

The Asymptotic Local Approach to Change Detection and Model Validation

ALBERT BENVENISTE, MEMBER, IEEE, MICHÈLE BASSEVILLE, AND GEORGES V. MOUSTAKIDES, MEMBER, IEEE

Abstract—We present a systematic approach for the design of change detection and model validation algorithms for dynamical systems. We show how to associate to any identification algorithm a change detection and a model validation procedure, which are optimal in some asymptotic sense. The foundations of our method go back to the asymptotic local approach in statistics, and our method generalizes this approach.

INTRODUCTION

THE problem of detecting changes in dynamical properties of signals and systems has received growing attention in the last 15 years, as can be seen from the survey papers [21] and [1], and the monograph [2]. Actually, this problem arises in several areas of automatic control and signal processing, which may be classified as follows:

- 1) failure detection in controlled systems,
- 2) segmentation of signals or images for the purpose of pattern recognition, and
- 3) gain updating in adaptive algorithms, for tracking quick variations of the parameters.

Many applied fields have been already concerned, as discussed, for example, in [1], and a significant amount of methodological tools are now available. See the above mentioned reference for an extensive bibliography on this subject.

On the other hand, the areas of system identification and system monitoring are primarily concerned with the problem of model validation in the following cases.

- 1) Check whether a given model set fits the considered system (identify the best model within the chosen model set, and perform model validation to ultimately accept or reject the selected model set).
- 2) Check whether a given nominal model (intended, for instance, to describe the ideal behavior of a given system) fits the considered system.

Most of the control softwares provide routines to perform model validation; the usual way is to monitor prediction errors, equation errors, etc.,... (see [11] for example). However, most of the model validation techniques are rather ad hoc from the statistical viewpoint.

The purpose of this paper is to present a fairly general methodology to associate closely to any identification procedure, and, more generally, to any adaptive algorithm, a change detection and a model validation procedure. The foundations of our approach are found in Le Cam's work on contiguity of probability measures, which lead to the so-called asymptotic local point of view in statistics, see the book [19], and also the fundamental papers [18] and [9]. This approach provides an effective way to design or analyze likelihood ratio based testing procedures when the alternative hypotheses become closer as the length of the

record goes to infinity. Starting from this idea, [3] and [16] studied a situation in which no likelihood ratio approach could be effective due to the presence of nonstationary nuisance parameters; hence, starting from the well-known instrumental variable method, they derived a closely related testing procedure using again a local asymptotic approach. The present paper shows that this situation is indeed general: *the asymptotic local approach provides us with a general methodology to associate to any adaptive algorithm an optimal testing procedure* for both change detection and model validation problems.

Finally, the problem of diagnosing the origin of the detected changes has been mainly addressed via the multiple model approach [21], [23]; this approach is, for example, used in aeronautics. We shall show that our approach trivially extends to this problem, thanks to a sensitivity method suited to the identification of the origin of small changes. Moreover, as we shall see, this will allow us to recognize the origin of changes in terms of *nonidentifiable* models (think of a complex system modeled on one hand by a large physical, often nonidentifiable model, and on the other by some smaller black-box identifiable model, and try to understand the origin of the changes in terms of the physical model).

The paper is organized as follows. The problem is stated in Section I. The mathematical background of the asymptotic local approach is presented in Section II. The local procedures of change detection and model validation are, respectively, presented in Sections III and IV; the methods are illustrated on examples. Finally, the diagnosis problem is investigated in Section V.

I. PROBLEM STATEMENT

The problem will be stated for the case of change detection; the problem of model validation will be directly stated in Section IV. Consider a dynamical system subject to sudden changes. Our purpose is

- 1) to decide on-line whether a change occurred or not,
- 2) if a change occurred, to estimate the change time,
- 3) to identify the origin and the magnitude of the change.

Let us first investigate some examples.

A. Examples

1) *Jump in the Mean of a Signal*: Consider a signal of the form

$$y_n = \theta_*(n) + v_n \quad (1.1.1)$$

where (v_n) is a sequence of i.i.d. random variables with distribution μ , and θ_* is a piecewise constant function. The problem is to detect the changes in θ_* , and to estimate the magnitude and the location of the jumps.

2) *Changes in an AR Process*: Consider an AR process of the form

$$y_n = \sum_{i=1}^p a_i y_{n-i} + \sigma v_n \quad (1.1.2)$$

Manuscript received September 15, 1986; revised February 11, 1987. Paper recommended by Associate Editor, H. L. Weinert.

A. Benveniste and G. Moustakides are with IRISA/INRIA, Rennes Cedex, France.

M. Basseville is with IRISA/CNRS, Rennes, Cedex, France.

IEEE Log Number 8714797.

where (v_n) is a zero mean i.i.d. sequence of unit variance. The model (1.1.2) is summarized in the parameter

$$\theta_*^T := (a_1, \dots, a_p; \sigma). \quad (1.1.3)$$

Setting

$$\phi_{n-1}^T := (y_{n-1}, \dots, y_{n-p}) \quad (1.1.4)$$

(1.1.2) can be rewritten under the following state-space form:

$$\phi_n = A(\theta_*)\phi_{n-1} + B(\theta_*)v_n$$

$$y_n = (1, 0, \dots, 0)\phi_n$$

$$A(\theta_*) = \begin{bmatrix} a_1 & & a_p \\ 1 & & 0 \\ & 1 & 0 \end{bmatrix}, \quad B(\theta_*) = \begin{bmatrix} \sigma \\ 0 \\ 0 \end{bmatrix} \quad (1.1.5)$$

where the unspecified entries of the matrix $A(\theta_*)$ are equal to zero. The formulas (1.1.5) express the fact that (ϕ_n) is a *controlled Markov chain* with control parameter θ_* . This means that (ϕ_n) is a Markov chain the transition matrix of which depends on the parameter θ_* . Assuming θ_* to be piecewise constant, it is desired 1) to detect its jumps, 2) to estimate the change times, 3) to estimate the magnitude of the jumps.

These two basic examples motivate the following general setting, which uses the framework of *controlled semi-Markov processes*, as used, for example, in [7].

B. Detection of Changes in a Controlled Semi-Markov Process

We shall say that (X_n) is a *controlled semi-Markov process with control parameter θ_** if (X_n) is of the form

$$P\{\xi_n \in G | \xi_{n-1}, \xi_{n-2}, \dots\} = \int_G \pi_{\theta_*}(\xi_{n-1}, dx) \\ X_n = f(\xi_n) \quad (1.2.1)$$

where $\pi_{\theta_*}(\xi, dx)$ is the transition probability of a Markov chain (ξ_n) depending on a parameter θ_* . The model (1.2.1) represents the true system. Accordingly, the *sequential change detection problem* in the system (1.2.1) is formulated as follows.

DS: There exists an instant $r: 0 < r \leq +\infty$, such that (X_n) is controlled by the parameter

$$\theta_* = \theta_0 \quad \text{for } n < r$$

$$\theta_* = \theta_1 \quad \text{for } n \geq r.$$

The questions we want to answer are then the following. Given a record X_1, \dots, X_n ,

- 1) detection decide between $n < r$ (no change occurred before n) and $r \leq n$ (a change occurred before n);
- 2) estimation when $r \leq n$ has been decided, estimate the change time r ;
- 3) identification if anyone is unknown, identify θ_0 and/or θ_1 .

Of course, only a subset of these problems is of interest in some cases. For example, only the problem 1) has to be investigated in failure detection when no diagnosis is required.

