



# The asymptotic local approach to change detection and model validation

Albert Benveniste, Michèle Basseville, George V. Moustakides

## ► To cite this version:

Albert Benveniste, Michèle Basseville, George V. Moustakides. The asymptotic local approach to change detection and model validation. [Research Report] RR-0564, INRIA. 1986. inria-00075990

**HAL Id: inria-00075990**

**<https://inria.hal.science/inria-00075990>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CENTRE DE RENNES  
**IRISA**

Institut National  
de Recherche  
en Informatique  
et en Automatique

Domaine de Voluceau  
Rocquencourt  
BP 105  
78153 Le Chesnay Cedex  
France  
Tél. (1) 39 63 55 11

Rapports de Recherche

N° 564

**THE ASYMPTOTIC LOCAL  
APPROACH  
TO CHANGE DETECTION  
AND MODEL VALIDATION**

Albert BENVENISTE  
Michèle BASSEVILLE  
Georges MOUSTAKIDES

Septembre 1986

Campus Universitaire de Beaulieu  
35042 - RENNES CÉDEX  
FRANCE  
Téléphone: 99 36 20 00  
Télex: UNIRISA 950 473 F  
Télécopie: 99 38 38 32

Publication Interne n° 307

Aout 1986.

36 pages.

### THE ASYMPTOTIC LOCAL APPROACH TO CHANGE DETECTION AND MODEL VALIDATION

Albert Benveniste(IRISA/INRIA), Michèle Basseville(IRISA/CNRS),  
Georges Moustakides(IRISA/INRIA).

(L'approche asymptotique locale pour les problèmes de détection de changements et de validation de modèles).

RESUME On présente une approche systématique pour la conception d'algorithmes de détection de changements et de validation de modèles pour les systèmes dynamiques. On montre comment associer à tout algorithme d'identification des procédures de détection de changements et de validation de modèles qui sont optimales asymptotiquement en un certain sens. Les fondements de notre méthode s'appuient sur l'approche asymptotique locale des Statisticiens, et notre méthode généralise cette approche.

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 procedures, which are optimal in some asymptotic meaning. The foundations of our method go back to the "asymptotic local" approach in Statistics, and our method generalizes this approach.

# THE ASYMPTOTIC LOCAL APPROACH TO CHANGE DETECTION AND MODEL VALIDATION.

Albert Benveniste (\*), Michele Basseville (\*\*), Georges Moustakides (\*)  
(\*) IRISA/INRIA, (\*\*) IRISA/CNRS  
Campus de Beaulieu  
35042 RENNES CEDEX, FRANCE

**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 meaning. 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 a growing attention these last fifteen years, as can be seen from the survey papers [Willsky 1976 – a], [Basseville 1986], and the monography [Basseville & Benveniste 1986]. 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,
3. gains updating in adaptive algorithms, for tracking quick variations of the parameters.

Many applied fields have been already concerned, as discussed for example in [Basseville 1986], 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 by 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);

---

Keywords: adaptive algorithms, change detection, model validation, statistical methods, stochastic processes.

2. check whether a given nominal model (intended, for instance, to describe the ideal behaviour 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 [Ljung & Soderstrom 1983] 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 [Roussas 1972], and also the fundamental papers [Nikiforov 1986] and [Deshayes & Picard 1986]. This approach provides an effective way to design or analyse likelihood ratio based testing procedures when the alternative hypotheses become closer as the length of the record goes to infinity. Starting from this idea, [Basseville & al. 1986] and [Moustakides & Benveniste 1986] studied a situation in which no likelihood ratio approach could be effective due to the presence of nonstationary nuisance parameters; hence, starting from the wellknown 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 the 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 [Willsky 1976 – a], [Willsky 1986]; this approach is for example used in the 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 *non identifiable* models (think of a complex system modelled on one hand by a large physical, often non identifiable model, and on the other hand 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 1. The mathematical background of the asymptotic local approach is presented in section 2. The local procedures of change detection and model validation are respectively presented in section 3 and section 4; the methods are illustrated on examples. Finally, the diagnosis problem is investigated in section 5.

## Chapter One

### PROBLEM STATEMENT

The problem will be stated for the case of change detection; the problem of model validation will be directly stated in section 4. 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.

#### 1.1 Examples.

##### 1.1.1 Jump in the mean of a signal.

Consider a signal of the form

$$y_n = \theta_*(n) + v_n \quad (1-1)$$

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

##### 1.1.2 Changes in an AR process.

Consider an AR process of the form

$$y_n = \sum_{i=1}^p a_i y_{n-i} + \sigma v_n \quad (1-2)$$

where  $(v_n)$  is a zero mean i.i.d. sequence of unit variance. The model (1-2) is summarized in the parameter

$$\theta_*^T := (a_1, \dots, a_p; \sigma) \quad (1-3)$$

Setting

$$\Phi_{n-1}^T := (y_{n-1}, \dots, y_{n-p}) \quad (1-4)$$

(1-2) can be rewritten under the following state space form

$$\begin{aligned} \Phi_n &= A(\theta_*)\Phi_{n-1} + B(\theta_*)v_n \\ y_n &= (1, 0, \dots, 0)\Phi_n \end{aligned}$$

