A mathematical framework for new fault detection schemes in nonlinear stochastic continuous-time dynamical systems

Pedro J. Zufiria

ABSTRACT

In this work, a mathematical unifying framework for designing new fault detection schemes in nonlinear stochastic continuous-time dynamical systems is developed. These schemes are based on a stochastic process, called the residual, which reflects the system behavior and whose changes are to be detected. A quickest detection scheme for the residual is proposed, which is based on the computed likelihood ratios for time-varying statistical changes in the Ornstein–Uhlenbeck process. Several expressions are provided, depending on a priori knowledge of the fault, which can be employed in a proposed CUSUM-type approximated scheme. This general setting gathers different existing fault detection schemes within a unifying framework, and allows for the definition of new ones. A comparative simulation example illustrates the behavior of the proposed schemes.

1. Introduction

Model-based schemes for dynamical system fault detection make use of an explicit analytical system representation for redundancy checking, so that they generate the residual as a measure of discrepancy between the model and real system behavior. Residual generation and analysis techniques have been extensively developed in two directions. On the one hand, stochastic discrete-time models combine statistical hypothesis tests with geometrical tools in the design and characterization of detection algorithms for linear systems [5,15,20,23]. On the other hand, deterministic continuous-time models, which make use of identification and control theory tools, have shown to be suitable for nonlinear system fault detection [1,15,17,19,46,47,60].

Recently, continuous-time nonlinear stochastic models have been employed in system fault diagnosis [12,13,41–43,50]. These models can characterize system and sensor noises as well as disturbances in very elegant and analytically concise formulations; in addition, they can exploit the advantages of statistical tools for evaluating the quality of the detection schemes. These new detection and isolation proposed algorithms are based on the generation of a residual modelled as a continuous-time stochastic process. The residual known statistics under the no-fault hypothesis, do change when a fault occurs; hence, these schemes rely on the detection of statistical changes in such residual stochastic process.

In this paper, a general framework for quickest detection of changes in a continuous time residual stochastic process is presented. A detection set up is defined where the quickest detection problem shows to be appropriate for modelling fault detection. The usual known solutions of the quickest detection problem require the computation of the appropriate log-likelihood ratio (LR). Therefore, several log-likelihood ratios are provided for such framework, depending on the *priori* knowledge on the fault. These LRs are employed in a newly proposed CUSUM-type detection formulation, and some existing schemes are framed as approximators within this general setting. Simulations illustrate their comparative performance.

The paper is organized as follows. In Section 2, the design of system fault detection schemes is presented as a residual change detection problem. The steps proposed for a quickest detection scheme are presented in Section 3. Different log-likelihood ratios are computed in Section 4 and new CUSUM-type approximated algorithms are proposed in Section 5. A comparative simulation example is developed in Section 6 to illustrate the theory. Finally, some concluding remarks are summarized in Section 7.

2. Problem statement

Let us consider the following nonlinear time-varying dynamical system

$$\dot{x}(t) = f(x(t), u(t), \theta_0, t) + \eta(t) + \phi(t),$$ (1)
$$y(t) = h(x(t), u(t), t),$$
$$x(0) = x_0,$$

where $x(t) \in \mathbb{R}^n$ is the system state, which has known initial value $x_0 \in \mathbb{R}^n$; $u(t) \in \mathbb{R}^m$ is the control input; the known function $f \in \mathcal{C}^1(\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^+, \mathbb{R}^n)$ represents the dynamics of the nominal model; the random vector $\eta : \mathbb{R}^+ \to \mathbb{R}^n$, which gathers external disturbances and modelling errors, corresponds to an n-dimensional stochastic generalized process whose components are zero mean white Gaussian noise (WGN) with correlation matrix function $R_\eta(t, \tau) = \delta(t - \tau) \cdot \Sigma$, where the $\delta(t - \tau)$ distribution multiplies the components of the $n \times n$ correlation matrix Σ. Fault process $\phi : \mathbb{R}^+ \to \mathbb{R}^n$ represents the changes in the system dynamics, starting at *unknown time* T_0, and it is assumed to be an n-dimensional stochastic process whose components are Gaussian generalized processes, given by a linear combination of a Mean Squared (MS) continuous Gaussian random process and white Gaussian noise. Different types of faults are also gathered in this model, including both additive faults and parametric faults (2) (see [41–43,50] for more details). The derivatives of stochastic system (1) are interpreted as MS derivatives. Finally, $y(t) \in \mathbb{R}^l$ is the measurable output, and the nonlinear mapping $h : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^l$ can represent different output availability situations.

Applying a Luenberger observer-type consistency checker to the system, the following residual is obtained (see [12–14,41–43,46,50]):

$$\epsilon(t) = \int_0^t e^{-\Lambda(t-\tau)} \eta(\tau) d\tau + \int_{T_0}^t e^{-\Lambda(t-\tau)} \phi(\tau) d\tau = \epsilon_\eta(t) + \epsilon_\phi(t),$$ (2)

where the matrix $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_i > 0$, $i = 1, \ldots, n$ is a design set of parameters; for the sake of simplicity $\Lambda = \lambda \cdot I_n$ with $\lambda > 0$ multiplying all elements of the $n \times n$ identity matrix I_n will be considered. Under such assumptions, the residual vector process $\epsilon(t)$ defined via the MS integral (or Itô integral for generalized processes) does exist, it is sample continuous with continuous mean and variance–covariance matrix, and it is formed by Gaussian components [33].

In all cases, the analysis of this residual vector $\epsilon(t)$ is the key procedure to conclude if a fault has occurred or not in the system.

2.1. Residual evaluation

When the stochastic process $\epsilon(t)$ is considered for intervals of time *before* T_0 (i.e., when the fault has not shown up yet) the corresponding filtration of sigma-algebras $(\mathcal{F})_{t \geqslant 0}$ and the associated probability measure P_0, supporting $\epsilon = \epsilon_\eta(t)$, will be labeled with H_0. Hence $\epsilon_\eta(t) = \epsilon_{H_0}(t)$ stands for the residual under hypothesis H_0 (no fault) and it is an Ornstein–Uhlenbeck process with $E[\epsilon_\eta(t)] = 0$. On the other hand, when we consider intervals of time *after* T_0 (i.e., after the fault has occurred) the corresponding filtration of sigma-algebras and the associated probability measure P_1 supporting $\epsilon = \epsilon_\eta(t) + \epsilon_\phi(t)$, are labeled with H_1. Hence $\epsilon_\eta(t) + \epsilon_\phi(t) = \epsilon_{H_1}(t)$, where $\epsilon_\phi(t)$ gathers all the information concerning the fault, its properties depending on $\phi(t)$.

Fault detection schemes are based on the study of the properties of the residual $\epsilon(t)$ in order to detect the merging of some "signal" $\epsilon_\phi(t)$ added to the "background noise" $\epsilon_\eta(t)$. Therefore, the implementation of such detection schemes is determined by the priori available information about $\epsilon_\eta(t)$ and $\epsilon_\phi(t)$ (which strongly depends on the unknown fault time T_0), as well as the range of time in which the observation of the residual is performed.

3. Detection setup

3.1. Classical detection theory

Classical signal detection theory is aimed to choose between the hypotheses H_0 and H_1 working in a predefined observation period of time $I = [T_1, T_2]$ under the assumption that only one of the two probability measures P_i, $i = 0, 1$, is fully applicable in the whole I. With respect to our fault detection problem, this would require to assume that $T_0 \notin I$ so that either $T_0 > T_2$ (meaning $\epsilon(t) = \epsilon_{H_0}(t)$), or $T_0 < T_1$ (meaning $\epsilon(t) = \epsilon_{H_1}(t)$) can happen; therefore, the value of T_0 would not play any role in this setting. Well known classical detection schemes have been developed, for scalar $\epsilon(t)$, under some strong

assumptions on the statistics of $\epsilon_\phi(t)$ and $\epsilon_\eta(t)$ [26]. For instance, if $\epsilon_\phi(t)$ is deterministic with known profile in $[T_1, T_2]$ (note that in our problem this would require the knowledge of T_0) and $\epsilon_\eta(t)$ is Gaussian white noise, the detection (and classification of the signal) can be easily implemented via a matched filter detector structure [48,57]. Some generalizations of these results for Gaussian $\epsilon_\phi(t)$ and colored $\epsilon_\eta(t)$ will be addressed below because, although these fixed interval detection results cannot be directly applied to our fault detection problem (2), they establish the foundations to build up on-line schemes for such problem.

3.2. Sequential detection

Sequential detection theory allows for the possibility of observing along open ended intervals $[T_1, t]$, so that one can choose on-line the time $t = T_d$ at which the selection between H_0 and H_1 is performed. Such stopping time T_d adapted to the observations $\epsilon(\tau)$, $T_1 \leqslant \tau \leqslant T_d$ determines the smallest instant of time at which the selection can be performed satisfying certain quality criteria (low error probabilities). As in fixed interval detection, $T_0 \notin [T_1, t]$ is assumed so that either $T_0 = \infty$ meaning $\epsilon(t) = \epsilon_{H_0}(t)$, or $T_0 < T_1$ meaning $\epsilon(t) = \epsilon_{H_1}(t)$ needs to be assumed. Again, optimal well known solutions are available, for instance, when $\epsilon_\phi(t)$ is a known constant and $\epsilon_\eta(t)$ is a Gaussian white noise [26,21,22,56,58]; these results also set a good basis for dealing with the problem defined in Section 2.