C. A Basic Problem and Its Solution: Change in the Mean of Independent Gaussian Vector Random Variables

This problem is the easiest change detection problem, and will illustrate our purpose. As a matter of fact, its solution will appear

as a basic component of the general change detection problems we shall investigate subsequently. Consider a sequence of independent Gaussian vector random variables (Y_n) with constant covariance matrix R , and with mean equal to 0 until time $r - 1$, and equal to θ from time r , where θ is an unknown parameter. The well-known solution to this change detection problem is the GLR test (generalized likelihood ratio), see [22]. Recall briefly how this test is obtained. First, fix r and θ . Given the record Y_1, \dots, Y_n , the log likelihood ratio between the hypotheses

- H_0 : there is no change until n
- H_1 : there is a change at time r of magnitude θ

is given by

$$S_r^n(\theta) = \sum_{k=r}^n Y_k^T R^{-1} Y_k - \sum_{k=r}^n (Y_k - \theta)^T R^{-1} (Y_k - \theta) \\ = 2 \sum_{k=r}^n Y_k^T R^{-1} \theta - (n - r + 1) \theta^T R^{-1} \theta. \quad (1.3.1)$$

Replacing θ by its most likely value under the hypothesis of change (with r still fixed), we get

$$S_r^n := \max_{\theta} S_r^n(\theta) = (\Delta_r^n)^T R^{-1} \Delta_r^n \quad (1.3.2)$$

where

$$\Delta_r^n := (n - r + 1)^{-1/2} \sum_{k=r}^n Y_k \quad (1.3.3)$$

and

$$\hat{\theta}(n, r) := \arg \max_{\theta} S_r^n(\theta) = (n - r + 1)^{-1/2} \Delta_r^n. \quad (1.3.4)$$

Taking in (1.3.2) the maximum with respect to r yields

$$G_n := \max_r S_r^n, \quad r_n = \arg \max_r S_r^n. \quad (1.3.5)$$

Finally, the stopping rule to decide that a change occurred is given by

$$\nu = \min \{n : G_n \geq \lambda\} \quad (1.3.6)$$

while the estimates of the instant of change and the magnitude of the jump are, respectively, given by

$$\hat{r} = \hat{r}_\nu, \quad \hat{\theta} = \hat{\theta}(\nu, \hat{r}). \quad (1.3.7)$$

The formulas (1.3.3)–(1.3.7) define the complete change detection test procedure for this case.

II. MATHEMATICAL BACKGROUND

In this section, we shall associate to any adaptive algorithm a fundamental invariance principle which will be the basis of our method.

A. Some Useful Background on Adaptive Algorithms, and Problem Statement

We shall first introduce the kind of adaptive algorithms we shall consider; we shall use the form and related assumptions of [7]; see also [15] or [10] for slightly different assumptions.

1) *Some Background on Adaptive Algorithms:* The adaptive algorithms we shall consider are of the form

$$\theta_n = \theta_{n-1} + \gamma_n H(\theta_{n-1}, X_n) \quad (2.1.1)$$

where θ belongs to R^d or to some submanifold of R^d , and the state X_n belongs to R^k . The gain γ_n can decrease to 0, or converge to a positive constant limit. The state vector X_n is generally a *semi-Markov process controlled by the parameter* θ ; this means that

$$P(\xi_n \in d\xi | \xi_{n-1}, \xi_{n-2}, \dots; \theta_{n-1}, \theta_{n-2}, \dots) = \pi_{\theta_{n-1}}(\xi_{n-1}, d\xi) \\ X_n = f(\xi_n) \quad (2.1.2)$$

where the extended state (ξ_n) is, for θ fixed, a Markov chain with transition probability $\pi_\theta(\xi, dx)$ which depends on the parameter θ . We assume that, for every θ belonging to the domain of the algorithm, the Markov chain (ξ_n) admits a unique invariant probability measure (i.e., is ergodic). This framework includes the case where the state (X_n) itself is a stationary semi-Markov process with distribution independent of θ . It includes also the case of *conditionally linear dynamics*, such as used in [11], i.e.,

$$X_n = A(\theta_{n-1})X_{n-1} + B(\theta_{n-1})W_n$$

where $A(\theta)$ and $B(\theta)$ are matrices, and (W_n) is an i.i.d. zero mean sequence (in this case, (X_n) is asymptotically ergodic for θ fixed if and only if the matrix $A(\theta)$ is asymptotically stable). The function $H(\theta, X)$ can be discontinuous, but we shall assume that the following *mean vector field* is smooth

$$h(\theta) := \lim_{n \rightarrow \infty} E_\theta(H(\theta, X_n)) \quad (2.1.3)$$

where E_θ denotes the expectation under the law P_θ of the process $(X_n)_{n \geq 0}$ for θ fixed. The ODE associated to the algorithm is then

$$\dot{\theta} = h(\theta), \quad \theta(0) = z \quad (2.1.4)$$

the solution of which will be denoted by $(\theta(t))_{t \geq 0}$ or $(\theta(z, t))_{t \geq 0}$ accordingly. We are now ready to introduce the framework we shall use for the change detection problem.

2) Problem Statement

a) Investigation of the Least-Squares Algorithm for AR Identification: The identification of θ_* in (1.1.5) can be, for example, performed via the least-squares stochastic gradient algorithm

$$\theta_n = \theta_{n-1} + \gamma \phi_n[\theta_*] e_n[\theta_{n-1}, \theta_*] \\ e_n(\theta, \theta_*) := y_n(\theta_*) - \phi_n^T(\theta_*) \cdot \theta \quad (2.1.5)$$

where $\phi_n[\theta_*]$ is defined in (1.1.5). In (2.1.5), the dependence on the true parameter θ_* has been made explicit, to introduce the convenient form for the change detection problem. In this form, the true system θ_* is obviously not available to the user, but clearly influences the conditional distribution of the pair (ϕ_n, y_n) . Hence, the investigation of this example motivates the following form we shall use for the adaptive algorithms instead of (2.1.1), (2.1.2) (we shall always in the sequel write z instead of θ_* to denote the true system):

$$\theta_n = \theta_{n-1} + \gamma H(\theta_{n-1}; X_n) \\ P(\xi_n \in d\xi | \xi_{n-1}, \xi_{n-2}, \dots; \theta_{n-1}, \theta_{n-2}, \dots) \\ = \pi_{\theta_{n-1}, z}(\xi_{n-1}, d\xi) \\ X_n = f(\xi_n) \quad (2.1.6)$$

where the parameter z represents the *true system*. The second equation of (2.1.6) expresses the fact that the true system does influence the law of the state ξ_n ; note that the form of the algorithm, as available to the user, has been kept unchanged.

As usual, more sophisticated gain strategies can be used; for example, the classical least-squares algorithm makes use of a recursively updated matrix gain instead of the crude constant scalar one used in (2.1.5). But the gain strategy is irrelevant for

our purpose, only the random vector field $H(\theta_{n-1}; X_n)$ will be relevant.

b) Problem Statement and Assumptions: Our starting point is now the triple introduced in (2.1.6), namely:

$$H(\theta; X_n), \pi_{\theta, z}(\xi, dx), f \quad (2.1.7)$$

where

- θ is the adjustable parameter available to the user
- z is the parameter which represents the true system; z is not available to the user
- the state X_n is a semi-Markov process controlled by the pair (θ, z) , the law of which will be denoted by $P_{\theta, z}$.

