

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**

Open access books available

**122,000**

International authors and editors

**135M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Nonlinear Filtering of Weak Chaotic Signals

---

Valeri Kontorovich, Zinaida Lovtchikova and  
Fernando Ramos-Alarcon

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.70717>

---

## Abstract

In recent years, the application of nonlinear filtering for processing chaotic signals has become relevant. A common factor in all nonlinear filtering algorithms is that they operate in an instantaneous fashion, that is, at each cycle, a one moment of time magnitude of the signal of interest is processed. This operation regime yields good performance metrics, in terms of mean squared error (MSE) when the signal-to-noise ratio (SNR) is greater than one and shows moderate degradation for SNR values no smaller than  $-3$  dB. Many practical applications require detection for smaller SNR values (weak signals). This chapter presents the theoretical tools and developments that allow nonlinear filtering of weak chaotic signals, avoiding the degradation of the MSE when the SNR is rather small. The innovation introduced through this approach is that the nonlinear filtering becomes multimoment, that is, the influence of more than one moment of time magnitudes is involved in the processing. Some other approaches are also presented.

**Keywords:** nonlinear filtering, chaotic systems, Rossler attractor, Lorenz attractor, Chua attractor, Kalman filter, weak signals, mean squared error

---

## 1. Introduction

The detection of chaotic (stochastic) weak signals is relevant (among others) for applications such as biomedical telemetry [1, 2], seismological signal processing [3], underwater signal processing [4], interference modeling [5], etc. Effective detection of weak and rather weak chaotic signals ( $-3$  dB or less) is a challenge whose solution can improve, for example, the link budget (communication distance). Among different approaches to this problem, one can mention techniques such as stochastic resonance [4], instantaneous spectral cloning [6], etc. The problem in this chapter is addressed from the standpoint of nonlinear filtering techniques which earlier was designed to operate with signal-to-noise ratio (SNR) values bigger than one

or at least rather close to one (with an acceptable slight degradation as the SNR approaches  $-3$  dB [7]). Far down  $-3$  dB, the performance of the available filtering methods drops down sharply and becomes ineffective. One of the possible explanations for this issue is that current nonlinear filtering algorithms can be considered as one moment in the sense that they operate in an instantaneous fashion, that is, during each operation cycle, they process an instantaneous one moment of time magnitude of the received aggregate signal; in the next cycle, a new instantaneous one moment of time magnitude is processed and so on. This is precisely the operation rule for all known optimum algorithms and their quasi-optimum versions as well, for instance, the extended Kalman filter (EKF) [7], but it can also be found in strategies such as unscented Kalman filter (UKF), Gauss-Hermite filter (GHF), and quadrature Kalman filter (QKF), among others. One of the goals of this chapter is to describe the detection of weak chaotic signals applying the principles of noninstantaneous filtering in a block way, that is, multimoment filtering theory [8], through a real-time implementation in a digital signal processing (DSP) block. Moreover, some space of this chapter will be dedicated to the conditionally optimum approach for the nonlinear filtering methods as well, together with some asymptotic methods.

Theoretically, for many cases, the chaos might be represented as an output signal of dissipative continuous dynamic systems (strange attractors) [9]:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t)), \quad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad (1)$$

where  $\mathbf{f}(\bullet) = [f_1(\mathbf{x}), \dots, f_n(\mathbf{x})]^T$  is a differentiable vector function.

According to the idea of Kolmogorov, the equations for strange attractors (1) can be successfully transformed in the equivalent stochastic form as a stochastic differential equation (SDE) [9, 10]:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t)) + \boldsymbol{\varepsilon}\boldsymbol{\xi}(t). \quad (2)$$

The influence of a weak external source of white noise is denoted by  $\boldsymbol{\xi}(t)$ , and the noise intensities are given in a matrix form  $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_{ij}]^{n \times n}$ .

Note that a stationary distribution  $W_{st}(\mathbf{x})$  exists even when the weak white noise component is tending to zero [11–13].

Nonlinear filtering of chaotic desired signals comes up naturally when SDE (2) is used as model of chaos. This follows straight from the classical theory of nonlinear filtering for Markov processes, proposed more than 50 years ago [14, 15] and extensively developed in subsequent studies [16–21], although those methods are still under development.

From the practical implementation point of view, the nonlinear filtering strategies are approximate (see the references above). This follows from the fact that, in general, there is no analytical solution for the a posteriori probability density functions when one attempts solving the Stratonovich-Kushner equations (SKE).

In the following, some of the numerous nonlinear filtering approximate approaches that have been developed will be presented.

## 2. Nonlinear filtering for Markovian processes

Let assume that filtering of the following received signal is required:

$$\mathbf{y}(t) = \mathbf{s}(t, \mathbf{x}(t)) + \mathbf{n}_0(t), \quad (3)$$

where  $\mathbf{s}(\cdot)$  is a vector function of the “message dependent” desired signal (which is subject of filtering) of dimension “ $m$ ,” the received signal is denoted by the vector  $\mathbf{y}(t)$  (also of dimension “ $m$ ”), and  $\mathbf{n}_0$  is a vector of the white additive noises characterized by the intensity matrix  $N_0(m \times m)$ . The following SDE is used to model the signal  $\mathbf{s}(\cdot)$  as an  $n$ -dimensional Markov diffusion process [22]:

$$\dot{\mathbf{x}} = \mathbf{g}(t, \mathbf{x}) + \boldsymbol{\xi}(t). \quad (4)$$

Strictly speaking, Eqs. (4) and (2) are the same SDE, and the vector function  $\mathbf{g}(\cdot)$  substitutes  $\mathbf{f}(\cdot)$  in (2); for (4),  $\mathbf{D}$  denotes the correspondent matrix of intensities for  $\boldsymbol{\xi}(\cdot)$ .

Under this assumption ([14, 22] and so on), one can use the so-called Fokker-Planck-Kolmogorov (FPK) equation in order to solve the a priori probability density function (a priori PDF), for  $\mathbf{x}(t)$ :

$$\frac{\partial W_{PR}(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^n \frac{\partial}{\partial x_i} [g_i(t, \mathbf{x}) W_{PR}(\mathbf{x}, t)] + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial x_i \partial x_j} [D_{ij} W_{PR}(\mathbf{x}, t)], \quad (5)$$

where  $W_{PR}(\mathbf{x}, t_0) = W_0(\mathbf{x})$ .