3.3. Quickest detection

Finally, *quickest detection* theory addresses the construction of on-line sequential schemes along open ended intervals $[T_1, t]$, considering the fact that $T_0 \in [T_1, t]$ is unknown [26,48,49,54]. A quickest detection scheme is defined as a stopping time T_d adapted to the observations $\epsilon(\tau)$, $0 \leqslant \tau \leqslant T_d$ which announces that the fault has occurred at or before T_d. It is desirable that $T_d \geqslant T_0$ and it should be as close as possible to T_0. Hence, quickest detection does precisely correspond with the fault detection problem formulation in Section 2.

Quickest detection schemes can be grounded on different settings of optimality where the Bayesian an Min–max criteria are the most extended. The Bayesian approach requires the knowledge (or assumption) of a priori distribution of T_0 considered as a random variable [53]. In the Min–max approach, the expected value of the detection delay $T_d - T_0$ is selected for a minimization goal, traded-off with the false alarm rate; several criteria can be selected depending of the conditioning of such expected value. Among them, the *worst case detection delay* is the most extended:

$$\inf_{T_d \in \mathcal{T}} \left\{ \operatorname*{supess\,sup}_{T_0 \geqslant 0} E_{H_1}[T_d - T_0/T_d \geqslant T_0; \epsilon(\tau), 0 \leqslant \tau \leqslant T_0] \right\}, \quad \text{subject to } E_{H_0}[T_d] \geqslant \gamma,$$

where $\mathcal{T}$ stands for the set of detection times associated with all the considered detection schemes. Note that this criterion defines a trade-off between the detection delay and γ (associated with the false alarm rate).

In general, the application of this optimality criterion to (2) is not straightforward. Exact results were first obtained for a discrete time Min–max formulation in the specific problem of a *known constant* change of mean in a WG process. Such Min–max formulation was first presented in [36] where the asymptotic optimality of the discrete time CUSUM solution was proved (the CUSUM algorithm is based on the decomposition of the log-likelihood ratio of independent and identically distributed – i.i.d. – samples into *cumulative sums*, and had been first exposed in [45]). Non-asymptotic aspects of the optimality of such CUSUM solution for the Min–max problem were proved in [38], and later extended to a specific class of non i.i.d. processes in [39].

In the continuous time setting, optimal solutions are also known for the case of a *known constant* change in the mean of a WG process. For that simple situation, defining the log-likelihood ratio $LR(T_0, t) = \log \frac{l(t)}{l(T_0)} = L(t) - L(T_0)$, the scheme

$$T_d = \inf \left\{ t > 0 : \max_{T_0 \in [0,t]} LR(T_0, t) \geqslant h \right\} \tag{3}$$

is optimal [7,55] (note that the value of h is determined by the restriction $E_{H_0}[T_d] \geqslant \gamma$). Again, in the context of a WG process, (3) leads to a CUSUM (in a proper sense, a cumulative integral) formulation since $LR(t_1, t_2) + LR(t_2, t_3) = LR(t_1, t_3)$. Recently, in a bit more general setting (allowing drift changes in the process) the optimality of such continuous-time CUSUM scheme has been proven for a criterion based on the Kullback–Leibler divergence [40].

The optimality of the CUSUM for the simplest case of change in mean μ makes use of $LR(0, t) = \mu \int_0^t \epsilon(\tau)d\tau - \frac{1}{2}\mu^2 t$. This likelihood ratio can also be seen as optimal for the classical detection problem in the fixed interval $[0, t]$. In this framework of classical detection theory, considering the typical generalization for time-varying $\mu(t)$ we obtain $LR(0, t) = \int_0^t \mu(\tau)\epsilon(\tau)d\tau - \frac{1}{2}\int_0^t \mu^2(\tau)d\tau$ [26]. Such expression can also be extended to a vector form $LR(t) = \int_0^t \mu^T(\tau)\Sigma^{-1}\epsilon(\tau)d\tau - \frac{1}{2}\int_0^t \mu^T(\tau)\Sigma^{-1}\mu(\tau)d\tau$ by taking, for instance, appropriate limits in the discrete time forms provided in [5].

Motivated by these possible generalizations within the framework of classical detection theory, in this paper we propose the use of generalized $LR(0, t)$ expressions into the scheme (3) for quickest detection purposes. Nevertheless, some fundamental issues must be considered. First, the above mentioned classical $LR(0, t)$ functions do not apply to the case of (2) since $\epsilon_\eta(t)$ is not WG. Secondly, the scheme in (3) needs not be optimal for solving the quickest detection problem since $\epsilon_\phi(t)$ will

in general be different from $\mu \cdot s(t - T_0)$ (with $s(t)$ being the step function). Third, the scheme in (3) is computationally cumbersome; some approximations to reduce this computational cost are useful.

In the following Sections, we address these issues.

4. Likelihood ratios for different types of faults

4.1. The Ornstein–Uhlenbeck process

As mentioned above, existing optimal results do not directly apply to (2) since $\epsilon_\eta(t)$ is an Ornstein–Uhlenbeck (hence colored) process and $\epsilon_\phi(t)$ may respond to a diverse set of fault situations.

Concerning the Ornstein–Uhlenbeck process, there exists a large amount of literature addressing parameter estimation in the scalar case. Most of the works provide estimators for the drift parameter λ (see [10,18,35]) which in our case is known by design; in [3,30,52] maximum likelihood (ML) estimators for the mean parameter μ are provided whereas the estimation of σ is usually referred to the computation of a limit [6]. In general, these results do not apply to the case of time-varying vector faults. In [4] a ML estimator is proposed for the amplitude of a scalar known-shape additive function. As a complement to these existing results, we will now provide the log-likelihood ratio function for the vector case in different situations.

4.2. Known-profile deterministic faults

The following theorem provides the log-likelihood ratio function (i.e., the Radon–Nikodym derivative $\frac{dP_1}{dP_0}$) associated with a known time-varying change in the vector mean:

Theorem 4.1. Let $\epsilon(t)$ be defined in (2) such that $\eta(t)$ is a white noise zero mean vector with $E[\eta(t)\eta(\tau)] = \delta(t - \tau) \cdot \Sigma$, and $\phi(t) \in C^2$ is a known deterministic function. For stationary $\epsilon(t)$, the corresponding log-likelihood ratio function $\frac{dP_1}{dP_0}$ on an interval $[0, t]$ is given by:

$$LR_\epsilon(0, t) = \left(\epsilon(0) - \frac{1}{2}\epsilon_\phi(0) \right)^T \Sigma^{-1} \left(\lambda\epsilon_\phi(0) - \dot{\epsilon}_\phi(0) \right) + \left(\epsilon(t) - \frac{1}{2}\epsilon_\phi(t) \right)^T \Sigma^{-1} \left(\lambda\epsilon_\phi(t) + \dot{\epsilon}_\phi(t) \right)$$
$$+ \int_0^t \left(\epsilon(\tau) - \frac{1}{2}\epsilon_\phi(\tau) \right)^T \Sigma^{-1} \left(\lambda^2\epsilon_\phi(\tau) - \ddot{\epsilon}_\phi(\tau) \right) d\tau, \tag{4}$$

where integrals are considered in the Mean Square (MS) sense.

Proof. See Appendix A. □

Remark. For the ease of notation we consider here the interval $[0, t]$; all the results can be directly translated into a general interval $[t_1, t_2]$, which will be employed later in Section 5.

If we consider the scalar case, with $\Sigma = \sigma^2$, a similar result is provided in [4], and an equivalent result is provided in [26] for a colored scalar process (similar to the Ornstein–Uhlenbeck process) in the context of Karhunen–Loève expansions, via the resolution of an integral equation.

An interesting alternative form of (4) is given by:

Corollary 4.1. Formally (4) can be written as

$$LR_\epsilon(0, t) = <\epsilon(\cdot), \epsilon_\phi(\cdot)>_{[0,t]} - \frac{1}{2}\|\epsilon_\phi(\cdot)\|^2_{[0,t]}, \tag{5}$$

where

$$<\epsilon(\cdot), \epsilon_\phi(\cdot)>_{[0,t]} = \int_0^t [\dot{\epsilon}(\tau) + \lambda\epsilon(\tau)]\Sigma^{-1}[\dot{\epsilon}_\phi(\tau) + \lambda\epsilon_\phi(\tau)] \, d\tau + 2\lambda\epsilon(0)\Sigma^{-1}\epsilon_\phi(0) \tag{6}$$

defines an inner product in $C^2([0, t], \mathbb{R}^n)$.

Proof. It is obtained applying integration by parts to (4), and rearranging terms. □

Note that since $\epsilon(t)$ is an Ornstein–Uhlenbeck process, it is only MS-differentiable in a generic sense; hence, the integral must be considered in the Stratonovich sense (compatible with the usual calculus rules, see [27]), which in this case is equivalent to the Itô integral because ϵ_ϕ is deterministic.

In [25], a particularization of (6) to the scalar case defines *formally* a Reproducing Kernel Hilbert Space (RKHS) associated with $R(t, s) = \frac{e^{\lambda|t-s|}}{2\lambda}$ in $[0, t]$ (for a rigorous definition of RKHSs generated by stochastic processes see [8]). This perspective allows for a geometric interpretation of the detection scheme: ϵ is projected onto the one-dimensional subspace spanned by ϵ_ϕ

which is statistically uncorrelated to its orthogonal complement. Hence, the problem reduces to a simple detection rule based on these projected one-dimensional sufficient statistics.