This triple will be simply referred to in the sequel as the *random vector field*. As usual, we shall assume that the transition probability $\pi_{\theta, z}(\xi, dx)$ is ergodic and positively recurrent, and hence admits a unique invariant probability. To simplify our notations, we shall denote, respectively, by

$$\hat{P}_{\theta, z} \text{ and } \hat{E}_{\theta, z} \quad (2.1.8)$$

the steady-state distribution of the process (X_n) under (θ, z) and corresponding expectation. As usual for adaptive algorithms, the following mean vector field is associated to (2.1.7), according to the notation (2.1.8)

$$h(\theta, z) := \hat{E}_{\theta, z}(H(\theta; X_n)) (= \lim_{n \rightarrow \infty} E_{\theta, z}[H(\theta; X_n)]) \quad (2.1.9)$$

where $\hat{E}_{\theta, z}$ denotes the expectation with respect to $\hat{P}_{\theta, z}$. This is nothing but the usual mean vector field of the associated ODE, where the dependence of the true system z has been made explicit.

From now on, we shall distinguish a *nominal model*

$$\theta = \theta_0$$

chosen by the user. The problem is to *detect small deviations of the true system z from the nominal model θ_0 by only monitoring the random vector field $H(\theta_0; X_n)$* . The following assumptions will be in force in the sequel, and we shall denote by h_θ and h_z , respectively, the first and second partial derivatives of h .

Assumption NS: The model set matches the true system structure in the following sense: for every z ,

$$h(\theta, z) = 0 \Leftrightarrow \theta = z. \quad (2.1.10)$$

Consequence: The following relationship holds:

$$h_\theta(z, z) = -h_z(z, z) \quad (2.1.11)$$

the proof of which is obvious and left to the reader. We are now ready to present our problem statement.

Change Detection Problem: Given a nominal model θ_0 chosen by the user, and a record X_1, \dots, X_n of length n of the state vector; test between the following hypotheses by using the random vector field trajectory $\{H(\theta_0; X_k)\}_{1 \leq k \leq n}$

- H_0 : $z \equiv \theta_0$;
- H_1 : $z = \theta_0 + \theta/\sqrt{n}$, where $\theta \neq 0$ is an unknown change;
- H'_1 : there exists $\tau \in]0, 1[$, such that

$$z = \theta_0 \text{ for } k < \tau n$$

$$z = \theta_0 + \frac{\theta}{\sqrt{n}} \text{ for } \tau n \leq k \leq n,$$

where $\theta \neq 0$ is an unknown change.

Comment: The hypothesis H_0 expresses that the nominal model is identical to the true system; the hypothesis H_1 corresponds to a constant deviation between the nominal model and the true system of magnitude order $n^{-1/2}$; finally, the hypothesis H'_1 corresponds to the occurrence of a change of magnitude order

$n^{-1/2}$ inside the record. Introducing the scaling factor \sqrt{n} is classical in statistics, and is known as the *asymptotic local approach*. The interested reader is referred to [18], [19], and [9] for further information on the asymptotic local approach for the likelihood ratio testing methods.

B. Main Results

Assumption NS is in force throughout this section. Fix a nominal model θ_0 , and consider the following cumulative sum, where $m \leq n$

$$D_{n,m}(\theta_0, \theta) := \frac{1}{\sqrt{n}} \sum_{k=1}^m H(\theta_0; X_k), \quad z = \theta_0 + \frac{\theta}{\sqrt{n}} \text{ in } \pi_{\theta_0, z}$$

$$D_n(\theta_0, \theta) := D_{n,n}(\theta_0, \theta) \quad (2.2.1)$$

[cf. (2.1.7)]. In other words, the dependence on the true system z has been made explicit in the definition of $D_{n,m}(\theta_0, \theta)$. We shall now describe the asymptotic behavior, when the length n of the record tends to infinity, of this cumulative sum under the three hypotheses H_0 , H_1 , and H_1' .

1) *Behavior of the Cumulative Sum Under the Hypothesis of No Change:* This behavior is described in the following theorem.

Theorem 1: i) Behavior of the marginal distribution:

$$D_n(\theta_0, 0) \xrightarrow[n \rightarrow \infty]{L} N(0, R(\theta_0)) \quad (2.2.2)$$

where $R(\theta)$ is given by

$$R(\theta) := \sum_{n=-\infty}^{+\infty} \text{cov}_{\theta, \theta} [H(\theta; X_n), H(\theta; X_0)]. \quad (2.2.3)$$

$\hat{P}_{\theta, \theta}$ denotes the steady-state law of X_n when the adjustable parameter and the true system are both equal to θ , and $\text{cov}_{\theta, \theta}$ is the covariance with respect to $\hat{P}_{\theta, \theta}$.

ii) *Invariance Principle:* For $t \in [0, 1]$, set

$$D_{n,t}(\theta_0, \theta) := D_{n,m}(\theta_0, \theta) \text{ where } m = [nt]. \quad (2.2.4)$$

Then,

$$[R(\theta_0)]^{-1/2} \cdot \{D_{n,t}(\theta_0, 0)\}_{0 \leq t \leq 1} \rightarrow (W_t)_{0 \leq t \leq 1} \quad (2.2.5)$$

when n tends to infinity, where (W_t) is a *Brownian motion*, and \rightarrow denotes the weak convergence of processes. ■

Proof: Of course i) is a consequence of ii). See the Appendix for a proof of this theorem and precise assumptions on the random vector field (2.1.7).

2) *Behavior of the Cumulative Sum Under the Hypothesis of Change:* We shall directly investigate H_1' , since H_1 is a subcase of the former hypothesis. Consider $\tau \in [0, 1]$, and let us introduce for $m \leq n$ the following cumulative sum:

$$D_{n,m}(\theta_0, \theta, \tau) := \frac{1}{\sqrt{n}} \sum_{k=1}^m H(\theta_0; X_k),$$

$$z = \theta_0 \text{ for } k \leq \min(m, [n\tau])$$

$$z = \theta_0 + n^{-1/2}\theta \text{ for } k > \min(m, [n\tau]) \quad (2.2.6)$$

in $\pi_{\theta_0, z}$, cf. (2.1.7). This cumulative sum reflects the effect on the vector field H of a deviation of magnitude order $n^{-1/2}$ between the nominal model and the true system, which occurred at time $n\tau$. The behavior of this cumulative sum is described in the following theorem.

Theorem 2: Behavior of the cumulative sum under the hypothesis H_1' . Let $\tau \in [0, 1]$ be given. Set

$$D_{n,t}(\theta_0, \theta, \tau) := D_{n,m}(\theta_0, \theta, \tau) \text{ where } m = [nt]. \quad (2.2.7)$$

Then, when n tends to infinity, the process $\{D_{n,t}(\theta_0, \theta, \tau)\}_{0 \leq t \leq 1}$ converges weakly towards the process $\{D_t(\theta_0, \theta, \tau)\}_{0 \leq t \leq 1}$, solution of the linear stochastic differential equation

$$dD_t = -1_{\{t \geq \tau\}} \cdot h_\theta \cdot \theta dt + R^{1/2} \cdot dW_t \quad (2.2.8)$$

where $R = R(\theta_0)$ is given in (2.2.3), while $h_\theta = h_\theta(\theta_0, \theta_0)$ is defined in (2.1.11).

Corollary 3: Hypothesis H_1

$$D_n(\theta_0, \theta) \xrightarrow[n \rightarrow \infty]{L} N[-h_\theta \cdot \theta, R(\theta_0)]. \quad (2.2.9)$$

This corollary is directly carried out from Theorem 2 by taking $\tau = 0$.

Proof: See the Appendix.

