Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Raising the noise to improve performance in optimal processing

François Chapeau-Blondeau and David Rousseau

Laboratoire d'Ingénierie des Systèmes Automatisés (LISA), Université d'Angers, 62 avenue Notre Dame du Lac, F-49000 Angers, France
E-mail: chapeau@univ-angers.fr and david.rousseau@univ-angers.fr

Received 10 June 2008
Accepted 1 August 2008
Published 5 January 2009

Online at stacks.iop.org/JSTAT/2009/P01003
doi:10.1088/1742-5468/2009/01/P01003

Abstract. We formulate, in general terms, the classical theory of optimal detection and optimal estimation of signal in noise. In this framework, we exhibit specific examples of optimal detectors and optimal estimators endowed with a performance which can be improved by injecting more noise. From this proof of feasibility by examples, we suggest a general mechanism by which noise improvement of optimal processing, although seemingly paradoxical, may indeed occur. Beyond specific examples, this leads us to the formulation of open problems concerning the general characterization, including the conditions of formal feasibility and of practical realizability, of such situations of optimal processing improved by noise.

Keywords: analysis of algorithms

J. Stat. Mech. (2009) P01003

 1742-5468/09/P01003+15$30.00

Contents

1. Introduction **2**

2. Optimal detection **3**
- 2.1. Classical theory of optimal detection 3
- 2.2. A classical detection example 4
- 2.3. Beneficial role of noise 5

3. Optimal estimation **8**
- 3.1. Classical theory of optimal estimation 8
- 3.2. Estimation with phase noise 9

4. The basic mechanism **11**

5. Open problems of noise **12**

References **14**

1. Introduction

Signal and information processing very often has to cope with noise. Noise commonly acts as a nuisance. However, specific phenomena such as those related to stochastic resonance, and currently under investigation, tend to show that noise can sometimes play a beneficial role [1]–[3]. Many situations of stochastic resonance or useful-noise effects have been reported for signal and information processing. This included demonstration of the possibilities of improvement by noise in standard signal processing operations like detection [4]–[13] or estimation [14]–[22] of signal in noise. However, most of these studies have focused on improvement by noise of suboptimal signal processors. By contrast, the present paper will focus on optimal processors. Examples of optimal processing improved by noise will be described, so as to exhibit some concrete proofs of feasibility. Next, a general mechanism will be uncovered which explains how improvement by noise of optimal processing can indeed occur, however paradoxical it may seem at first sight. Open questions will then be formulated concerning the general characterization of the optimal processing problems and their solutions, that can take advantage of improvement by noise.

This paper takes place within the classical frameworks of optimal detection and optimal estimation of signal in noise. For self-completeness of the paper, the classical theory of these frameworks will be briefly recalled. This will also serve to explicitly visualize the place of the classical derivations where the (unexpected) possibility of improvement by noise can make its way in. We will be considering a general optimal processing situation under the classical form as follows: an input signal $s(t)$ is coupled to a random noise $\xi(t)$ by some physical process, so as to produce an observable signal $x(t)$. At N distinct times t_k which are given, N observations are collected $x(t_k) = x_k$, for $k = 1$ to N. From the N observations $(x_1, \ldots, x_N) = \boldsymbol{x}$, one wants to perform, about the input signal $s(t)$, some inference that would be optimal in the sense of a meaningful criterion of performance.

J. Stat. Mech. (2009) P01003

2. Optimal detection

2.1. Classical theory of optimal detection

As an embodiment of the general situation of section 1, we consider a standard two-hypotheses detection problem, where the input signal $s(t)$ can be any one of two known signals, i.e. $s(t) \equiv s_0(t)$ with known prior probability P_0 or $s(t) \equiv s_1(t)$ with prior probability $P_1 = 1 - P_0$. The input signal $s(t)$ is mixed in some way with the 'corrupting' noise $\xi(t)$ to yield the observable signal $x(t)$. From the observations $\boldsymbol{x} = (x_1, \ldots, x_N)$ one has then to detect whether $s(t) \equiv s_0(t)$ (hypothesis H_0) or $s(t) \equiv s_1(t)$ (hypothesis H_1) holds.

Following classical detection theory [23, 24], any detection procedure can be formalized by specifying that the detector will decide $s(t) \equiv s_0(t)$ whenever the data $\boldsymbol{x} = (x_1, \ldots, x_N)$ falls in the region $\mathcal{R}_0$ of $\mathbb{R}^N$, and it will decide $s(t) \equiv s_1(t)$ when $\boldsymbol{x}$ falls in the complementary region $\mathcal{R}_1$ of $\mathbb{R}^N$. In this context, a meaningful criterion of performance is (other criteria of Neyman–Pearson or minimax types are also possible) the probability of detection error $P_{\mathrm{er}} = \Pr\{s_1 \,\text{decided}|\mathrm{H}_0 \,\text{true}\}P_0 + \Pr\{s_0 \,\text{decided}|\mathrm{H}_1 \,\text{true}\}P_1$, also expressible as

$$P_{\mathrm{er}} = P_1 \int_{\mathcal{R}_0} p(\boldsymbol{x}|\mathrm{H}_1)\,\mathrm{d}\boldsymbol{x} + P_0 \int_{\mathcal{R}_1} p(\boldsymbol{x}|\mathrm{H}_0)\,\mathrm{d}\boldsymbol{x}, \tag{1}$$

where $p(\boldsymbol{x}|\mathrm{H}_j)$ is the probability density for observing $\boldsymbol{x}$ when hypothesis H_j holds, with $j \in \{0, 1\}$, and the notation $\int .\,\mathrm{d}\boldsymbol{x}$ stands for the N-dimensional integral $\int \cdots \int .\,\mathrm{d}x_1 \cdots \mathrm{d}x_N$.

Since $\mathcal{R}_0$ and $\mathcal{R}_1$ are complementary in $\mathbb{R}^N$, one has

$$\int_{\mathcal{R}_0} p(\boldsymbol{x}|\mathrm{H}_1)\,\mathrm{d}\boldsymbol{x} = 1 - \int_{\mathcal{R}_1} p(\boldsymbol{x}|\mathrm{H}_1)\,\mathrm{d}\boldsymbol{x}, \tag{2}$$

which, substituted in equation (1), yields

$$P_{\mathrm{er}} = P_1 + \int_{\mathcal{R}_1} [P_0 p(\boldsymbol{x}|\mathrm{H}_0) - P_1 p(\boldsymbol{x}|\mathrm{H}_1)]\,\mathrm{d}\boldsymbol{x}. \tag{3}$$

