Optimal sequential deduction and isolation of changes in stochastic systems

Igor Nikiforov

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Optimal sequential detection and isolation of changes in stochastic systems

Igor V. Nikiforov

N° 2063

Septembre 1993

PROGRAMME 5

Traitement du signal,
automatique
et productique

Rapport
de recherche

1993

ISSN 0249-6399

Optimal sequential detection and isolation of changes in stochastic systems

Igor V. Nikiforov *

Programme 5 — Traitement du signal, automatique et productique
Projet AS

Rapport de recherche n°2063 — Septembre 1993 — 34 pages

Abstract: The purpose of this paper is to give a new statistical approach to the change diagnosis (detection/isolation) problem. The change detection problem has received extensive research attention. On the contrary, change isolation is mainly an unsolved problem. We consider a stochastic dynamical system with abrupt changes and investigate the multihypothesis extension of Lorden's results. We introduce a joint criterion of optimality for the detection/isolation problem and then design a change detection/isolation algorithm. We also investigate the statistical properties of this algorithm. We prove a *lower bound for the criterion* in a class of sequential change detection/isolation algorithms. It is shown that the proposed algorithm is *asymptotically optimal* in this class. The theoretical results are applied to the case of additive changes in linear stochastic models.

Key-words: sequential change detection and isolation, generalized change detection problem, asymptotic optimality, linear stochastic models with additive changes.

(Résumé : tsvp)

*I.N. is INRIA and CNRS visitor, on leave from Institute of Control Sciences, Moscow, Russia. E-mail: nikiforov@irisa.fr

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : (33) 99 84 71 00 – Télécopie : (33) 99 38 38 32

Détection et isolation séquentielles optimales de changements dans les système stochastiques

Résumé : L'objectif de cette article est la détection et isolation de changements dans les système stochastiques.

Mots-clé : détection et isolation séquentielle de rupture, optimalité asymptotique, model stochastique linear avec des changements additifs.

Contents

I. Introduction

Statistical decision tools for detecting and isolating abrupt changes in the properties of stochastic signals and dynamical systems have numerous applications, from on-line fault diagnosis in complex technical systems to edge detection in images and detection of signals with unknown arrival time in geophysics, radar and sonar signal processing. For example, the early on-line fault diagnosis (change detection and isolation) in industrial processes helps in preventing these processes from more catastrophic failures. As another example, let us consider the problem of targets detection and identification. Assume that there are several types of targets. Each of them can appear at an unknown moment of time. The problem is to detect the fact that a target has arrived and classify the type of target as soon as possible.

The solution of the change diagnosis problem which is almost traditional now consists of subdividing this problem into two stages: change (fault) detection and change (fault) isolation, which are executed sequentially. As mentioned in pioneer paper [24], the *fault detection* (or *alarm task* by A.Willsky) consists of "making a binary decision - either that something has gone wrong or that everything is fine". The *fault isolation* "is that of determining the source of the fault".

The change detection problem has received extensive research attention. Mathematically the change detection problem can be formulated as that of the *quickest detection of abrupt changes in stochastic systems*. Recent results and references can be found in book [3]. The two main classes of quickest detection problems are the Bayesian and the non-Bayesian approaches. The first optimality results for the Bayesian approach were obtained in [18], [19]. More recent results can be found in [16]. The first algorithm for the non-Bayesian approach was suggested by E.S.Page in [15]. It was the cumulative sum algorithm (CUSUM). The asymptotic minimax ("worst case") optimality of the CUSUM was proved in [10], where G.Lorden has given a *lower bound* for the *worst mean delay for detection* and has proven that the CUSUM algorithm reaches this lower bound. Recently, nonasymptotic aspects of optimality for non-Bayesian algorithms were investigated in [12], [17]. The asymptotic minimax optimality of the CUSUM algorithm in the case of dependent random processes was obtained in [2].

The change isolation problem has been investigated much less. Up to our knowledge no proof of optimality in a mathematically precise sense exists. Moreover, because the quickest detection criterion of optimality does not take into account the isolation problem, the way to combine the detection and isolation algorithms together is not obvious. Therefore the following problems remain unsolved :

- What is a convenient criterion of optimality for the fault isolation problems?

- How may we avoid contradictions between the criteria of the detection and isolation stages? Typically a short mean detection delay is desirable, but a longer decision delay can improve the result of the isolation stage. We shall prove that this is not always necessary.
- What is a lower bound for the performance index in a class of detection/isolation algorithms?
- What is an optimal (or asymptotically optimal) algorithm which reaches this lower bound?

The goal of this paper is to present a new statistical method for *joint detection and isolation of changes* in the properties of stochastic systems, and to prove the *asymptotic optimality* of this method. The criterion of optimality of this "*generalized change detection problem*" consists in minimizing the worst mean detection/isolation delay for a given mean time before a false alarm or a false isolation.

The paper is the first attempt to solve this new problem. For this reason we assume that the statistical models before and after changes are known *exactly*. This means that we assume the case of *simple hypotheses*. The case of *composite hypotheses* will be discussed elsewhere. The paper is organized as follows.

First, we give the generalized change detection problem statement, the *basic model* which is a finite parametric family of distributions, and the criteria of optimality in section II. We also give an intuitive definition of the criteria and discuss some features of the proposed criteria.

Next, we design the change detection/isolation algorithm for the basic model in section III. We also investigate the statistical properties of this algorithm. The main results are stated in Theorem 1.

In section IV. we investigate a lower bound for the worst mean detection/isolation delay in a class of sequential change detection/isolation algorithms. The main result is established in Theorem 2.

Finally, in section V. we introduce two types of linear stochastic models with additive abrupt changes. The first type of stochastic models is a regression model with redundancy. The second type is a stochastic dynamical model. In these two cases we show how the *new* change detection/isolation problems can be reduced to the *basic problem* and we discuss some new features which play a key role in these new models. We also investigate the statistical properties of the change detection/isolation algorithm. The results are given in Theorem 3 and Theorem 4.

II. Problem statement

In this section we give an intuitive formulation of the change detection/isolation criteria and discuss some features of the proposed criteria. Next, we define the basic model which is a finite parametric family of distributions and give a formal definition of the criteria.