$$A(\theta_*) = \begin{bmatrix} a_1 & & a_p \\ 1 & & \\ & & 1 & 0 \end{bmatrix}, \quad B(\theta_*) = \begin{bmatrix} \sigma \\ 0 \\ 0 \end{bmatrix} \quad (1-5)$$

where the unspecified entries of the matrix  $A(\theta_*)$  are equal to zero. The formulas (1-5) express the fact that  $(\Phi_n)$  is a **controlled Markov chain** with control parameter  $\theta_*$ . This means that  $(\Phi_n)$  is a Markov chain the transition matrix of which depends on the parameter  $\theta_*$ . Assuming  $\theta_*$  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 [Benveniste, Metivier & Priouret 1986].

## 1.2 Detection of changes in a controlled semi-Markov process.

We shall say that  $(X_n)$  is a **Controlled Semi-Markov Process** with control parameter  $\theta_*$  if  $(X_n)$  is of the form:

$$\begin{aligned} P\{\xi_n \in G \mid \xi_{n-1}, \xi_{n-2}, \dots\} &= \int_G \pi_{\theta_*}(\xi_{n-1}, dx) \\ X_n &= f(\xi_n) \end{aligned} \quad (1-6)$$

where  $\pi_{\theta_*}(\xi, dx)$  is the transition probability of a Markov chain  $(\xi_n)$  depending on a parameter  $\theta_*$ . The model (1-6) represents the «true system». Accordingly, the **sequential change detection problem** in the system (1-6) is formulated as follows:

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

$$\begin{aligned} \theta_* &= \theta_0 \text{ for } n < r \\ \theta_* &= \theta_1 \text{ for } n \geq r \end{aligned}$$

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  $\theta_0$  and/or  $\theta_1$ .

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.

### 1.3 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  $\theta$  from time  $r$ , where  $\theta$  is an unknown parameter. The wellknown solution to this change detection problem is the GLR test (« Generalized Likelihood Ratio »), see [Willsky & Jones 1976 - b]. Recall briefly how this test is obtained. First, fix  $r$  and  $\theta$ . Given the record  $Y_1, \dots, Y_n$ , the loglikelihood ratio between the hypotheses

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

is given by

$$\begin{aligned}
 S_r^n(\theta) &= \sum_{k=r}^n Y_k^T R^{-1} Y_k - \sum_{k=r}^n (Y_k - \theta)^T R^{-1} (Y_k - \theta) \\
 &= 2 \sum_{k=r}^n Y_k^T R^{-1} \theta - (n - r + 1) \theta^T R^{-1} \theta
 \end{aligned} \tag{1-7}$$

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



$$S_r^n := \max_{\theta} S_r^n(\theta) = (\Delta_r^n)^T R^{-1} \Delta_r^n \quad (1-8)$$

where

$$\Delta_r^n := (n - r + 1)^{-1/2} \sum_{k=r}^n Y_k \quad (1-9)$$

and

$$\hat{\theta}(n, r) := \arg \max_{\theta} S_r^n(\theta) = (n - r + 1)^{-1/2} \Delta_r^n \quad (1-10)$$

Taking in (1-8) the maximum with respect to  $r$  yields

$$G_n := \max_r S_r^n, \quad r_n = \arg \max_r S_r^n \quad (1-11)$$

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

$$v = \min\{n : G_n \geq \lambda\} \quad (1-12)$$

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

$$\hat{r} = \hat{r}_v, \quad \hat{\theta} = \hat{\theta}(v, \hat{r}) \quad (1-13)$$

The formulas (1-9 to 1-13) define the complete change detection test procedure for this case.

## Chapter Two

### MATHEMATICAL BACKGROUND.

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

#### 2.1 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 [Benveniste & al. 1986], see also [Metivier & Priouret 1984] or [Kushner & Schwartz 1984] for slightly different assumptions.

##### 2.1.1 Some background on adaptive algorithms.

The adaptive algorithms we shall consider are of the form

$$\theta_n = \theta_{n-1} + \gamma_n H(\theta_{n-1}, X_n) \quad (2-1)$$

where  $\theta$  belongs to  $R^d$  or to some submanifold of  $R^d$ , and the state  $X_n$  belongs to  $R^k$ . The gain  $\gamma_n$  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  $\theta$* ; this means that

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

$$X_n = f(\xi_n) \quad (2-2)$$

where the extended state  $(\xi_n)$  is, for  $\theta$  fixed, a Markov chain with transition probability  $\pi_\theta(\xi, dx)$  which depends on the parameter  $\theta$ . We assume that, for every  $\theta$  belonging to the domain of the algorithm, the Markov chain  $(\xi_n)$  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  $\theta$ . It includes also the case of *conditionnally linear dynamics*, such as used in [Ljung & Soderstrom 1983], 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  $\theta$  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

$$h(\theta) := \lim_{n \rightarrow \infty} E_\theta(H(\theta, X_n)) \quad (2-3)$$

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

$$\dot{\theta} = h(\theta), \quad \theta(0) = z \quad (2-4)$$

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.