III. LOCAL CHANGE DETECTION TECHNIQUES

A. The Local Test

From the user's point of view, the cumulative sums $D_{n,m}$ given by the formulas (2.2.1) or (2.2.6) are identical, since they differ only via a change on the true parameter z , which is not observed by the user. These cumulative sums will be from now on denoted as follows:

$$D_{n,m}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{k=1}^m Y_k(\theta_0), \quad Y_k(\theta_0) := H(\theta_0; X_k). \quad (3.1.1)$$

We shall interpret Theorem 2 as follows. Assume a change of magnitude order $n^{-1/2}$ occurred at time r in the direction of change θ , and n is large enough. Then, *considering the random variables $Y_k(\theta_0)$ as independent, and distributed as follows:*

$$Y_k(\theta_0) \approx N[0, R(\theta_0)], \quad k < r$$

$$Y_k(\theta_0) \approx N[-h_\theta(\theta_0) \cdot \theta, R(\theta_0)], \quad k \geq r \quad (3.1.2)$$

would exactly result in the asymptotic behavior described by (2.2.8). Hence, we shall replace the original testing problem (H_0 against H_1') by the asymptotically equivalent problem of detecting changes in the mean of the independent Gaussian variables according to (3.1.1), (3.1.2). Restricting the study to the case where the direction of change θ is unknown, we shall apply the formulas (1.3.3)-(1.3.7) to the detection of a change like (3.1.2). This gives the formulas of the *local change detection procedure*:

$$S_r^n(\theta) = 2 \sum_{k=r}^n [Y_k^T R^{-1} h_\theta \cdot \theta] - (n-r+1)\theta^T \cdot h_\theta^T R^{-1} h_\theta \cdot \theta \quad (3.1.3)$$

where θ_0 has been deleted for simplicity. Fixing r and maximizing with respect to θ yields

$$S_r^n := \max_{\theta} S_r^n(\theta) = (\Delta_r^n)^T R^{-1} \Delta_r^n,$$

$$\Delta_r^n := (n-r+1)^{-1/2} \sum_{k=r}^n Y_k,$$

$$\hat{\theta}(n, r) := \arg \max_{\theta} S_r^n(\theta) = -(n-r+1)^{-1/2} h_\theta^{-1} \Delta_r^n. \quad (3.1.4)$$

The stopping rule and the estimates of the change time and magnitude are given by the formulas

$$\begin{aligned} \text{i) } G_n &:= \max_r S_r^n, \nu = \min \{n : G_n \geq \lambda\} \\ \text{ii) } r_n &= \arg \max_r S_r^n, \hat{r} = \hat{r}_r, \hat{\theta} = \hat{\theta}(v, \hat{r}). \end{aligned} \quad (3.1.5)$$

Note that (3.1.5-i) is sufficient if the change detection only is of interest. The local test is given by the formulas (3.1.1), (3.1.4), (3.1.5). The threshold λ is easily selected by knowing that, under the no change hypothesis, we have

$$E_{\theta_0}(S_r^n) = d, \quad (3.1.6)$$

since $(n - r + 1)S_r^n$ is (approximately) a central χ^2 with $d(n - r + 1)$ degrees of freedom.

Comments: 1) A probably more commonly used method is the following one.

i) Run the adaptive identification algorithm with constant gain

$$\theta_n = \theta_{n-1} + \gamma [H(\theta_{n-1}, X_n)].$$

ii) Use a χ^2 -test of the form

$$(\theta_n - \theta_0)^T \Sigma^{-1} (\theta_n - \theta_0) \geq \lambda \quad (3.1.7)$$

with a suitable matrix Σ using the fact that $\theta_n - \theta_0$ is approximately Gaussian and zero mean for γ small in the hypothesis of no change (use a central limit theorem for stochastic approximations to select the proper matrix Σ ; see, for example [10] and [7]). This latter method is in fact far from being as efficient as our method. It is known in fact that the deviation $\theta_n - \theta_0$ is quite complex: this deviation behaves like a first-order Gaussian Markov process [10]. But in this case, it is known [22] that the best local test involves the innovations of this Markov process, which gives something different from (3.1.7). Our method is precisely the right way to test for small changes in the true system. We shall now illustrate this method on two nontrivial typical examples, and show that it is the *convenient generalization of the local likelihood ratio tests* introduced by Le Cam ([18], [19], [9]); other examples can be found in [7].

2) The formulas (3.1.1), (3.1.4), (3.1.5) *do define the local test*: these formulas have to be applied in practice *regardless of any asymptotic consideration* (i.e., no small change is needed, nor a large delay for detection). The only purpose of the asymptotic framework was to provide guidelines for the design of local tests. On the other hand, this framework is also used to show that (at least for local likelihood ratio model validation techniques, see [19]) local methods are asymptotically¹ uniformly most powerful, which justify their use from the theoretical viewpoint. But let us again emphasize that these assumptions no longer have to be taken into account in practice.

B. Examples

1) *Change Detection in AR Processes:* The objective is to detect changes in the parameter θ in the system

$$y_n = \phi_n^T \theta + v_n, \phi_n^T = (y_{n-1}, \dots, y_{n-p}). \quad (3.2.1)$$

We apply our method with the random vector field of the classical least-squares algorithm, namely

$$H(\theta, y_n, \phi_n) := \phi_n e_n(\theta), e_n(\theta) := y_n - \phi_n^T \theta. \quad (3.2.2)$$

The matrix $R(\theta_0)$ corresponding to (2.2.3) is given by

$$R(\theta_0) = E_{\theta_0}(v_n \phi_n \phi_n^T v_n) = \sigma^2 \Sigma(\theta_0) \quad (3.2.3)$$

¹ Here, "asymptotically" means for long sample and small change.

where σ^2 is the variance of v_n and $\Sigma(\theta_0)$ the covariance of the regression vector ϕ_n for the nominal model θ_0 . This gives

$$\Delta_r^n(\theta_0) = (n - r + 1)^{-1/2} \sum_{k=r}^n \phi_k e_k(\theta_0). \quad (3.2.4)$$

It is easy to verify that $\sigma^{-2} \Delta_r^n(\theta_0)$ is the derivative with respect to θ of the log likelihood of the sample y_1, \dots, y_n under H_0 , while $\sigma^{-2} \Sigma(\theta_0)$ is the Fisher information matrix. Comparing the obtained procedure with [18] and [8] reveals that (3.1.1), (3.1.4), (3.1.5), (3.2.3), (3.2.4) yields the so-called *local likelihood ratio test*, which is the convenient procedure to detect small changes in the parameters of an AR process.