A. Intuitive formulation of the change detection/isolation problem

Let us assume that there exists a discrete time stochastic dynamical model $\mathcal{F}_n(\theta)$, $n = 1, 2, \ldots$. The vector $\theta \in \mathbf{R}^r$ is the parameter of interest. Let $\mathcal{F} = \{\mathcal{F}(\theta) : \theta \subset \Omega, \quad \Omega = \bigcup_{l=0}^{K-1} \theta_l\}$ be a finite family (K members) of this model. Until the unknown time k the vector is $\theta = \theta_0$ and from $k+1$ it becomes $\theta = \theta_l$ for some l, $l = 1, \ldots, K-1$. Therefore $\mathcal{F}_n(\theta)$ is the model with *abrupt changes* :

$$\mathcal{F}_n(\theta) = \begin{cases} \mathcal{F}(\theta_0) & \text{if} \quad n \leq k \\ \mathcal{F}(\theta_l) & \text{if} \quad n \geq k+1 \end{cases}, \text{for some } l = 1, \ldots, K-1, \text{and } k = 0,1,2, \ldots, \quad (1)$$

where $\mathcal{F}(\theta_0)$ is the normal operation mode of the model $\mathcal{F}$ and $\mathcal{F}(\theta_l)$ is the mode with fault number $l \geq 1$. We assume that the values of θ_l are known *a priori*. The change time $k+1$ and number l are unknown.

Let $(Y_n)_{n \geq 1}$ be a sequence of observations, which are coming from system (1). The problem is to *detect* and *isolate* the change in θ. In other words we have to determine the type of fault (number l) as soon as possible. The change detection/isolation algorithm has to compute a *pair* (N, ν) based on the observations $Y_1, Y_2, \ldots$, where N is the *alarm time* at which the ν-type change is detected/isolated and ν, $\nu = 1, \ldots, K-1$, is the *final decision*. In other words, at time N the hypothesis $\mathcal{H}_\nu : \{\theta = \theta_\nu\}$ is accepted.

The following three situations can occur :

- If the change is detected/isolated *after* time k ($N > k$ is true), then the delay for detection/isolation of l-type change is

$$\tau_l = N - k.$$

- On the contrary, if the changes in θ are detected *before* time k or if the final decision is *false* ($\nu \neq l$), then these are *false alarms or false isolations* which we charactterize in the following manner :

  - **False alarms.** Let the observations $(Y_n)_{n \geq 1}$ come from the normal mode system $\mathcal{F}(\theta_0)$. Consider the following sequence of alarm times

$$N_0 = 0 < N_1 < N_2 < \cdots < N_r < \cdots,$$

  where N_r is the alarm time of the detection/isolation algorithm which is applied to $Y_{N_{r-1}+1}, Y_{N_{r-1}+2}, \ldots$. Define the *first* false alarm time $N^{\nu=j}$ of j-type in this sequence :

$$N^{\nu=j} = \inf_{r \geq 1} \{N_r \ : \ \nu_r = j\}, \quad 1 \leq j \leq K-1,$$

where $\inf\{\emptyset\} = \infty$ as usual.

– **False isolations.** In order to avoid uncertainties in initial conditions let us assume that $k = 0$. In other words we assume that the observations $(Y_n)_{n\geq 1}$ come from the mode $\mathcal{F}(\theta_l)$ with fault number $l \geq 1$. Define the *first* false isolation time $N^{\nu=j}$ of j-type in this sequence :

$$N^{\nu=j} = \inf_{r\geq 1}\{N_r \; : \; \nu_r = j\}, \quad 1 \leq j \neq l \leq K-1.$$

It is intuitively obvious that the optimality criterion must favor fast detection/isolation with few false alarms and few false isolations. In other words the delay $\tau_l = N - k$ given that $N > k$ should be stochastically small for each $l = 1, \ldots, K-1$, and $N^{\nu=j} = \inf_{r\geq 1}\{N_r \; : \; \nu_r = j\}$ should be stochastically large for each combinations of numbers $j \neq l$.

B. The basic model

We consider a finite family of distributions $\mathcal{P} = \{P_i, \; i = 0, \ldots, K-1\}$ with densities $\{p_i, \; i = 0, \ldots, K-1\}$ with respect to a measure μ. In the parametric case, we assume that $\mathcal{P} = \{P_\theta, \; \theta \in \Omega\}$, where $\Omega = \bigcup_{i=0}^{K-1} \theta_i, \;\; \Omega \subset \mathbf{R}^r$ and we denote the density function of this family by $p_\theta(X)$.

Let $(X_n)_{n\geq 1}$ be an independent random sequence observed *sequentially* and $X_1, \ldots, X_k$ have distribution P_0 while $X_{k+1}, X_{k+2}, \ldots$ have distribution $P_l, l = 1, \ldots, K-1$:

$$\mathcal{L}(X_n) = \begin{cases} P_0 & \text{if} \quad n \leq k \\ P_l & \text{if} \quad n \geq k+1 \end{cases}, \quad k = 0, 1, 2, \ldots \tag{2}$$

The change time $k+1$ and number l are unknown. We assume that for this family of distributions the following inequality is true :

$$0 < \rho_{ij} = \int p_i \ln \frac{p_i}{p_j} d\mu < \infty, \quad 0 \leq i \neq j \leq K-1, \tag{3}$$

where ρ_{ij} is the Kullback-Leibler information.