Once $LR_\epsilon(0, t)$ is defined, its distribution is required in order to establish the scheme thresholds of (3). In can be easily proved that the distribution of $LR_\epsilon(0, t)$ under hypothesis H_0 is Gaussian with:

$$E[LR_\epsilon(0, t)/H_0] = -\frac{1}{2}\left[\epsilon_\phi^T(0)\Sigma^{-1}(\lambda\epsilon_\phi(0) - \dot{\epsilon}_\phi(0)) + \epsilon_\phi^T(t)\Sigma^{-1}(\lambda\epsilon_\phi(t) + \dot{\epsilon}_\phi(t)) + \int_0^t \epsilon_\phi^T(\tau)\Sigma^{-1}(\lambda^2\epsilon_\phi(\tau) - \ddot{\epsilon}_\phi(\tau))d\tau\right]$$

$$Var[LR_\epsilon(0, t)/H_0] = \frac{1}{2\lambda}\left[(\lambda\epsilon_\phi(0) - \dot{\epsilon}_\phi(0))^T\Sigma^{-1}(\lambda\epsilon_\phi(0) - \dot{\epsilon}_\phi(0)) + (\lambda\epsilon_\phi(t) + \dot{\epsilon}_\phi(t))^T\Sigma^{-1}(\lambda\epsilon_\phi(t) + \dot{\epsilon}_\phi(t)) + e^{-\lambda t}(\lambda\epsilon_\phi(0)\right.$$

$$-\dot{\epsilon}_\phi(0))^T\Sigma^{-1}(\lambda\epsilon_\phi(t) + \dot{\epsilon}_\phi(t)) + (\lambda\epsilon_\phi(0) - \dot{\epsilon}_\phi(0))^T\Sigma^{-1}\int_0^t e^{-\lambda\tau}(\lambda^2\epsilon_\phi(\tau) - \ddot{\epsilon}_\phi(\tau))d\tau$$

$$+ (\lambda\epsilon_\phi(t) + \dot{\epsilon}_\phi(t))\Sigma^{-1}\int_0^t e^{-\lambda(t-\tau)}(\lambda^2\epsilon_\phi(\tau) - \ddot{\epsilon}_\phi(\tau))d\tau + \int_0^t\int_0^t e^{-\lambda|\tau_1 - \tau_2|}(\lambda^2\epsilon_\phi(\tau_1)$$

$$\left.-\ddot{\epsilon}_\phi(\tau_1))^T\Sigma^{-1}(\lambda^2\epsilon_\phi(\tau_2) - \ddot{\epsilon}_\phi(\tau_2))d\tau_1 d\tau_2\right].$$

Hence, the false alarm probabilities can be tuned by selecting the corresponding threshold values in the tests.

4.3. Unknown deterministic faults

In case that the change ϵ_ϕ is deterministic but unknown, the usual approaches rely on the computation of a Weighted Likelihood Ratio (WLR) or a Generalized Likelihood Ratio (GLR) (see [5]). Here, since no a priori distribution is considered on the fault, we propose one-sided tests via the GLR. We begin by considering the case of $\epsilon_\phi(t) = \mu \cdot m(t)$ with known shape $m(t)$ but unknown size μ:

Corollary 4.2. *Let us consider the same hypotheses as in Theorem 4.1 where $\epsilon_\phi(t) = \mu \cdot m(t)$ with known $m(t) \in \mathbb{R}^n$ and unknown $\mu \in \mathbb{R}$. Then $\max_\mu LR_\epsilon(0, t) = GLR_\epsilon(0, t)$ takes the form*

$$GLR_\epsilon(0, t) = \frac{\left[\epsilon^T(0)\Sigma^{-1}(\lambda m(0) - \dot{m}(0)) + \epsilon(t)^T\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + \int_0^t \epsilon(\tau)\Sigma^{-1}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau\right]^2}{2\left[m^T(0)\Sigma^{-1}(\lambda m(0) - \dot{m}(0)) + m^T(t)\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + \int_0^t m^T(\tau)\Sigma^{-1}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau\right]}, \quad (7)$$

where the maximum is attained at

$$\mu^* = \frac{\epsilon^T(0)\Sigma^{-1}(\lambda m(0) - \dot{m}(0)) + \epsilon(t)^T\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + \int_0^t \epsilon(\tau)\Sigma^{-1}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau}{m^T(0)\Sigma^{-1}(\lambda m(0) - \dot{m}(0)) + m^T(t)\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + \int_0^t m^T(\tau)\Sigma^{-1}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau}.$$

Proof. It is the result of a scalar maximization problem derived from (4). □

Note that μ^* is the maximum likelihood (ML) estimator of μ (a version of this result, for scalar $m(t)$, is given in [4]). Since μ^* is a sufficient statistic, its value provides the same information as $GLR_\epsilon(0, t)$; this log-likelihood ratio follows a χ^2 distribution, whereas μ^* has a Gaussian distribution with $E[\mu^*/H_0] = 0$ and

$$Var[\mu^*/H_0] = \frac{1}{2\lambda}\left[(\lambda m(0) - \dot{m}(0))^T\Sigma^{-1}(\lambda m(0) - \dot{m}(0)) + (\lambda m(t) + \dot{m}(t))^T\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + e^{-\lambda t}(\lambda m(0)\right.$$

$$-\dot{m}(0))^T\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + (\lambda m(0) - \dot{m}(0))^T\Sigma^{-1}\int_0^t e^{-\lambda\tau}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau + (\lambda m(t) + \dot{m}(t))\Sigma^{-1}$$

$$\times \int_0^t e^{-\lambda(t-\tau)}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau + \int_0^t\int_0^t e^{-\lambda|\tau_1 - \tau_2|}(\lambda^2 m(\tau_1) - \ddot{m}(\tau_1))^T\Sigma^{-1}(\lambda^2 m(\tau_2) - \ddot{m}(\tau_2))d\tau_1 d\tau_2\bigg]$$

$$\cdot\left[m^T(0)\Sigma^{-1}(\lambda m(0) - \dot{m}(0)) + m^T(t)\Sigma^{-1}(\lambda m(t) + \dot{m}(t)) + \int_0^t m^T(\tau)\Sigma^{-1}(\lambda^2 m(\tau) - \ddot{m}(\tau))d\tau\right]^{-2}.$$

Hence, in (3) the LR can be replaced by μ^*, its corresponding threshold h' being selected according to its distribution. We now consider the case of constant unknown vector fault.

Corollary 4.3. *Let us consider the same hypotheses as in Theorem 4.1 where $\epsilon_\phi(t) = \mu \in \mathbb{R}^n$ is unknown. The GLR is given by:*

$$GLR_\epsilon(0, t) = \frac{\lambda}{2(2 + \lambda t)}\left[\epsilon(0) + \epsilon(t) + \lambda\int_0^t \epsilon(\tau)d\tau\right]^T\Sigma^{-1}\left[\epsilon(0) + \epsilon(t) + \lambda\int_0^t \epsilon(\tau)d\tau\right],$$

where the maximum is attained at

$$\mu^* = \frac{\epsilon(0) + \epsilon(t) + \lambda \int_0^t \epsilon(\tau)d\tau}{2 + \lambda t}. \tag{8}$$

Proof. See Appendix B. □

It can be easily shown that μ^* has a Gaussian distribution such that $E[\mu^*/H_0] = 0$ and $Var[\mu^*/H_0] = \frac{1}{\lambda(2+\lambda t)}\Sigma$. The scalar version of this result can be found in [2].

Finally, in case that no a priori knowledge on ϵ_ϕ is available, we consider the GLR in the most general framework:

Lemma 4.1. *Let us consider the same hypotheses as in Theorem* 4.1 *where $\phi(t)$ is an unknown deterministic function. The corresponding generalized log-likelihood ratio function is formally given by:*

$$GLR_\epsilon(0,t) = \sup_{\epsilon_\phi} LR_\epsilon(0,t) = \frac{1}{2}\left[\lambda\epsilon^T(0)\Sigma^{-1}\epsilon(0) + \lambda\epsilon^T(t)\Sigma^{-1}\epsilon(t) + \lambda^2 \int_0^t \epsilon^T(\tau)\Sigma^{-1}\epsilon(\tau)d\tau + \int_0^t \dot{\epsilon}^T(\tau)\Sigma^{-1}\dot{\epsilon}(\tau)d\tau\right]. \tag{9}$$

Proof. See Appendix C. □

Note that since $\epsilon(t)$ is MS-differentiable only in a generalized sense, the last integral involving its squared derivatives will not be properly defined in the MS, Stratonovich or Itô sense. (See [26] for some practical comments on the treatment of first or second derivatives of signals containing white noise.)

4.4. Stochastic faults

When ϵ_ϕ is a stochastic process, it may change other moments of the distribution of the residual under H_1, such as the correlation function. Due to the colored nature of $\epsilon(t)$ it is not easy to implement schemes for detecting changes in such correlation (remember that the existing scalar estimators for σ do not employ a ML approach but rely on the computation of a limit [6]).