2) *Detecting Changes in the Poles of an ARMA Process with the Instrumental Local Test:* This example is much more interesting, since we shall derive with our method a new test, which is nonclassical, and has been proposed and analyzed in detail in [3] and [16]. Consider an ARMA process of the form

$$y_n = \sum_{i=1}^p a_i y_{n-i} + \sum_{j=1}^q b_j v_{n-j} + v_n \quad (3.2.5)$$

where (v_n) is a white noise. *Our purpose is to monitor possible changes in the AR parameters, while considering the MA parameters as nuisance parameters.* This is recognized as a difficult problem, since the poles and zeros of an ARMA process are tightly coupled (the Fisher information matrix exhibits coupling between AR and MA parameters). However, the instrumental variable (I.V.) method is known to be an identification procedure which satisfies our robustness requirements; for example, it is shown in [6] that the AR parameters can be consistently identified with the I.V. method even if the MA parameters are time-varying. Recall briefly this method [20]. Setting

$$\begin{aligned} \theta^T &:= (a_1, \dots, a_p) \\ \phi_n^T &:= (y_{n-1}, \dots, y_{n-p}) \\ \psi_n^T &:= (y_{n-q-1}, \dots, y_{n-q-p}) \end{aligned} \quad (3.2.6)$$

where ψ_n is the instrument, the I.V. method is given by

$$\begin{aligned} \theta_n &= \theta_{n-1} + \frac{1}{n} \Gamma_n^{-1} \psi_n (y_n - \phi_n^T \theta_{n-1}) \\ \Gamma_n &= \Gamma_{n-1} + \frac{1}{n} (\psi_n \phi_n^T - \Gamma_{n-1}). \end{aligned} \quad (3.2.7)$$

The random vector field of interest is here equal to

$$H(\theta_0; \psi_n, \phi_n, y_n) = \psi_n (y_n - \phi_n^T \theta_0) := Y_n(\theta_0) \quad (3.2.8)$$

where θ_0 is the nominal model. To apply our method, we must calculate the matrices

$$\begin{aligned} R(\theta_0) &:= \sum_{n \in Z} E_0 [Y_n(\theta_0) Y_0(\theta_0)^T] \\ &= \sum_{n=-q}^{+q} E_0 [\psi_n \psi_0^T (y_n - \phi_n^T \theta_0) (y_0 - \phi_0^T \theta_0)] \\ h_\theta(\theta_0) &= -E_0(\psi_n \phi_n^T) \end{aligned} \quad (3.2.9)$$

where E_0 is a shorthand for E_{θ_0} . The *instrumental test* is obtained by combining the formulas (3.1.1), (3.1.4), (3.1.5), (3.2.8), (3.2.9). As expected, this test exhibits very pleasant robustness properties with respect to the nuisance MA parameters: for instance, it is proved in [16] that the instrumental test does

effectively detect changes in the AR parameters, while ignoring possible changes in the MA parameters, a property which is certainly not satisfied by the likelihood ratio tests associated to ARMA processes! Hence, our general method allowed us to derive a new, nonclassical method of change detection. It is not our purpose here to discuss the details of practical implementations. The interested reader is referred to [3] for further details.

IV. LOCAL MODEL VALIDATION TECHNIQUES

Model validation is often considered as a must to verify the relevance of a given identification procedure for a given plant or system; but in fact model validation is also useful for other purposes. The main purposes of model validation can be indeed classified in the two following classes.

1) Verify that the selected model set is convenient to capture the behavior of a given plant or system. The usual procedure is then as follows: first identify the best model within the considered model set, second check (via model validation techniques) if the so obtained model describes the true system in a satisfactory way. For example, model validation is certainly needed when oversimple model sets are used to identify a given system.

2) Try to detect a possible inconsistency of a given system with respect to a given prespecified behavior; this *a priori* behavior could be a specification of the designer, or could also be the result of a previous identification performed when the system was supposed to be safe. Furthermore, perform a diagnosis on the origin of the possible changes. This second point of view was for example taken in [3], [4] for monitoring offshore structures subject to vibrations.

We shall now show that our local framework does apply for model validation as well, and further leads to theoretically sound model validation techniques, compared to the classically used ones.

A. Problem Statement

We use the same framework as for change detection. Consider a semi-Markov process (X_n) with conditional density parametrized by θ_* according to the formulas

$$P\{\xi_n \in G | \xi_{n-1}, \xi_{n-2}, \dots\} = \int_G \pi_{\theta_*}(\xi_{n-1}, dx)$$

$$X_n = f(\xi_n) \tag{4.1.1}$$

where $\pi_{\theta_*}(\xi, dx)$ is the transition probability parametrized by θ_* of the Markov chain (ξ_n) . The model validation problem is then simply formulated as the following hypothesis testing problem.

VM: Given a *nominal model* θ_0 , decide between the hypotheses

- $\theta_* = \theta_0$: the nominal model is valid
- $\theta_* \neq \theta_0$: the nominal model is not valid.

Here, θ_0 denotes the model to be validated (nominal model), while θ_* denotes the (*unknown*) true system corresponding to the observed record X_1, \dots, X_N .

B. Local Validation Method Associated to an Adaptive Algorithm

The problem VM is formulated in the local framework as follows. We consider again as in (2.1.7) the random vector field associated to the adaptive algorithm:

$$H(\theta; X_n), \pi_{\theta, z}(\xi, dx), f \tag{4.2.1}$$

where

- θ is the nominal model, chosen by the user;

- z denotes the true system, which is not available to the user;
- X_n , the state at time n , is a random semi-Markov process with transition matrix controlled by the pair (θ, z) .

Again Assumption NS is in force throughout this section. The problem VM is then formulated as follows: assume that a nominal model θ_0 and a record X_1, \dots, X_n of the state vector are given. It is desired, by monitoring the random vector field $(H(\theta_0; X_k))_{1 \leq k \leq n}$, to decide between the hypotheses:

$$H_0 : z = \theta_0$$

$$H_1 : z = \theta_0 + \frac{\theta}{\sqrt{n}} \text{ where } \theta \neq 0 \text{ is a fixed but unknown change.}$$

It is then clear that the problem VM is a particular case of the change detection problem we have studied before: just set $r = 0$ and delete the maximization with respect to r in the formulas (3.1.3), (3.1.4). This gives the following formulas, where the explicit dependence on θ_0 was deleted for the sake of simplicity. For a fixed θ , the log likelihood ratio between the hypotheses H_1 and H_0 is given by

$$S_n(\theta) = 2 \sum_{k=1}^n [Y_k^T R^{-1} h_\theta \cdot \theta] - n \theta^T \cdot h_\theta^T R^{-1} h_\theta \cdot \theta \tag{4.2.2}$$

where Y is defined in (3.1.1). Finally, the following χ^2 test is derived

$$\Delta_n^T R^{-1} \Delta_n \geq \lambda, \Delta_n = \frac{1}{n} \sum_{k=1}^n Y_k. \tag{4.2.3}$$

Again, the threshold λ can be chosen by using the fact that, under the hypothesis H_0 of validity of the nominal model θ_0 , $n \Delta_n^T R^{-1} \Delta_n$ is a χ^2 with n degrees of freedom. The same examples as before can be used to illustrate our method. See [4] for an extensive application of the instrumental test as a model validation method to check fatigues or failures in a vibrating structure. Furthermore, other examples are proposed as exercises in the book [7].

V. DIAGNOSIS

The diagnosis problem can be formulated for change detection as well as for model validation. Its goal is to identify the most likely origin of the change among a list of *a priori* fixed directions of change. For example, it can be desired to monitor a specific pole of a high order ARMA process. Furthermore, it can be desired to recognize the origin of the changes in terms of parametrizations which are not identifiable (this is, for example, the case when nonidentifiable physical models are referred to for the diagnosis). Let us first investigate this problem on a simplified example, borrowed from [4] and [17].