C. Formal definition of criteria

Let us pursue our discussion of the criteria of optimality. The change detection/isolation algorithm consists in computing a *pair* (N, ν) based on the observations $X_1, X_2, \ldots$. Let P_{k+1}^l be the distribution of observations $X_1, X_2, \ldots$ (2) when $k = 0, 1, 2, \ldots$ and E_{k+1}^l be the expectation under P_{k+1}^l. Therefore, the mean delay for the detection/isolation of l-type change is

$$\overline{\tau}_l = E_{k+1}^l(N - k | N > k, X_1, \ldots, X_k), \tag{4}$$

where $E(\ |\ )$ is the conditional expectation. It is obvious that, without knowing *a priori* the distribution of the change time $k+1$, the mean decision delay defined in (4) is a function of k and the past "trajectory" of the random sequence $X_1, \ldots X_k$. In many practical cases, it is useful to have an algorithm which is independent of the distribution of the change time $k+1$ and of the sample path of the observations $X_1, \ldots X_k$. For this reason we use another *minimax* performance index, which has been introduced in [10]. Hence, the *worst mean detection/isolation delay* is[1]

$$\overline{\tau}_l^* = \sup_{k \geq 0} \operatorname{esssup} E_{k+1}^l (N-k | N > k, X_1, \ldots, X_k). \tag{5}$$

On the other hand the mean time before the first false alarm $N^{\nu=j}$ of j-type is defined by the following formula :

$$E_0(N^{\nu=j}) = E_0 \left(\inf_{r \geq 1} \{ N_r \ :\ \nu_r = j \} \right),$$

where $E_0(\ ) = E_\infty^l(\ )$. Analogously, the mean time before the first false isolation $N^{\nu=j}$ of j-type is :

$$E_l(N^{\nu=j}) = E_l \left(\inf_{r \geq 1} \{ N_r \ :\ \nu_r = j \} \right),$$

where $E_l(\ ) = E_1^l(\ )$.

Let us consider the following *minimax* criterion. We require that the worst mean detection/isolation delay :

$$\overline{\tau}^* = \sup_{k \geq 0, 1 \leq l \leq K-1} \operatorname{esssup} E_{k+1}^l (N-k | N > k, X_1, \ldots, X_k) \tag{6}$$

should be *as small as possible* for a given minimum $\overline{T}$ of the mean times before a false alarm or a false isolation :

$$\min_{0 \leq i \leq K-1} \min_{1 \leq j \neq i \leq K-1} E_i \left(N^{\nu=j} \right) \geq \overline{T}. \tag{7}$$

Remark 1 *Usually, in the classical change detection problem the mean time before false alarm is equal to the mean time* between *false alarms. It follows from the fact that the system inspection and repair times are not of interest of as and thus are assumed to be zero. In other words we assume that the process of observation is restarted immediately as at the beginning. In the problem discussed we can assume that the $\overline{T}$ is equal to the minimum mean time between false alarms under the same assumptions.*

[1] Let us assume that $y, \acute{y}, x$ are the random values. We say that the $y = \operatorname{esssup} x$ if : 1) $\mathbf{P}(x \leq y) = 1$; 2) if $\mathbf{P}(x \leq \acute{y}) = 1$ then $\mathbf{P}(y \leq \acute{y}) = 1$, where $\mathbf{P}(A)$ is the probability of the event A.

Discussion. Let us explain criterion (6) - (7). Originally, this type of criterion was introduced for the quickest change detection problem in [10]. For a given change time $k+1$, the conditional expectation

$$\overline{\tau}_l = E^l_{k+1}(N-k|N>k, X_1,\ldots,X_k)$$

is a random value. For this reason we have to use the *essential supremum* $\operatorname{ess\,sup} \overline{\tau}_l(X_1,\ldots,X_k)$ in order to reach the smallest $A_l(k)$ such that

$$\overline{\tau}_l(X_1,\ldots,X_k) \leq A_l(k)$$

almost surely under the probability measure μ of the observations $X_1,\ldots,X_k$. After this we have to use the supremum $\sup_{k\geq 0, 1\leq l\leq K-1} A_l(k)$ in order to guarantee the worst mean detection/isolation delay with respect to the unknown change time and number l of the hypothesis $\mathcal{H}_l$. On the other hand we constrain the minimum value $\overline{T}$ of the mean time before a false alarm or a false isolation in order to guarantee some acceptable level of false solutions.

III. The basic algorithm and its properties

This section is organized in the following manner. First, we design the joint detection/isolation algorithm. Then we investigate the asymptotic statistical properties of this algorithm using criterion (6) - (7).

A. The change detection/isolation algorithm.

Sequential hypotheses testing problem. Let us start with the Armitage sequential probability ratio test (SPRT) for statistical multihypothesis testing problem (see [1] or handbook [9, p.237]). Again we suppose that $\mathcal{P} = \{P_i : i = 0,\ldots,K-1\}$ is a finite family of distributions.

Hence, we consider K simple hypotheses :

$$\mathcal{H}_i : \{\mathcal{L}(X_n)_{n\geq 1} = P_i\}, \;\; i = 0,\ldots,K-1.$$

We are observing sequentially an independent random sequence $(X_n)_{n\geq 1}$ with density p_i, where p_i is the density of an unknown member P_i of the family $\mathcal{P}$. The multihypothesis SPRT is nothing but the following pair (M,d), where M is the *stopping time* and d is the *final decision* which are defined in the following manner :

$$M = \min_{l=0,\ldots,K-1} M_l \tag{8}$$

$$M_l = \inf\left\{n\geq 1 : \min_{0\leq j\neq l\leq K-1} [S_1^n(l,j) - h_{lj}] \geq 0\right\} \tag{9}$$

$$d = \arg\min_{l=0,\ldots,K-1} M_l,$$

where

$$S_k^n(l,j) = \sum_{i=k}^{n} \ln \frac{p_{\theta_l}(X_i)}{p_{\theta_j}(X_i)}$$

is the *log likelihood ratio* (LR) between hypotheses $\mathcal{H}_l$ and $\mathcal{H}_j$ and h_{lj} are chosen thresholds.

The interpretation of SPRT (8) is very simple : it stops the first time n at which there exists some hypothesis $\mathcal{H}_l$ for which each log likelihood ratio $S_k^t(l,j)$ between $\mathcal{H}_l$ and $\mathcal{H}_j$ is greater than or equal to a chosen threshold h_{lj}, $0 \leq j \neq l \leq m-1$.

Design of the detection/isolation algorithm. The idea of the proposed detection/isolation algorithm is based on a class of "extended stopping variables" which was introduced in [10] originally. Because we are not interested in the detection of hypothesis $\mathcal{H}_0$ we will assume that $l = 1, \ldots, m-1$.