An alternative way of analysis relies on filtering the residual process $\epsilon(t)$; it can be estimated (in the least mean squared error sense) using the information on the previous interval $[0, t']$, with $t' < t$, via:

$$\hat{\epsilon}(t) = E[\epsilon(t)/\epsilon(\tau), 0 \leqslant \tau < t'],$$

so that computing $\xi(t) = \epsilon(t) - \hat{\epsilon}(t)$ we obtain an associated innovations process [34]. These conditional means can be easily derived:

Lemma 4.2. *Let us consider the same hypotheses as in Theorem* 4.1. *Then for $t' < t$:*

$$E_{H_0}[\epsilon(t)/\epsilon(\tau), 0 \leqslant \tau \leqslant t'] = E_{H_0}[\epsilon(t)/\epsilon(t')] = e^{-\lambda(t-t')}\epsilon(t'). \tag{10}$$

Proof. See Appendix D. (It is obtained by following a similar approach to the one in previous proofs.) □

This result confirms the Markov property of the Ornstein–Uhlenbeck process and the fact that it is a supermartingale [9]. From (10) one can easily prove that, fixing $t - t' = \Delta t$, the correlation function of process $\xi(t)$ is $R_\xi(t_1, t_2) = \frac{\Sigma}{2\lambda}e^{-\lambda|t_2-t_1|}(1 - e^{-2\lambda(\Delta t - |t_2-t_1|)})$ for $|t_1 - t_2| < \Delta t$, and $R_\xi(t_1, t_2) = 0$ for $|t_1 - t_2| \geqslant \Delta t$. Hence, for small Δt we generate an almost uncorrelated (and independent, since Gaussian) sequence $\xi(t)$. Since this sequence will also be affected by system faults, it can be employed as a new residual where to apply other alternative detection schemes. For instance, to detect changes in variance σ_ξ^2 for the scalar case, one can simply particularize above results for the limiting white noise case $\lim_{\lambda\to+\infty} \frac{\sigma^2}{2\lambda} = \sigma_\xi^2$. Applying a sampling procedure similar to previous proofs, the resulting likelihood ratio versus the null hypothesis (μ_0, σ_0) tends to be unbounded (due to the information coming from infinite independent samples) and takes the form:

$$LR_\xi(\sigma^*, \xi) = \lim_{N\to+\infty} A(N, \xi)^{\frac{N}{2}}, \quad \text{with} \quad \lim_{N\to+\infty} A(N, \xi) = \frac{e^{\frac{\frac{1}{t}\int_0^t(\xi(\tau)-\mu_0)^2 d\tau}{\sigma_0^2} - 1}}{\frac{\frac{1}{t}\int_0^t(\xi(\tau)-\mu_0)^2 d\tau}{\sigma_0^2}},$$

so that $\frac{\frac{1}{t}\int_0^t(\xi(\tau)-\mu_0)^2 d\tau}{\sigma_0^2}$ is the distance measure to be evaluated as a sufficient statistic, and A represents a transformed measure of such distance via the transformation $\frac{e^x}{x}$ which reaches its single minimum (with value 1) at $x = 1$.

When considering changes at the same time in both mean and variance, for the scalar case ($E[\epsilon_\phi] = \mu_1$ and $Var[\epsilon_\phi] = \sigma_1$, both unknown) we obtain:

$$LR_\xi(\mu^*, \sigma^*, \xi) = \lim_{N \to +\infty} B(N, \xi)^{\frac{N}{2}}, \quad \text{with} \quad \lim_{N \to +\infty} B(N, \xi) = \frac{e^{\frac{\frac{1}{t}\int_0^t (\xi(\tau)-\mu_0)^2 d\tau}{\sigma_0^2} - 1}}{\frac{\frac{1}{t}\int_0^t (\xi(\tau)-\mu_0)^2 d\tau}{\sigma_0^2} - \frac{(\mu_0-\bar{\xi})^2}{\sigma_0^2}},$$

so that $\bar{\xi} = \frac{1}{t}\int_0^t \xi(\tau)d\tau$ and $\frac{1}{t}\int_0^t(\xi(\tau)-\mu_0)^2 d\tau$ are sufficient statistics (note that $\frac{1}{t}\int_0^t(\xi(\tau)-\bar{\xi})^2 d\tau$ can be derived from them). Such statistics provide $x = \frac{\frac{1}{t}\int_0^t(\xi(\tau)-\mu_0)^2 d\tau}{\sigma_0^2}$ and $y = \frac{(\mu_0-\bar{\xi})^2}{\sigma_0^2} < x$, so that B represents a transformed measure via the transformation $\frac{e^x}{x-y}$, which reaches its single minimum (with value 1) at $x = 1$, $y = 0$.

Once the different log-likelihood functions have been presented as fundamental ingredients for building the hypothesis tests, we now reconsider the on-line formulation for the quickest detection problem.

5. CUSUM-type on-line algorithms

The CUSUM algorithm form of (3) can be interpreted as an off-line multiple hypotheses testing scheme or as a set of parallel open-ended tests (see [5] for a detailed discussion in the discrete time context). Alternatively, if we define $u(t) = LR(0, t)$, then (3) can be written as:

$$y(t) = u(t) - \inf_{\tau \in [0,t]} u(\tau),$$

$$T_d = \inf\{t > 0 : y(t) \geq h'\},$$

which allows for an adaptive threshold or repeated sequential probability ratio test (SPRT) interpretation.

Going back to the discrete-time formulations, in [44] the performance of a non-sequential *fixed-size sample (FSS)* algorithm is evaluated as compared to the optimal sequential CUSUM solution for the worst case criterion in the basic problem (change of mean of a white noise). Such FSS scheme would correspond, in a continuous time setting, to the above mentioned fixed interval schemes in classical detection, where successive intervals of the form $I_k = [kT, (k+1)T]$, $k = 1, 2, \ldots$ are considered for testing between H_0 and H_1; this successive testing procedure is stopped at the first value of k for which H_1 is selected. Suboptimal asymptotic properties of this FSS scheme are proved in [44] for the discrete time case.

Alternatively, the optimal scheme of type (3) can be also approximated using a sliding window, so that the search of the maximum over T_0 is restricted to fixed size intervals of the form $[t - T, t]$:

$$T_d^a = \inf\left\{t \geq T : \max_{T_0 \in [t-T,t]} LR(T_0, t) \geq h\right\}. \tag{11}$$

This type of window-limited approximations were proposed in [59] and later analyzed in [31] within a discrete time framework. In [13,42] alternative algorithms, also based on a sliding window, were employed to construct some fault detection schemes for stochastic continuous-time dynamical systems. In the following we develop this perspective.

5.1. Simplified sliding window scheme

Keeping in mind the above mentioned approximations of the CUSUM algorithm, here we present a simplified sliding window scheme which avoids the maximization step:

$$T_d^b = \inf\{t \geq T : LR(t - T, t) \geq h\}, \tag{12}$$

so that it furtherly reduces the computational cost. Hence, the algorithm only requires the successive computation of the $LR(t - T, t)$ function. In addition, we can easily estimate an approximated threshold value h (to bound the false alarm rate) using the known distribution of $LR_\epsilon(t - T, t)$. As mentioned above, the standard CUSUM algorithm for the basic problem relies on the cumulative computation of the log-likelihood, which can be recursively implemented. In our case, the formulation in (6) proves that:

$$LR_\epsilon(t_1, t_2) + LR_\epsilon(t_2, t_3) = LR_\epsilon(t_1, t_3) + 2\lambda\left(\epsilon(t_2) - \frac{1}{2}\epsilon_\phi(t_2)\right)^T \Sigma^{-1}\epsilon_\phi(t_2) \neq LR_\epsilon(t_1, t_3),$$

so that $LR_\epsilon(t_1, t_2)$ cannot be directly computed via a cumulative recursive procedure. Fortunately, the decomposition $LR(t_1, t_2) = X(t_1, t_2) + 2\lambda(\epsilon(t_1) - \frac{1}{2}\epsilon_\phi(t_1))^T \Sigma^{-1}\epsilon_\phi(t_1)$ allows for a cumulative computation of $X(t_1, t_2)$ and hence of $LR(t_1, t_2)$ since the additional term is directly available.

Note that a proper comparative analysis of the quality of the scheme (3) versus the approximations (11) and (12) would require the study of the corresponding Average Run Length (ARL) function [5]. This function provides a way to quantify the false alarm rate versus the sensitivity and detection delay of the schemes; unfortunately, it is difficult to compute the ARL in general, so that one must rely again on approximate results. Although we believe that known level crossing results for Gaussian processes [11,16,37] can be of help for approximating the ARL function, such task goes beyond the scope of this paper.