The Eq. (5) can be rewritten in another form [21, 23] as well:

$$\frac{\partial W_{PR}(\mathbf{x}, t)}{\partial t} = -\text{div} \boldsymbol{\pi}(\mathbf{x}, t), \quad (6)$$

or

$$\frac{\partial W_{PR}(\mathbf{x}, t)}{\partial t} = L_{PR}\{W_{PR}(\mathbf{x}, t)\}, \quad (7)$$

where  $\boldsymbol{\pi}(\mathbf{x}, t)$  is a probabilistic “flow” with components:

$$\boldsymbol{\pi}(\mathbf{x}, t) = g_i(\mathbf{x}, t) W_{PR}(\mathbf{x}, t) - \frac{1}{2} \sum_{j=1}^n \frac{\partial}{\partial x_j} [D_{ij} W_{PR}(\mathbf{x}, t)] . \quad (8)$$

In Eqs. (5)–(8),  $\{D_{ij}\}$  denote diffusion coefficients of the Markov process and  $\{g_i(\mathbf{x}, t)\}_1^n$  are the correspondent drift coefficients, and both of them will be used in the Stratonovich sense [14, 22];  $L_{PR}\{\cdot\}$  denotes a FPK linear operator.

The integrodifferential equation for the a posteriori probability density function  $W_{PS}(\mathbf{x}, t)$  is given by any of the two equivalent expressions (see [14]):

$$\frac{\partial W_{PS}(\mathbf{x}, t)}{\partial t} = L_{PR}\{W_{PS}(\mathbf{x}, t)\} + \frac{1}{2} \left[ F(\mathbf{x}, t) - \int_{-\infty}^{\infty} F(\mathbf{x}, t) W_{PS}(\mathbf{x}, t) d\mathbf{x} \right] W_{PS}(\mathbf{x}, t) \quad (9)$$

or

$$\frac{\partial W_{PS}(\mathbf{x}, t)}{\partial t} = -\text{div} \hat{\boldsymbol{\pi}}(\mathbf{x}, t) + \frac{1}{2} [F(\mathbf{x}, t) - \langle F(\mathbf{x}, t) \rangle] W_{PS}(\mathbf{x}, t), \quad (10)$$

where  $\langle F(\mathbf{x}, t) \rangle$  denotes the averaging of  $F(\mathbf{x}, t)$  given by  $\langle F(\mathbf{x}, t) \rangle = \int_{-\infty}^{\infty} F(\mathbf{x}, t) W_{PS}(\mathbf{x}, t) d\mathbf{x}$   
 $\hat{\boldsymbol{\pi}}(\mathbf{x}, t)$  is (5),  $W_{PR}(\mathbf{x}, t)$  is substituted by  $W_{PS}(\mathbf{x}, t)$ , and

$$F(\mathbf{x}, t) = \left[ \mathbf{y}(t) - \frac{1}{2} s(\mathbf{x}, t) \right]^T N_0^{-1} \left[ \mathbf{y}(t) - \frac{1}{2} s(\mathbf{x}, t) \right]. \quad (11)$$

The combination of Eqs. (9)–(11) is known as the Stratonovich-Kushner nonlinear equations (SKE), and they have an appealing physical sense: the first term in (9) represents the dynamics of the a priori data of  $\mathbf{x}(t)$ . For the second term, the analysis of observations is used to drive the innovation of the a priori data.

Using any optimization criteria, one can get  $\hat{\mathbf{x}}(t)$  (the optimum estimation of  $\mathbf{x}(t)$ ) which comes as a solution of (9), when  $\mathbf{y}(t)$  is the input signal, that is, filtering of  $\mathbf{x}(t)$ .

Here, one has to note that Eq. (9) turns into FPK (6) if the intensity of the additive noises  $N_0$  grows (the first term in (9) is dominant), and as a consequence, the filtering accuracy diminishes drastically. In the opposed scenario (large signal-to-noise ratio), the  $W_{PS}(\mathbf{x}, t)$  tends to the unimodal Gaussian PDF [14, 20].

Note that the time evolution of  $W_{PS}(\mathbf{x}, t)$  is completely described by the SKE but, as it was mentioned earlier, does not provide exact analytical solutions. There are very few exceptions: linear SDE (4) which yields the well-known Kalman filtering algorithm [14–24], the Zakai approach [25], and so on. Due to this, the nonlinear filtering algorithms are practically always approximate. As it was mentioned before, during almost 50 years of intensive research, the bibliography for nonlinear filtering algorithms has become enormous; in the next section, we will consider only few of those works taking into account the following considerations:

- the models applied for filtering of chaos correspond to the equations for Rössler, Chua, and Lorenz strange attractors with  $n = 3$ , that is, low dimensional;
- the algorithms for nonlinear filtering have to be of reduced computational complexity in order to satisfy real-time application requirements;
- the algorithms for nonlinear filtering, according to the aim of the material of the chapter, have to be able to perform satisfactorily in scenarios with low or very low signal-to-noise ratios (SNR), although the Gaussian assumption for  $W_{PS}(\mathbf{x})$  is not always valid;

•

$$s(\mathbf{x}(t), t) \cong \mathbf{x}(t); \quad (12)$$

- All  $D_{ij}$  are equal to zero, except  $D_{11} \cong D_1$  [11].

## 2.1. Approximate approaches for nonlinear filtering

For the sake of simplicity, it is “easier” to approximate the a posteriori PDF  $W_{PS}(\mathbf{x}, t)$  than the nonlinearity at (4) and (9) [16, 17, 19]. In this sense, let us just list some of the approximate approaches for  $W_{PS}(\mathbf{x}, t)$ :

- Integral or global approximations for  $W_{PS}(\mathbf{x}, t)$  [20];
- Functional approximations for  $W_{PS}(\mathbf{x}, t)$  [16, 21];
- Higher Order Statistics (HOS) approximations for  $W_{PS}(\mathbf{x}, t)$ , and so on;
- Gaussian approximations: extended Kalman filter (EKF) [14–24]; unscented Kalman filter (UKF) [19]; quadrature Kalman filter (QKF) [17]; iterated Kalman filter (IKF), etc.

It is hardly feasible to give a complete overview of all those methods; moreover, not all of them are adequate, taking into account the observations introduced at the end of the previous section.

Let us start with the extended Kalman filter (EKF): considering  $W_{PS}(\mathbf{x}, t)$  as a three-dimensional Gaussian PDF-  $\widehat{W}_G(\mathbf{x}, t)$ , from (9), it is possible to obtain the following equations for per-component of the mean estimates  $\{\hat{x}_i\}_1^3$  and for estimates of the elements of the a posteriori covariance matrix  $\{\hat{R}_{ij}\}_{i,j=1}^3$ :

$$\begin{aligned} \hat{x}_i &= \int_{-\infty}^{\infty} \left( \hat{\boldsymbol{\pi}}^T(\mathbf{x}, t) \text{grad } x_i \right) d\mathbf{x} + \frac{1}{2} \left[ \int_{-\infty}^{\infty} x_i \mathbf{F}(\mathbf{x}, t) \widehat{W}_G(\mathbf{x}, t) d\mathbf{x} - \hat{x}_i \int_{-\infty}^{\infty} \mathbf{F}(\mathbf{x}, t) \widehat{W}_G(\mathbf{x}, t) d\mathbf{x} \right] \\ \hat{R}_{ij} &= \int_{-\infty}^{\infty} \left( \hat{\boldsymbol{\pi}}^T(\mathbf{x}, t) \text{grad } \dot{x}_i \dot{x}_j \right) d\mathbf{x} + \frac{1}{2} \left[ \int_{-\infty}^{\infty} \dot{x}_i \dot{x}_j \mathbf{F}(\mathbf{x}, t) \widehat{W}_G(\mathbf{x}, t) d\mathbf{x} - \hat{R}_{ij} \int_{-\infty}^{\infty} \mathbf{F}(\mathbf{x}, t) \widehat{W}_G(\mathbf{x}, t) d\mathbf{x} \right], \end{aligned} \quad (13)$$

where  $\dot{x}_i = x_i - \hat{x}_i$  and  $\dot{x}_j = x_j - \hat{x}_j$ .