Following classical detection theory [23, 24], the detector that minimizes P_{er} can be obtained by making the integral over $\mathcal{R}_1$ on the right-hand side of equation (3) the more negative possible. This is realized by including in $\mathcal{R}_1$ all and only those points $\boldsymbol{x}$ for which the integrand $P_0 p(\boldsymbol{x}|\mathrm{H}_0) - P_1 p(\boldsymbol{x}|\mathrm{H}_1)$ is negative. This yields the optimal detector, which tests the likelihood ratio $L(\boldsymbol{x}) = p(\boldsymbol{x}|\mathrm{H}_1)/p(\boldsymbol{x}|\mathrm{H}_0)$ according to

$$L(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathrm{H}_1)}{p(\boldsymbol{x}|\mathrm{H}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{P_0}{P_1}. \tag{4}$$

When the decision regions $\mathcal{R}_0$ and $\mathcal{R}_1$ are defined according to the optimal test of equation (4), then on the right-hand side of P_{er} in equation (1), the two quantities to be integrated over $\mathcal{R}_0$ or $\mathcal{R}_1$ can be uniformly expressed, simultaneously over $\mathcal{R}_0$ and $\mathcal{R}_1$, as $\min[P_0 p(\boldsymbol{x}|\mathrm{H}_0), P_1 p(\boldsymbol{x}|\mathrm{H}_1)]$. It results that the minimal P_{er} reached by the optimal detector of equation (4) is expressible as

$$P_{\mathrm{er}}^{\min} = \int_{\mathbb{R}^N} \min[P_0 p(\boldsymbol{x}|\mathrm{H}_0), P_1 p(\boldsymbol{x}|\mathrm{H}_1)]\,\mathrm{d}\boldsymbol{x}. \tag{5}$$

J. Stat. Mech. (2009) P01003

Since $\min(a, b) = (a + b - |a - b|)/2$, the minimal probability of error of equation (5) reduces to

$$P_{\mathrm{er}}^{\min} = \tfrac{1}{2} - \tfrac{1}{2} \int_{\mathbb{R}^N} |P_1 p(\boldsymbol{x}|\mathrm{H}_1) - P_0 p(\boldsymbol{x}|\mathrm{H}_0)| \, \mathrm{d}\boldsymbol{x}. \tag{6}$$

The classical theory of optimal detection, as reviewed in this section 2.1, thus specifies, through equation (4), the best exploitation of the data $\boldsymbol{x}$ in order to reach the minimal probability of detection error $P_{\mathrm{er}}^{\min}$ given by equation (6). This theory is general in the sense that it applies for the detection of any two arbitrary known signals $s(t) \equiv s_0(t)$ and $s(t) \equiv s_1(t)$, mixed in any way, to any definite noise $\xi(t)$. The specificity of each problem is essentially coded in the two conditional probability densities $p(\boldsymbol{x}|\mathrm{H}_j)$, for $j \in \{0, 1\}$, which express the probabilization induced by the noise $\xi(t)$ once defined.

2.2. A classical detection example

For an application of the optimal detection procedure of section 2.1, we now consider that the signal–noise mixture $x(t)$ is the additive mixture

$$x(t) = s(t) + \xi(t), \tag{7}$$

with $\xi(t)$ a stationary white noise of cumulative distribution function $F_\xi(u)$ and probability density function $f_\xi(u) = \mathrm{d}F_\xi/\mathrm{d}u$. The level of the noise $\xi(t)$ is quantified by its root mean squared (rms) amplitude σ. The white noise assumption here means that, at any distinct observation times t_k, the noise samples $\xi(t_k)$, and consequently the observations $x_k = x(t_k)$, are statistically independent. It then follows that the conditional densities factorize as $p(\boldsymbol{x}|\mathrm{H}_j) = \prod_{k=1}^{N} p(x_k|\mathrm{H}_j)$, with

$$p(x_k|\mathrm{H}_j) = f_\xi[x_k - s_j(t_k)], \tag{8}$$

for $j \in \{0, 1\}$. We further consider the simple situation where the signals to be detected are the constant signals $s_0(t) = s_0$ and $s_1(t) = s_1$, for all t, with two constants $s_0 < s_1$.

In the common case where the white noise $\xi(t)$ in equation (7) is zero-mean Gaussian, it is well known that the optimal detector of equation (4) reduces to

$$\frac{1}{N} \sum_{k=1}^{N} x_k \underset{H_0}{\overset{H_1}{\gtrless}} \frac{s_0 + s_1}{2} + \frac{\sigma^2/N}{s_1 - s_0} \ln\left(\frac{P_0}{P_1}\right) = x_T. \tag{9}$$

This optimal test of equation (9) achieves the probability of error of equation (6) which is also

$$P_{\mathrm{er}}^{\min} = \frac{1}{2}\left[1 + P_1 \,\mathrm{erf}\left(\sqrt{N}\frac{x_T - s_1}{\sqrt{2}\sigma}\right) - P_0 \,\mathrm{erf}\left(\sqrt{N}\frac{x_T - s_0}{\sqrt{2}\sigma}\right)\right]. \tag{10}$$

It is easy to verify that this minimal probability of detection error $P_{\mathrm{er}}^{\min}$ of equation (10) monotonically increases when the noise level σ increases. This is depicted in some illustrative conditions by figure 1.

Figure 1 illustrates a common behavior which can be intuitively expected: the performance $P_{\mathrm{er}}^{\min}$ of the optimal detector monotonically degrades as the noise level σ increases. Such an expectation matches the *a priori* intuition that noise usually has a detrimental effect on information processing. However, this may not be the rule in general, and we show next that improvement by noise can sometimes apply to optimal detectors.

J. Stat. Mech. (2009) P01003

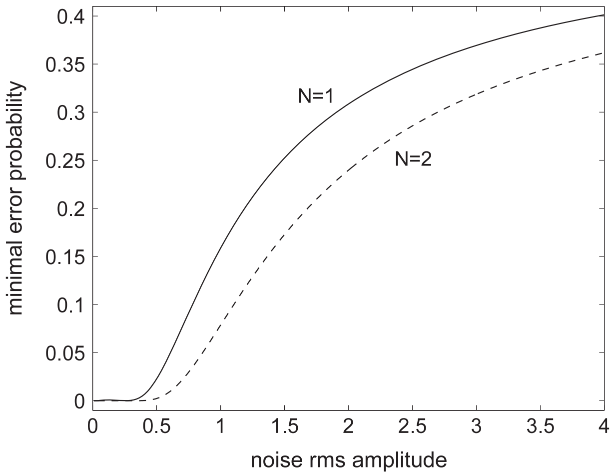

Figure 1. Minimal probability of error $P_{\rm er}^{\rm min}$ of equation (10) for the optimal detector of equation (9), as a function of the rms amplitude σ of the zero-mean Gaussian noise $\xi(t)$ with $N=1$ and $N=2$ data points. Also, $s_0(t) \equiv s_0 = -1$, $s_1(t) \equiv s_1 = 1$ and $P_0 = 1/2$.