Let us introduce the following stopping time

$$\tilde{N} = \min\{\tilde{N}^1, \ldots, \tilde{N}^{K-1}\} \tag{10}$$

and final decision

$$\tilde{\nu} = \text{argmin}\{\tilde{N}^1, \ldots, \tilde{N}^{K-1}\} \tag{11}$$

of the detection/isolation algorithm. Hence, the stopping time $\tilde{N}^l$ is responsible for the detection of hypothesis $\mathcal{H}_l$. Define $\tilde{N}^l$ by the following formula :

$$\begin{aligned} \tilde{N}^l &= \inf_{k \geq 1} \tilde{N}^l(k) \qquad (12) \\ \tilde{N}^l(k) &= \inf \left\{ n \geq k : \min_{0 \leq j \neq l \leq K-1} S_k^n(l,j) \geq h \right\} \end{aligned}$$

Discussion. Let us discuss the design of algorithm (10) - (12). It follows from (12) that the algorithm is based on the concept that is very important in detection theory, namely the log likelihood ratio

$$S_k^n(l,j) = \sum_{i=k}^{n} \ln \frac{p_l(X_i)}{p_j(X_i)} \tag{13}$$

between hypotheses $\mathcal{H}_l$ and $\mathcal{H}_j$. The key statistical properties of this ratio is as follows :

$$E_j(S_k^n) < 0 \text{ and } E_l(S_k^n) > 0.$$

In other words, a change in statistical model (2) is reflected as a change in the sign of the log LR mean.

If we have to *detect* a change in a distribution then the classical optimal solution of this problem is the *cumulative sum* (CUSUM) algorithm [15]. The

CUSUM algorithm is based on the comparison of the difference between the value of the log LR and its current minimum value :

$$g_n = S_1^n - \min_{0 \le k < n} S_1^k = \max_{1 \le k \le n} S_k^n, \quad S_1^0 = 0$$

with a given threshold. In other words the CUSUM stops at time N_c if, for some $k < N_c$, the observations $X_k, \ldots, X_{N_c}$ are *significant* for accepting the hypothesis about change :

$$N_c = \inf \left\{ n \ge 1 : \max_{1 \le k \le n} S_k^n \ge h \right\}.$$

If we have to *detect and isolate* l-type change in the model then, generally speaking, we may exploit the previous idea with some modifications. Now we have the set $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_{K-1}$ of alternatives. For this reason $\tilde{N}^l$ stops if, for some $k < \tilde{N}^l$, the observations $X_k, \ldots, X_{\tilde{N}^l}$ are *significant* for accepting the hypothesis $\mathcal{H}_l$ with respect to this set of alternatives

$$\tilde{N}^l = \inf \left\{ n \ge 1 : \max_{1 \le k \le n} \min_{0 \le j \ne l \le K-1} S_k^n(l,j) \ge h \right\}. \tag{14}$$

On the other hand stopping time (14) can be interpreted as the *generalized likelihood ratio* (GLR) algorithm [10], [11], [24]. The GLR algorithm for testing between two *composite* hypotheses $\mathbf{H}_0 = \{\theta \in \Theta_0\}$ and $\mathbf{H}_1 = \{\theta \in \Theta_1\}$ is based on the following statistics

$$\Lambda_k^n = \frac{\sup_{\theta \in \Theta_1} p_\theta(X_1^n)}{\sup_{\theta \in \Theta_0} p_\theta(X_1^n)}.$$

In our case the hypothesis $\mathbf{H}_1 = \mathcal{H}_l$ is simple and the hypothesis $\mathbf{H}_0 = \bigcup_{0 \le j \ne l \le K-1} \mathcal{H}_j$ is composite, but finite. Hence,

$$\begin{aligned} \ln \Lambda_k^n &= \ln \frac{p_l(X_k^n)}{\max_{0 \le j \ne l \le K-1} \{p_j(X_k^n)\}} \\ &= \ln \min_{0 \le j \ne l \le K-1} \left\{ \frac{p_l(X_k^n)}{p_j(X_k^n)} \right\} = \min_{0 \le j \ne l \le K-1} S_k^n(i,j). \end{aligned}$$

Let us add a comment on the threshold issue. In algorithm (10) - (12) we use the threshold h instead of h_{lj} in SPRT (8). The main idea of this choise is the fact that the level of false alarms (or false isolations) is a function of the thresholds (see Lemma 1 for details). Therefore, to have the same level of false decisions for the separate stopping times $\tilde{N}^l$ we choose the same level of the thresholds $h_{lj} = h$.

B. Statistical properties

Now we investigate the statistical properties, namely the relation between the worst mean detection/isolation delay $\overline{\tau}^*$ given in (6) and the mean time $\overline{T}$ before a false alarm or a false isolation given in (7).

The main result is stated in the following theorem :

Theorem 1 *Let $(\tilde{N}, \tilde{\nu})$ be detection/isolation algorithm (10) - (12). Then :*

$$\begin{aligned} \overline{\tau}^* &\leq \max_{1 \leq l \leq K-1} E_l(\tilde{N}) \sim \frac{\ln \overline{T}}{\rho^*} \quad \textit{as } \overline{T} \to \infty \\ \rho^* &= \min_{1 \leq l \leq K-1} \min_{0 \leq j \neq l \leq K-1} \rho_{lj} \, . \end{aligned} \tag{15}$$

The proof uses the following three results.