A. The Example of Vibrating Systems

Consider a vibrating system subject to external disturbances, and monitored via accelerometers or strain gauges. A relevant finite element approximation yields the continuous-time linear model

$$MZ + CZ + KZ = E \quad Z, E \in R^D$$

$$Y = LZ \quad Y \in R^d, d \ll D \tag{5.1.1}$$

where Z is the vector of the displacements (or accelerations), M is the mass matrix, C specifies the damping, and K is the stiffness matrix; E is the excitation, and the second equation expresses the fact that only a few components of the state are observed. We assume that the excitation is a white noise with a given unknown covariance matrix. The discrete-time version of (5.1.1) is, in

state-space form

$$\begin{aligned} X_{n+1} &= FX_n + V_{n+1} \\ Y_n &= HX_n \end{aligned} \quad (5.1.2)$$

(5.1.1) and (5.1.2) are related as follows:

$$\lambda(F) = \exp[\tau \cdot \lambda(Ms^2 + Cs + K)], \quad H\phi_\lambda = L\psi_\lambda \quad (5.1.3)$$

where τ is the discretization step, $\lambda(\cdot)$ denotes an eigenvalue, ϕ_λ the corresponding eigenvector of F , and ψ_λ the eigenvector of the polynomial matrix $(Ms^2 + Cs + K)$ associated to the eigenvalue $\exp(\tau\lambda)$. Finally, the eigenstructure of (5.1.2) reflects the eigenstructure of the vibrating system (5.1.1), while the excitation noise V_n is a byproduct of the excitation E . The problem is to diagnose the origin of changes in the vibrating characteristics of the system by simply monitoring the signal Y . Such a diagnosis can be formulated

- in the *modal* domain: the corresponding parametrization is in terms of the observed eigenstructure of the system, and is identifiable [because of (5.1.3)] if the model (5.1.2) is minimal.

- in the *physical* domain: the corresponding parametrization is in terms of the physical parameters M, C, K (or of related ones), and is generally not identifiable.

The two problems are not of the same difficulty. In the first case, it is for example possible to use the formula (3.1.4) in order to estimate the most likely change θ , since the corresponding parametrization is identifiable, while it is not possible to do it in the second case, since the corresponding model set is not identifiable. This is even more true if a model reduction is performed when replacing (5.1.1) by (5.1.2), which is always the case in practice. We shall thus propose a suitable general approach to solve this problem.

B. Some Prerequisites

Recall the following elementary result in Gaussian hypothesis testing. Let U be a random variable distributed as $N(\mu, \Sigma)$. Consider a full column-rank matrix M ; for testing $\mu = 0$ against $\mu = M\nu$, where $\nu \neq 0$, the log likelihood ratio is

$$\begin{aligned} T &= -\frac{1}{2} (U - M\nu)^T \Sigma^{-1} (U - M\nu) + \frac{1}{2} U^T \Sigma^{-1} U \\ &= U^T \Sigma^{-1} M\nu - \frac{1}{2} \nu^T M^T \Sigma^{-1} M\nu. \end{aligned}$$

The maximum likelihood estimate of ν is

$$\hat{\nu} = (M^T \Sigma^{-1} M)^{-1} M^T \Sigma^{-1} U.$$

So that we get the following χ^2 statistics to test $\mu = 0$ against $\mu = M\nu$, $\nu \neq 0$

$$\chi = U^T \Sigma^{-1} M [M^T \Sigma^{-1} M]^{-1} M^T \Sigma^{-1} U. \quad (5.2.1)$$

We shall now apply these elementary formulas to our asymptotic local approach. For simplicity, we shall investigate the case of model validation only, but it should be clear that the whole analysis extends to change detection as well.

C. Diagnosis of Changes on Identifiable Model Sets

Going back to the statement of the local method for model validation, the general diagnosis problem is formulated as follows. Assume that a nominal model θ_0 and a record X_1, \dots, X_n of the state vector are given. It is desired, by monitoring the random vector field $(H(\theta_0; X_k))_{1 \leq k \leq n}$, to decide between the

hypotheses:

$$H_0 : z = \theta_0$$

$$H_1 : z = \theta_0 + \frac{\theta}{\sqrt{n}}, \text{ where } \theta \neq 0 \text{ is a fixed but unknown change}$$

inside some specified subspace Θ .

Choose a full column rank matrix M_Θ such that $\Theta = \text{range}(M_\Theta)$; keeping (3.1.2) in mind, we can directly use the formula (5.2.1) to get the suitable version of the formula (4.2.3)

$$\begin{aligned} \Delta_n^T R^{-1} M [M^T R^{-1} M]^{-1} M^T R^{-1} \Delta_n &\cong \lambda \\ M &= h_\theta(\theta_0) \cdot M_\Theta \end{aligned} \quad (5.3.1)$$

where Δ_n is defined in (4.2.3). This can be used as follows. If Θ is the subspace spanned by a subset of the coordinates of the parameter θ , choose M_Θ as being the corresponding projection matrix. If Θ is the subspace spanned by a subset of the coordinates of another identifiable parameter set $\Psi = \{\psi\}$ such that there exists a change of coordinates

$$\theta = F(\psi) \quad (5.3.2)$$

with F being a local diffeomorphism, build M_Θ as being the corresponding subset of the columns of the Jacobian matrix

$$F'(\psi_0) \quad (5.3.3)$$

where ψ_0 is the nominal model in the parameter set Ψ . This was, for example, the approach taken in [3], [4] for diagnosing changes in specified poles of a vector ARMA process.

D. Diagnosis on Nonidentifiable Model Sets

1) A Rough Procedure: Assume now the parameter set Ψ to be monitored is not identifiable, which means that F in (5.3.2) is C^1 and onto, but is *not a diffeomorphism* (since it is not bijective!). Then, it is no more possible to apply the preceding method, since the nominal model ψ_0 used in (5.3.3) to compute the Jacobian cannot be derived from θ_0 . The suitable modification of the method is then the following.

Step 1: Select a nominal model θ_0 in the (identifiable) parameter set of the adaptive algorithm. For example, the nominal model can be simply identified, say, on a previous reference record.

Step 2: We assume that a coarse nominal ψ_0 is available:

$$\theta_0 \approx F(\psi_0) \quad (5.4.1)$$

(this is, for example, the case when ψ_0 is an approximate physical model of the system provided by the designer). We compute the Jacobian $F'(\psi_0)$ at this coarse nominal model, and we select the proper columns. This provides us with coarse directions of change to be monitored, which are sufficiently accurate in practice.

Since the parameter space Ψ is assumed to be much larger than the identifiable parameter space Θ , the selection of proper subspaces in Ψ must be taken with some care; otherwise the selection rule (5.3.2) could lead to subspaces $\Theta_1, \Theta_2, \dots$ which might not be distinguished by our testing method. To overcome this difficulty, the following method is proposed in [17] and currently implemented for the application to vibration monitoring.

2) An Improved Procedure: Step 1 of the former procedure is kept. Step 2 is refined as follows.

Step 2.1: Assume a coarse nominal model ψ_0 in the sense of (5.4.1) is available; select candidate changes $\delta\psi_1, \delta\psi_2, \dots$ to be monitored in the large parameter space Ψ and compute the corresponding changes

$$\delta\theta_i = F'(\Psi_0) \cdot \delta\psi_i \quad (5.4.2)$$

in the identifiable parameter space Θ . Let us emphasize that the magnitude of these changes is of importance: some realistic changes in the parameter set Ψ may result in too small changes in the identifiable parameter set Θ , so that there is no hope to detect them. Consequently, such changes have to be removed from the monitoring procedure, this is the purpose of the next step.

Step 2.2: Selection of the monitorable changes. Given a change $\delta\theta$ obtained via the Step 2.1, the noncentrality parameter of the χ^2 statistics (5.3.1) corresponding to $M_\Theta = \delta\theta$ is equal to

$$\delta\theta^T \cdot h_\theta^T(\theta_0) \cdot R^{-1} \cdot h_\theta(\theta_0) \cdot \delta\theta. \tag{5.4.3}$$

Compare to a chosen threshold this noncentrality parameter for all the $\delta\theta_i$'s obtained in Step 2.1, and keep only those changes for which the threshold is exceeded: these are the monitorable changes.