2.3. Beneficial role of noise

We now turn for the additive white noise $\xi(t)$ in equation (7) to a non-Gaussian case by way of the family of zero-mean Gaussian mixture with standardized probability density (with $0 < m < 1$):

$$f_{\rm gm}(u) = \frac{1}{2\sqrt{2\pi}\sqrt{1-m^2}} \left\{ \exp\left[-\frac{(u+m)^2}{2(1-m^2)}\right] + \exp\left[-\frac{(u-m)^2}{2(1-m^2)}\right] \right\}, \quad (11)$$

and cumulative distribution function:

$$F_{\rm gm}(u) = \frac{1}{2} + \frac{1}{4}\left[\mathrm{erf}\left(\frac{u+m}{\sqrt{2}\sqrt{1-m^2}}\right) + \mathrm{erf}\left(\frac{u-m}{\sqrt{2}\sqrt{1-m^2}}\right)\right]. \quad (12)$$

As $m \to 0$, equation (11) approaches the zero-mean unit-variance Gaussian density; as $m \to 1$, equation (11) approaches the zero-mean unit-variance dichotomic density at ± 1, as in [25]. We consider for $\xi(t)$ the density $f_\xi(u) = f_{\rm gm}(u/\sigma)/\sigma$ which is a zero-mean Gaussian-mixture density with standard deviation σ. This density $f_\xi(u)$ is plugged into equation (8), and then via equation (6) it yields the performance $P_{\rm er}^{\rm min}$ of the optimal detection with Gaussian-mixture noise. Figure 2 represents different evolutions of the performance $P_{\rm er}^{\rm min}$ in equation (6) of the optimal detector as the noise rms amplitude σ increases.

Figure 2 exhibits the possibility, as also found in [26], of nonmonotonic evolutions of the performance $P_{\rm er}^{\rm min}$ of the optimal detector, as the level σ of the Gaussian-mixture noise is raised. When the noise level σ starts to rise above zero in figure 2, the probability of error $P_{\rm er}^{\rm min}$ starts to gradually degrade (to increase), manifesting here a detrimental action of the noise. However, this degradation of $P_{\rm er}^{\rm min}$ does not always proceed monotonically as σ is further increased. Conditions exist in figure 2, where the probability of error $P_{\rm er}^{\rm min}$ improves (decreases) when the noise level σ is further raised over some ranges. At even

J. Stat. Mech. (2009) P01003

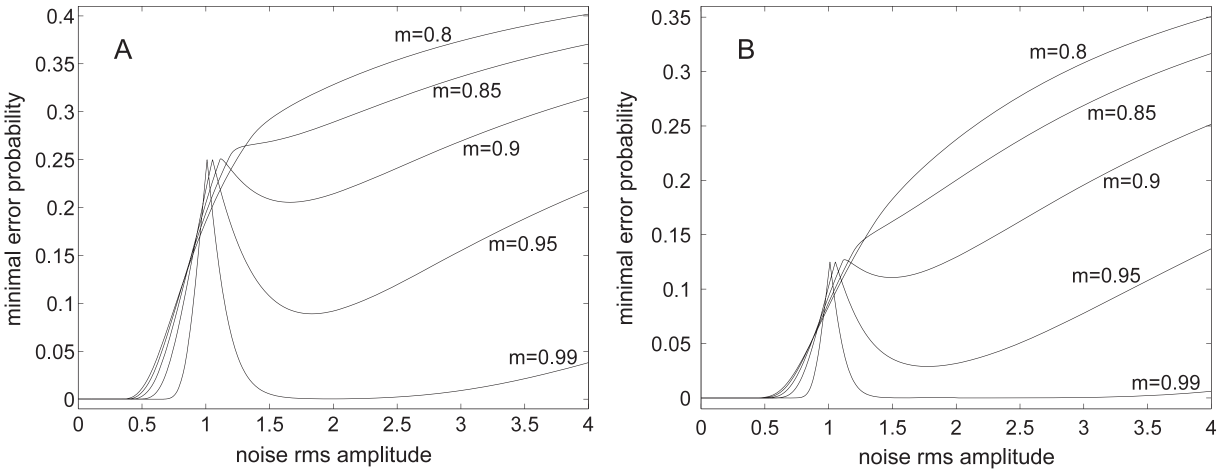

Figure 2. Minimal probability of error $P_{\text{er}}^{\min}$ of equation (6) for the optimal detector of equation (4), as a function of the rms amplitude σ of the zero-mean Gaussian-mixture noise $\xi(t)$ from equation (11) at different m. Also, $s_0(t) \equiv s_0 = -1$, $s_1(t) \equiv s_1 = 1$ and $P_0 = 1/2$; $N = 1$ (panel A) or $N = 2$ (panel B).

larger levels of σ, the detrimental action of the noise resumes and $P_{\text{er}}^{\min}$ degrades again by increasing towards the least favorable value of $\min(P_0, P_1)$ which is $1/2$ in figure 2.

The results of figure 2 demonstrate by example that the performance $P_{\text{er}}^{\min}$ of an optimal detector does not necessarily degrade as the noise level increases. On the contrary, figure 2 shows conditions where, for an optimal detector initially operating at a noise level $\sigma \approx 1$, the optimal performance $P_{\text{er}}^{\min}$ improves if the optimal detector is taken to operate at a higher noise level $\sigma \approx 1.5$. In the example of figure 2, the beneficial action of noise occurs when the noise $\xi(t)$ departs sufficiently from a Gaussian noise, i.e. when m in equation (11) is sufficiently close to 1. In contrast, values of m approaching zero in figure 2 lead to the Gaussian case of figure 1, where increase of the noise level σ monotonically degrades the performance $P_{\text{er}}^{\min}$. Qualitatively, it can be realized that the non-Gaussian density $f_\xi(\cdot)$ at m close to 1 has two peaks which make the two noisy constants s_0 and s_1 more distinguishable as the noise level σ is raised over some range, as depicted in figure 3. Quantitatively, this translates into the improvement by noise of the performance $P_{\text{er}}^{\min}$ in optimal detection, as visible in figure 2.