Lemma 1 *Let $\tilde{N}^j$ be stopping variables (12) with respect to $X_1, X_2, \ldots \sim P_l$. Then*

$$E_l(\tilde{N}^j) \geq e^h \quad \textit{for } l = 0, \ldots, K-1, \;\; 1 \leq j \neq l \leq K-1. \tag{16}$$

Proof of Lemma 1:

It is known from Lorden's Theorem 2 [10] that the expectation of the stopping variable $N = \inf_{k \geq 1}\{n(k)\}$, where $n(k)$ is the stopping time of the open-ended SPRT which is activated at time k, satisfies the following inequality

$$E_l(N) \geq \frac{1}{\alpha}$$

when $\mathbf{P}_l(n(1) < \infty) \leq \alpha$. Hence, it will suffice to show that

$$\mathbf{P}_l(\tilde{N}^j(1) < \infty) \leq e^{-h}.$$

Let us consider the following stopping variables :

$$\tilde{N}^j(1) = \inf \left\{ n \geq 1 : \min_{0 \leq i \neq j \leq K-1} S_1^n(j,i) \geq h \right\}$$

and let B denote the event $\{\tilde{N}^j(1) < \infty\}$, then

$$B = \bigcap_{i \neq j} A_i,$$

where $A_i = \{\tilde{N}^{ji}(1) < \infty\}$ and

$$\tilde{N}^{ji}(1) = \inf \left\{ n \geq 1 : S_1^n(j,i) \geq h \right\}.$$

It is easy to show that

$$\mathbf{P}_l(B) \leq \mathbf{P}_l(A_l).$$

Hence, from the above formula and Wald's inequality $\mathbf{P}_l(A_l) \leq e^{-h}$ (see [22, pp.40-44]) it results that

$$\mathbf{P}_l(\tilde{N}^j(1) < \infty) \leq e^{-h}. \tag{17}$$

The proof of Lemma 1 is complete.

Lemma 2 *Let*
$\tilde{N}^j$ be stopping variables (12) with respect to $X_1, \ldots, X_k, X_{k+1}, \ldots \sim P^j_{k+1}$. Then

$$\overline{\tau}^*_j \leq E_j[\tilde{N}^j(1)] \sim \frac{h}{\min_{0 \leq i \neq j \leq K-1} \rho_{ji}} \quad \textit{as} \;\; h \to \infty. \tag{18}$$

Proof of Lemma 2: The first part of the proof follows from Lorden's Theorem 2 [10]. Namely, note here that the event $\{\tilde{N}^j \leq k\}$, where :

$$\tilde{N}^j = \inf_{k \geq 1} \tilde{N}^j(k)$$

is the union of $\{\tilde{N}^j(1) \leq k\}, \{\tilde{N}^j(2) \leq k-1\}, \ldots, \{\tilde{N}^j(k) \leq 1\}$. From this results that the worst mean detection/isolation delay satisfies the following inequality

$$\overline{\tau}^*_j \leq E_j(\tilde{N}^j(1)).$$

Let us define the following stopping variable $M(h)$:

$$M = \inf \left\{ n \geq 1 : \min \left(\sum_{i=1}^{n} y^1_i, \ldots, \sum_{i=1}^{n} y^{K-1}_i \right) \geq h \right\},$$

where $Y_1, \ldots, Y_n$, $Y_i = (y^1_i, \ldots, y^{K-1}_i)^T$ is a sequence of independent identically distributed (i.i.d.) random vectors. Moreover, we assume that

$$0 < \min(m_1, , \ldots, m_{K-1}), \;\; \text{where} \;\; m_j = E(y^j).$$

From Farrell's Theorem 3 [7] we know the following properties of M :

$$\mathbf{P}(M < \infty) = 1; \;\; \lim_{h \to \infty} h^{-1} E(M(h)) = \frac{1}{\min(m_1, m_2 \ldots, m_{K-1})}.$$

In the case of stopping variable $\tilde{N}^j$ we have $m_i = \rho_{ji} > 0$. Therefore,

$$E_j(\tilde{N}^j(1)) \sim \frac{h}{\min_{0 \leq i \neq j \leq K-1} \rho_{ji}}.$$

The proof of Lemma 2 is complete.

Corollary 1 *Let $\tilde{N}$ be detection/isolation algorithm (10) - (12). Then*

$$\overline{\tau}^* \leq \max_{1 \leq j \leq K-1} E_j[\tilde{N}^j(1)] \sim \frac{h}{\rho^*} \quad \textit{as} \;\; h \to \infty. \tag{19}$$

Proof: Formula (19) follows at once from the definition of $\tilde{N}$ (10) and formula (18) (Lemma 2).

Proof of Theorem 1:

First, let us show that

$$E_0(\tilde{N}^{\nu=j}) \geq e^h \quad j = 1, \ldots, K-1.$$

Define the following two events: $\{\tilde{N}^{\nu=j} \leq n\}$, where $\tilde{N}^{\nu=j} = \inf_{r\geq 1}\{\tilde{N}_r : \nu_r = j\}$, and $\{\tilde{N}^j \leq n\}$. Denote by r_0^j the argument of the above minimum. It is obvious that $\{\tilde{N}^{\nu=j} \leq n | r_0^j\} = \{\tilde{N}^j \leq n\}$ when $r_0^j = 1$ and $\{\tilde{N}^{\nu=j} \leq n | r_0^j\} \subset \{\tilde{N}^j \leq n\}$ when $r_0^j > 1$. Therefore,

$$\mathbf{P}_0(\tilde{N}^{\nu=j} \leq n | r_0^j) \leq \mathbf{P}_0(\tilde{N}^j \leq n) \;\text{ when }\; r_0^j \geq 1$$

and

$$\mathbf{P}_0(\tilde{N}^{\nu=j} > n | r_0^j) \geq \mathbf{P}_0(\tilde{N}^j > n) \;\text{ when }\; r_0^j \geq 1.$$

Now, since

$$E_0(\tilde{N}^{\nu=j} | r_0^j) = \sum_{n=0}^{\infty} \mathbf{P}_0(\tilde{N}^{\nu=j} > n | r_0^j) \geq \sum_{n=0}^{\infty} \mathbf{P}_0(\tilde{N}^j > n) = E_0(\tilde{N}^j)$$

we have from Lemma 1 that

$$E_0(\tilde{N}^j) \geq e^h.$$

Finally, we get

$$E_0(\tilde{N}^{\nu=j}) = E_0\left[E_0(\tilde{N}^{\nu=j} | r_0^j)\right] \geq e^h. \tag{20}$$

Second, let us show that

$$E_l(\tilde{N}^{\nu=j}) \geq e^h, \;\; l = 1, \ldots, K-1, \;\; 1 \leq j \neq l \leq K-1. \tag{21}$$

The proof of this step of Theorem 1 is much the same as the previous step. It will suffice to use $\mathbf{P}_l(\;)(E_l(\;))$ instead of $\mathbf{P}_0(\;)(E_0(\;))$.