Step 2.3: Perform any clustering method to replace by a single unit vector $\tilde{\theta}$ each subset of $\delta\theta_i$'s that are close to each other in the sense of the following distance measure (we write θ for short instead of $\delta\theta$, and we remove the dependence on the nominal model θ_0):

$$d(\theta_i, \theta_j) = 1 - \rho_{i,j}$$

$$\rho_{i,j}^2 = \frac{(\theta_i^T h_\theta^T R^{-1} h_\theta \theta_j)^2}{(\theta_i^T h_\theta^T R^{-1} h_\theta \theta_i)(\theta_j^T h_\theta^T R^{-1} h_\theta \theta_j)}$$

Note that $\rho_{i,j}$ is nothing but the cosine of the angle between the subspaces spanned by the vectors θ_i and θ_j . The final subspaces to be monitored are the spaces spanned by the unit vectors $\tilde{\theta}$ obtained in Step 2.3: denote these subspaces by $\tilde{\Theta}_1, \dots, \tilde{\Theta}_N$, they are guaranteed to correspond to hypotheses which are both monitorable and distinguishable from each other.

Step 2.4: Monitor in parallel all the corresponding χ^2 -statistics to decide the origin of the possible changes in terms of the large, nonidentifiable, parameter set Ψ . The resulting procedure can allow the diagnosis of small changes with a good accuracy.

This method is currently implemented on the vibrating system example mentioned above. Primary results show that² 1) the clustering step produces classes that are coherent from the mechanical viewpoint, 2) the local tests built according to the described procedure do allow the isolation of changes occurring in the considered classes.

VI. CONCLUSION

We have introduced a general method to associate to any identification procedure a change detection and a model validation procedure. This general approach is based on the so-called asymptotic local approach used in the area of statistics as a tool to analyze or design likelihood ratio testing procedures. Our method extends the former one to procedures which are no more based on likelihood ratios. Furthermore, this method provides as a direct byproduct correctly sounded procedures for the diagnosis of the origin of the changes, even when those changes are formulated in terms of (nonidentifiable) larger models. The method was illustrated on two typical examples: the least-squares algorithm, where the classical local likelihood ratio approach was rederived in this way, and the instrumental test, a procedure, recently proposed by the present authors, which is associated to the well-known instrumental variable method, and was used on a significant practical application in vibration mechanics.

APPENDIX A

ASSUMPTIONS AND PROOFS

In this Appendix, we give the assumptions and proofs of Theorems 1 and 2. The basic references are the three papers by Mac Leish given hereafter.

² When no numerical ill conditioning occurs, such as encountered on a single one among several simulation examples.

A. Mixingales, and a Theorem of Mac Leish

Mixingales: Mixingales have been introduced in [12] and are probably the largest class of dependent random variables. Let $\{\Omega, F, F_n, P\}$ a probability space endowed with an increasing discrete-time indexed family of σ -algebras. The conditional expectation $E(U/F_n)$ will be denoted for short by $E_n U$, and $\|\cdot\|_2$ will denote the L^2 -norm.

Definition A.1: The sequence of random variables Z_n is a *mixingale* if, for sequences of finite nonnegative constants c_n and ψ_m where $\psi_m \rightarrow 0$ as $m \rightarrow \infty$, we have for all $n \geq 1, m \geq 0$

$$\begin{aligned} \text{a) } & \|E_{n-m} Z_n\|_2 \leq \psi_m c_n \\ \text{b) } & \|Z_n - E_{n+m} Z_n\|_2 \leq \psi_{m+1} c_n. \end{aligned} \tag{6.4.1}$$

For example, $\psi_m \equiv 0$ corresponds to martingales. Strong laws and invariance principles are given in [12], [13] for mixingales with exponentially vanishing ψ_m 's. The class of mixingales contains all the classes of mixing processes, and also the functions of mixing processes introduced by Billingsley. However, we shall need stronger results, namely invariance principles for *triangular arrays of mixingales*; such results are found in [14] and will be stated now.

Triangular Arrays of Mixingales: Let $\{Z_{n,i}\}_{i=1,2,\dots,n=1,2,\dots}$ be a double array of zero mean random variables defined on a probability space $\{\Omega, F, P\}$. Let $k_n(t)$ be a sequence of nonrandom integer valued, nondecreasing, right continuous functions on $[0, \infty)$. We form a random function

$$W_n(t) = \sum_{i=1}^{k_n(t)} Z_{n,i} \tag{6.4.2}$$

and we wish to show the weak convergence of $W_n(t)$ to a standard Wiener process W . Suppose

$$\sigma_{n,i}^2 = \sum_j E(Z_{n,i} Z_{n,j}) \tag{6.4.3}$$

does exist, and is such that the following conditions hold for each $T < \infty$:

$$\begin{aligned} \text{a) } & \sup_{s < t < T} \limsup_{n \rightarrow \infty} \frac{\sum_{i=1}^{k_n(t)} \sigma_{ni}^2}{t-s} < \infty \\ \text{b) } & \left\{ \frac{Z_{ni}^2}{\sigma_{ni}^2} \right\}_{n=1,2,\dots,i \leq k_n(T)} \text{ is uniformly integrable} \\ \text{c) } & \max_{i \leq k_n(T)} \sigma_{n,i} \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned} \tag{6.4.4}$$

These conditions are sufficient for the Lindeberg condition to hold for the considered array. We also require that the sequence $\{Z_{n,i}\}_{i=1,2,\dots}$ constitutes a mixingale with respect to some double array of increasing σ -algebras $F_{n,i}$, namely

$$\begin{aligned} \|E(Z_{ni} | F_{n,i-k})\|_2 & \leq \psi_k \sigma_{ni} \\ \|Z_{ni} - E(Z_{ni} | F_{n,i+k})\|_2 & \leq \psi_{k+1} \sigma_{ni}. \end{aligned} \tag{6.4.5}$$

Theorem [14]: Suppose conditions (6.4.4), (6.4.5) are in force, $\psi_k \downarrow 0$ exponentially fast, and, for each $s < t < u$

$$E \left| E \left\{ \left(\sum_{i=k_n(s)}^{k_n(u)} Z_{ni} \right)^2 \middle| F_{n,k_n(s)} \right\} - (u-s) \right| \rightarrow 0 \tag{6.4.6}$$

as $n \rightarrow \infty$, then W_n converges weakly to a standard Wiener process.

B. Assumptions for Theorems 1 and 2

To apply the theorem of Mac Leish, we choose

$$\begin{aligned}
 z_n &= \theta_0 + n^{-1/2}\theta \\
 P &= P_{\theta_0, z_n} \\
 Z_{n,i} &= n^{-1/2}\lambda^T [R(\theta_0, z_n)]^{-1/2} [H(\theta_0, X_i) - E[H(\theta_0, X_i)]] \\
 F_{n,i} &= \sigma \{ \xi_k : k \leq i \} \\
 T &= 1, k_n(t) = t
 \end{aligned} \tag{6.4.7}$$

where λ is an arbitrary unit vector, and

$$R(\theta, z) := \sum_{n=-\infty}^{+\infty} c\hat{v}_{\theta, z}[H(\theta; X_n), H(\theta; X_0)].$$

The conditions (6.4.4) are obviously satisfied. We shall furthermore assume that the conditions (6.4.5), (6.4.6) hold, with constants ψ_k that are uniform in λ and n large enough, and that the function $z \rightarrow R(\theta_0, z)$ is continuous and bounded from below. These are very weak conditions: stationarity of (X_n) is not required, discontinuities are allowed on the random vector field $\theta \rightarrow H(\theta, X_n)$, and the mixingale condition is satisfied for every reasonable Markov chain ξ_n . Verification of the assumptions (6.4.5), (6.4.6) can be a heavy task on difficult examples, but such a verification is beyond the scope of the present paper. See, for example, [5] for the verification of the stronger function of ϕ -mixing condition on nontrivial examples.

