett.* **2**, L169–L181. (doi:10.1142/S0219477502000774)
- Oliaei O. 2003 Stochastic resonance in sigma-delta modulators. *Electron. Lett.* **39**, 1–2. (doi:10.1049/el:20030128)
- McDonnell MD, Stocks NG, Pearce CEM, Abbott D. 2005 Analog-to-digital conversion using suprathreshold stochastic resonance. In *Proc. SPIE smart structures, devices, and systems II, Sydney, Australia, 13–15 December*, vol. 5649 (ed. SF Al-Sarawi), pp. 75–84. SPIE. (doi:10.1117/12.582493)
- Nguyen T. 2007 Robust data-optimized stochastic analog-to-digital converters. *IEEE Trans. Signal Process.* **55**, 2735–2740. (doi:10.1109/TSP.2007.893938)
- Zozor S, Amblard PO, Duchêne C. 2007 On pooling networks and fluctuation in suboptimal detection framework. *Fluct. Noise Lett.* **7**, L39–L60. (doi:10.1142/S0219477507003684)
- Hari VN, Anand GV, Premkumar AB, Madhukumar AS. 2012 Design and performance analysis of a signal detector based on suprathreshold stochastic resonance. *Signal Process.* **92**, 1745–1757. (doi:10.1016/j.sigpro.2012.01.013)
- Olson RH, Carr DW. 2004 A digital accelerometer array utilizing suprathreshold stochastic resonance for detection of sub-Brownian noise floor accelerations. Sandia National Laboratories, Sandia Report SAND2004-6441.
- McDonnell MD, Stocks NG, Pearce CEM, Abbott D. 2005 Quantization in the presence of large amplitude threshold noise. *Fluct. Noise Lett.* **5**, L457–L468. (doi:10.1142/S0219477505002884)
- McDonnell MD, Abbott D. 2004 Signal reconstruction via noise through a system of parallel threshold nonlinearities. In *IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Montreal, Canada, 17–21 May*, vol. 2, pp. 809–812. IEEE. (doi:10.1109/ICASSP.2004.1326381)
- McDonnell MD, Stocks NG, Pearce CEM, Abbott D. 2008 *Stochastic resonance: from suprathreshold stochastic resonance to stochastic signal quantization*. Cambridge, UK: Cambridge University Press.
- Xu L, Vladusich T, Duan F, Gunn LJ, Abbott D, McDonnell MD. 2015 Decoding suprathreshold stochastic resonance with optimal weights. *Phys. Lett. A* **379**, 2277–2283. (doi:10.1016/j.physleta.2015.05.032)
- Xu L, Duan F, Abbott D, McDonnell MD. 2016 Optimal weighted suprathreshold stochastic resonance with multigroup saturating sensors. *Phys. A* **457**, 348–355. (doi:10.1016/j.physa.2016.03.064)
- Hänggi P, Jung P, Zerbe C, Moss F. 1993 Can colored noise improve stochastic resonance? *J. Stat. Phys.* **70**, 25–47. (doi:10.1007/BF01053952)
- Widrow B, Glover JR, McCool JM, Kaunitz J, Williams CS, Hearn RH, Zeidler JR, Dong JRE, Goodlin RC. 1975 Adaptive noise cancelling: principles and applications. *Proc. IEEE* **63**, 1692–1716. (doi:10.1109/PROC.1975.10036)
- Widrow B, Stearns SD. 1985 *Adaptive signal processing*. Englewood Cliffs, NJ: Prentice-Hall.
- Sayed AH, Kailath T. 1994 A state-space approach to adaptive RLS filtering. *IEEE Signal Process. Mag.* **11**, 18–60. (doi:10.1109/79.295229)
- Douglas SC. 1995 Generalized gradient adaptive step sizes for stochastic gradient adaptive filters. In *Int. Conf. Acoustics, Speech, and Signal Processing, 1995 (ICASSP-95), Detroit, MI, 9–12 May*, vol. 2, pp. 1396–1399. IEEE.
- Haykin SS. 2008 *Adaptive filter theory*. Upper Saddle River, NJ: Pearson Education.
- Widrow B, McCool JM, Larimore MG, Johnson CR. 1976 Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proc. IEEE* **64**, 1151–1162. (doi:10.1109/PROC.1976.10286)
- Widrow B, Walach E. 1984 On the statistical efficiency of the LMS algorithm with nonstationary inputs. *IEEE Trans. Info. Theory* **30**, 211–221. (doi:10.1109/TIT.1984.1056892)
- Simon D. 2006 *Optimal state estimation: Kalman, H ∞ and nonlinear approaches*. New York, NY: John Wiley and Sons.
- McDonnell MD, Grant AJ, Land I, Vellambi BN, Abbott D, Lever K. 2011 Gain from the two-envelope problem via information asymmetry: on the suboptimality of randomized switching. *Proc.*

- R. Soc. A* **467**, 2825–2851. (doi:10.1098/rspa.2010.0541)
29. Mandic DP, Kanna S, Constantinides AG. 2015 On the intrinsic relationship between the least mean square and Kalman filters. *IEEE Signal Process. Mag.* **32**, 117–122. (doi:10.1109/MSP.2015.2461733)
30. Zhang L, Song A, He J. 2010 Effect of colored noise on logical stochastic resonance in bistable dynamics. *Phys. Rev. E* **82**, 051106. (doi:10.1103/PhysRevE.82.051106)
31. Ma J, Xiao T, Hou Z, Xin H. 2008 Coherence resonance induced by colored noise near Hopf bifurcation. *Chaos* **18**, 043116. (doi:10.1063/1.3013178)
32. Bordet M, Morfu S. 2013 Experimental and numerical study of noise effects in a Fitzhugh–Nagumo system driven by a biharmonic signal. *Chaos Soliton. Fract.* **54**, 82–89. (doi:10.1016/j.chaos.2013.05.020)
33. Zeng L, Li J, Shi J. 2012 M-ary signal detection via a bistable system in the presence of Lévy noise. *Chaos. Soliton. Fract.* **45**, 378–382. (doi:10.1016/j.chaos.2011.10.012)
34. Nozaki D, Mar DJ, Grigg P, Collins JJ. 1999 Effects of colored noise on stochastic resonance in sensory neurons. *Phys. Rev. Lett.* **82**, 2402–2405. (doi:10.1103/PhysRevLett.82.2402)
35. Xu B, Li J, Duan F, Zheng J. 2003 Effects of colored noise on multi-frequency signal processing via stochastic resonance with tuning system parameters. *Chaos Soliton. Fract.* **16**, 93–106. (doi:10.1016/S0960-0779(02)00201-1)
36. Duan F, Chapeau-Blondeau F, Abbott D. 2014 Stochastic resonance with colored noise for neural signal detection. *PLoS ONE* **9**, e91345. (doi:10.1371/journal.pone.0091345)
37. Duan F, Chapeau-Blondeau F, Abbott D. 2016 Capacity of very noisy communication channels based on Fisher information. *Sci. Rep.* **6**, 27946. (doi:10.1038/srep27946)
38. Collins JJ, Priplata AA, Gravelle DC, Niemi J, Harry J, Lipsitz LA. 2003 Noise-enhanced human sensorimotor function. *IEEE Eng. Med. Biol. Mag.* **22**, 76–83. (doi:10.1109/MEMB.2003.1195700)