5.2. Unification of previous schemes proposed in the literature

Previous works in the continuous time system fault detection literature can be framed as special cases of the above general setting. In [41–43,50] a heuristic setting for scalar signals is proposed, via the definition of a *moving angle* between $\epsilon(t)$ and the known profile $\epsilon_\phi(t)$ (resulting from a parametric fault), so that its conditional sample realization:

$$\cos(\epsilon, \epsilon_\phi)_T^S(t) = \frac{<\epsilon, \epsilon_\phi>_T^S(t)}{||\epsilon||_T^S(t) \cdot ||\epsilon_\phi||_T^S(t)}, \quad \text{where} \quad <\epsilon, \epsilon_\phi>_T^S(t) = \frac{1}{T} \int_{t-T}^{t} \epsilon(\tau)\epsilon_\phi(\tau)d\tau \tag{13}$$

is employed as a residual testing measure. It can be proved that such scheme reduces to (12) where $LR(t - T, t)$ is given by (13) as an approximation of (5). Note that since ϵ_ϕ is assumed to be known, the measure proposed in (5) can be normalized and approximated as:

$$\frac{LR_\epsilon(0, t)}{||\epsilon_\phi(\cdot)||_{[0,t]}^2} = \frac{<\epsilon(\cdot), \epsilon_\phi(\cdot)>_{[0,t]}}{||\epsilon_\phi(\cdot)||_{[0,t]}^2} - \frac{1}{2} \approx \frac{<\epsilon(\cdot), \epsilon_\phi(\cdot)>_{[0,t]}}{||\epsilon(\cdot)||_{[0,t]}||\epsilon_\phi(\cdot)||_{[0,t]}} - \frac{1}{2} = \cos(\epsilon, \epsilon_\phi) - \frac{1}{2}, \tag{14}$$

where $\cos(\epsilon, \epsilon_\phi)$ defines the angle induced by the inner product in (6), and $||\epsilon_\phi(\cdot)||_{[0,t]} \gg ||\epsilon_\eta(\cdot)||_{[0,t]}$ has been assumed for the approximation. Considering on the other hand an approximation of (6) for large values of λ we obtain:

$$<\epsilon(\cdot), \epsilon_\phi(\cdot)>_{[t-T,t]} \approx \frac{\lambda^2}{\sigma^2} \int_{t-T}^{t} \epsilon(\tau)\epsilon_\phi(\tau)d\tau, \tag{15}$$

so that $<\epsilon, \epsilon_\phi>_T^S(t)$ is approximately equivalent to $<\epsilon(\cdot), \epsilon_\phi(\cdot)>_{[t-T,t]}$; gathering (14) and (15) we obtain (13).

Alternatively, in [12–14] different detection schemes are considered componentwise, named $\mu_a(t) = \epsilon(t)$, $\mu_b(t) = \frac{1}{t} \int_0^t \epsilon(\tau)d\tau$, and $\mu_c(t) = \frac{1}{T} \int_{t-T}^{t} \epsilon(\tau)d\tau$, under the framework of constant mean estimation. These schemes can also be seen as simplifications of (12) where LR is replaced by the corresponding approximations of the sufficient statistic (8), again for large values of λ:

$$\mu^*(t_1, t_2) = \frac{\epsilon(t_1) + \epsilon(t_2) + \lambda \int_{t_1}^{t_2} \epsilon(\tau)d\tau}{2 + \lambda(t_2 - t_1)} \xrightarrow[\lambda \to +\infty]{} \mu_{app}^*(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \epsilon(\tau)d\tau,$$

so that $\mu_a = \lim_{T \to 0} \mu_{app}^*(t - T, t)$, $\mu_b = \mu_{app}^*(0, t)$ and $\mu_c = \mu_{app}^*(t - T, t)$.

6. Simulation example

In this section we comparatively illustrate the application of some presented detection schemes to the bidimensional system:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} + \begin{bmatrix} \eta_1(t) \\ \eta_2(t) \end{bmatrix} + s(t - T_0) \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \end{bmatrix}, \quad \text{where} \quad \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} x_2 \\ 2(1 - x_1^2)x_2 - x_1 \end{bmatrix},$$

with initial condition $x(0) = [x_1(0), x_2(0)]^T = [0, 1]^T$. This system characterizes the dynamics of the van der Pol oscillator, taking into account external disturbances via the vector random process $\eta(t) = [\eta_1(t), \eta_2(t)]^T$. Such component processes are assumed to be mutually independent, and distributed as white Gaussian noise (WGN), with zero mean and autocorrelation function $R_{WGN}(t - 1, t_2) = \sigma_{WGN}^2 \delta(t_1 - t_2)$ where $\sigma_{\eta_1}^2 = \sigma_{\eta_2}^2 = 0.1$. With these parameter values, the trajectories of the system in normal operation tend on the average to a limit cycle (corresponding to the periodic solution of the deterministic van der Pol oscillator).

We have considered that an abrupt fault occurs in the system at time $T_0 = 50$, whose consequences in the oscillator evolution are modelled by:
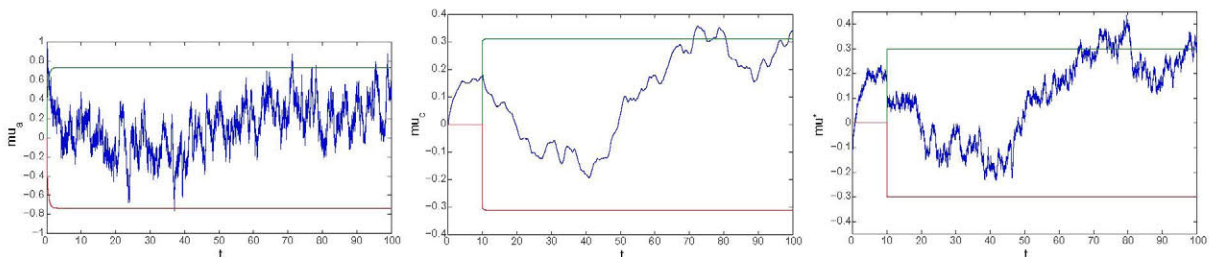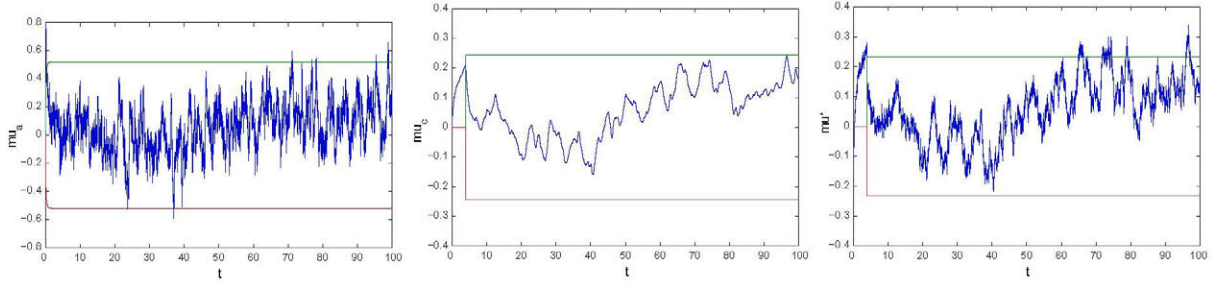
$$\phi(t) = \begin{bmatrix} 0 \\ 0.3 \end{bmatrix}.$$

Fig. 1. Time evolution of detection schemes based on $\mu_a(t)$, $\mu_c(t)$ and $\mu^*(t)$, for $T = 10$ and $\lambda = 1$.

Fig. 2. Time evolution of detection schemes based on $\mu_a(t)$, $\mu_c(t)$ and $\mu^*(t)$, for $T = 4$ and $\lambda = 2$.

We consider that the detection schemes can measure the system state space variables; hence, defining:

$$\dot{\hat{x}}(t) = -\Lambda(\hat{x}(t) - x(t)) + f(x(t)), \quad \hat{x}(0) = x(0); \quad f(x) = [f_1(x_1, x_2), f_2(x_1, x_2)]^T, \quad \Lambda = -5I_2,$$

the residual $\epsilon(t) = x(t) - \hat{x}(t)$ follows the stochastic differential equation:

$$\dot{\epsilon}(t) = -\Lambda\epsilon(t) + \eta(t) + s(t - T_0)\phi(t), \quad \epsilon(0) = 0,$$

whose solution takes the form of (2). The conditional expectations are:

$$E[\epsilon(t)/H_0] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad E[\epsilon(t)/H_1] = \begin{bmatrix} 0 \\ 0.3 \int_0^t e^{-\lambda(t-\tau)} d\tau \end{bmatrix};$$

since the first residual component has zero mean before and after the fault, only the second component will be employed.

We apply the sliding window scheme proposed in (12) with μ^* of (8) as a sufficient estimator measure. Different values of the window size T and the drift parameter λ are considered in order to compare this scheme (based on μ^*) with the detection schemes proposed in [13,14] (based on estimators μ_a and μ_c). The selected test sizes have been $\gamma_1 = \gamma_2 = 0.001$ providing the tests acceptance regions $(-3.29\sqrt{Var_{\mu/H_0}(t)}, 3.29\sqrt{Var_{\mu/H_0}(t)})$, for the corresponding $\mu \in [\mu_a, \mu_c, \mu^*]$ (stationary variance values have been considered). When any of the estimators crosses the frontier of its corresponding acceptance region an alarm is triggered, indicating that it is likely a fault has occurred in the system.

Fig. 1 shows a sample realization of mean estimators $\mu_a(t)$, $\mu_c(t)$ and $\mu^*(t)$ (as well as their corresponding bounds) for a window size $T = 10$ and drift $\lambda = 1$. Note that the computations for $\mu_c(t)$ and $\mu^*(t)$ have only meaning for $t \geqslant T$. Scheme μ_a is the only one presenting false alarm (at $(T_d^a)_{H_0} = 37,08$) before the occurrence of the fault. After T_0, the schemes trigger the alarm at $T_d^a = 70,99$, $T_d^c = 71,23$, and $T_d^* = 65,46$ respectively. Clearly, μ^* outperforms the other detection schemes: it combines the fast dynamics of $\mu_a(t)$ and the filtered tendency of $\mu_c(t)$, to provide an good trade-of between robustness against false alarms and sensitivity to faults. This behavior is typical in a wide range of values of $\lambda \in [1, 10]$. For smaller values of λ, the scheme μ^* may trigger some false alarms (its behavior approaching that of μ_a); for very large values of λ, the scheme μ^* approaches the conservative behavior of μ_c, but it is a bit more sensitive to faults.