Relation (15) follows at once from (20), (21) and Corollary 1. The proof of Theorem 1 is complete.

IV. Asymptotic theory

In this section we prove a lower bound for the worst mean detection/isolation delay over the class $\mathcal{K}_\gamma$ of sequential change detection/isolation algorithms. First, we give some technical results on sequential multihypothesis tests and then we prove this asymptotic lower bound for $\overline{\tau}^*$. Finally, we compare this new lower bound with the lower bound for the worst mean delay of the classical change detection (without isolation !) algorithms which have been mentioned before.

Lemma 3 *Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables. Let $\mathcal{H}_0, \ldots, \mathcal{H}_{K-1}$ be $K \geq 2$ hypotheses, where $\mathcal{H}_i$ is the hypothesis that X has density p_i with respect to some probability measure μ, for $i = 0, \ldots, K-1$ and assume inequality (3) to be true. Let $E_i(N)$ be the average sample number (ASN) in a sequential test which chooses one of the K hypotheses subject to a $K \times K$ error matrix $A = \|a_{ij}\|$, where $a_{ij} = \mathbf{P}_i(\text{accepting } \mathcal{H}_j)$, $i, j = 0, \ldots, K-1$. Let us reparameterize matrix A in the following form :*

$$A = \begin{pmatrix} 1-\sum_{l=1}^{K-1}\alpha_l & \alpha_1 & \ldots & \alpha_{K-1} \\ \gamma_1 & 1-\sum_{l=2}^{K-1}\beta_{1l}-\gamma_1 & \ldots & \beta_{1,K-1} \\ \gamma_2 & \beta_{21} & \ldots & \beta_{2,K-1} \\ \ldots & \ldots & \ldots & \ldots \\ \gamma_i & \beta_{i1} & \ldots & \beta_{i,K-1} \\ \ldots & \ldots & \ldots & \ldots \\ \gamma_{K-1} & \beta_{K-1,1} & \ldots & 1-\sum_{l=1}^{K-2}\beta_{K-1,l}-\gamma_{K-1} \end{pmatrix}. \quad (22)$$

Then a lower bound for $E_i(N)$ is given by the following formula :

$$E_i(N) \geq \max\left\{\frac{(1-\tilde{\gamma}_i)\ln\alpha_i^{-1}-\ln 2}{\rho_{i0}}, \max_{1\leq j\neq i\leq K-1}\left(\frac{(1-\tilde{\gamma}_i)\ln\beta_{ji}^{-1}-\ln 2}{\rho_{ij}}\right)\right\} \quad (23)$$

for $i = 1, \ldots, K-1$, where $\tilde{\gamma}_i = \gamma_i + \sum_{l=1, l\neq i}^{K-1}\beta_{il}$.

Proof:

Let us proof the first part of inequality (23). The following inequality appears in Theorem 3.1 [20] as a *generalized Wald lower bound* for ASN :

$$E_i(N)\rho_{ij} \geq \sum_{l=0}^{K-1} a_{il} \ln \frac{a_{il}}{a_{jl}} \quad \text{for} \quad i = 1, \ldots, K-1,$$

where index j is arbitrary except for $j \neq i$. Let us assume that $j = 0$ and $1 \leq i \leq K-1$. In accordance with the notations (22), (23) and Lemma 6 [20] we can write

$$\sum_{l=0}^{K-1} a_{il} \ln \frac{a_{il}}{a_{jl}} = \gamma_i \ln \frac{\gamma_i}{1-\sum_{l=1}^{K-1}\alpha_l} + \left(1 - \sum_{l=1, l\neq i}^{K-1}\beta_{il} - \gamma_i\right) \cdot$$

$$
\begin{aligned}
&\cdot \ln \frac{\left(1-\sum_{l=1,l\neq i}^{K-1}\beta_{il}-\gamma_i\right)}{\alpha_i} + \sum_{l=1,l\neq i}^{K-1} \beta_{il} \ln \frac{\beta_{il}}{\alpha_l} \\
\geq\ & \tilde{\gamma}_i \ln \frac{\tilde{\gamma}_i}{1-\alpha_i} + (1-\tilde{\gamma}_i) \ln \frac{1-\tilde{\gamma}_i}{\alpha_i},
\end{aligned}
$$

where $\tilde{\gamma}_i = \gamma_i + \sum_{l=1,l\neq i}^{K-1} \beta_{il}$. Finally, the following inequality follows from the fact that $\tilde{\gamma}_i \ln(1-\alpha_i)$ is nonnegative and the minimum value of $\tilde{\gamma}_i \ln \tilde{\gamma}_i + (1-\tilde{\gamma}_i)\ln(1-\tilde{\gamma}_i)$ is equal to $-\ln 2$:

$$
E_i(N) \geq \frac{(1-\tilde{\gamma}_i)\ln \alpha_i^{-1} - \ln 2}{\rho_{i0}} \quad \text{for } i = 1, \ldots, K-1. \tag{24}
$$

Let us proof the second part of inequality (23). Let $1 \leq i \leq K-1$ and $1 \leq j \neq i \leq K-1$. And again we can write by the same arguments :

$$
\begin{aligned}
\sum_{l=0}^{K-1} a_{il} \ln \frac{a_{il}}{a_{jl}} &= \gamma_i \ln \frac{\gamma_i}{\gamma_j} + \left(1 - \sum_{l=1,l\neq i}^{K-1} \beta_{il} - \gamma_i\right) \ln \frac{\left(1-\sum_{l=1,l\neq i}^{K-1}\beta_{il}-\gamma_i\right)}{\beta_{ji}} \\
&+ \beta_{ij} \ln \frac{\beta_{ij}}{1-\sum_{l=1,l\neq i}^{K-1}\beta_{il}-\gamma_i} + \sum_{l=1,l\neq i,j}^{K-1} \beta_{il} \ln \frac{\beta_{il}}{\beta_{jl}} \\
&\geq \tilde{\gamma}_i \ln \frac{\tilde{\gamma}_i}{1-\beta_{ji}} + (1-\tilde{\gamma}_i) \ln \frac{1-\tilde{\gamma}_i}{\beta_{ji}} \\
&\geq (1-\tilde{\gamma}_i) \ln \beta_{ji}^{-1} - \ln 2
\end{aligned} \tag{25}
$$