## 2.1.2 Problem statement.

### 2.1.2.1 Investigation of the least squares algorithm for AR identification.

The identification of  $\theta_*$  in (1-5) can be for example performed via the least squares stochastic gradient algorithm

$$\begin{aligned} \theta_n &= \theta_{n-1} + \gamma \Phi_n[\theta_*] e_n[\theta_{n-1}, \theta_*] \\ e_n(\theta, \theta_*) &:= y_n(\theta_*) - \Phi_n^T(\theta) \cdot \theta \end{aligned} \quad (2-5)$$

where  $\Phi_n[\theta_*]$  is defined in (1-5). In (2-5), the dependence on the true parameter  $\theta_*$  has been made explicit, to introduce the convenient form for the change detection problem. In this form, the true system  $\theta_*$  is obviously not available to the user, but clearly influences the conditional distribution of the pair  $(\Phi_n, y_n)$ . Hence the investigation of this example motivates the following form we shall use for the adaptive algorithms instead of (2-1, 2-2) (we shall always in the sequel write  $z$  instead of  $\theta_*$  to denote the true system):

$$\theta_n = \theta_{n-1} + \gamma H(\theta_{n-1}; X_n)$$

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

$$X_n = f(\xi_n) \quad (2-6)$$

where the parameter  $z$  represents the *true system*. The second equation of (2-6) expresses the

fact that the true system does influence the law of the state  $\xi_n$ ; note that the form of the algorithm, as available to the user, has been kept unchanged.

As usually, 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 (2-5). But the gain strategy is irrelevant for our purpose, only the random vector field  $H(\theta_{n-1}; X_n)$  will be relevant.

### 2.1.2.2 Problem statement and assumptions.

Our starting point is now the triple introduced in (2-6), namely:

$$H(\theta; X_n), \pi_{\theta,z}(\xi, dx), f \quad (2-7)$$

where

- $\theta$  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  $(\theta, z)$ , 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 for adaptive algorithms, the following mean vector field is associated to (2-7)

$$h(\theta, z) := \lim_{n \rightarrow \infty} E_{\theta,z}(H(\theta; X_n)) \quad (2-8)$$

where  $E_{\theta,z}$  denotes the expectation with respect to  $P_{\theta,z}$ . This is nothing but the usual mean vector field of the associated ODE, where the dependence on 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  $\theta_0$  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  $h_\theta$  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  $z$ ,

$$h(\theta, z) = 0 \Leftrightarrow \theta = z \quad (2-9)$$

**CONSEQUENCE :** The following relationship holds :

$$h_{\theta}(z, z) = -h_z(z, z) \quad [1] \quad (2-10)$$

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  $\theta_0$  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  $[H(\theta_0; X_k)]_{1 \leq k \leq n}$

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

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

$$z = \theta_0 + \frac{\theta}{\sqrt{n}} \text{ for } \tau n \leq k \leq n, \text{ where } \theta \neq 0 \text{ is an unknown change.}$$

**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 [Nikiforov 1986], [Roussas 1972] and [Deshayes et Picard 1986] for further information on the asymptotic local approach for the likelihood ratio testing methods.

## 2.2 Main results.

The assumption NS is in force throughout this section. Fix a nominal model  $\theta_0$ , and consider the following cumulative sum, where  $m \leq n$

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

---

[1]:  $h_{\theta}[\theta, z]$  and  $h_z[\theta, z]$  denote respectively the first and second partial derivatives of the function  $h[\theta, z]$

$$D_n(\theta_0, \theta) := D_{n,n}(\theta_0, \theta) \quad (2-11)$$

(cf. (2-7)). 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 behaviour, when the length  $n$  of the record tends to infinity, of this cumulative sum under the three hypotheses  $H_0, H_1$  and  $H'_1$ .

### 2.2.1 Behaviour of the cumulative sum under the hypothesis of no change.

This behaviour is described in the following theorem:

#### THEOREM 1 :

(i) *behaviour of the marginal distribution:*

$$D_n(\theta_0, 0) \xrightarrow[n \rightarrow \infty]{L} N(0, R(\theta_0)) \quad (2-12)$$

where  $R(\theta)$  is given by

$$R(\theta) := \sum_{n=-\infty}^{+\infty} \text{cov}_{\theta, \theta} [H(\theta; X_n), H(\theta; X_0)] \quad (2-13)$$

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

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

$$D_{n,t}(\theta_0, \theta) := D_{n,m}(\theta_0, \theta) \text{ where } m = [nt] \quad (2-14)$$

Then,

$$[R(\theta_0)]^{-1/2} \cdot \left\{ D_{n,t}(\theta_0, 0) \right\}_{0 \leq t \leq 1} \rightarrow (W_t)_{0 \leq t \leq 1} \quad (2-15)$$

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 (2-7).