Fig. 2 shows a sample realization of mean estimators $\mu_a(t)$, $\mu_c(t)$ and $\mu^*(t)$ (as well as their corresponding bounds) for a window size $T = 4$ and drift $\lambda = 2$. Scheme μ_a is the only one presenting several false alarms (the first one at $(T_d^a)_{H_0} = 21,2$) before the occurrence of the fault. After T_0, the schemes μ_a and μ^* trigger the alarm at $T_d^a = 71.01$, and $T_d^* = 64,97$, whereas μ_c does not even fire the alarm in the considered range of time. Again, μ^* outperforms the other detection schemes. This behavior for $T = 4$ is typical in a wide range of values of $\lambda \in [1, 10]$ (in cases where μ_c detects the fault it does so with $T_d^c > 90$). Note that with this smaller value of T, the scheme μ^* provides an excellent trade-of between robustness and sensitivity.

7. Concluding remarks

A unified setting has been defined for quickest detection of changes in continuous-time residual stochastic processes. Log-likelihood ratios have been provided for the Ornstein–Uhlenbeck process under different fault situations. A CUSUM-type approximation algorithm has also been proposed which makes use of the log-likelihood functions into a simplifying sliding window scheme. The unified setting, successfully gathers, depending on different types of simplifications, the existing detection schemes for continuous-time stochastic dynamical systems [12–14,41–43,50]. This general formulation allows for the definition of new detection schemes and for a comparative analysis among them. A simulation example has illustrated also these facts.

The analytical results presented in this paper may serve as a firm foundation for further research in more general dynamical systems models, including, for instance, differential–algebraic equations [51] as well as delays in the state or control variables [28,29].

Acknowledgement

Appendix A. Proof of Theorem 4.1

Given the deterministic vector function $\epsilon_\phi(t)$, the Gaussian distribution of $\epsilon(t)$ is totally determined by the distribution of $\epsilon_\eta(t) = [\epsilon_{\eta_1}(t), \ldots, \epsilon_{\eta_n}(t)]$. Hence, we only need to compute $R(t_1, t_2) = E[\epsilon_\eta(t_1)\epsilon_\eta^T(t_2)]$:

$$R(t_1, t_2) = E\left[\int_0^{t_1} e^{-A(t_1-\tau_1)}\eta(\tau_1)d\tau_1 \int_0^{t_2} \eta^T(\tau_2)e^{-A(t_2-\tau_2)}d\tau_2\right] = \int_0^{t_1}\int_0^{t_2} e^{-A(t_1-\tau_1)}E\left[\eta(\tau_1)\eta^T(\tau_2)\right]e^{-A(t_2-\tau_2)}d\tau_1 d\tau_2$$

$$= \int_0^{\min(t_1,t_2)} e^{-A(t_1-\tau)}\Sigma e^{-A(t_2-\tau)}d\tau = \frac{e^{-\lambda|t_1-t_2|} - e^{-\lambda(t_1+t_2)}}{2\lambda}\Sigma \xrightarrow[\min(t_1,t_2)\to+\infty]{} \frac{e^{-\lambda|t_1-t_2|}}{2\lambda}\Sigma,$$

since $A = \lambda I_n$. Let us now consider a sample of vector $\epsilon = [\epsilon(t_1), \ldots, \epsilon(t_N)]$, with $0 = t_1 < t_2 < \ldots < t_N = t$, which can be reordered as the vector of $N \cdot n$ components $\vec{\epsilon} = [\epsilon_1(t_1), \ldots, \epsilon_1(t_N), \epsilon_2(t_1), \ldots, \epsilon_2(t_N), \ldots, \epsilon_n(t_1), \ldots, \epsilon_n(t_N)]$. Then

$$f_0(\vec{\epsilon}) = \frac{1}{(2\pi)^{\frac{Nn}{2}}\sqrt{|C|}}e^{-\frac{1}{2}\vec{\epsilon}^T C^{-1}\vec{\epsilon}},$$

with

$$C = C_N \otimes \Sigma, \quad \text{and} \quad C_N = \frac{1}{2\lambda}\cdot\begin{bmatrix} 1 & e^{-\lambda|t_1-t_2|} & e^{-\lambda|t_1-t_3|} & \cdots \\ e^{-\lambda|t_2-t_1|} & 1 & e^{-\lambda|t_2-t_3|} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ e^{-\lambda|t-t_1|} & \cdots & \cdots & 1 \end{bmatrix} = \frac{1}{2\lambda}\cdot\begin{bmatrix} 1 & e^{-\lambda\delta t} & e^{-2\lambda\delta t} & \cdots \\ e^{-\lambda\delta t} & 1 & e^{-\lambda\delta t} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ e^{-(N-1)\lambda\delta t} & \cdots & \cdots & 1 \end{bmatrix},$$

where $\otimes$ stands for the Kronecker product of matrices [32], and uniform sampling $t_{i+1} - t_i = \delta t$ has been considered in the last equality. The inverse takes the form

$$C^{-1} = C_N^{-1} \otimes \Sigma^{-1}, \quad \text{with} \quad C_N^{-1} = \frac{2\lambda}{e^{\lambda\delta t} - e^{-\lambda\delta t}}\cdot\begin{bmatrix} e^{\lambda\delta t} & -1 & 0 & 0 & \cdots & 0 \\ -1 & (e^{\lambda\delta t}+e^{-\lambda\delta t}) & -1 & 0 & \cdots & 0 \\ 0 & -1 & (e^{\lambda\delta t}+e^{-\lambda\delta t}) & -1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & -1 & (e^{\lambda\delta t}+e^{-\lambda\delta t}) & -1 \\ 0 & \cdots & 0 & 0 & -1 & e^{\lambda\delta t} \end{bmatrix}.$$

Considering a similar sampling procedure for the known vector sequence $\epsilon_\phi(t)$, the log-likelihood ratio versus the zero-mean hypothesis $(\epsilon_\phi(t) = 0)$ becomes

$$LR(\vec{\epsilon}) = \ln\left[\frac{f_1(\vec{\epsilon})}{f_0(\vec{\epsilon})}\right] = \ln\left[\frac{e^{-\frac{1}{2}(\vec{\epsilon}-\vec{\epsilon}_\phi)^T C^{-1}(\vec{\epsilon}-\vec{\epsilon}_\phi)}}{e^{-\frac{1}{2}\vec{\epsilon}^T C^{-1}\vec{\epsilon}}}\right] = \vec{\epsilon}^T C^{-1}\vec{\epsilon}_\phi - \frac{1}{2}\vec{\epsilon}_\phi^T C^{-1}\vec{\epsilon}_\phi = (\vec{\epsilon} - \frac{1}{2}\vec{\epsilon}_\phi)^T C^{-1}\vec{\epsilon}_\phi = \frac{2\lambda}{e^{\lambda\delta t}-e^{-\lambda\delta t}}$$

$$\times \left[(\epsilon(0) - \frac{1}{2}\epsilon_\phi(0))^T\Sigma^{-1}(e^{\lambda\delta t}\epsilon_\phi(0) - \epsilon_\phi(t_2)) + (\epsilon(t) - \frac{1}{2}\epsilon_\phi(t))^T\Sigma^{-1}(e^{\lambda\delta t}\epsilon_\phi(t) - \epsilon_\phi(t_{N-1})) + \sum_{i=2}^{N-1}(\epsilon(t_i)\right.$$

$$\left. -\frac{1}{2}\epsilon_\phi(t_i))^T\Sigma^{-1}((e^{-\lambda\delta t}+e^{\lambda\delta t})\epsilon_\phi(t_i) - \epsilon_\phi(t_{i-1}) - \epsilon_\phi(t_{i+1}))\right].$$

We now consider the limit when $N \to \infty$ for a fixed interval $[0, t]$ (i.e., $\delta t = \frac{t}{N} \to 0$). The limit of the first terms is straightforward and the remaining sum can be written as

$$S(\delta t) = 2\lambda\sum_{i=2}^{N-1}(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i))^T\Sigma^{-1}\frac{(e^{-\lambda\delta t}+e^{\lambda\delta t})\epsilon_\phi(t_i) - \epsilon_\phi(t_i-\delta t) - \epsilon_\phi(t_i+\delta t)}{(e^{\lambda\delta t}-e^{-\lambda\delta t})\delta t}\delta t$$

$$= 2\lambda\sum_{i=2}^{N-1}\left(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i)\right)^T\Sigma^{-1}f(t_i, \delta t)\delta t$$

and it can be proved that

$$\lim_{(t,\delta t)\to(t_i,0)} f(t_i, \delta t) = \frac{\lambda^2\epsilon_\phi(t_i) - \ddot{\epsilon}_\phi(t_i)}{2\lambda} = g(t_i), \quad \forall t_i \in [0, t],$$

by using polar (r, θ) incremental coordinates around $(t_i, 0)$ and applying L'Hôpital's rule to the $r \to 0$ limit. Hence, $f(t_i, \delta t)$ can be continuously defined in a compact set $[0, t] \times [-a, a]$, $a > 0$, which concludes its uniform continuity. Therefore, $f(t_i, \delta t)$ converges to $g(t_i)$ when $\delta t \to 0$ uniformly in $[0, t]$. We can then define:

$$M_{\delta t} = \max_{t_i \in [0, t]} ||f(t_i, \delta t) - g(t_i)||,$$

such that $\lim_{\delta t \to 0} M_{\delta t} = 0$.