The matrix form [14–16, 20] can be used to represent Eq. (13); however, for some specific applications, per-component representation (13) could be more adequate (see the following).

It is reasonable to assume convergence to the stationary values  $\overline{R}_{ij}$  for  $\forall \widehat{R}_{ij}(t)$  when  $t \rightarrow \infty$ , and as a result, the second equation in (13) can be expressed as a system of nonlinear algebraic equations, with standard numerical solutions. This consideration is relevant for real-time scenarios, as it significantly simplifies the implementation of the related EKF algorithms.

Functional approximation for  $W_{PS}(\mathbf{x}, t)$  is, as it was described in [16, 21],

$$W_{PS}(\mathbf{x}, t) = \prod_{i=1}^3 W_{PS}(x_i) \left[ 1 + \sum_{q=2}^3 \sum_{j=1}^{q-1} \frac{R_{qj}}{R_{qq}R_{ji}} (x_q - \hat{x}_q) (x_j - \hat{x}_j) \right]. \quad (14)$$

From (14), we see that the functional approximation for the PDF is sufficiently non-Gaussian (marginal  $W_{PS}(x_i)$  is arbitrary), but for “joint” characterization of the vector  $\hat{\mathbf{x}}$ , only elements of the a posteriori covariance matrix  $\hat{R}_{ij}$  are considered.

It can be shown that the equations for  $\{\widehat{x}_i\}_1^n$  and  $\{\widehat{R}_{ij}\}$  coincide with those in (13), and the unique difference would be that one has to apply in (13) the approximation for  $W_{PS}(\mathbf{x}, t)$  instead of  $\widehat{W}_G(\mathbf{x}, t)$ . The resulting integrals can be solved either through the Gauss-Hermit quadrature formula [17, 18] or analytically.

The integral or Global approximation for  $W_{PS}(\mathbf{x}, t)$  is another approach for approximate solution. Maybe the experienced reader already noticed that the last two approximations for  $W_{PS}(\mathbf{x}, t)$  can be considered as “local” as they offer maximum of  $W_{PS}(\mathbf{x}, t)$ , estimation of  $\{\widehat{x}_i\}$ , and  $\{\widehat{R}_{ij}\}$ .

For conditions of significantly large SNR, this is sufficient, but for low SNR, one has to find a different approach, known as integral approximation. This strategy was suggested as an adequate approximation of  $W_{PS}(\mathbf{x}, t)$  together with the PDF’s “tails,” that is, for the whole span of  $\mathbf{x}$ .

Let us suppose that  $W_{PS}(\mathbf{x}, t)$  can be characterized as:

$$W_{PS}(\mathbf{x}, t) = W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t)). \quad (15)$$

Here  $\boldsymbol{\alpha}$  is an unknown vector of approximation parameters. As an approximation criterion for PDF, it is possible to use the Kullback measure; thus, one might obtain the following equation for the unknown vector  $\boldsymbol{\alpha}$ :

$$\dot{\boldsymbol{\alpha}} = \langle L_{PR}^+ \{ \mathbf{h}(\mathbf{x}, t) \} \rangle + \mathbf{V}^{-1}(t) \langle \mathbf{h}(\mathbf{x}, t) \mathbf{F}(\mathbf{x}, t) \rangle, \quad (16)$$

where  $\mathbf{h}(\mathbf{x}, t) = \frac{\partial \ln W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))}{\partial \boldsymbol{\alpha}}$ ,  $\mathbf{V}(t) = \int_{-\infty}^{\infty} \left[ \frac{\partial \ln W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))}{\partial \boldsymbol{\alpha}} \right]^T W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t)) d\mathbf{x} = \frac{\partial^2 W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T}$ ,  $L_{PR}^+ \{ \bullet \}$  is a self-adjoint operator to the FPK operator [22].

Now, as an integral approximation of  $W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))$ , let us choose the so-called “Dynkin PDF” with  $\boldsymbol{\alpha}(t)$  is the vector of sufficient statistics for  $W_{PS}(\cdot)$ :

$$W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t)) = C \exp \left\{ \sum_{p=1}^K \alpha_p(t) \phi_p(\mathbf{x}) + \phi_0(\mathbf{x}) \right\}, \quad (17)$$

where  $\{\phi_p(\mathbf{x})\}$  are orthogonal multidimensional operators: Laguerre, Hermite, and so on.

One can notice that there is a significant coincidence between (17) and the orthogonal series characterization of  $W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))$  [22]: even though both apply series of orthogonal functions, in (17), it is not used for  $W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))$  but for its monotonical transform  $\ln\{W_{PS}(\mathbf{x}, \boldsymbol{\alpha}(t))\}$ . So, the coefficients  $\{\alpha_p(t)\}$  can be expressed by means of the cumulants of  $W_{PS}(\mathbf{x})$  [22]. Thanks to this, instead of searching for a solution of (17), hardly possible in an analytically way, one can search directly equations for the cumulants (HOS) of  $W_{PS}(\mathbf{x}, t)$  [16, 26].

Here, the HOS approach will be presented because the last problem was addressed in the cited references. It is worth noticing the following: for real-time scenarios when  $n > 1$ , equations for HOS and Eq. (16) are significantly complex; for  $n = 1$ , both strategies are equivalent [26].

### 3. Multimoment filtering of chaos

As it follows from the material of Section 2, all the algorithms are “one-moment” in the sense that they are operating only with the data at each time instant, that is, they are tracking instantaneously one moment magnitude of the received aggregate signal. As it was shown at [27], the adequate filtering algorithm (for the one-moment case) is an Extended Kalman Filter (EKF).

This choice is more or less expected, due to the experience which is already known from the available references (see above). EKF shows rather good performance for the filtering of chaotic signals: the mean squared error (MSE) is less than 1% when SNR is about  $-3$  dB, and for SNR bigger than  $-3$  dB, the results are much better.

In this regard, a question arises: is it possible to improve this approach in the sense of getting still rather good MSE's for successively lower thresholds of the SNR with an algorithm of reasonable complexity? The following material attempts to prove that the answer is “yes,” if one can apply some additional information from the received aggregate signal taken on several sequential time instants.