From a practical point of view, if one wants to take advantage of a beneficial increase in the noise level as it exists in figure 2, one cannot increase the noise level simply by adding an independent white noise $\eta(t)$ to the observation signal $x(t)$ of equation (7) so as to realize $x(t) + \eta(t) = s(t) + \xi(t) + \eta(t)$. In this way, the probability density of the augmented noise $\xi(t) + \eta(t)$ would no longer follow the initial non-Gaussian density $f_\xi(\cdot)$. The process would no longer adhere to the conditions of figure 2 which assume an invariant non-Gaussian density as the noise level is raised. The theoretical analysis of figure 2 should be extended to replace the initial density $f_\xi(\cdot)$ by the composite convolved density $f_\xi(\cdot) * f_\eta(\cdot)$ as the noise $\eta(t)$ with density $f_\eta(\cdot)$ is added. An alternative though, to raise the noise $\xi(t)$ while adhering to the conditions of figure 2, is to assume the possibility of a more internal physical parameter, like a temperature, which would allow to increase

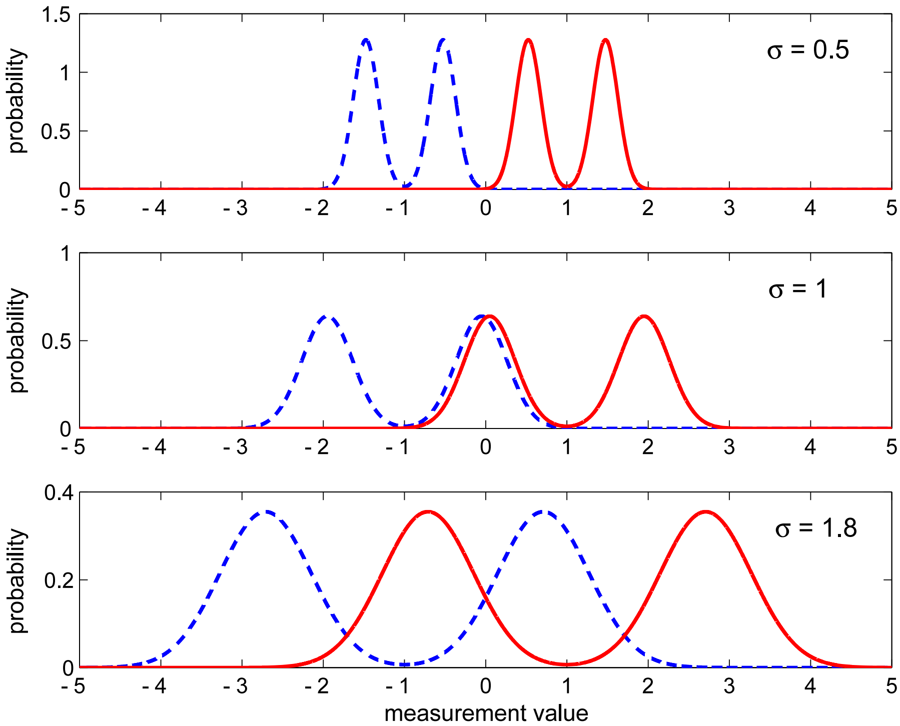

Figure 3. For a measurement x_k, probability densities from equation (8) under both hypotheses: $p(x_k|\mathrm{H}_0) = f_\xi(x_k - s_0)$ (blue dashed line) and $p(x_k|\mathrm{H}_1) = f_\xi(x_k - s_1)$ (red solid line). The noise probability density $f_\xi(u) = f_{\mathrm{gm}}(u/\sigma)/\sigma$ is the zero-mean Gaussian mixture from equation (11) at $m = 0.95$ with standard deviation σ. The signals to be detected are the constant $s_0(t) \equiv s_0 = -1$ and $s_1(t) \equiv s_1 = 1$; and $P_0 = 1/2$. At $\sigma = 1$ (middle panel), the densities $p(x_k|\mathrm{H}_0)$ and $p(x_k|\mathrm{H}_1)$ have a relatively strong overlap, and consequently, based on the measurement x_k, the signals $s_0 = -1$ and $s_1 = 1$ are more likely to be confused. Henceforth, this noise level $\sigma = 1$ is associated with a large probability of detection error $P_{\mathrm{er}}^{\mathrm{min}}$ in figure 2. By contrast, at $\sigma = 0.5$ (upper panel) and at $\sigma = 1.8$ (lower panel), the densities $p(x_k|\mathrm{H}_0)$ and $p(x_k|\mathrm{H}_1)$ have smaller overlap, and consequently, based on the measurement x_k, the signals $s_0 = -1$ and $s_1 = 1$ are less likely to be confused. This is associated with a smaller probability of detection error $P_{\mathrm{er}}^{\mathrm{min}}$ in figure 2 at these noise levels $\sigma = 0.5$ and 1.8. This illustrates the nonmonotonic action of an increase of the noise level σ on the performance $P_{\mathrm{er}}^{\mathrm{min}}$.

the noise level σ while maintaining the non-Gaussian density $f_\xi(\cdot)$ invariant in shape. The noise $\xi(t)$ at the level σ where it is, is certainly ruled by some definite physical process fixing σ, and a control is assumed in this process allowing us to raise σ. Also, to complement this practical perspective, noise bearing some similarity with the bimodal noise of this section 2.3 could be found in practice with a 'logical' noise formed as follows. A logical device or a random telegraphic signal would randomly switch between two fixed values coding the logical states 0 and 1; in addition, each of these two states would be corrupted by an additive Gaussian noise. The result then would be a bimodal noise with two Gaussian peaks, as could exist in the environment of logical or telegraphic devices.

Aside from these practical aspects and the specific mechanism depicted in figure 3, the main message we want to retain here from the results of figure 2 is in principle: the performance of an optimal detector can sometimes improve when the noise level increases. This property is demonstrated by an example in this section 2.3, which involves an additive signal–noise mixture with non-Gaussian noise. Other examples further demonstrate the same property for detection on non-additive signal–noise mixture [27, 28] with Gaussian noise [29]. Later in this paper, beyond demonstration by examples, we will address the open problem of a general characterization of such optimal detection tasks which can benefit from an increase in the noise. Before, we show next that improvement by noise of optimal processing can also be observed in optimal estimation.

3. Optimal estimation

3.1. Classical theory of optimal estimation

Another embodiment of the general situation of section 1 is a standard parameter estimation problem, where the input signal $s(t)$ is dependent upon an unknown parameter ν, i.e. $s(t) \equiv s_\nu(t)$. The input signal $s_\nu(t)$ is mixed to the noise $\xi(t)$ to yield the observable signal $x(t)$. From the observations $\boldsymbol{x} = (x_1, \ldots, x_N)$ one has then to estimate a value $\widehat{\nu}(\boldsymbol{x})$ for the unknown parameter. In this context [23, 30], a meaningful criterion of performance can be the rms estimation error

$$\mathcal{E} = \sqrt{\mathrm{E}\{[\widehat{\nu}(\boldsymbol{x}) - \nu]^2\}}. \tag{13}$$

When ν is a deterministic unknown parameter, the random noise $\xi(t)$ mixed to the input signal $s_\nu(t)$ induces a probability density $p(\boldsymbol{x};\nu)$ for the data $\boldsymbol{x}$. The expectation $\mathrm{E}(\cdot)$, defining in equation (13) the rms estimation error $\mathcal{E}$ of estimator $\widehat{\nu}(\boldsymbol{x})$, then comes out as

$$\mathcal{E} = \sqrt{\int_{\mathbb{R}^N} [\widehat{\nu}(\boldsymbol{x}) - \nu]^2 \; p(\boldsymbol{x};\nu)\,\mathrm{d}\boldsymbol{x}}. \tag{14}$$