### 2.2.2 Behaviour 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

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

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

$$z = \theta_0 + n^{-1/2} \theta \text{ for } k > \min(m, [n\tau]) \quad (2-16)$$

in  $\pi_{\theta_0, z}$ , cf. (2-7). 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 behaviour of this cumulative sum is described in the following theorem:

**THEOREM 2 :** *Behaviour of the cumulative sum under the hypothesis  $H'_1$ . Let  $\tau \in [0, 1]$  be given. Set*

$$D_{n,t}(\theta_0, \theta, \tau) := D_{n,m}(\theta_0, \theta, \tau) \text{ where } m = [nt] \quad (2-17)$$

*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*

$$dD_t = -1_{\{t \geq \tau\}} \cdot h_{\theta} \cdot \theta dt + R^{1/2} \cdot dW_t \quad (2-18)$$

where  $R = R(\theta_0)$  is given in (2-13), while  $h_{\theta} = h_{\theta}(\theta_0, \theta_0)$  is defined in (2-10).

**COROLLARY 3 :** *Hypothesis  $H_1$*

$$D_n(\theta_0, \theta) \xrightarrow[n \rightarrow \infty]{L} N[-h_{\theta} \cdot \theta, R(\theta_0)] \quad (2-19)$$

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

**PROOF:** See the appendix.

## Chapter Three

### LOCAL CHANGE DETECTION TECHNIQUES.

#### 3.1 The local test.

From the user's point of view, the cumulative sums  $D_{n,m}$  given by the formulas (2-11) or (2-16) 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:

$$D_{n,m}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{k=1}^m Y_k(\theta_0), \quad Y_k(\theta_0) := H(\theta_0; X_k) \quad (3-1)$$

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  $\theta$ , and  $n$  is large enough. Then, *considering the random variables  $Y_k(\theta_0)$  as independent, and distributed as follows*

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

$$Y_k(\theta_0) \approx N[-h_\theta(\theta_0) \cdot \theta, R(\theta_0)], \quad k \geq r \quad (3-2)$$

would exactly result in the asymptotic behaviour described by (2-18). 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, 3-2). Restricting the study to the case where the direction of change  $\theta$  is unknown, we shall apply the formulas (1-9 to 1-13) to the detection of a change like (3-2). This gives the formulas of the *local change detection procedure*:

$$S_r^n(\theta) = 2 \sum_{k=r}^n [Y_k^T R^{-1} h_\theta \cdot \theta] - (n - r + 1) \theta^T \cdot h_\theta^T R^{-1} h_\theta \cdot \theta \quad (3-3)$$

where  $\theta_0$  has been deleted for simplicity. Fixing  $r$  and maximizing with respect to  $\theta$  yields

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



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

$$\hat{\theta}(n, r) := \arg \max_{\theta} S_r^n(\theta) = -(n - r + 1)^{-1/2} h_{\hat{\theta}}^{-1} \Delta_r^n \quad (3-4)$$

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

$$(i) \quad G_n := \max_r S_r^n, \quad v = \min\{n : G_n \geq \lambda\}$$

$$(ii) \quad r_n = \arg \max_r S_r^n, \quad \hat{r} = \hat{r}_v, \quad \hat{\theta} = \hat{\theta}(v, \hat{r}) \quad (3-5)$$

Note that (3-5-i) is sufficient if the change detection only is of interest. The local test is given by the formulas (3-1, 3-4, 3-5). The threshold  $\lambda$  is easily selected by knowing that, under the no change hypothesis, we have

$$E_{\theta_0}(S_r^n) = d, \quad (3-6)$$

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

**COMMENT :** A probably more commonly used method is the following one.

1) run the adaptive identification algorithm with constant gain

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

2) use a  $\chi^2$ -test of the form

$$(\theta_n - \theta_0)^T \Sigma^{-1} (\theta_n - \theta_0) \geq \lambda \quad (3-7)$$

with a suitable matrix  $\Sigma$  using the fact that  $\theta_n - \theta_0$  is approximately Gaussian and zero mean for  $\gamma$  small in the hypothesis of no change (use a Central Limit theorem for stochastic approximations to select the proper matrix  $\Sigma$ , see for example [Kushner 1984], [Benveniste & al. 1986]). 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 [Kushner 1984]. But in this case, it is known [Willisky & Jones 1976] that the best local test involves the innovations of this Markov process, which gives something different from (3-7). Our method is precisely the right way to test for small changes in the true system. We shall now illustrate this method on two non trivial typical examples, and show that it is *the convenient generalization of the «local likelihood ratio tests»* introduced by Le Cam ([Nikiforov 1986], [Roussas 1972], [Deshayes & Picard 1986]); other examples can be found in [Benveniste, Metivier & Priouret 1986].

