Proceedings of 25th Conference
on Decision and Control
Athens, Greece • December 1986

MODELLING AND MONITORING OF CHANGES IN DYNAMICAL SYSTEMS

A. Benveniste *, M. Basseville **, G. 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); 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 & Soderström 1983] for example). However, most of the model validation techniques are rather ad hoc from the statistical viewpoint.

The purpose of this article 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 foudations 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 identifying 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 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).

I PROBLEM STATEMENT

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

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

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 of the jumps.

1.1.2 Changes in an AR process.

Consider an AR process of the form

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

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

$$\theta_*^T := (a_1, \ldots, a_p ; \sigma) \quad (1\text{-}3)$$

Setting

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

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

$$\Phi_n = A(\theta_*)\Phi_{n-1} + B(\theta_*) v_n$$
$$y_n = (1,\ 0, \ldots,\ 0)\Phi_n$$

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

where the unspecified entries of the matrix $A(\theta_*)$ are equal to zero. The formulas (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.

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** θ_* if (X_n) is of the form:

$$P\{\xi_n \in G \mid \xi_{n-1}, \xi_{n-2}, \ldots\} = \int_G \pi_{\theta_*}(\xi_{n-1}, dx)$$
$$X_n = f(\xi_n) \quad (1\text{-}6)$$

where $\pi_{\theta_*}(\xi, dx)$ is the transition probability of a Markov chain (ξ_n) depending on a parameter θ_*. 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*

$\theta_* = \theta_0$ *for* $n < r$
$\theta_* = \theta_1$ *for* $n \geq r$

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

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.

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 θ from time r, where θ is an unknown parameter. The wellknown solution of this problem is the GLR test (« Generalized Likelihood Ratio »), see [Willsky & Jones 1976-b]. Recall briefly how this test is obtained. First, fix r and θ. Given the record $Y_1, \ldots, Y_n$, 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 θ

is given by

$$S_r^n(\theta) = \sum_{k=r}^{n} Y_k^T R^{-1} Y_k - \sum_{k=r}^{n} (Y_k - \theta)^T R^{-1} (Y_k - \theta)$$
$$= 2\sum_{k=r}^{n} Y_k^T R^{-1}\theta - (n-r+1)\theta^T R^{-1}\theta \quad (1\text{-}7)$$

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

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

where

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

and

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

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

$$G_n := \max_r S_r^n, \quad r_n = \arg\max_r S_r^n \tag{1-11}$$

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

$$\nu = \min\{n : G_n \geq \lambda\} \tag{1-12}$$

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

$$\hat{r} = \hat{r}_\nu, \quad \hat{\theta} = \hat{\theta}(\nu, \hat{r}) \tag{1-13}$$

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

II FOUNDATIONS OF THE ASYMPTOTIC LOCAL POINT OF VIEW.

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] for slightly different assumptions.

2.1.1 Some background on adaptive algorithms.

The adaptive algorithms we shall consider are of the form

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

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

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

$$X_n = f(\xi_n) \tag{2-2}$$

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 *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. For θ fixed, (X_n) is asymptotically ergodic 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

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

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

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

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 θ_* in (1-5) can be for example performed via the least squares stochastic gradient algorithm

$$\theta_n = \theta_{n-1} + \gamma\, \Phi_n[\theta_*]\, e_n[\theta_{n-1}, \theta_*]$$
$$e_n(\theta, \theta_*) := y_n(\theta_*) - \Phi_n^T(\theta_*).\theta \tag{2-5}$$

where $\Phi_n[\theta_*]$ is defined in (1-5). In (2-5), the dependence on the true parameter θ_* has been made explicit, although this parameter is obviously unknown to the user. The motivation for introducing the true parameter will be made clear in the sequel. This example motivates the following form we shall use for the adaptive algorithms

$$\theta_n = \theta_{n-1} + \gamma\, H(\theta_{n-1}, z\,;\, X_n) \tag{2-6}$$

where the parameter z represents the *true system*. 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 $H(\theta_{n-1}, z\,;\, X_n)$ will be relevant.

2.1.2.2 Problem statement and assumptions.

Our starting point is now the random vector field

$$H(\theta, z\,; X_n) \tag{2-7}$$

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).

As usual for adaptive algorithms, the following mean vector field is associated to (2-7)

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

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.

WARNING: Let us emphasize that z is a parameter which is not available to the user, and has been introduced only for the sake of the theoretical analysis. The user only knows the form

$$H(\theta\ ;X_n) \tag{2-9}$$

which is directly borrowed from the usual form of adaptive algorithm. □

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 $H(\theta_0, z\ ;X_n)$*. The following assumptions will be inforce in the sequel, and we shall denote by h_θ and h_z respectively the first and second partial derivatives of h.