An estimator with interesting properties is the maximum likelihood estimator defined as [23, 30]

$$\widehat{\nu}_{\mathrm{ML}}(\boldsymbol{x}) = \arg\max_{\nu}\; p(\boldsymbol{x};\nu). \tag{15}$$

In the asymptotic regime $N \to \infty$ of a large data set, the maximum likelihood estimator $\widehat{\nu}_{\mathrm{ML}}(\boldsymbol{x})$ is the optimal estimator that minimizes the rms estimation error of equation (14), achieving a minimal rms error expressible as

$$\mathcal{E}_{\min} = \sqrt{\frac{1}{J(\boldsymbol{x})}}, \tag{16}$$

where $J(\boldsymbol{x})$ is the Fisher information contained in the data $\boldsymbol{x}$ about the unknown parameter ν and is defined as

$$J(\boldsymbol{x}) = \int_{\mathbb{R}^N} \frac{1}{p(\boldsymbol{x};\nu)} \left[\frac{\partial}{\partial\nu} p(\boldsymbol{x};\nu)\right]^2 \,\mathrm{d}\boldsymbol{x}. \tag{17}$$

The classical theory of optimal estimation, as reviewed in this section 3.1, applies for parameter estimation on any parametric signal $s_\nu(t)$, mixed in any way, to any definite noise $\xi(t)$. The specificity of each problem is essentially coded in the probability density $p(\boldsymbol{x};\nu)$, which expresses the probabilization induced by the noise $\xi(t)$ once defined.

J. Stat. Mech. (2009) P01003

3.2. Estimation with phase noise

For an application of the optimal estimation procedure of section 3.1, we now consider our input signal $s_\nu(t)$ under the form of a periodic wave $s_\nu(t) = w(\nu t)$ of unknown frequency ν, where $w(t)$ is a known periodic 'mother' waveform of period unity. The noise $\xi(t)$ acts on the phase of the wave so as to form the observable signal

$$x(t) = w[\nu t + \xi(t)]. \tag{18}$$

Such a periodic signal corrupted by a phase noise will be seen, for instance, by a sensor receiving a periodic wave which traveled through a fluctuating or turbulent propagation medium producing the phase noise. Based on the data $\boldsymbol{x} = (x_1, \ldots, x_N)$ observed on the noisy signal $x(t)$, the frequency ν is to be estimated.

We assume a white noise $\xi(t)$, meaning that at distinct times t_k the noise samples $\xi(t_k)$, and therefore the data $x_k = x(t_k)$, are statistically independent, so that the probability density $p(\boldsymbol{x};\nu)$ factorizes as $p(\boldsymbol{x};\nu) = \prod_{k=1}^N p(x_k;\nu)$. Also, the samples $\xi(t_k)$ are identically distributed, with cumulative distribution function $F_\xi(u)$ and probability density function $f_\xi(u) = \mathrm{d}F_\xi/\mathrm{d}u$. We further consider the simple situation where $w(t)$ is a square wave of period 1 with $w(t) = 1$ when $t \in [0, 1/2)$ and $w(t) = -1$ when $t \in [1/2, 1)$. With $\delta(\cdot)$ the Dirac delta function, we have the density

$$p(x_k;\nu) = \Pr\{x_k = -1;\nu\}\delta(x_k + 1) + \Pr\{x_k = 1;\nu\}\delta(x_k - 1), \tag{19}$$

with the probability

$$\Pr\{x_k = 1;\nu\} = \Pr\{w[\nu t_k + \xi(t_k)] = 1\} \tag{20}$$

$$= \Pr\left\{\nu t_k + \xi(t_k) \in \bigcup_\ell [\ell, \ell + 1/2)\right\} \tag{21}$$

$$= \Pr\left\{\xi(t_k) \in \bigcup_\ell [\ell - \nu t_k, \ell - \nu t_k + 1/2)\right\} \tag{22}$$

$$= \sum_{\ell=-\infty}^{+\infty} \int_{\ell-\nu t_k}^{\ell-\nu t_k+1/2} f_\xi(u)\,\mathrm{d}u \tag{23}$$

$$= \sum_{\ell=-\infty}^{+\infty} [F_\xi(\ell - \nu t_k + 1/2) - F_\xi(\ell - \nu t_k)], \tag{24}$$

ℓ integer, and the probability

$$\Pr\{x_k = -1;\nu\} = 1 - \Pr\{x_k = 1;\nu\}. \tag{25}$$

Under the white noise assumption, Fisher information is additive and one has $J(\boldsymbol{x}) = \sum_{k=1}^N J(x_k)$, with

$$J(x_k) = \sum_{x_k=-1,1} \frac{1}{\Pr\{x_k;\nu\}} \left[\frac{\partial}{\partial\nu} \Pr\{x_k;\nu\}\right]^2 \tag{26}$$

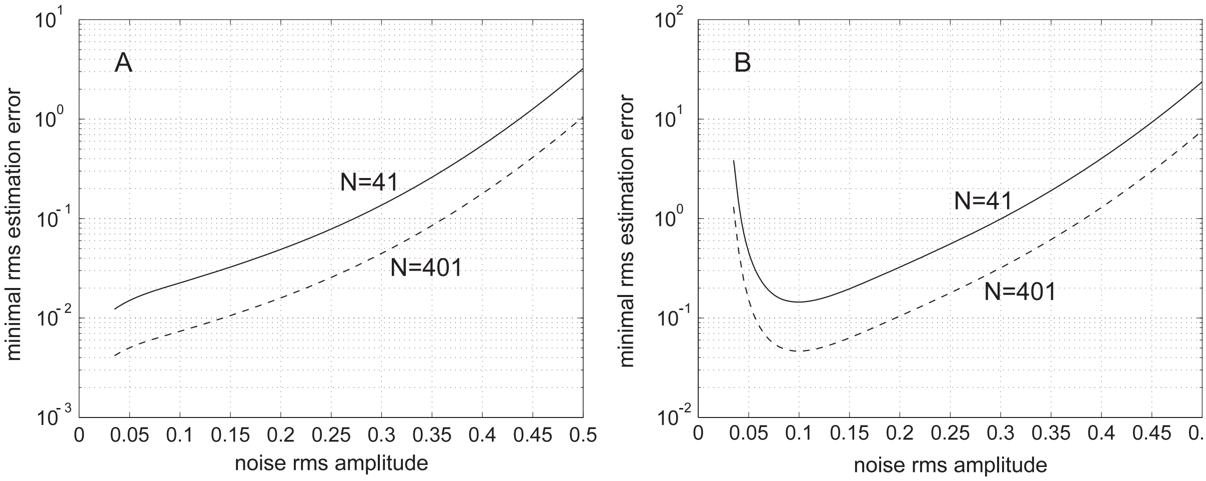

Figure 4. Minimal rms estimation error $\mathcal{E}_{\min}$ from equation (16) for the asymptotically optimal estimator formed by the maximum likelihood estimator of equation (15) at large N as a function of the rms amplitude σ of the zero-mean Gaussian noise $\xi(t)$. Estimation of the frequency $\nu = 1$ of a square wave is performed from N observations at times $\boldsymbol{t} = (t_1, t_2, \ldots, t_N)$ selected between t_1 and t_N with step Δt, which we denote $\boldsymbol{t} = [t_1 : \Delta t : t_N]$. Panel A: $\boldsymbol{t} = [0 : 5 \times 10^{-2} : 2]$ for $N = 41$ (solid line), $\boldsymbol{t} = [0 : 5 \times 10^{-3} : 2]$ for $N = 401$ (dashed line). Panel B: $\boldsymbol{t} = [0.25 : 2.5 \times 10^{-3} : 0.35]$ for $N = 41$ (solid line), $\boldsymbol{t} = [0.25 : 2.5 \times 10^{-4} : 0.35]$ for $N = 401$ (dashed line).