### 3.2 Examples.

#### 3.2.1 Change detection in AR processes.

The objective is to detect changes in the parameter  $\theta$  in the system

$$y_n = \Phi_n^T \theta + v_n, \quad \Phi_n^T = (y_{n-1}, \dots, y_{n-p}) \quad (3-8)$$

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

$$H(\theta, y_n, \Phi_n) := \Phi_n e_n(\theta), \quad e_n(\theta) := y_n - \Phi_n^T \theta \quad (3-9)$$

The matrix  $R(\theta_0)$  corresponding to (2-13) is given by

$$R(\theta_0) = E_{\theta_0}(v_n \Phi_n \Phi_n^T v_n) = \sigma^2 \Sigma(\theta_0) \quad (3-10)$$

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

$$\Delta_r^n(\theta_0) = (n - r + 1)^{-1/2} \sum_{k=r}^n \Phi_k e_k(\theta_0) \quad (3-11)$$

It is easy to verify that  $\sigma^{-2} \Delta_0^n(\theta_0)$  is the derivative with respect to  $\theta$  of the loglikelihood 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 [Nikiforov 1986] and [Davies 1973] reveals that (3-1, 3-4, 3-5, 3-10, 3-11) yields the so called local likelihood ratio test, which is the convenient procedure to detect small changes in the parameters of an AR process.

#### 3.2.2 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 non classical, and has been proposed and analysed in details in [Basseville & al. 1986], [Moustakides & Benveniste 1986]. Consider an ARMA process of the form

$$y_n = \sum_{i=1}^p a_i y_{n-i} + \sum_{j=1}^q b_j v_{n-j} + v_n \quad (3-12)$$

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 [Benveniste & Fuchs 1985] 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 [Stoica & al. 1985]. Setting

$$\begin{aligned}\theta^T &:= (a_1, \dots, a_p) \\ \Phi_n^T &:= (y_{n-1}, \dots, y_{n-p}) \\ \Psi_n^T &:= (y_{n-q-1}, \dots, y_{n-q-p})\end{aligned}\quad (3-13)$$

where  $\Psi_n$  is the instrument, the I.V. method is given by

$$\begin{aligned}\theta_n &= \theta_{n-1} + \frac{1}{n} \Gamma_n^{-1} \Psi_n (y_n - \Phi_n^T \theta_{n-1}) \\ \Gamma_n &= \Gamma_{n-1} + \frac{1}{n} (\Psi_n \Phi_n^T - \Gamma_{n-1})\end{aligned}\quad (3-14)$$

The random vector field of interest is here equal to

$$H(\theta_0; \Psi_n, \Phi_n, y_n) = \Psi_n (y_n - \Phi_n^T \theta_0) =: Y_n(\theta_0) \quad (3-15)$$

where  $\theta_0$  is the nominal model. To apply our method, we must calculate the matrices

$$\begin{aligned}R(\theta_0) &:= \sum_{n \in \mathbb{Z}} E_0 [Y_n(\theta_0) Y_0(\theta_0)^T] \\ &= \sum_{n=-q}^{+q} E_0 [\Psi_n \Psi_0^T (y_n - \Phi_n^T \theta_0)(y_0 - \Phi_0^T \theta_0)] \\ h_\theta(\theta_0) &= -E_0(\Psi_n \Phi_n^T),\end{aligned}\quad (3-16)$$

where  $E_0$  is a shorthand for  $E_{\theta_0, \theta_0}$ . The *instrumental test* is obtained by combining the formulas (3-1, 3-4, 3-5, 3-15, 3-16). As expected, this test exhibits very pleasant robustness properties with respect to the nuisance MA parameters: for instance, it is proved in [Moustakides & Benveniste 1986] 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, non classical method of change detection. It is not our purpose here to discuss the details of practical implementations. The interested reader is referred to [Basseville & Benveniste 1986] for further details.

## Chapter Four

### 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 behaviour 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 pre-specified behaviour; this a priori behaviour 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 [Basseville & al. 1985, 1986] 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 sounded model validation techniques, compared to the classically used ones.

#### 4.1 Problem statement.

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

$$\begin{aligned} P\{\xi_n \in G | \xi_{n-1}, \xi_{n-2}, \dots\} &= \int_G \pi_{\theta_*}(\xi_{n-1}, dx) \\ X_n &= f(\xi_n) \end{aligned} \tag{4-1}$$

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

**[VM]** *Given a nominal model  $\theta_0$ , decide between the hypotheses*

$\theta_* = \theta_0$  : *the nominal model is valid*

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

Here,  $\theta_0$  denotes the model to be validated (nominal model), while  $\theta_*$  denotes the (unknown) true system corresponding to the observed record  $X_1, \dots, X_N$ .

## 4.2 Local validation method associated to an adaptive algorithm.

The problem [VM] is formulated in the local framework as follows. We consider again as in (2-7) the random vector field associated to the adaptive algorithm:

$$H(\theta; X_n), \pi_{\theta, z}(\xi, dx), f \quad (4-2)$$

where

- $\theta$  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  $(\theta, z)$ .