Therefore, we have the following formula

$$
E_i(N) \geq \max_{1\leq j\neq i\leq K-1} \left\{ \frac{(1-\tilde{\gamma}_i)\ln \beta_{ji}^{-1} - \ln 2}{\rho_{ij}} \right\} \quad \text{for } i = 1, \ldots, K-1. \tag{26}
$$

Definition 1 *Let $\mathcal{K}_\gamma$ be the* class *of all sequential detection/isolation algorithms (N,ν), where N is the extended stopping variable and ν is the final decision, that satisfy the following inequalities*

$$
\min_{0\leq i\leq m-1} \min_{1\leq j\neq i\leq m-1} E_i\left(\min_{r\geq 1}\{N_r \ : \ \nu_r = j\}\right) \geq \gamma. \tag{27}
$$

Theorem 2 *Consider class (27). Let us define the lower bound $n(\gamma)$ as the infimum of the worst mean detection/isolation delay in the class $\mathcal{K}_\gamma$*

$$
n(\gamma) = \inf_{(N,\nu)\in\mathcal{K}_\gamma} (\overline{\tau}^*).
$$

Let inequality (3) be true. Then

$$
n(\gamma) \sim \frac{\ln \gamma}{\rho^*} \quad as \ \gamma \to \infty, \tag{28}
$$

where

$$\rho^* = \min_{1 \le l \le K-1} \min_{0 \le j \ne l \le K-1} \rho_{lj}.$$

Proof:

The proof consists of two parts. The first part includes the derivation of the asymptotic relation between the worst mean detection/isolation delay and the mean time before false alarms. The second part includes the derivation of an analogous result for false isolations.

Note that the scheme of our proof is as in Lorden's Theorem 3 (see details in [10]). The novelty is the extension of Lorden's results to the case of $K > 2$ hypotheses.

The first part:

It is sufficient to show that for every $\epsilon_l \in (0,1)$ there exists $C_1(\epsilon_l) < \infty$ $l = 1, \ldots, K-1$ such that for all $(N, \nu) \in \mathcal{K}_\gamma$ the following inequality is true

$$\overline{\tau}_l^* \ge \frac{(1-\epsilon_l)\ln\gamma + C_1(\epsilon_l)}{\rho_{l0}}, \quad l = 1, \ldots, K-1. \tag{29}$$

As in [10] let us introduce the following "additional" stopping variables :

$$T_0 = 0 < T_1 < T_2 \cdots,$$

where $T_{r+1} (r = 0, 1, \ldots)$ is defined by

$$\begin{aligned} T_{r+1} &= \max_{1 \le l \le K-1} T_{r+1}^l \\ T_{r+1}^l &= \inf\left\{ n \ge T_r + 1 : \frac{p_l(X_{T_r+1}) \cdots p_l(X_n)}{p_0(X_{T_r+1}) \cdots p_0(X_n)} \le \epsilon_l \right\} \end{aligned}$$

Note here that $\mathbf{P}_l(T_r^l < \infty) \le \epsilon_l$ (see [22, pp.40-44]). Let us assume that $X_1, \ldots, X_k \sim P_0$ and $X_{k+1}, X_{k+2}, \ldots \sim P_l$. In other words P_{k+1}^l is the distribution of observations $X_1, \ldots, X_k, X_{k+1}, \ldots$. It follows from the same Wald's arguments that

$$\mathbf{P}_{k+1}^l(T_r^l < \infty | T_{r-1} = k < N) \le \epsilon_l$$

provided that

$$\mathbf{P}_0(T_{r-1} = k < N) > 0.$$

Moreover, it is obvious that

$$\{T_r < \infty\} = \bigcap_{l=1}^{K-1} \{T_r^l < \infty\}.$$

For this reason the following inequality holds

$$\mathbf{P}_{k+1}^l(T_r < \infty | T_{r-1} = k < N) \le \epsilon_l.$$

We denote all subsets $D_{rk} = \{T_{r-1} = k < N\}$, for which $\mathbf{P}_0(D_{rk}) > 0$ and, hence, also $\mathbf{P}^l_{k+1}(D_{rk}) > 0$. Let us define the following sequential test (N^*, d^*) on the subset D_{rk} by using the stopping variables T_r,N and the final decision ν :

$$\begin{aligned} N^* &= \min\{N, T_r\} \\ d^* &= \begin{cases} \nu & \text{if} \quad N \leq T_r \\ 0 & \text{if} \quad N > T_r \end{cases} \end{aligned} \tag{30}$$

In other words, at time N^* one of the following hypotheses $\mathcal{H}_{0 \leq d^* \leq K-1}$ is accepted :

$$\begin{cases} \mathcal{H}_0 & : \quad X_{k+1}, X_{k+2}, \ldots, X_{N^*} & \sim P_0 & \text{no change} \\ \mathcal{H}_1 & : \quad X_{k+1}, X_{k+2}, \ldots X_{N^*} & \sim P_1 & \text{1-type of change} \\ \ldots & : \quad \ldots & \ldots & \ldots \\ \mathcal{H}_{K-1} & : \quad X_{k+1}, X_{k+2}, \ldots X_{N^*} & \sim P_{K-1} & K-\text{1-type of change} \end{cases} .$$

Let us consider the statistical properties of this sequential test. The conditional expectation of the sample number of observations taken for the test is

$$E^l_{k+1}(N^* - k | D_{rk}).$$

It results from the definition of the worst mean detection/isolation delay that

$$\begin{aligned} \overline{\tau}^*_l &= \sup_{k \geq 0} \operatorname{esssup} E^l_{k+1}(N - k | N > k, X_1, \ldots, X_k) \\ &\geq \sup_{k \geq 0} \operatorname{esssup} E^l_{k+1}(\min\{N, T_r\} - k | D_{rk}) \\ &\geq E^l_{k+1}(N^* - k | D_{rk}) \end{aligned}$$