Hence, coming back to the sum:

$$S(\delta t) = 2\lambda \sum_{i=2}^{N-1} \left(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i) \right)^T \Sigma^{-1} g(t_i)\delta t + 2\lambda \sum_{i=2}^{N-1} \left(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i) \right)^T \Sigma^{-1} (f(t_i, \delta t) - g(t_i))\delta t = S_1(\delta t) + S_2(\delta t).$$

Since $\epsilon(t)$ is a finite covariance MS-continuous process, the first term converges in Mean Square sense to:

$$S_1(\delta t) \underset{\delta t \to 0}{\overset{MS}{\to}} \int_0^t \left(\epsilon(\tau) - \frac{1}{2}\epsilon_\phi(\tau) \right)^T \Sigma^{-1} (\lambda^2 \epsilon_\phi(\tau) - \ddot{\epsilon}_\phi(\tau))d\tau,$$

whereas the second term can be bounded by:

$$|S_2(\delta t)| = 2\lambda \left| \sum_{i=2}^{N-1} \left(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i) \right)^T \Sigma^{-1} (f_{\delta t}(t_i) - f(t_i))\delta t \right| \leqslant 2\lambda \sum_{i=2}^{N-1} \left\| \left(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i) \right)^T \Sigma^{-1} \right\| ||f_{\delta t}(t_i) - g(t_i))||\delta t$$

$$\leqslant 2\lambda M_{\delta t} \sum_{i=2}^{N-1} \left\| \left(\epsilon(t_i) - \frac{1}{2}\epsilon_\phi(t_i) \right)^T \Sigma^{-1} \right\| \delta t = 2\lambda M_{\delta t} S_3(\delta t).$$

Taking the limit on the bound:

$$|S_2(\delta t)| \leqslant 2\lambda M_{\delta t} S_3(\delta t) \overset{\delta t \to 0}{\underset{MS}{\to}} 0,$$

because $\lim_{\delta t \to 0} S_3(\delta t)$ is a bounded integral due to the finite covariance and MS-continuity of $\epsilon(t)$. Therefore $S_2(\delta) \overset{\delta t \to 0}{\underset{MS}{\to}} 0$ implying that $S(\delta t)$ has the same MS-integral limit as $S_1(\delta)$, and we conclude (4).

Appendix B. Proof of Corollary 4.3

Making $\epsilon_\phi = \mu$ in (4) we obtain:

$$LR_\epsilon(0, t) = \left(\epsilon(0) - \frac{1}{2}\mu \right)^T \Sigma^{-1} \lambda\mu + \left(\epsilon(t) - \frac{1}{2}\mu \right)^T \Sigma^{-1} \lambda\mu + \lambda^2 \int_0^t \left(\epsilon(\tau) - \frac{1}{2}\mu \right)^T \Sigma^{-1} \mu d\tau$$

$$= \lambda \left(\epsilon(0) + \epsilon(t) + \lambda \int_0^t \epsilon(\tau)d\tau \right)^T \Sigma^{-1}\mu - \left(1 + \frac{\lambda t}{2} \right) \mu^T \Sigma^{-1} \mu.$$

Taking the maximum:

$$GLR_\epsilon(0, t) = \max_\mu LR_\epsilon(0, t) = \frac{\lambda}{2(2 + \lambda t)} \left[\epsilon(0) + \epsilon(t) + \lambda \int_0^t \epsilon(\tau)d\tau \right]^T \Sigma^{-1} \left[\epsilon(0) + \epsilon(t) + \lambda \int_0^t \epsilon(\tau)d\tau \right],$$

where

$$\mu^* = \frac{\epsilon(0) + \epsilon(t) + \lambda \int_0^t \epsilon(\tau)d\tau}{2 + \lambda t}$$

is the maximum likelihood estimator of μ.

Appendix C. Proof of Lemma 4.1

Defining the generalized log-likelihood ratio:

$$GLR(\epsilon) = \ln \left[\max_{\epsilon_\phi} \frac{f_1(\epsilon(t_1), \ldots, \epsilon(t_N))}{f_0(\epsilon(t_1), \ldots, \epsilon(t_N))} \right] = \ln \left[\max_{\epsilon_\phi} \frac{e^{-\frac{1}{2}(\epsilon - \epsilon_\phi)^T C^{-1}(\epsilon - \epsilon_\phi)}}{e^{-\frac{1}{2}\epsilon^T C^{-1}\epsilon}} \right] = \ln \left[\frac{1}{e^{-\frac{1}{2}\epsilon^T C^{-1}\epsilon}} \right] = \frac{1}{2}\epsilon^T C^{-1}\epsilon,$$

which after some calculations:

$$GLR(\epsilon) = \frac{\lambda}{e^{\lambda\delta t} - e^{-\lambda\delta t}} \left[\epsilon(t_1)\Sigma^{-1}(e^{\lambda\delta t}\epsilon(t_1) - \epsilon(t_2)) + \epsilon(t_N)\Sigma^{-1}(e^{\lambda\delta t}\epsilon(t_N) - \epsilon(t_{N-1})) + \sum_{i=2}^{N-1}\epsilon(t_i)\Sigma^{-1}\left((e^{-\lambda\delta t} + e^{\lambda\delta t})\epsilon(t_i)\right.\right.$$

$$\left.\left. -\epsilon(t_{i-1}) - \epsilon(t_{i+1})\right) \right].$$

Taking the limit when $N \to \infty$ for a fixed interval (or, equivalently, making $\delta t = \frac{T}{N} \to 0$) we can *formally* proceed as in the proof of Appendix A; we can then *formally* integrate the result by parts to obtain (9). Note that such stochastic integral involving the squared derivative of $\epsilon(t)$ will not exist in the usual MS, Stratonovich or Itô sense.

Appendix D. Proof of Lemma 4.2

Note that the joint distribution of $[\epsilon_j(t_i), \ldots, \epsilon_j(t_1)]$ whose covariance matrix C_i is the same as for $[\epsilon_j(t_1), \ldots, \epsilon_j(t_i)]$ where C_i and C_i^{-1} are known. Let us note that:

$$C_i = \frac{1}{2\lambda} \cdot \begin{bmatrix} 1 & e^{-\lambda|t_1-t_2|} & e^{-\lambda|t_1-t_3|} & \ldots \\ e^{-\lambda|t_2-t_1|} & 1 & e^{-\lambda|t_2-t_3|} & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ e^{-\lambda|t-t_1|} & \ldots & \ldots & 1 \end{bmatrix} = \left[\begin{array}{c|c} 1 & C_{i,i-1}^T \\ \hline C_{i,i-1} & C_{i-1} \end{array} \right]$$

so that $C_{i,i-1}^T = \frac{e^{\lambda(t_i-t_{i-1})}}{2\lambda}\left[1, e^{-\lambda(t_{i-1}-t_{i-2})}, e^{-\lambda(t_{i-1}-t_{i-3})}, \ldots, e^{-\lambda(t_{i-1}-t_1)}\right]$. If we make $t_j - t_{j-1} = \delta t$, $j = 2, \ldots, i-1$, this expression simplifies to $C_{i,i-1}^T = \frac{e^{\lambda(t_i-t_{i-1})}}{2\lambda}\left[1, e^{-\lambda\delta t}, e^{-2\lambda\delta t}, \ldots, e^{-(i-2)\lambda\delta t}\right]$.

Considering all the components of $\epsilon(t)$, we define the vector $\vec{\epsilon} = [\epsilon_1(t_i), \epsilon_2(t_i), \ldots, \epsilon_n(t_i), \epsilon_1(t_{i-1}), \ldots, \epsilon_n(t_1)]$, whose covariance matrix is given by:

$$\Sigma \otimes C_i = \left[\begin{array}{c|c} \Sigma & \Sigma \otimes C_{i,i-1}^T \\ \hline \Sigma \otimes C_{i,i-1} & \Sigma \otimes C_{i-1} \end{array} \right].$$

It is a known result [24] that:

$$E[\epsilon(t_i)/\epsilon(t_{i-1}), \ldots, \epsilon(t_1)] = E[\epsilon(t_i)] + \left(\Sigma \otimes C_{i,i-1}^T\right)\left(\Sigma^{-1} \otimes C_{i-1}^{-1}\right) \begin{bmatrix} \epsilon(t_{i-1}) - E[\epsilon(t_{i-1})] \\ \vdots \\ \epsilon(t_1) - E[\epsilon(t_1)] \end{bmatrix}.$$

Hence, if we consider $E[\epsilon(t_j)/H_0] = 0$, $\forall j = 1, \ldots, i$, and after some calculations the expression simplifies to:

$$E_{H_0}[\epsilon(t_i)/\epsilon(t_{i-1}), \ldots, \epsilon(t_1)] = \frac{e^{-\lambda(t_i-t_{i-1})}}{e^{\lambda\delta t} - e^{-\lambda\delta t}}\left[(e^{\lambda\delta t} - e^{-\lambda\delta t})\epsilon(t_{i-1})\right] = e^{-\lambda(t_i-t_{i-1})}\epsilon(t_{i-1}),$$

which confirms the Markov property of $\epsilon(t)$. Hence, for the continuous time version, if we take the limit when $i - 1 \to \infty$ for a fixed interval (or, equivalently, making $\delta t = \frac{t_{i-1}-t_1}{i-1} \to 0$):

$$\lim_{\delta t \to 0} E_{H_0}[\epsilon(t_i)/\epsilon(t_{i-1}), \ldots, \epsilon(t_1)] = E_{H_0}[\epsilon(t_i)/\epsilon(\tau), 0 \leqslant \tau \leqslant t_{i-1}] = e^{-\lambda(t_i-t_{i-1})}\epsilon(t_{i-1}),$$

we directly obtain (10).