Again the assumption NS is in force throughout this section. The problem [VM] is then formulated as follows: assume that a nominal model  $\theta_0$  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 maximisation with respect to  $r$  in the formulas (3-3, 3-4). This gives the following formulas, where the explicit dependence on  $\theta_0$  was deleted for the sake of simplicity. For a fixed  $\theta$ , the loglikelihood ratio between the hypotheses  $H_1$  and  $H_0$  is given by

$$S_n(\theta) = 2 \sum_{k=1}^n [Y_k^T R^{-1} h_{\theta} \cdot \theta] - n \theta^T \cdot h_{\theta}^T R^{-1} h_{\theta} \cdot \theta \quad (4-3)$$

where  $Y$  is defined in (3-1). Finally, the following  $\chi^2$  test is derived

$$\Delta_n^T R^{-1} \Delta_n \geq \lambda, \quad \Delta_n = \frac{1}{n} \sum_{k=1}^n Y_k \quad (4-4)$$

Again, the threshold  $\lambda$  can be chosen by using the fact that, under the hypothesis  $H_0$  of validity of the nominal model  $\theta_0$ ,  $n \Delta_n^T R^{-1} \Delta_n$  is a  $\chi^2$  with  $n$  degrees of freedom. The same examples as before can be used to illustrate our method. See [Basseville & al. 1985] 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 [Benveniste, Metivier & Priouret 1986].

## Chapter Five

### 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 non identifiable physical models are referred to for the diagnosis). Let us first investigate this problem on a simplified example, borrowed from [Basseville & al. 1985] and [Moustakides 1986].

#### 5.1 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

$$\begin{aligned} M\ddot{Z} + C\dot{Z} + KZ &= E \quad Z, E \in R^D \\ Y &= LZ \quad Y \in R^d, d \ll D \end{aligned} \quad (5-1)$$

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) is, in state space form

$$\begin{aligned} X_{n+1} &= FX_n + V_{n+1} \\ Y_n &= HX_n \end{aligned} \quad (5-2)$$

(5-1) and (5-2) are related as follows

$$\lambda(F) = \exp[\tau \lambda(Ms^2 + Cs + K)], \quad H\Phi_\lambda = L\Psi_\lambda \quad (5-3)$$

where  $\tau$  is the discretization step,  $\lambda(\cdot)$  denotes an eigenvalue,  $\Phi_\lambda$  the corresponding eigenvector of  $F$ , and  $\Psi_\lambda$  the eigenvector of the polynomial matrix  $(Ms^2 + Cs + K)$  associated to the eigenvalue  $\exp(\tau \lambda)$ . Finally, the eigenstructure of (5-2) reflects the eigenstructure of the vibrating system (5-1), 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 (5-3)) if the model (5-2) 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 (3-4) in order to estimate the most likely change  $\theta$ , 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) by (5-2), which is always the case in practice. We shall thus propose a suitable general approach to solve this problem.

## 5.2 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  $\mu = Mv$ , where  $v \neq 0$ , the loglikelihood ratio is

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

The maximum likelihood estimate of  $v$  is

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

So that we get the following  $\chi^2$  statistics to test  $\mu = 0$  against  $\mu = Mv$ ,  $v \neq 0$

$$\chi = U^T \Sigma^{-1} M \left[ M^T \Sigma^{-1} M \right]^{-1} M^T \Sigma^{-1} U \quad (5-4)$$

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.



### 5.3 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  $\theta_0$  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  $\Theta$

Choose a full column rank matrix  $M_\Theta$  such that  $\Theta = \text{range}(M_\Theta)$ ; keeping (3-2) in mind, we can directly use the formula (5-4) to get the suitable version of the formula (4-4)

$$\Delta_n^T R^{-1} M [M^T R^{-1} M]^{-1} M^T R^{-1} \Delta_n \geq \lambda$$

$$M = h_\theta(\theta_0) \cdot M_\Theta \quad (5-5)$$

where  $\Delta_n$  is defined in (4-4). This can be used as follows. If  $\Theta$  is the subspace spanned by a subset of the coordinates of the parameter  $\theta$ , choose  $M_\Theta$  as being the corresponding projection matrix. If  $\Theta$  is the subspace spanned by a subset of the coordinates of an other identifiable parameter set  $\Psi = \{\psi\}$  such that there exists a change of coordinates

$$\theta = F(\psi) \quad (5-6)$$

with  $F$  being a local diffeomorphism, build  $M_\Theta$  as being the corresponding subset of the columns of the Jacobian matrix

$$F'(\psi_0) \quad (5-7)$$

where  $\psi_0$  is the nominal model in the parameter set  $\Psi$ . This was for example the approach taken in [Basseville & al. 1986] for diagnosing changes in specified poles of a vector ARMA process.

## 5.4 Diagnosis on non identifiable model sets.

### 5.4.1 A rough procedure.

Assume now the parameter set  $\Psi$  to be monitored is not identifiable, which means that  $F$  in (5-6) 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  $\psi_0$  used in (5-7) to compute the Jacobian cannot be derived from  $\theta_0$ . The suitable modification of the method is then the following.