It means that the information has to be considered in the block manner by aggregating data, in our case, for several time instants ([8, 16, 27], and so on.). The difference between the following approach and that from the cited references is precisely the aggregated data obtained for many time instants: multimoment algorithms are carried out through the generalization of the Stratonovich-Kushner equations (SKE) for the corresponding multimoment data, and therefore, in the following, all heuristics for the simplification, considered as Generalized SKE (GSKE), are not arbitrary but can be taken as generalized heuristics from the “standard” one-moment SKE (see below). This gives a “hope” to achieve the abovementioned improvement for the SNR threshold with less complex tools.

It follows from the fact that, as it was shown in [8] (see also the references therein), the GSKE comes from the same structure as its one-moment prototype. So the way of its simplification (except for the limiting of the number of time instants) in order to get a quasi-optimum algorithm, could be done in a similar way as for the one-moment case: approximation of the a posteriori PDF (characteristic function) in SKE with a minimum set of significant parameters. Moreover, there is an additional way to improve the accuracy of the quasi-optimum solution for the GSKE: assume this quasi-optimum algorithm as a “given structure,” as it was proposed in [16] and also considered in the following.

#### 3.1. Generalization of SKE for the multimoment case

In the same way, as it was underlined earlier, the chaos is “generated” by the equation:

$$\dot{x} = f(x), \quad x \in R^n, \quad x(t_0) = x_0, \quad (18)$$

where  $f(\bullet) = [f_1(x), \dots, f_n(x)]^T$  is a differentiable vector function and it can be considered as a degenerated Markov process from the following stochastic equation:



$$\dot{x} = f(x) + \varepsilon \xi(t), \quad (19)$$

where  $\xi(t)$  is a vector of “weak” external white noise with the related positively defined matrix of “intensities”  $\varepsilon = [\varepsilon_{ij}]^{n \times n}$ .

In the following, one can consider both the ordinary differential equation (ODE) (18) and the stochastic differential equation (SDE) (19) when the noise intensities tend equally to zero. Adding the  $\varepsilon$  term in (19) guarantees the existence of a stationary PDF for  $x(t)$  as well, no matter how small the elements of  $\varepsilon$  might be [28]. So, one can suppose that this stationary PDF,  $W_{ST}(x)$ , is known a priori. For our case in practical sense, one can deal actually only with the stationary PDF, which we assumed is modeled by means of a chaotic process (concretely let us say the first component,  $x_1(t)$ , of certain attractor model). Certainly  $W_{ST}(x_1)$  can easily be obtained from  $W_{ST}(x)$ . If the two PDFs coincide in terms of certain fitness criteria, then only for simplicity in the subsequent developments, the SDE (19) can be substituted by its statistically equivalent one-dimensional SDE with the same  $W_{ST}(x_1)$ :

$$\dot{x}_1 = f(x_1) + \sqrt{\varepsilon} \xi(t), \quad (20)$$

where  $f(x_1) = \frac{\varepsilon}{2} \frac{d}{dx_1} \ln W_{ST}(x_1)$  and  $\varepsilon$  in (20) can be considered here as a “scale factor” and can be chosen by equalizing the average powers of real  $x_1(t)$  and solution of (20). Formally, there is no need for all those operations, but then the reader has to be extremely concentrated with “multiindex” definitions: one index for the number of applied components of the attractor and another index for the time instant, that is,  $x^m(t_i)$ , which might cause confusion in further developments, as  $x_1(t)$  is an observable component whose dynamics depends on other “nonobservable” components. For those reasons, in the following, (20) will be considered as a model of the desired signal for filtering.

Let us introduce the following notation for the time instants (time moments):  $t_1 < t_2 < t_3 \dots < t_n$  and  $x_i = x(t_i)$ ,  $i = \overline{1, n}$ . Then,  $\{x(t_i)\}_1^n$  forms a vector  $x(t) = [x(t_1), \dots, x(t_n)]^T$  and  $W_n(x, t) \cong W_n(x_1, \dots, x_n; t_1, \dots, t_n)$ ;  $W_n(x, t)$  is an a priori PDF for  $x(t)$ . As it follows from ([16], ch. 5):

$$\frac{\partial W_n(x, t)}{\partial t_i} = L_i \{W_n(x, t)\} \quad (21)$$

where  $L_i \{\bullet\} = -\frac{\partial}{\partial x_i} K_1(x_i) + \frac{1}{2} \frac{\partial^2}{\partial x_i^2} K_2(x_i)$  is the FPK operator [16] with  $K_1(x_i) = f_1(x_i)$ ,  $K_2(x_i) = \varepsilon^2$ . It is easy to show that by consecutive differentiation one can obtain:

$$\frac{\partial^n W_n(x, t)}{\partial t_1 \dots \partial t_n} = \prod_{i=1}^n L_i \{W_n(x, t)\}, \quad (22)$$

$$L_{PR} \{\bullet\} = \prod_{i=1}^n L_i \{\bullet\}. \quad (23)$$

Certainly, the adjoint operator [16, 22] for the multimoment case is:

$$\widehat{L}_{PR}\{\bullet\} = \prod_{i=1}^n \widehat{L}_i\{\bullet\}, \quad (24)$$

where  $\widehat{L}_i = K_1(x_i) \frac{\partial}{\partial x_i} + \frac{K_2(x_i)}{2} \frac{\partial^2}{\partial x_i^2}$  is a Kolmogorov operator [22].

Let us then introduce the a posteriori  $n$ -dimensional PDF  $W_{ps}(\mathbf{y}|\mathbf{x},t) \cong W_{ps}(\mathbf{x},t)$  for the multimoment case. Then, repeating formally the development for the SKE, but in this case generalized for the “ $n$ ” time case (in the same way as it was done at [27]), one can get:

$$\frac{\partial^n W_{ps}(\mathbf{x},t)}{\partial t_1 \dots \partial t_n} = L_{PR}\{W_{ps}(\mathbf{x},t)\} + \frac{1}{2} \left[ \mathbf{F}(\mathbf{x},t) - \int_{R^n} \mathbf{F}(\mathbf{x},t) W_{ps}(\mathbf{x},t) d\mathbf{x} \right] W_{ps}(\mathbf{x},t) \quad (25)$$

with  $t = [t_1, \dots, t_n]^T$ ,

$$\mathbf{F}(\mathbf{x},t) = \frac{[\mathbf{y}(t) - \frac{1}{2}\mathbf{x}(t)]^T}{N_0} \left[ \mathbf{y}(t) - \frac{1}{2}\mathbf{x}(t) \right], \quad (26)$$

where  $\mathbf{y}(t) = [y(t_1), \dots, y(t_n)]^T$  is the vector of  $\{x(t_i)\}_1^n$  taken from  $y(t) = x(t) + n(t)$  and  $n(t)$  is the AWGN with intensity  $N_0$ .