References

[1] E. Alcorta-García, P.M. Frank, Deterministic nonlinear observer-based approaches to fault diagnosis: a survey, IFAC Control Engineering Practice 5 (1997) 663–670.

[2] M. Arató, Linear Stochastic Systems with Constant Coefficients LNCIS, vol. 45, Springer Verlag, 1982.

[3] M. Arató, S. Fegyverneki, New statistical investigations on the Ornstein–Uhlenbeck process, Computers and Mathematics with Applications 44 (2002) 677–692.

[4] S. Baran, G. Gap, M.C.A. van Zuijlen, Estimation of the mean of stationary and nonstationary Ornstein–Uhlenbeck processes and sheets, Computers and Mathematics with Applications 45 (2003) 563–579.

[5] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes, Theory and Application, Prentice Hall, 1993.

[6] G. Baxter, A strong limit theorem for Gaussian processes, Proceedings of the American Mathematical Society 7 (1956) 522–525.

[7] M. Beibel, A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure, Annals of Statistics 24 (1996) 1804–1812.

[8] A. Berlinet, C. Thomas-Agnan, Reproducing Kernel Hilbert Spaces in Probability and Statistics, Kluwer Academic Publishers, 2004.

[9] P. Billingsley, Probability and Measure, John Wiley & Sons, 1995.

[10] J.P.N. Bishwal, A. Bose, Rates of convergence of approximate maximum-likelihood estimators in the Orstein–Uhlenbeck process, Computers and Mathematics with Applications 42 (2001) 23–38.

[11] I.F. Blake, W.C. Lindsey, Level-crossing problem for random processes, IEEE Transactions on Information Theory 19 (35) (1973) 295–315. May.

[12] Á. Castillo, P. Zufiria, M.M. Polycarpou, F. Previdi, T. Parisini, Fault detection and isolation scheme in continuous time nonlinear stochastic systems, in: Proceedings of the 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes SAFEPROCESS 2003, 2003, pp. 651–656.

[13] Á. Castillo, Fault Detection and Isolation via Continuous Time Statistics, Ph.D. thesis, E.T.S. Ingenieros Industriales (Universidad Politécnica de Madrid), 2006.

[14] Á. Castillo, P. Zufiria, Fault detection schemes for continuous-time stochastic dynamical systems, IEEE Transactions on Automatic Control 54 (8) (2009) 1820–1836.
[15] J. Chen, R. Patton, Robust Model-Based Fault Diagnosis for Dynamic Systems, Kluwer Academic Publishers, 1999.
[16] E. Csáki, D. Khoshnevisan, Capacity estimates boundary crossings and the Ornstein–Uhlenbeck process in Wiener space, Eletronic Communications in Probability 4 (1999) 103–109.
[17] C. De Persis, A. Isidori, A geometric approach to nonlinear fault detection and isolation, IEEE Transactions on Automatic Control 46 (2001) 853–865.
[18] D. Florens-Landais, H. Pham, Large deviations in estimation of an Ornstein–Uhlenbeck model, Journal of Applied Probability 36 (1999) 60–77.
[19] P.M. Frank, Analytical and qualitative model-based fault diagnosis – a survey and some new results, European Journal of Control 11 (2) (1996) 26–28.
[20] J.J. Gertler, Fault Detection and Diagnosis in Engineering Systems, Marcel Dekker, New York, 1998.
[21] M. Ghosh, N. Mukhopadhyay, P.K. Sen, Sequential Estimation, Wiley Series in Probability and Statistics (1997).
[22] Z. Govindarajulu, Sequential Statistics, World Scientific, 2004.
[23] R. Iserman, Fault-Diagnosis Systems, An Introduction to Fault Detection and Fault Tolerance, Springer Verlag, 2006.
[24] R.A. Johnson, D.W. Wichern, Applied Multivariable Statistical Analysis, Prentice Hall, 1992.
[25] T. Kailath, A RKHS approach to detection and estimation problems—part I: Deterministic signals in gaussian noise, IEEE Transactions on Information Theory 17 (5) (1971) 530–549.
[26] T. Kailath, H.V. Poor, Detection of stochastic processes, IEEE Transactions on Information Theory 44 (6) (1998) 2230–2259.
[27] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer Verlag, 1988.
[28] J. Klamka, Stochastic controllability of linear systems with state delays, International Journal of Applied Mathematics and Computer Science 17 (1) (2007) 5–13.
[29] J. Klamka, Stochastic controllability and minimum energy control of systems with multiple delays in control, Applied Mathematics and Computation 206 (2) (2008) 704–715.
[30] M.L. Kleptsyna, A. Le Breton, Statistical analysis of the fractional Ornstein–Uhlenbeck type process, Statistical Inference for Stochastic Processes 5 (2002) 229–248.
[31] T.L. Lai, Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems, IEEE Transactions on Information Theory 46 (2) (2000) 595–608.
[32] P. Lancaster, M. Tismenetsky, The Theory of Matrices, second ed., Academic Press, 1985.
[33] H.M. Larson, B.O. Shubert, Probabilistic Models in Engineering Sciences, vol. I, John Wiley and Sons, 1979.
[34] R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes. I. General Theory, second ed., Springer Verlag, Kluwer Academic Publishers, 2001.
[35] R.S. Liptser, A.N. Shiryaev, Statistics of Random Processes. II. Applications, second ed., Springer Verlag, Kluwer Academic Publishers, 2001.
[36] G. Lorden, Procedures for reacting to a change in distribution, Annals of Mathematical Statistics 42 (1971) 1897–1908.
[37] D.R. Morgan, On level-crossing excursions of Gaussian low-pass random processes, IEEE Transactions on Signal Processing 55 (7) (2007) 3623–3632.
[38] G.V. Moustakides, Optimal stopping times for detecting changes in distributions, Annals of Statistics 14 (1986) 1379–1387.
[39] G.V. Moustakides, Quickest detection of abrupt changes for a class of random processes, IEEE Transactions on Information Theory 44 (5) (1998) 1965–1968.
[40] G.V. Moustakides, Optimality of the CUSUM procedure in continuous time, Annals of Statistics 32 (1) (2004) 302–315.
[41] U. Münz, Parametric Fault Diagnosis in Stochastic Systems, Masters thesis, E.T.S. Ingenieros Telecomunicación (Universidad Politcnica de Madrid), 2005.
[42] U. Münz, P. Zufiria, Parametric Fault Diagnosis in Stochastic Dynamical Systems, in: Proceedings of the 19th CEDYA, Madrid, Spain, 2005.
[43] U. Münz, P. Zufiria, Diagnosis of unknown parametric faults in non-linear stochastic dynamical systems, International Journal of Control 82 (4) (2009) 603–619.
[44] I.V. Nikiforov, Two strategies in the problem of change detection and isolation, IEEE Transactions on Information Theory 43 (2) (1997) 770–776.
[45] E.S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115.
[46] M.M. Polycarpou, A.T. Vemuri, Learning Methodology for Failure Detection and Accomodation, IEEE Control Systems (1995) 16–24.
[47] M.M. Polycarpou, A.B. Trunov, Learning approach to nonlinear fault diagnosis: detectability analysis, IEEE Transactions on Automatic Control 45 (2000) 806–812.
[48] H.V. Poor, An Introduction to Signal Detection and Estimation, second ed., Springer Verlag, 1994.
[49] H.V. Poor, O. Hadjiliadis, Quickest Detection, Cambridge University Press, 2009.
[50] M. Reble, U. Münz, F. Allgower, Diagnosis of parametric faults in multivariable nonlinear systems, in: Proceedings of the 46th IEEE Conference on Decision and Control Conference 2007, 2007, pp. 336–371.
[51] R. Riaza, P.J. Zufiria, Differential-algebraic equations and singular perturbation methods in recurrent neural learning, Dynamical Systems 18 (2003) 89–105.
[52] A. Szimayer, R. Maller, Testing for mean reversion in processes of Ornstein–Uhlenbeck type, Statistical Inference for Stochastic Processes 7 (2004) 95–113.
[53] A.N. Shiryaev, On optimum methods in quickest detection problems, Theory of Probability and its Applications VIII (1) (1963) 22–46.
[54] A.N. Shiryaev, Optimal Stopping Rules, Springer Verlag, 1978.
[55] A.N. Shiryaev, Minimax optimality of the method of cumulative sums (cusum) in the case of continuous time, Communications of the Moscow Mathematical Society (1996) 750–751.
[56] D. Siegmund, Sequential Analysis, Tests and Confidence Intervals, Springer Verlag, 1985.
[57] H.L. Van Trees, Detection, Estimation and Modulation Theory: Part I, Wiley, 1968.
[58] A. Wald, Sequential Analysis, John Wiley & Sons, 1947.
[59] A.S. Willsky, H.L. Jones, A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems, IEEE Transactions on Automatic Control AC-21 (1976) 108–112.
[60] X. Zhang, T. Parisini, M.M. Polycarpou, Sensor bias fault isolation in a class of nonlinear systems, IEEE Transactions on Automatic Control 50 (3) (2005) 370–376.