the Fisher information contained in the observation $x_k = x(t_k)$. In addition, from equation (24) one has the derivative

$$\frac{\partial}{\partial \nu} \Pr\{x_k = 1; \nu\} = -t_k \sum_{\ell=-\infty}^{+\infty} [f_\xi(\ell - \nu t_k + 1/2) - f_\xi(\ell - \nu t_k)]. \tag{27}$$

To complete the specification of the problem, we choose $\xi(t)$ as a zero-mean Gaussian noise with standard deviation σ. Figure 4 then represents the evolution of the performance $\mathcal{E}_{\min}$ from equation (16), at large N, in different conditions of estimation.

Figure 4(A) corresponds to a favorable configuration of the N given observation times $(t_1, t_2, \ldots, t_N) = \boldsymbol{t}$ that are well distributed in relation to the period $1/\nu$ of the wave. In this case, the phase noise $\xi(t)$ is felt as a nuisance and the optimal estimation error $\mathcal{E}_{\min}$ monotonically degrades (increases) as the noise level σ increases. By contrast, figure 4(B) corresponds to a less favorable configuration of the N observation times $\boldsymbol{t} = (t_1, t_2, \ldots, t_N)$ that concentrate over a duration less than one period $1/\nu$ of the wave. In this case, qualitatively, the phase noise $\xi(t)$ plays a constructive role as it allows more variability in the values accessible to the data $\boldsymbol{x} = (x_1, \ldots, x_N)$ observed from the noisy signal of equation (18). This is manifested quantitatively in figure 4(B) by an optimal estimation error $\mathcal{E}_{\min}$ experiencing a nonmonotonic evolution as the level σ of the phase noise grows, with ranges where the error $\mathcal{E}_{\min}$ decreases when the noise level σ increases.

The beneficial action in figure 4(B) is obtained with Gaussian noise. This means that the noise level σ can be increased by addition of an independent Gaussian noise $\eta(t)$ to a pre-existing initial Gaussian phase noise $\xi(t)$. This, from equation (18), realizes the observable signal $x(t) = w[\nu t + \xi(t) + \eta(t)]$, with the augmented noise $\xi(t) + \eta(t)$ which remains Gaussian, as in figure 4, while its rms amplitude increases. In practice, for the periodic wave traveling through a fluctuating medium, as evoked just after equation (18), the initial phase noise $\xi(t)$ in equation (18) is due to random fluctuations in the propagating medium. Then, a receiving sensor subjected to random vibrations according to the additional noise $\eta(t)$ will produce the observable signal $x(t) = w[\nu t + \xi(t) + \eta(t)]$. In outline, optimal estimation on a periodic wave traveling through a fluctuating medium could be improved by randomly shaking the receiver in an appropriate way.

The results of this section 3.2 demonstrate, with an example, the possibility of improving the performance of an optimal estimator when the noise level increases. Section 3.2 addresses the estimation of a deterministic unknown parameter ν and it gives new results on noise-aided optimal estimation. A comparable property of improvement by noise in optimal estimation was demonstrated with another example in [31]. Reference [31] addresses estimation of a stochastic unknown parameter ν, with a classic Bayesian estimator minimizing the rms error of equation (13), when the expectation $\mathrm{E}(\cdot)$ in equation (13) is according to the probabilization established in conjunction by the noise $\xi(t)$ and the prior probability on ν.

4. The basic mechanism

Sections 2.3 and 3.2, as well as [27]–[29], [31], report examples of improvement by noise in optimal processing. We are dealing here with optimal processing (optimal detection and optimal estimation) in a classical sense, as defined by classical optimal detection and estimation theories [23, 24, 30]. The measures of performance which are analyzed are standard measures for detection and estimation, i.e. the probability of detection error P_{er} defined in equation (1) and the rms estimation error $\mathcal{E}$ of equation (13). These quantities are measures commonly used for performance evaluation in the classical theories of statistical detection and estimation, and they represent the performance that is reached in practice through ensemble average over a large number of realizations of the random signal $x(t)$ which is optimally processed. The results of figures 2 and 4 are ensemble averages in this sense, and they demonstrate in concrete examples the possibility of improvement by noise of the performance of optimal processors. In practice, as we briefly indicated, the example of figure 2 could be relevant, for instance, to the detection of signals in bimodal 'logical' noise, while the example of figure 4 could be relevant, for instance, to the estimation on periodic waves traveling through fluctuating media. However, our main purpose here is not to argue about the practical usefulness of these examples, but rather we want to focus on their meaning in principle. These examples stand as proofs of feasibility in principle that it is possible to improve the performance of optimal processors by increasing the noise. It may then seem paradoxical that optimal processors, in the classical sense, can be improved by raising the noise. If they are truly optimal, how can they be improved?

The point is that these processors are optimal in the sense that they represent the best possible deterministic processing that can be done on the data $\boldsymbol{x}$ to optimize a fixed given measure of performance Q (for instance, P_{er} of equation (1) or $\mathcal{E}$ of equation (13)). In the classical theory of these optimal processors, the performance Q is a functional of the probabilization established by the initial noise $\xi(t)$. This probabilization is expressed by

$p(\boldsymbol{x}|\mathrm{H}_j)$ in P_{er} of equation (1), and by $p(\boldsymbol{x};\nu)$ in $\mathcal{E}$ of equation (14). In the optimization of the performance Q, this probabilization is kept fixed: $p(\boldsymbol{x}|\mathrm{H}_j)$ and $p(\boldsymbol{x};\nu)$ are fixed given functions of the variable $\boldsymbol{x}$. Classical optimal theory then derives the best deterministic processing of the data $\boldsymbol{x}$ to optimize the performance Q at a level Q_1. This level Q_1 is therefore the best value of the performance that can be achieved by deterministic processing of the data $\boldsymbol{x}$ in the presence of a fixed probabilization of the problem and of the functional Q. What is realized by injection of more noise is a change of this probabilization of the problem. If the probabilization in the functional Q is changed, then the optimal processor, which is now optimal in the presence of the new probabilization of the functional Q, may achieve an improved performance Q_2 strictly better than Q_1. This is what happens in the examples of sections 2.3 and 3.2 (and in [27]–[29], [31]). It is even possible that a suboptimal processor in the sense of Q based on the new probabilization achieves a performance strictly better than the performance Q_1 optimal in the sense of the initial probabilization. An important point is that, even when the probabilization is changed by the injection of noise, it is the same detection or estimation problem which is addressed at the root: which signal $s(t)$ is hidden in the noise? And also, the measure of performance keeps the same physical signification and quantifies the same thing: the fraction of error in detection, or the mean squared difference between the estimate and the true value of the parameter. It is only the functional form of the measure of performance, as a function of the data, which is changed, not the signification of it.