Analyzing (25) by comparing it with the standard form of the SKE (see Eqs. (9) and (10) in part II), one can see that there is a total “structural” identity! The same matter takes place for the a posteriori cumulants [16, 27], that is:

$$\frac{\partial \kappa_{r_1, \dots, r_n}^{ps}(t)}{\partial t_1 \dots \partial t_n} = (-j)^k \frac{\partial^k}{\partial \lambda_1^{r_1} \dots \partial \lambda_n^{r_n}} \left\{ \left[ \frac{M\{\widehat{L} \exp\{j\boldsymbol{\lambda}^T \mathbf{x}\} \mathbf{F}(\mathbf{x},t)\}}{M\{\exp\{j\boldsymbol{\lambda}^T \mathbf{x}\}\}} \right]_{\lambda=0} + \left[ \frac{M\{\exp\{j\boldsymbol{\lambda}^T \mathbf{x}\} \mathbf{F}(\mathbf{x},t)\}}{M\{\exp\{j\boldsymbol{\lambda}^T \mathbf{x}\}\}} \right]_{\lambda=0} \right\} \quad (27)$$

where  $r_1 + r_2 + \dots + r_n = k, k = 1, 2, \dots$

One can see from (25) and (26) that those algorithms are rather complex for implementation in real-time regime. So, in addition to the one-moment SKE, they have to be modified in order to get the quasi-optimum solution.

### 3.2. Quasi-optimum solutions. Generalized EKF

One has to know that “quasi-optimum” solutions (for any problem) are based on some heuristics and those heuristics have to be reasonable and based on previous experience in solving similar problems. In the case of multimoment filtering, the analogies can be the following (of course implicit considerations for complexity have always to be taken into account):

- The priority will be given to the quasi-linear approximation for nonlinear functions in the same way as it was assumed for the “standard” one-moment filtering.

- All algorithms for block processing show that there is in some sense a reasonable block length for the processed data. Taking into account the complexity limits and that the covariance function of the chaos initially drops rather fast [29], let us take first  $n = 2$ .
- The approximation of the a posteriori PDF (characteristic function) has to apply the minimum set of first cumulants; one has to remind that, as the order of cumulants grows, their significance for PDF approximation vanishes [22];

Taking these observations into account, let us take  $n = 2$ , that is, two-moment filtering case, then [16, 22]:

$$\theta_{ps}(\boldsymbol{\lambda}) \cong \exp \left\{ \sum_{s=1}^2 \frac{j^s}{s!} \sum_{r_1, r_2}^s \kappa_s(t_1, t_2) \boldsymbol{\lambda}^{r_1} \boldsymbol{\lambda}^{r_2} \right\}, \quad (28)$$

and cumulants are:

$$\kappa_{r_1, \dots, r_n}^{(ps)}(t) = (-j)^k \left[ \frac{\partial^k}{\partial \lambda_1^{r_1} \dots \partial \lambda_n^{r_n}} \ln \theta_{ps}(\boldsymbol{\lambda}) \right]_{\lambda=0}.$$

Another assumption is that the a posteriori process is supposed to be stationary; then, the one-moment cumulants for  $t_1$  and  $t_2$  have to be the same and the only mutual cumulant taken into account might be  $\kappa_{11}(t_1, t_2)$ . Next, for each moment “ $t_1$ ” and “ $t_2$ ” one-moment cumulants can be calculated applying Gaussian approximation for the a posteriori PDF, and for the two-moment case, the “functional approximation” could be applied. In a rigorous sense, the a posteriori variance  $\kappa_2^{ps}$  has to be evaluated as  $\kappa_2^{ps}(t_1, t_2)$ , considering the covariance among time instants “ $t_1$ ” and “ $t_2$ ”; in the following, the heuristic strategy will be introduced, which avoids the cumbersome calculations.

One can obtain the first two-moment cumulants:

$$\begin{aligned} \dot{\kappa}_1 &= \langle K_1(x) \rangle + \frac{1}{2} \langle xF(x, t) \rangle - \frac{\kappa_1}{2} \langle xF(x, t) \rangle \\ \dot{\kappa}_2 &= \langle 2xK_1(x) \rangle - \kappa_1 \langle K_1(x) \rangle + \varepsilon + \frac{1}{2} \langle (x - \kappa_1)^2 F(x, t) \rangle - \frac{\kappa_2}{2} \langle xF(x, t) \rangle, \end{aligned} \quad (29)$$

where  $\langle \rangle$  is a symbol for the averaging procedure,  $F(x, t) = \frac{1}{N_0} [y(t) - \frac{1}{2}x(t)]$ , and  $K_1(x)$  is the drift coefficient for (19).

One has to notice that at (29)  $\kappa_1(t)$  is an estimation of the filtered signal (in our case, it is a chaotic signal);  $\kappa_2(t)$  is a measure of the filtering accuracy. As it can be seen from (29), those equations were written without any intention for linearization, that is, they are presented in a generalized form. For the quasi-linear algorithms, it is well known [27] that  $\kappa_2(t)/N_0$  is the main part for the “averaging coefficient” of the second element in the first quasi-linear equation of (29), that is, it is an averaging value for the instantaneous information actualization from the entering desired signal plus noise. Thus, if one can reduce  $\kappa_2$  through the two-moment processing, the accuracy of the quasi-linear method will grow and the challenge stated before

will be almost solved. To achieve the latter, one can take into account that  $\kappa_2(t)$  in the stationary regime is oscillating around its stationary value  $\overline{\kappa_2(t)} = \lim_{t \rightarrow \infty} \kappa_2(t)$  which is commonly assumed as an accuracy measure in the one-moment case.

The value of  $\overline{\kappa_2(t)}$  can be diminished applying the information from  $\kappa_{11}(t)$  also in the stationary case, that is,  $\overline{\kappa_{11}} = \lim_{t_1, t_2 \rightarrow \infty} \kappa_{11}(t_1, t_2)$ ; then, it is known that  $\widehat{\kappa}_2 = \overline{\kappa_2}(1 - \overline{\kappa_{11}}^2)$  and it is always less than  $\overline{\kappa_2}$ , if and only if the  $\overline{\kappa_{11}} \geq 0$ ; In this way,  $\widehat{\kappa}_2$  can be used as a new weighting coefficient in (29). To find  $\kappa_{11}(t_1, t_2)$  from (27), some cumbersome developments are required which finally yield to:

$$\frac{\partial \kappa_{11}(t_1, t_2)}{\partial t_1 \partial t_2} = \langle K_1(x_1)K_1(x_2) \rangle + \langle \frac{x_1 x_2}{2} F(x, t) \rangle - \frac{\kappa_{11}(t_1, t_2)}{2} \langle F(x, t) \rangle \quad (30)$$

and

$$\overline{\kappa_{11}} = \frac{2 \left[ \langle \frac{x_1 x_2 F(x, t)}{2} + 2K_1(x_1)K_1(x_2) \rangle \right]}{\langle F(x, t) \rangle}. \quad (31)$$

First we would like to stress here that, as we are interested in covariance calculation, it is necessary to preserve the notations  $x(t_1) = x_1$  and  $x(t_2) = x_2$ . Second, we want to “improve” the stationary value  $\overline{\kappa_2}$  evaluated for the one-moment case through its indirect dependence on  $\overline{\kappa_{11}}$  as if it was “evaluated” for the two-moment case.