STEP 1: select a nominal model  $\theta_0$  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 model  $\psi_0$  is available :

$$\theta_0 \approx F(\psi_0) \quad (5-8)$$

(this is for example the case when  $\psi_0$  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  $\Psi$  is assumed to be much larger than the identifiable parameter space  $\Theta$ , the selection of proper subspaces in  $\Psi$  must be taken with some care; otherwise the selection rule (5-6) 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 [Moustakides 1986] and currently implemented for the application to vibration monitoring.

### 5.4.2 An improved procedure.

The step 1 of the former procedure is kept. The step 2 is refined as follows.

STEP 2.1: Assume a coarse nominal model  $\psi_0$  in the sense of (5-8) is available; select candidate changes  $\delta\psi_1, \delta\psi_2, \dots$  to be monitored in the large parameter space  $\Psi$  and compute the corresponding changes

$$\delta\theta_i = F'(\psi_0) \cdot \delta\psi_i \quad (5-9)$$

in the identifiable parameter space  $\Theta$ . Let us emphasise that the *magnitude* of these changes are

of importance: some realistic changes in the parameter set  $\Psi$  may result in too small changes in the identifiable parameter set  $\Theta$ , 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  $\chi^2$  statistics (5-5) corresponding to  $M_\Theta = \delta\theta$  is equal to

$$\delta\theta^T \cdot h_\Theta^T(\theta_0) \cdot R^{-1} \cdot h_\Theta(\theta_0) \cdot \delta\theta \quad (5-10)$$

Compare to a chosen threshold this noncentrality parameter for all the  $\delta\theta_i$ 's obtained in the 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  $\theta$  for short instead of  $\delta\theta$ , and we remove the dependence on the nominal model  $\theta_0$ ):

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

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

Note that  $\rho_{i,j}$  is nothing but the cosine of the angle between the subspaces spanned by the vectors  $\theta_i$  and  $\theta_j$ . 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  $\chi^2$ -statistics to decide the origin of the possible changes in terms of the large, non identifiable, parameter set  $\Psi$ . The resulting procedure can allow the diagnosis of small changes with a good accuracy.

## **Chapter Six**

### **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 analyse 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 (non identifiable) 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 wellknown instrumental variable method.

## 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.

### A.1 Mixingales, and a theorem of Mac Leish.

#### Mixingales

Mixingales have been introduced in [Mac Leish 1975-a], 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  $\sigma$ -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  $\psi_m$  where  $\psi_m \rightarrow 0$  as  $m \rightarrow \infty$ , we have for all  $n \geq 1, m \geq 0$

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

For example,  $\psi_m \equiv 0$  corresponds to martingales. Strong laws and invariance principles are given in [Mac Leish 1975-a,b] for mixingales with exponentially vanishing  $\psi_m$ '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 [Mac Leish 1977], 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

$$W_n(t) = \sum_{i=1}^{k_n(t)} Z_{n,i} \tag{a-2}$$

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

$$\sigma_{n,i}^2 = \sum_j E(Z_{n,i} Z_{n,j}) \quad (\text{a-3})$$

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

$$(a) \sup_{s < t < T} \limsup_{n \rightarrow \infty} \frac{\sum_{k_n(s)}^{k_n(t)} \sigma_{ni}^2}{t-s} < \infty$$

$$(b) \left\{ \frac{Z_{ni}^2}{\sigma_{ni}^2} \right\}_{n=1,2,\dots, i \leq k_n(T)} \text{ is uniformly integrable}$$

$$(c) \max_{i \leq k_n(T)} \sigma_{n,i} \rightarrow 0 \text{ as } n \rightarrow \infty \quad (\text{a-4})$$

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  $\sigma$ -algebras  $F_{n,i}$ , namely

$$\begin{aligned} \|E(Z_{ni} | F_{n,i-k})\|_2 &\leq \psi_k \sigma_{ni} \\ \|Z_{ni} - E(Z_{ni} | F_{n,i+k})\|_2 &\leq \psi_{k+1} \sigma_{ni} \end{aligned} \quad (\text{a-5})$$

**THEOREM [Mac Leish 1977]:** Suppose conditions (a-4,a-5) are in force,  $\psi_k \downarrow 0$  exponentially fast, and, for each  $s < t < u$

$$E \left| E \left\{ \left( \sum_{i=k_n(t)}^{k_n(u)} Z_{ni} \right)^2 \middle| F_{n,k_n(s)} \right\} - (u-t) \right| \rightarrow 0 \quad (\text{a-6})$$

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

## A.2 Assumptions for theorems 1 and 2.

To apply the theorem of Mac Leish, we choose

$$z_n = \theta_0 + n^{-1/2} \theta$$

$$Z_{n,i} = n^{-1/2} \lambda^T [R(\theta_0, z_n)]^{-1/2} [H(\theta_0, X_i) - h(\theta_0, z_n)]$$

$$F_{n,i} = \sigma \{ \xi_k : k \leq i \}$$

$$P = P_{\theta_0, z_n}$$

$$T = 1, \quad k_n(t) = t \quad (a-7)$$

where  $\lambda$  is an arbitrary unit vector, and

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

The conditions (a-4) are obviously satisfied. We shall furthermore assume that the conditions (a-5, a-6) hold, with constants  $\psi_k$  that are uniform in  $\lambda$  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: 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  $\xi_n$ ; see for example [Benveniste, Goursat & Ruget 1980] for the verification of the stronger «function of  $\phi$  - mixing» condition on non trivial examples.