5. Open problems of noise

We have described a general approach through which the possibility of improvement by noise of optimal processing can be analyzed. We have shown two specific examples concretely demonstrating situations of noise-improved optimal processing. We have argued that, at a general level, the basic mechanism possibly authorizing improvement by noise in optimal processing is a change of probabilization of the processing problem. To go further beyond the present proof by examples and the general mechanism we uncovered, an important step is now in making more explicit the favorable changes of probabilization that could possibly lead to improved optimal processing. The favorable changes of probabilization may be specific to any definite processing problems and need be explored separately. There are, however, several general open problems in this direction which can be formulated to serve as guidelines, and those we now discuss.

A favorable change of probabilization, to give way to what can be interpreted as a noise-improved performance, should be a change of probabilization that goes in the direction of raising the noise, something we can call an ‘overprobabilization’. A formal change of probabilization that would only amount to reducing the level of the initial noise $\xi(t)$ would in general trivially lead to an improved performance of the optimal processor. This is apparent with $P_{\mathrm{er}}^{\min}$ of equation (10) which is the best performance achieved by the optimal deterministic detector of a constant signal in Gaussian white noise. This $P_{\mathrm{er}}^{\min}$ is a function of the noise rms amplitude σ assumed fixed in the optimization process (fixed probabilization) leading to the optimal detector of equation (9). In this $P_{\mathrm{er}}^{\min}$ of equation (10), if now σ is reduced (a change in the probabilization), the performance $P_{\mathrm{er}}^{\min}$ of the optimal detector is improved, as can be seen in figure 1. Yet, this is a trivial improvement through a change of probabilization amounting to reducing the initial noise.

The direction which is interesting to explore is the opposite: an improved performance by raising the noise. The possibility thereof is exemplified in sections 2.3 and 3.2.

Beyond these proofs of feasibility by examples in sections 2.3 and 3.2, we are thus led to the following open problem: is it possible to obtain a general characterization of the optimal processing problems and their solutions for which the optimal processors achieve a performance improvable by overprobabilization (a change of probabilization by increasing the noise)?

In this respect, noise-improved optimal processings were obtained in section 2.3 on an *additive* signal–noise mixture with *non-Gaussian* white noise (see figure 2) and in section 3.2 on a *non-additive* signal–noise mixture with *Gaussian* white noise (see figure 4(B)). One is thus led to ask whether general conditions exist concerning the additive/non-additive and Gaussian/non-Gaussian characteristics of the noise, in order to authorize improvement by noise in optimal processing.

Also, an important reference in detection and estimation is provided by the case of an *additive* signal–noise mixture with *Gaussian* white noise. Then another specific question is: is it possible to improve the optimal detection or optimal estimation of a signal in an *additive* signal–noise mixture with *Gaussian* white noise by injecting more noise ? Formally, this would amount to finding a change in the forms of functions $p(\boldsymbol{x}|\mathrm{H}_j)$ in P_{er} of equation (1), or $p(\boldsymbol{x};\nu)$ in $\mathcal{E}$ of equation (14), through an overprobabilization associated with an improvement of the functional measuring the performance. Alternatively, another question is: is there a proof of principle that this is not possible? To answer in one way or the other, a difficulty is that there exists *a priori* a large (infinite) number of possible overprobabilizations which can be considered to change, at least formally, the forms of functions $p(\boldsymbol{x}|\mathrm{H}_j)$ in equation (1) or $p(\boldsymbol{x};\nu)$ in equation (14).

Another issue is to characterize the beneficial overprobabilizations that are compatible with the underlying physics of the problem. Not all formally conceivable changes of probabilization are physically realizable in a given process. This issue of the physical realizability of a beneficial increase of the noise has already been discussed above for both examples of sections 2.3 and 3.2. Usually, inference about the information signal $s(t)$ is performed from the processing of an observation signal $x(t)$ resulting from an arbitrary mixture with the corrupting noise $\xi(t)$. This mixture expresses the underlying physics realizing the signal–noise coupling, and we shall here formally denote this mixture as $x(t) = \mathcal{M}_1[s(t), \xi(t)]$. Equations (7) and (18) are understood here as two examples of this initial mixture operation $\mathcal{M}_1(\cdot)$. Overprobabilization then can be performed by increasing the noise in several ways. A first possibility is as in figure 2, where the initial signal–noise mixture $x(t) = \mathcal{M}_1[s(t), \xi(t) \equiv \xi_1(t)]$ is changed to $x(t) = \mathcal{M}_1[s(t), \xi(t) \equiv \xi_2(t)]$ by increasing the initial noise $\xi(t) \equiv \xi_1(t)$ to a higher level $\xi(t) \equiv \xi_2(t)$. A second possibility for overprobabilization is as in figure 4(B), when another independent noise $\eta(t)$ can be injected into the process to realize a new mixture $\mathcal{M}_2(\cdot)$ of the three ingredients $s(t)$, $\xi(t)$ and $\eta(t)$, yielding the new observation signal $x(t) = \mathcal{M}_2[s(t), \xi(t), \eta(t)]$. The example of figure 4(B) is interpretable as the special case where $\mathcal{M}_2[s(t), \xi(t), \eta(t)] = \mathcal{M}_1[s(t), \xi(t) + \eta(t)]$. The above two possibilities of overprobabilization increase the noise by acting at the level of the underlying physical process that produces the observable signal $x(t)$. Their practical realizability is dependent upon the specific structure of the underlying physical process, and the external control available upon it, to authorize or not the implementation of the intended increase in the noise which has been formally proved beneficial. A third possibility for overprobabilization does not assume action on the underlying physics but directly operates on the initial observation signal $\mathcal{M}_1[s(t), \xi(t)]$ to further mix it, by $\mathcal{M}_3(\cdot)$, to an independent noise $\eta(t)$ to realize the new observable mixture signal $\mathcal{M}_3\{\mathcal{M}_1[s(t), \xi(t)], \eta(t)\}$, with $\mathcal{M}_3(\cdot)$ a general mixing operation certainly not restricted to an additive mixing. Strictly speaking, we did not show any example of this third kind of improvement by noise in optimal processing. The question remains open of whether some exist or not.