Thus in doing so, the direct calculation of the quasi-linear algorithm for the two-moment case is bypassed (see (29) and (30)). For applications in real time, the formal calculus is almost impossible. Instead, we simplified it with a formal “ignorance” of the two-moment features. There might be for sure a compromise between the complexity and the improvement attempt for the “classic” EKF.

In order to avoid some additional complexities for the calculation of (31), let us make the following assumption: introduce the SNR of the filtering in the way:  $h^2 = \frac{\overline{\kappa_1}^2}{N_0} \ll 1$ , that is, weak signal case. In this regard [16, 27], the a priori data are the main influence, that is, approximately only  $\langle K_1(x_1) K_1(x_2) \rangle$  can be applied. Or one can simply apply a Gaussian approximation for the second equation in (29) for the stationary regime ( $\dot{\kappa}_2 \cong 0$ )

$$2 \left[ \overline{K_1'(\kappa_1)} + \overline{\kappa_{11}} \right] \frac{\overline{\kappa_2}}{2} + \varepsilon + \frac{1}{4} \overline{F''(\kappa_1)} \overline{\kappa_2}^2 = 0. \quad (32)$$

In the case  $h^2 < 1$ , it is possible to achieve:

$$\overline{\kappa_2} \sim \frac{1}{2 \left[ \overline{K'(\kappa_1)} + \overline{\kappa_{11}} \right]}, \quad (33)$$

and if  $\overline{\kappa_{11}} > 0$ , and  $K'(\kappa_1) \geq 0$ ,  $\kappa_2$  is always reduced compared with the one-moment approach. Formula (33) can be seen as another illustration about the usefulness of the heuristic

approximation proposed above. Then, to evaluate the order of the  $\bar{\kappa}_{11}$ , let us apply for averaging of  $\langle K_1(x_1) K_1(x_2) \rangle$  the functional approximation of  $W_{ps}(x_1, x_2)$  in the way:

$$W_{ps}(x_1, x_2) = \frac{1}{2\pi\bar{\kappa}_2} \exp\left[-\frac{(x_1 - \kappa_1)^2}{2\bar{\kappa}_2}\right] \exp\left[-\frac{(x_2 - \kappa_1)^2}{2\bar{\kappa}_2}\right] [1 + \bar{\kappa}_{11}(x_1 - \kappa_1)(x_2 - \kappa_1)]. \quad (34)$$

As an approximate result, one can substitute (34) in (33), assume  $h^2 < 1$  and see that the normalized value  $\bar{\kappa}_{11}$  has the same order as  $h^2$ , that is,  $\bar{\kappa}_{11} \sim O(h^2)$ . This is an important consideration because usually the pure chaos has a low covariance interval [29] and one can obtain a very small MSE for two time instants  $t_1$  and  $t_2$  arbitrarily close. In this sense and fixing  $\text{SNR} \sim 0.5$  and  $\text{MSE} \sim 0.1\%$ , an equivalent MSE can be reached using the two-moment approach but with an SNR threshold 30% lower than for the one-moment case. Let us be emphatic and say that the approximation  $\bar{\kappa}_{11} \sim O(h^2)$  is valid just for  $h^2 < 1$ , and calculation of  $\hat{\bar{\kappa}}_2 \sim \bar{\kappa}_2(1 - \bar{\kappa}_{11}^2)$  has to be updated instantaneously because  $h^2$  is varying in the interval  $0 \leq h^2 < 1$ .

Of course this calculation is quite approximated and true superiority for the two-moment case of the modified quasi-linear strategy has to be verified by computer experiments. Anyway it is a strong sign indicating that the use of the two-moment strategy can be very opportunistic if and only if one can find strategies to reduce the computational complexity, for example, the generalized extended Kalman filter (GEKF) algorithm.

Finally, let us reiterate that the GEKF is yet a one-moment strategy for quasi-optimum filtering, but internally makes processing of the statistical features of the chaotic data (input) through the multimoment (two-moment) apparatus. That is why this modified GEKF improved accuracy in comparison with the standard EKF. In the following in order to additionally improve the accuracy of this one-moment modified EKF, it is convenient to apply the principles of the theory of so-called "conditionally optimum filtering" proposed in ([16], ch. 9), taking this generalized EKF as the "tolerance" or "admitted" filter.

#### 4. Conditionally optimum filtering approach

The ideas and methods for conditionally optimum filtering are rather simple and are thoroughly described at ([16], ch. 9). So, let us first present the basic idea of this method. In the general case, the conditional optimum filter for the optimum estimation of the desired signal  $x(t)$  in presence of AWGN  $n(t)$  can be presented in the form [16]:

$$\dot{\kappa}_1 = \alpha\xi(y, \kappa_1, t) + \beta\eta(y, \kappa_1, t)y(t), \quad (35)$$

where  $\kappa_1(t)$  is a filtered signal;  $y(t) = x(t) + n(t)$ ;  $n(t)$  is the AWGN with intensity  $N_0$ ;  $\alpha, \beta$  are some time-dependent coefficients which have to be found.

The representation (35) is a generalized representation of the filtering algorithms where  $\dot{\kappa}_1$  is the expectation of the filtered signal. It is clear as well [16] that this form is valid also for the quasi-optimum nonlinear filtering algorithms. In the previous part, a modified EKF algorithm

was proposed for the two-time-moment case, which shows rather opportunistic improvement of the filtering accuracy, applying some heuristics related to the simplified implementation of the two-moment principle of filtering. Sure those simplifications do not allow taking full advantage of the application of the two-moment principle. Once again, this simplification is reasonable for diminishing the dimension of the filtering algorithm in order to make it practical for real-time applications. Therefore, the hope for further improvement of the characteristics of this modified EKF might be based on further optimization in the framework of conditional optimality [16].

In the theory of conditional optimality, the structure of the filter is already chosen (in our case, it is the GEKF) and the only chance for further accuracy improvement is to optimize the coefficients  $\alpha(t)$  and  $\beta(t)$  in order to minimize the MSE. The structure which was chosen initially is a so-called admitted structure which actually belongs to a class of the admitted filters. The next step is to minimize the MSE. The minimization of the MSE is a strategy in which the admitted filter makes an optimal transition at the moment "s" ( $s > t, s \rightarrow t$ ) from an initial stage, at moment "t," to a new stage at the moment "s" with the minimum MSE. The algorithm of such kind of filter is "conditionally optimum" according to Ref. [16].