ASSUMPTION NS.[1] *the model set matches the true system structure in the following sense: for every z,*

$$h(\theta, z) = 0 \Leftrightarrow \theta = z \tag{2-10}$$

CONSEQUENCE : The following relationship holds :

$$h_\theta(z,z) = -h_z(z,z) \tag{2-11}$$

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, ..., X_n of length n of the state vector; test the following hypotheses against each other by using the random vector field trajectory $[H(\theta_0\ ;X_k)]_{1 \leqq k \leqq n}$

- 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 : *there exists $\tau \in]0,1[$, such that*

$z = \theta_0$ for $k < \tau n$

$z = \theta_0 + \frac{\theta}{\sqrt{n}}$ *for $\tau n \leqq k \leqq n$, 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 [Nikiforov 1986], [Roussas 1972] and [Deshayes et Picard 1986] for further information in 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 θ_0, and consider the following cumulative sum, where $m \leqq n$

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

$$D_n(\theta_0, \theta) := D_{n,n}(\theta_0, \theta) \tag{2-12}$$

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:*

$$D_n(\theta_0, 0) \xrightarrow[n \to \infty]{L} N(0, R(\theta_0)) \tag{2-13}$$

where $R(\theta)$ is given by

$$R(\theta) := \sum_{n=-\infty}^{+\infty} cov_{\theta;\theta}\left[H(\theta,\theta\ ;X_n), H(\theta,\theta\ ;X_0)\right] \tag{2-14}$$

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

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

$$D_{n,t}(\theta_0, \theta) := D_{n,m}(\theta_0, \theta) \quad \text{where } m = [nt] \tag{2-15}$$

Then,

$$[R(\theta_0)]^{-1/2} \cdot \left\{D_{n,t}(\theta_0, 0)\right\}_{0 \leqq t \leqq 1} \to (W_t)_{0 \leqq t \leqq 1} \tag{2-16}$$

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

PROOF: Of course (i) is a consequence of (ii). On the other hand, (ii) is a classical invariance principle for dependent processes, see for example [Mac Leish 1975-b, main theorem].

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 $m \leqq n$ the following cumulative sum

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

$$+ \frac{1}{\sqrt{n}} \sum_{k=\min(m,[n\tau])+1}^{m} H(\theta_0, \theta_0 + \frac{\theta}{\sqrt{n}}\ ;X_k) \tag{2-17}$$

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 $\tau \in [0,1]$, Set

$$D_{n,t}(\theta_0, \theta, \tau) := D_{n,m}(\theta_0, \theta, \tau) \quad \text{where } m = [nt] \tag{2-18}$$

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

$$dD_t = -1_{\{t \geq \tau\}} \cdot h_\theta \cdot \theta \, dt + R^{1/2}(\theta_0) \cdot dW_t \tag{2-19}$$

where $R(\theta_0)$ is given in (2-14), while

$$h_\theta := h_\theta(\theta_0, \theta_0)$$

is defined in (2-11).

COROLLARY 3 : Hypothesis H_1

$$D_n(\theta_0, \theta) \xrightarrow[n \to \infty]{L} N[-h_\theta \cdot \theta, R(\theta_0)] \tag{2-20}$$

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

PROOF: A first order Taylor expansion yields

$$H(\theta_0, \theta_0 + \frac{\theta}{\sqrt{n}} ; X_k)$$

$$= H(\theta_0, \theta_0 ; X_k) + \frac{1}{\sqrt{n}} H_z(\theta_0, \theta_0 ; X_k) \cdot \theta + o(\frac{1}{\sqrt{n}})$$

Summing over the index k yields

$$D_{n,t}(\theta_0, \theta, \tau)$$

$$= D_{n,t}(\theta_0, 0) + \frac{1}{n} \sum_{k=\min([nt],[n\tau])+1}^{[nt]} H_z(\theta_0, \theta_0 ; X_k) \cdot \theta + o(1)$$

$$= [1] + [2] + o(1) \tag{2-21}$$

The behaviour of [1] is described in the theorem 1 - ii. To analyse [2], we refer again to [Mac Leish 1975 - a] for a suitable law of large numbers for dependent processes, which yields

$$[2] \to (t - \tau)_+ \cdot h_z(\theta_0, \theta_0) \cdot \theta \tag{2-22}$$

where $x_+ = \max(0, x)$. But, the theorem 1 and the formulas (2-11), (2-21), (2-22) give together the theorem 2.

III ASYMPTOTIC LOCAL APPROACH TO CHANGE DETECTION.