C. Proofs

Let us begin with the proof of Theorem 1. First, notice that

$$E_{\theta_0, z_n}[H(\theta_0, X_i)] \approx h(\theta_0, z_n) \tag{6.4.8}$$

for n and i large.

Using the notations of (2.2.4), the theorem of Mac Leish implies that

$$\lambda^T [R(\theta_0)]^{-1/2} \cdot \{ D_{n,i}(\theta_0, 0) \}_{0 \leq i \leq 1} \rightarrow \lambda^T (W_t)_{0 \leq t \leq 1}. \tag{6.4.9}$$

Since λ is arbitrary, this proves the theorem. In fact, a simple invariance principle for mixingales would have sufficed for Theorem 1.

The proof of Theorem 2 requires the stronger result on arrays of mixingales. It is obviously sufficient to prove Theorem 2 for $\tau = 0$. We shall write $D_{n,i}(\theta)$ for short instead of $D_{n,i}(\theta_0, \theta, \tau)$, and delete the dependency on θ_0 when no confusion can occur. Taking into account (6.4.8), the theorem of Mac Leish implies here that

$$\begin{aligned}
 \lambda^T [R(\theta_0)]^{-1/2} \left[D_{n,i}(\theta) - n^{-1/2} \sum_{k=1}^{[nt]} h(\theta_0, z_n) \right]_{0 \leq t \leq 1} \\
 \rightarrow \lambda^T (W_t)_{0 \leq t \leq 1}. \tag{6.4.10}
 \end{aligned}$$

Finally, a first-order Taylor expansion

$$\begin{aligned}
 h(\theta_0, z_n) &\approx h(\theta_0, \theta_0) + n^{-1/2} h_z(\theta_0 \cdot \theta_0), \theta \\
 &= -n^{-1/2} h_\theta(\theta_0, \theta_0) \cdot \theta
 \end{aligned}$$

gives Theorem 2.

REFERENCES

- [1] M. Basseville, "Detection of changes in signals and systems," presented at the 2nd IFAC Workshop on Adaptive Systems in Control and Signal Processing, Lund, Sweden, 1986.
- [2] M. Basseville and A. Benveniste, Eds., *Detection of Abrupt Changes in Signals and Dynamical Systems* (Lecture Notes in Control and Information Sciences, Vol. 77). New York: Springer-Verlag, 1986.
- [3] M. Basseville, A. Benveniste, and G. V. Moustakides, "Detection and diagnosis of abrupt changes in modal characteristics of nonstationary digital signals," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 412-417, 1986.
- [4] M. Basseville, A. Benveniste, G. V. Moustakides, A. Rougee, "Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems," Res. Rep. IRISA 276/INRIA, no. 477; also in *Automatica*, to be published.
- [5] A. Benveniste, M. Goursat, G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms, with applications to data communication algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 1042-1058, 1980.
- [6] A. Benveniste and J. J. Fuchs, "Single sample modal identification of a nonstationary stochastic process," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 66-74, 1985.
- [7] A. Benveniste, M. Metivier, and P. Priouret, *Algorithmes Adaptatifs et Approximations Stochastiques* (Techniques Stochastiques). Paris: Masson, 1987.
- [8] R. B. Davies, "Asymptotic inference in stationary Gaussian time series," *Adv. Appl. Prob.*, vol. 5, no. 3, pp. 469-497, 1973.
- [9] J. Deshayes and D. Picard, *Off-Line Analysis of Change-Point Models Using Nonparametric and Likelihood Methods* (Lecture Notes in Control and Information Sciences, Vol. 77). New York: Springer-Verlag, 1986.
- [10] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic System Theory*. Cambridge, MA: M.I.T. Press, 1984.
- [11] L. L. Ljung and T. Soderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: M.I.T. Press, 1983.
- [12] D. L. Mac Leish, "A maximal inequality and dependent strong laws," *Ann. Probability*, vol. 3, no. 5, pp. 829-839, 1975.
- [13] —, "Invariance principles for dependent variables," *Zeitschrift fur Wahrscheinlichkeitstheorie verw. Geb.* vol. 32, pp. 165-178, 1975.
- [14] —, "On the invariance principle for nonstationary mixingales," *Ann. Probability*, vol. 5, no. 4, pp. 616-621, 1977.
- [15] M. Metivier and P. Priouret, "Application of a lemma of Kushner and Clark to general classes of stochastic algorithms," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 140-150, 1984.
- [16] G. V. Moustakides and A. Benveniste, "Detecting changes in the AR parameters of a nonstationary ARMA process," *Stochastics*, vol. 16, pp. 137-155, 1986.
- [17] G. V. Moustakides, "The problem of diagnosis with respect to physical parameters for changes in structures," *IRISA Tech. Rep.* 295, 1986.
- [18] I. V. Nikiforov, *Sequential Detection of Changes in Stochastic Systems* (Lecture Notes in Control and Information Sciences, Vol. 77). New York: Springer-Verlag, 1986.
- [19] G. G. Roussas, *Contiguity of Probability Measures, Some Applications in Statistics*. Cambridge, MA: Cambridge University Press, 1972.
- [20] P. Stoica, T. Soderström, and B. Friedlander, "Optimal instrumental variable estimates of the AR parameters of an ARMA process," *IEEE Trans. Automat. Contr.*, vol. AC-30, no. 11, pp. 1066-1074, 1985.
- [21] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601-611, 1976.
- [22] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to detection and estimation of jumps in linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 108-112, 1976.
- [23] A. S. Willsky, *Detection of Abrupt Changes in Dynamic Systems* (Lecture Notes in Control and Information Sciences, Vol. 77). New York: Springer-Verlag, 1986.



Albert Benveniste (M'81) was born in Paris, France, in 1949. He graduated from the Ecole des Mines de Paris in 1971.

From 1971 to 1973, he was with the Centre d'Automatique de l'Ecole des Mines, Fontainebleau. From 1974 to 1976, he was with INRIA, Rocquencourt. Since 1976, he has been with IRISA, Rennes. After some work in probability theory (Markov processes and ergodic theory) for his "thèse d'Etat" in 1975, his interest moved towards the area of applied mathematics

(signal processing, identification and adaptive algorithms, speech and image coding, data communication systems). He is the author of numerous papers in

identification, adaptive algorithms, data transmission systems, image coding, and change detection.



Michèle Basseville was born in Paris, France, in 1952. She graduated from the Ecole Normale Supérieure, Fontenay-aux-Roses, France.

Since 1976, she has been with IRISA, Rennes. Until her "thèse d'Etat" in 1982, her main interest has been on-line segmentation of digital signals, and then has moved towards change detection and diagnosis in dynamical systems. She has coedited (with A. Benveniste) a Springer-Verlag LNCIS issue devoted to the design of statistical algorithms for change detection.



Georges V. Moustakides (S'79-M'82-SM'82-M'83) was born in Drama, Greece, on April 16, 1955. He received the diploma in electrical engineering from the National Technical University, Athens, Greece, the M.Sc. degree in systems engineering from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1979, 1980, and 1983, respectively.

From 1983 to 1986, he worked in the Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, France. His interests include detection of signals in dependent noise, detection of changes in systems, and theory of optimal stopping rules.