### A.3 Proofs.

Let us begin with the proof of theorem 1. Using the notations of (2-14), the theorem of Mac Leish implies that

$$\lambda^T [R(\theta_0)]^{-1/2} \cdot \left\{ D_{n,t}(\theta_0, 0) \right\}_{0 \leq t \leq 1} \rightarrow \lambda^T (W_t)_{0 \leq t \leq 1} \quad (a-8)$$

Since  $\lambda$  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  $D_{n,t}(\theta)$  for short instead of  $D_{n,t}(\theta_0, \theta, \tau)$ , and delete the dependency on  $\theta_0$  when no confusion can occur. The theorem of Mac Leish implies here that

$$\lambda^T [R(\theta_0)]^{-1/2} \left[ D_{n,t}(\theta) - n^{-1/2} \sum_{k=1}^{[n\eta]} h(\theta_0, z_k) \right]_{0 \leq t \leq 1} \rightarrow \lambda^T (W_t)_{0 \leq t \leq 1} \quad (a-9)$$

Finally, a first order Taylor expansion

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

gives the theorem 2.

## REFERENCES

- [Basseville 1986]: "Detection of changes in signals and systems", 2nd IFAC Workshop on adaptive systems in control and signal processing, Lund, Sweden.
- [Basseville & Benveniste Ed. 1986]: *Detection of abrupt changes in signals and dynamical systems*, LNCIS 77, Springer Verlag.
- [Basseville, Benveniste, & Moustakides 1986]: "Detection and diagnosis of abrupt changes in modal characteristics of nonstationary digital signals", *IEEE-IT-32* No 3, 412-417.
- [Basseville, Benveniste, Moustakides & Rougee 1985]: "Detection and diagnosis of changes in the eigenstructure of nonstationary multivariable systems", research rep. IRISA No 276/INRIA No 477, submitted for publication to Automatica.
- [Benveniste, Goursat & Ruget 1980]: "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms, with applications to data communication algorithms", *IEEE-AC-25* No 6, 1042-1058.
- [Benveniste & Fuchs 1985]: "Single Sample modal identification of a nonstationary stochastic process", *IEEE-AC-30* No 1, 66-74.
- [Benveniste, Metivier & Priouret 1986]: *Algorithmes adaptatifs et approximations stochastiques*, Techniques stochastiques, Masson, Paris.
- [Davies 1973]: "Asymptotic inference in stationary Gaussian time series", *Adv. Appl. Prob.* vol 5 No 3, 469-497.
- [Deshayes & Picard 1986]: "Off-line analysis of change-point models using non parametric and likelihood methods", in [Basseville & Benveniste Ed. 1986].



[Kushner 1984]: *Approximation and weak convergence methods for random processes, with applications to stochastic system theory*, MIT Press.

[Ljung & Soderström 1983]: *Theory and practice of recursive identification*, MIT Press.

[Mac Leish 1975 - a]: "A maximal inequality and dependent strong laws", *Annals of Probability*, Vol 3 No 5, 829 - 839

[Mac Leish 1975 - b]: "Invariance principles for dependent variables", *Zeitschrift für Wahrscheinlichkeitstheorie verw. Geb.* 32, 165 - 178.

[Mac Leish 1977]: "On the invariance principle for nonstationary mixingales", *The Annals of Probability*, vol 5 No 4, 616 - 621.

[Metivier & Priouret 1984]: "Application of a lemma of Kushner and Clark to general classes of stochastic algorithms" *IEEE - IT - 30*, 140 - 150.

[Moustakides & Benveniste 1986]: "Detecting changes in the AR parameters of a nonstationary ARMA process", *Stochastics*, Vol 16, 137 - 155.

[Moustakides 1986]: "The problem of diagnosis with respect to physical parameters for changes in structures", IRISA tech. report 295.

[Nikiforov 1986]: "Sequential detection of changes in stochastic systems", in [Basseville & Benveniste Ed. 1986].

[Roussas 1972]: 7[Contiguity of probability measures, some applications in statistics], Cambridge University press.

[Stoica, Soderström & Friedlander 1985]: "Optimal Instrumental Variable estimates of the AR parameters of an ARMA process", *IEEE - AC - 30* No 11, 1066 - 1074 .

[Willsky 1976 - a]: "A survey of design methods for failure detection in dynamic systems", *Automatica*, 12, 601 - 611.

[Willsky & Jones 1976 - b]: "A generalized likelihood ratio approach to detection and estimation of jumps in linear systems" *IEEE - AC - 21* 108 - 112.

[Willsky 1986]: "Detection of abrupt changes in dynamic systems", in [Basseville & Benveniste Ed. 1986].