3.1 The local test.

From the user's point of view, the cumulative sums $D_{n,m}$ given by the formulas (2-12) or (2-17) are identical, since they differ only via a change on the «true» parameter z, which is not observed by the user. This common user's form is simply obtained by deleting the true parameter z in these formulas; in other words, the user knows the cumulative sum built on the standard random vector field $H(\theta, X_n)$ of the formula (2-9):

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

We shall interpret the 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 vairables $Y_k(\theta_0)$ as independent, and distributed as follows*

$$Y_k(\theta_0) \approx N[0, R(\theta_0)], \quad k < r$$
$$Y_k(\theta_0) \approx N[-h_\theta(\theta_0) \cdot \theta, \ R(\theta_0)], \quad k \geq r \tag{3-2}$$

would exactly result in the asymptotic behaviour described by (2-19). 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 θ 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:*

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

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

$$S_r^n := \max_\theta S_r^n(\theta) = (\Delta_r^n)^T R^{-1} \Delta_r^n,$$
$$\Delta_r^n := (n - r + 1)^{-1/2} \sum_{k=r}^{n} Y_k,$$
$$\hat{\theta}(n, r) := \arg\max_\theta S_r^n(\theta) = -(n - r + 1)^{-1/2} \, h_\theta^{-1} \Delta_r^n \tag{3-3}$$

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

$$G_n := \max_r S_r^n, \quad \nu = \min\{n : G_n \geq \lambda\}$$
$$r_n = \arg\max_r S_r^n, \quad \hat{r} = \hat{r}_\nu, \quad \hat{\theta} = \hat{\theta}(\nu, \hat{r}) \tag{3-4}$$

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

$$E_{\theta_0}(S_r^n) = d, \tag{3-5}$$

since $(n - r + 1) S_r^n$ is (approximately) a central χ^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 χ^2-test of the form

$$(\theta_n - \theta_0)^T \Sigma^{-1} (\theta_n - \theta_0) \gtreqless \lambda \tag{3-6}$$

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 [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 [Willsky & Jones 1976] that the best local test involves the innovations of this Markov process, which gives something different from (3-6). 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]).

3.2 Examples.

3.2.1 Change detection in AR processes.

The objective is to detect changes in the parameter θ in the system

$$y_n = \Phi_n^T \theta + v_n, \quad \Phi_n^T = (y_{n-1}, \dots, y_{n-p}) \tag{3-7}$$

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

$$H(\theta, y_n, \Phi_n) := \Phi_n e_n(\theta), \quad e_n(\theta) := y_n - \Phi_n^T \theta \tag{3-8}$$

The matrix $R(\theta_0)$ corresponding to (2-14) is given by

$$R(\theta_0) = E_{\theta_0}(v_n \Phi_n \Phi_n^T v_n) = \sigma^2 \Sigma(\theta_0) \tag{3-9}$$

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

$$\Delta_r^n(\theta_0) = (n - r + 1)^{-1/2} \sum_{k=r}^{n} \Phi_k e_k(\theta_0) \tag{3-10}$$

It is easy to verify that $\sigma^{-2} \Delta_0^n(\theta_0)$ is the derivative with respect to θ 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-ii, 3-3, 3-4, 3-9, 3-10) 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

$$y_n = \sum_{i=1}^{p} a_i y_{n-i} + \sum_{j=1}^{q} b_j v_{n-j} + v_n \tag{3-11}$$

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

$$\begin{aligned} \theta^T &:= (a_1, \dots, a_p) \\ \Phi_n^T &:= (y_{n-1}, \dots, y_{n-p}) \\ \Psi_n^T &:= (y_{n-q-1}, \dots, y_{n-q-p}) \end{aligned} \tag{3-12}$$

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

$$\begin{aligned} \theta_n &= \theta_{n-1} + \frac{1}{n} \Gamma_n^{-1} \Psi_n (y_n - \Phi_n^T \theta_{n-1}) \\ \Gamma_n &= \Gamma_{n-1} + \frac{1}{n} (\Psi_n \Phi_n^T - \Gamma_{n-1}) \end{aligned} \tag{3-13}$$

The random vector field of interest is here equal to

$$H(\theta_0 ; \Psi_n, \Phi_n, y_n) = \Psi_n (y_n - \Phi_n^T \theta_0) := Y_n(\theta_0) \tag{3-14}$$

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

$$R(\theta_0) := \sum_{n \in Z} E_0 \left[Y_n(\theta_0) Y_0(\theta_0)^T \right]$$

$$= \sum_{n=-q}^{+q} E_0 \left[\Psi_n \Psi_0^T (y_n - \Phi_n^T \theta_0)(y_0 - \Phi_0^T \theta_0) \right]$$

$$h_\theta(\theta_0) = -E_0(\Psi_n \Phi_n^T), \tag{3-15}$$

where E_0 is a shortage for E_{θ_0, θ_0}. The *instrumental test* is obtained by combining the formulas (3-1-ii, 3-3, 3-4, 3-14, 3-15). 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 intrumental 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.

CONCLUSION

We have introduced a general method to associate to any identification procedure a change detection 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 likekihood ratio testing procedures. Our method extends the former one to procedures which are no more based on likelihood ratios. 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. Furthermore, this method provides as a direct byproduct correctly sounded procedures for model validation, as well as for the diagnosis of the origin of the changes, even when those changes are formulated in terms of (non identifiable) larger models; these latter points will be reported elsewhere.

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.

[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 & Soderstr˜om 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 fur Wahrscheinlichkeitstheorie verw. Geb.* 32, 165-178.

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

[Stoica, Soderstr˜om & 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].