When considering improvement of optimal processing by mixing with an external noise $\eta(t)$, it may be helpful to have gone through the following argument. In optimal processing, if a mixing with an external noise $\eta(t)$ is found beneficial, on average, to improve the performance, then there should exist one specific realization of $\eta(t)$ which is especially beneficial. This realization can then be taken as a deterministic set of values which could be mixed with the data to realize a deterministic processing which would improve the optimal performance. But this should not be possible, since no deterministic processing can do better than the initial optimal processor to maximize the performance. The point is that the measure of performance invoked by this argument is the initial measure of performance. Yet the measure of performance now has changed: introduction of the external noise $\eta(t)$ changes the functional form of the measure of performance. And, as explained in section 4, nothing prohibits *a priori* a processing according to this new measure of performance to improve over the optimal processing according to the initial measure of performance.

Returning to open questions, for a given optimal processing problem, ultimately one would like to be able to characterize, when it exists and among those physically realizable, the optimal overprobabilization, i.e. that yielding the best improvement by raising the noise.

For a given optimal processing problem, it appears that questions can be posed at two levels: (i) is it formally possible to find a beneficial overprobabilization of the problem associated with an improvement of the functional measuring the performance? and (ii) Is this overprobabilization formally proved beneficial, physically realizable in practice? These may seem two independent levels, the first one related to the abstract structure of the processing operations on the signals and the second related to the concrete structure of the physical processes generating the signals. A final question arises: are these two levels really independent, or are there connections between these informational and physical levels, limiting what can be ultimately achieved in optimal processing in the presence of noise?

References

[1] Gammaitoni L, Hänggi P, Jung P and Marchesoni F, *Stochastic resonance*, 1998 *Rev. Mod. Phys.* **70** 223

[2] Chapeau-Blondeau F and Rousseau D, *Noise improvements in stochastic resonance: from signal amplification to optimal detection*, 2002 *Fluct. Noise Lett.* **2** L221

[3] McDonnell M D, Stocks N G, Pearce C E M and Abbott D, 2008 *Stochastic Resonance: From Suprathreshold Stochastic Resonance to Stochastic Signal Quantization* (Cambridge: Cambridge University Press)

[4] Inchiosa M E and Bulsara A R, *Signal detection statistics of stochastic resonators*, 1996 *Phys. Rev.* E **53** R2021

[5] Galdi V, Pierro V and Pinto I M, *Evaluation of stochastic-resonance-based detectors of weak harmonic signals in additive white Gaussian noise*, 1998 *Phys. Rev.* E **57** 6470

[6] Tougaard J, *Signal detection theory, detectability and stochastic resonance effects*, 2002 *Biol. Cybernet.* **87** 79

[7] Zozor S and Amblard P O, *On the use of stochastic resonance in sine detection*, 2002 *Signal Process.* **82** 353

[8] Saha A A and Anand G V, *Design of detectors based on stochastic resonance*, 2003 *Signal Process.* **83** 1193

[9] Zozor S and Amblard P O, *Stochastic resonance in locally optimal detectors*, 2003 *IEEE Trans. Signal Process.* **51** 3177

[10] Rousseau D and Chapeau-Blondeau F, *Constructive role of noise in signal detection from parallel arrays of quantizers*, 2005 *Signal Process.* **85** 571

[11] Duan F and Abbott D, *Signal detection for frequency-shift keying via short-time stochastic resonance*, 2005 *Phys. Lett.* A **344** 401

[12] Rousseau D, Anand G V and Chapeau-Blondeau F, *Noise-enhanced nonlinear detector to improve signal detection in non-Gaussian noise*, 2006 *Signal Process.* **86** 3456

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

[14] Greenwood P E, Ward L M and Wefelmeyer W, *Statistical analysis of stochastic resonance in a simple setting*, 1999 *Phys. Rev.* E **60** 4687

[15] Chapeau-Blondeau F, *Noise-aided nonlinear Bayesian estimation*, 2002 *Phys. Rev.* E **66** 032101

[16] McDonnell M D, Abbott D and Pearce C E M, *An analysis of noise enhanced information transmission in an array of comparators*, 2002 *Microelectron. J.* **33** 1079

[17] Rousseau D, Duan F and Chapeau-Blondeau F, *Suprathreshold stochastic resonance and noise-enhanced Fisher information in arrays of threshold devices*, 2003 *Phys. Rev.* E **68** 031107

[18] Greenwood P E, Müller U U and Ward L M, *Soft threshold stochastic resonance*, 2004 *Phys. Rev.* E **70** 051110

[19] Wang Y and Wu L, *Stochastic resonance and noise-enhanced Fisher information*, 2005 *Fluct. Noise Lett.* **5** L435

[20] Chapeau-Blondeau F, Blanchard S and Rousseau D, *Noise-enhanced Fisher information in parallel arrays of sensors with saturation*, 2006 *Phys. Rev.* E **74** 031102

[21] Rousseau D and Chapeau-Blondeau F, *Noise-improved Bayesian estimation with arrays of one-bit quantizers*, 2007 *IEEE Trans. Instrum. Meas.* **56** 2658

[22] Chapeau-Blondeau F, Blanchard S and Rousseau D, *Fisher information and noise-aided power estimation from one-bit quantizers*, 2008 *Digital Signal Process.* **18** 434

[23] Van Trees H L, 2001 *Detection, Estimation, and Modulation Theory, Part 1* (New York: Wiley)

[24] Kay S M, 1998 *Fundamentals of Statistical Signal Processing: Detection Theory* (Englewood Cliffs, NJ: Prentice-Hall)

[25] Barik D, Ghosh P K and Ray D S, *Langevin dynamics with dichotomous noise; direct simulation and applications*, 2006 *J. Stat. Mech.* P03010

[26] Chapeau-Blondeau F and Rousseau D, *Constructive action of additive noise in optimal detection*, 2005 *Int. J. Bifur. Chaos* **15** 2985

[27] Chapeau-Blondeau F, *Stochastic resonance for an optimal detector with phase noise*, 2003 *Signal Process.* **83** 665

[28] Rousseau D and Chapeau-Blondeau F, *Stochastic resonance and improvement by noise in optimal detection strategies*, 2005 *Digital Signal Process.* **15** 19

[29] Chapeau-Blondeau F and Rousseau D, *Injecting noise to improve performance of optimal detector*, 2007 *Electron. Lett.* **43** 897

[30] Kay S M, 1993 *Fundamentals of Statistical Signal Processing: Estimation Theory* (Englewood Cliffs, NJ: Prentice-Hall)

[31] Chapeau-Blondeau F and Rousseau D, *Noise-enhanced performance for an optimal Bayesian estimator*, 2004 *IEEE Trans. Signal Process.* **52** 1327

J. Stat. Mech. (2009) P01003