Hereafter we are not going to present all the material related to this approach as it was comprehensively described at ([16], ch. 9), we will only apply the necessary final formulas from there. Unfortunately, full use of the abovementioned approach is not possible (as we will see in the following), and so, we will present some developments that allow to obtain the coefficients  $\alpha(t)$  and  $\beta(t)$  successfully.

#### 4.1. Approach to find unknown coefficients $\alpha(t)$ and $\beta(t)$

It is possible to present an admitted structure of the conditionally optimum filter from (29) in two equivalent forms:

$$\dot{\kappa}_1 = \alpha \left[ K_1(\kappa_1) + \frac{\widehat{\kappa}_2}{2} K_1''(\kappa_1) \right] + \beta \frac{\widehat{\kappa}_2}{N_0} [y(t) - \kappa_1(t)] \quad (36)$$

$$\dot{\kappa}_1 = \alpha \left[ K_1(\kappa_1) + \frac{\widehat{\kappa}_2}{2} K_1''(\kappa_1) - \frac{\widehat{\kappa}_2 \kappa_1}{N_0} \right] + \beta \frac{\widehat{\kappa}_2}{N_0} y(t), \quad (37)$$

where, as it was proposed earlier,

$$\widehat{\kappa}_2 = \bar{\kappa}_2 (1 - \bar{\kappa}_{11}^2). \quad (38)$$

Then, from (36) and (37), one has

$$\xi(t) = K_1(\kappa_1) + \frac{\widehat{\kappa}_2}{2} K_1''(\kappa_1), \quad \eta(t) = \frac{\widehat{\kappa}_2}{N_0} [y(t) - \kappa_1(t)] \quad (39)$$

$$\xi(t) = K_1(\kappa_1) + \frac{\widehat{\kappa}_2}{2} K_1''(\kappa_1) - \frac{\widehat{\kappa}_2 \kappa_1}{N_0}, \quad \eta(t) = \frac{\widehat{\kappa}_2}{N_0} y(t). \quad (40)$$

One can see that in this regard,  $\alpha$  and  $\beta$  are weighting coefficients of a priori information related to the desired chaotic signal and a posteriori data. This issue was thoroughly commented in [27]. For  $\text{SNR} < 1$ , the weight of  $\xi(t)$  obviously prevails, because a posteriori data are strongly corrupted by the additive noise. Nevertheless, taking into account that  $\widehat{\kappa}_2$  is rather small for the modified EKF, in the following,  $\widehat{\kappa}_2$  (which is actually the MSE) will be considered as a “small parameter” in all the approximations.

In order to follow all definitions and notations from ([16], ch. 9), one has to use the Ito form in all the equations:

$$\begin{aligned} dy &= Xdt + dW_1 = \phi_1(y, x, t)dt + \psi_1(y, x, t)dW_1 \\ dx &= f(x)dt + dW_2 = \phi_1(x, t)dt + \psi(x, t)dW_2, \end{aligned} \quad (41)$$

where  $\{W_i(t)\}$  are independent Wiener processes,  $i = 1, 2$ . It is obvious that:

$$\begin{aligned} \phi_1(x, t) &= f(x) = \kappa_1(x) \\ \phi_1(y, x, t) &= x \\ \psi_1(y, x) &= 1 \\ \psi(x, t) &= 1 \end{aligned} \quad (42)$$

Then, from ([16], ch. 9)

$$\widehat{x}_s - \widehat{x}_t \cong \kappa_s - \kappa_t = \alpha \xi_t \Delta t + \beta \eta_t (\phi_{1_t} \Delta t + \phi_{1_t} \Delta W). \quad (43)$$

Unbiased conditions for the optimum estimation from (43) are [16]:

$$\alpha \langle \xi_t \rangle + \langle \eta_t \phi_{1_t} \rangle - \langle \phi_{1_t} \rangle = 0. \quad (44)$$

Taking  $\xi_t$  and  $\eta_t$  according to its definitions from (40), it is easy to get from (44):

$$\alpha m_1 + \beta m_2 = m_0, \quad (45)$$

where  $m_0 = \langle \phi_{1_t} \rangle$ ,  $m_1 = \langle \xi_t \rangle$ ,  $m_2 = \langle \eta_t \phi_{1_t} \rangle$ .

Taking into account (42) with conditions  $\widehat{\kappa}_2 < 1$  and assuming that  $K_1(\kappa_1) \approx K_1''(\kappa_1) \approx 0^1$ , finally one gets:

$$\frac{\beta}{\alpha} = \frac{\kappa_2}{\langle x^2 \rangle}. \quad (46)$$

The next step, as it was proposed in ([16], ch. 9), is focused on checking the correlation conditions for the error  $(\kappa_s - x_s)$  with the vector  $[\xi \Delta t, \eta \Delta y]$  which yields to [16]:

<sup>1</sup>This assumption follows from symmetry conditions for  $f(x)$ .

$$\beta = \kappa_{0_2} \kappa_{22}^{-1}, \quad (47)$$

where

$$\kappa_{0_2} = \langle (x_t - \kappa_1) x_t \frac{\widehat{\kappa}_2}{N_0} y_t \rangle + \langle \eta_t \cdot y_t \rangle \frac{\widehat{\kappa}_2}{N_0}, \quad \kappa_{22} = \left( \frac{\widehat{\kappa}_2}{N_0} \right)^2 \langle y_t^2 \cdot \eta_t \rangle. \quad (48)$$

From the second equation in (48), it follows that  $\beta \rightarrow \infty$  which is a clear absurd. So, why this happened and what is wrong? Is the approach in ([16], ch. 9) wrong? Definitely, no. It is possible to show that the estimate  $\kappa_1$  is unbiased and decorrelated with both components  $\xi(t)$  and  $\eta(t)$ , but for our special case, the condition that  $\kappa_{22}$  (a matrix in the general case) has to be invertible is violated. Opposed as it was stated in ([16], ch. 9), the approach is not working.

The solution might be found from direct calculation of  $(x - \kappa_1)$  from the SDE of chaos and (29) and by minimization of  $\langle (x - \kappa_1)^2 \rangle$  by  $\alpha$  or  $\beta$ .

#### 4.2. Direct evaluation of the MSE and its minimization

As a first step, let us calculate the difference between the solution of (20) and (39) by applying (46):

$$(x - \kappa_1) = \int_0^T \left\{ [K_1(x) - \alpha K_1(\kappa_1)] - \frac{\alpha \kappa_1 \widehat{\kappa}_2 n(t)}{\langle x^2 \rangle N_0} \right\} dt. \quad (49)$$

Let us take the second power of (49) and make a statistical average. One has to notice that the second power of (49) is a double integral and  $\langle n(t_1) n(t_2) \rangle = N_0 \delta(t_2 - t_1)$ . Then, applying finally the assumption  $\widehat{\kappa}_2 < 1$ , one can get for the MSE:

$$MSE \approx \langle K_1^2(x) \rangle + \alpha^2 \langle K_1^2(\kappa_1) \rangle - 2\alpha \langle K_1(x) K_1(\kappa_1) \rangle + \frac{\alpha^2 \widehat{\kappa}_2}{\langle x^2 \rangle N_0}. \quad (50)$$

Looking for the minimum of (50) in terms of “ $\alpha$ ”, one easily finds:

$$\alpha = \frac{\langle K_1(\kappa_1) K_1(x) \rangle}{\langle K_1^2(\kappa_1) \rangle + \frac{\widehat{\kappa}_2}{\langle x^2 \rangle}}. \quad (51)$$

Assuming that still  $\widehat{\kappa}_2$  is a “small parameter,” it follows that  $\alpha \approx 1$  and  $\beta \cong \frac{\kappa_1}{\langle x^2 \rangle} \cong O\left(\frac{1}{\kappa_1}\right)$ . In this regard,

$$MSE \sim \frac{\left(\frac{\widehat{\kappa}_2}{N_0}\right)^2}{\langle x^2 \rangle}. \quad (52)$$

Comparing Eq. (52) with the MSE of the one-moment filtering which is  $\kappa_2$ , one can see that the conditional optimum filtering might significantly improve the MSE with the same SNR or significantly diminish the SNR threshold for a fixed MSE.



The authors consider that the two-moment filtering of chaos together with the conditionally optimum principle is a very opportunistic approach to significantly improve the MSE for chaos filtering.

## Author details

Valeri Kontorovich<sup>1\*</sup>, Zinaida Lovtchikova<sup>2</sup> and Fernando Ramos-Alarcon<sup>1</sup>

\*Address all correspondence to: valeri@cinvestav.mx

1 Electrical Engineering Department, Communications Section, CINVESTAV-IPN, México D.F

2 Engineering and Advanced Technology Interdisciplinary Professional Unit, UPIITA-IPN, México D.F

## References

- [1] Rodriguez-Bermudez G, Garcia-Laencina PJ. Analysis of EEG signals using nonlinear dynamics and chaos: A review. *Applied Mathematics & Information Sciences*. 2015;**9**(5):2309-2321
- [2] Wu G-Q, et al. Chaotic signatures of heart rate variability and its power spectrum in health, Aging and Heart Failure. *PLoS ONE Journal*. 2009 Feb;**4**(2)
- [3] L. Yue, Baojun Y, et al. Ability to detect weak effective seismic signals by utilizing chaotic vibrator system, *Chinese Science Bulletin*, Vol. 51, No. 24, pp. 3010-3017, 2006 Dec
- [4] Shu-Yao J, Fei Y, et al. Application of stochastic resonance technology in underwater acoustic weak signal detection. *Conference OCEANS*, 10–13 April, Shangai; 2016
- [5] Eisencraft M, et al. White Gaussian chaos, *IEEE Communication Letters*, V. 21, No. 8, pp. 1719-1722, August, 2017
- [6] Ataie V, Esman D, et al. Subnoise detection of a fast random event. *Science Magazine*. 2015 Dec;**350**(6266):1343-1345
- [7] Arditti D, Alcocer A, Ramos F, Kontorovich V. Adaptive Mitigation of Platform-Generated Radio-Frequency Interference. Patent No. 14180219.9-1852; 2014
- [8] Kontorovich V, Lovtchikova Z. Multi-moment statistical characterization and non-linear filtering of chaos. *COMPEL The International Journal for Computation and Mathematics in Electrical Engineering*. 2013;**32**(3):885-900
- [9] Eckmann J, Ruelle D. Ergodic theory and strange attractors. *Reviews of Modern Physics*. July 1985;**57**:617-656
- [10] Kontorovich V. Applied statistical analysis for strange attractors and related problems. *Mathematical Methods in the Applied Sciences*. 2007;**30**:1705-1717
- [11] Anischenko VS, et al. Statistical properties of dynamical chaos. *Physics-Uspekhi*. 2005;**48**(2):151-166

- [12] Kontorovich V, et al. Analysis of Rössler attractor and its applications Special Issue on Non-linear Dynamics and Synchronization. The Open Cybernetics and Systemics Journal. 2009
- [13] Mijangos M, Kontorovich V, Aguilar-Torrentera J. Some statistical properties of strange attractors: Engineering view. Journal of Physics: Conference Series 012147. 2008 Mar;**96**:6
- [14] Stratonovich R. Topics of the Theory of Random Noise; V.1 and V.2. Gordon and Breach; 1963
- [15] Kushner H. Dynamic equations for optimal non-linear filtering. Journal of Differential Equations. 1971;**3**:179-190
- [16] Pugachev V, Sinitsyn I. Stochastic Differential Systems. Analysis and Filtering, John Wiley & Sons, Chichester, New York, 1987
- [17] Arasaratnan I, Haykin S, Elliot R. Discrete-time non-linear filtering algorithms using Gauss-Hermite quadrature. Proceedings of the IEEE. May, 2007;**96**(5):953-977
- [18] Ito K, Xiong K. Gaussian filters for non-linear filtering problems. IEEE Transactions on Automatic Control. 2000 May;**45**(5):910-927
- [19] Julier S, Uhlman J. Unscented filtering and non-linear estimation. Proceedings of the IEEE. 2000 Mar;**92**(3):401-422
- [20] Jazwinski A. Stochastic Processing and Filtering Theory. NY: Academic Press; 1970
- [21] Kazakov I, Artemiev V. Optimization of dynamic systems with random structure, Nauka, Moscow, Russia, 1980 (in Russian)
- [22] Primak S, Kontorovich V, Lyandres V. Stochastic Methods and Their Applications to Communications: Stochastic Differential Equations Approach. John Wiley & Sons; 2004
- [23] Van Trees H. Detection, Estimation and Modulation theory. John Wiley & Sons; 2001
- [24] Kushner H, Budhiraja A. A non-linear filtering algorithm based on an approximation of the conditional distribution. IEEE Transactions on Automatic Control. March 2000;**45**(3):580-585
- [25] Zakai M. On the optimal filtering of diffusion processes. Wahrscheinlichkeitstheorie-verngebiete. 1969;**11**:230-243
- [26] Kontorovich V. Non-linear filtering for Markov stochastic processes using high-order statistics (HOS) approach. Non-linear Analysis; Theory, Methods and Applications. 1997;**30**(5):3165-3170
- [27] Kontorovich V, Lovtchikova Z. Non-linear filtering of chaos for real time applications, Chapter 3 in the book *Selected topics in non-linear dynamics and theoretical electrical engineering*, Eds. K. Kyamanya, et al., Springer, Berlin Heidelberg 2013
- [28] Anischenko VS, et al. Non-linear Dynamics of chaotic and stochastic systems, Springer, Berlin Heidelberg, 2007
- [29] Kontorovich V, Lovtchikova Z, Ramos F. Correlation properties of chaos: Cumulant approach. Mathematical and Computational Applications. 2010;**15**(5):946-952