On the subset D_{rk} Lemma 3 holds with the probabilities

$$\begin{aligned} 1 - \tilde{\gamma}_l &= \mathbf{P}_l(\text{accepting } \mathcal{H}_l) = \mathbf{P}^l_{k+1}(N \leq T_r \cap \nu = l | D_{rk}) \\ \alpha_l &= \mathbf{P}_0(\text{accepting } \mathcal{H}_l) = \mathbf{P}_0(N \leq T_r \cap \nu = l | D_{rk}) \end{aligned}$$

So, we have the following lower bound for the ASN of test (30) on the subset D_{rk} :

$$\begin{aligned} \overline{\tau}^*_l &\geq E^l_{k+1}(N^* - k | D_{rk}) \\ &\geq \frac{\mathbf{P}^l_{k+1}(N \leq T_r \cap \nu = l | D_{rk}) | \ln \mathbf{P}_0(N \leq T_r \cap \nu = l | D_{rk})| - \ln 2}{\rho_{l0}} \end{aligned}$$

It is obvious that $\{T_r = \infty | D_{rk}\} \subset \{N \leq T_r \cap \nu = l | D_{rk}\}$. For this reason we get

$$\mathbf{P}^l_{k+1}(N \leq T_r \cap \nu = l | D_{rk}) \geq \mathbf{P}^l_{k+1}(T_r = \infty | D_{rk})$$

and

$$\mathbf{P}^l_{k+1}(N \leq T_r \cap \nu = l | D_{rk}) \geq 1 - \epsilon_l.$$

Hence, we have

$$\overline{\tau}^*_l \geq \frac{(1-\epsilon_l)|\ln \mathbf{P}_0(N \leq T_r \cap \nu = l|D_{rk})| - \ln 2}{\rho_{l0}}. \tag{31}$$

Let $R = \inf\{r \geq 1 : N \leq T_r\}$ and $R^l = \inf\{r \geq 1 : N \leq T_r \cap \nu = l\}$. If $\mathbf{P}_0(R^l \geq r) > 0$ then $\mathbf{P}_0(R^l < r+1|R^l \geq r)$ is well defined and we get the following formula

$$\mathbf{P}_0(R^l \geq r) = \frac{\mathbf{P}_0(R^l = r)}{\mathbf{P}_0(R^l < r+1|R^l \geq r)}.$$

Therefore

$$E_0(R^l) = \sum_{r=1}^{\infty} \mathbf{P}_0(R^l \geq r) = \sum_{r=1}^{\infty} \left[\frac{1}{\mathbf{P}_0(R^l < r+1|R^l \geq r)}\right] \mathbf{P}_0(R^l = r).$$

Let us consider the following events $\{R^l \geq r\} \subset \{R \geq r\}$. It follows from this assumption that

$$\mathbf{P}_0(R^l < r+1|R^l \geq r) = \frac{\mathbf{P}_0(R^l = r)}{\mathbf{P}_0(R^l \geq r)} \geq \frac{\mathbf{P}_0(R = r \cap \nu = l)}{\mathbf{P}_0(R \geq r)} = \mathbf{P}_0(R < r+1 \cap = l|R \geq r)$$

for $r = 1, 2, \ldots$. Denote by Q the following minimum

$$Q = \inf_{r\geq 1}\{\mathbf{P}_0(R < r+1 \cap \nu = l|R \geq r)\}.$$

Therefore

$$E_0(R^l) \leq Q^{-1}. \tag{32}$$

Under assumption that $\mathbf{P}_0(R \geq r) > 0$ the probability $\mathbf{P}_0(R < r+1|R \geq r)$ is well defined and

$$\mathbf{P}_0(R^l < r+1 \cap \nu = l|R^l \geq r) = \mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} < N).$$

It can be shown [10] that $\mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} < N)$ is an average over k of the probabilities $\mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} = k < N)$ satisfying (31) :

$$\begin{aligned}\mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} < N) &= E_0\left[\mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} = k < N)|T_{r-1} < N\right] \\ &= \sum_{k=0}^{\infty} \mathbf{P}_0(T_{r-1} = k)\mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} = k < N)\end{aligned}$$

Moreover, from Jensen inequality for convex functions it follows that for all values $r : \mathbf{P}_0(R \geq r) > 0$ we get

$$\overline{\tau}^*_l \geq \frac{(1-\epsilon_l)|\ln \mathbf{P}_0(N \leq T_r \cap \nu = l|T_{r-1} < N)| - \ln 2}{\rho_{l0}}.$$

The substitution of inequality (32) in the previous formula results in

$$\overline{\tau}_l^* \geq \frac{(1-\epsilon_l)\ln E_0(R^l) - \ln 2}{\rho_{l0}}. \tag{33}$$

Since $T_1, T_2, \ldots$ is a sequence of i.i.d. random variables and $E_0(R^l) < \infty$, Wald's identity holds and we have [22, pp.52-54; App. A.3]

$$E_0(T_{R^l}) = E_0(T_1)E_0(R^l).$$

Moreover, it is obvious that $E_0(T_{R^l}) \geq E_0(N^{\nu=l})$ and finally, we get the following lower bound :

$$\overline{\tau}_l^* \geq \frac{(1-\epsilon_l)\ln E_0(N^{\nu=l}) - (1-\epsilon_l)\ln E_0(T_1) - \ln 2}{\rho_{l0}}, \quad l = 1, \ldots, K-1. \tag{34}$$

The proof of the first part of the theorem is complete.

The second part:
It will suffice to show that for all $(N, \nu) \in \mathcal{K}_\gamma$ the following inequality is true

$$\overline{\tau}^*(1 + o(1)) \geq \frac{\ln\gamma - C_2(K)}{\rho} \quad \text{as} \quad \gamma \to \infty, \tag{35}$$

where $C_2(K) = (K-2)e^{-1} + \ln 2$ and

$$\rho = \min_{1 \leq l \leq K-1} \min_{1 \leq j \neq l \leq K-1} \rho_{lj}. \tag{36}$$

Let us assume that $X_1, X_2, \ldots \sim P_l$, $l = 1, \ldots, K-1$. Now we do not need to introduce the "artificial" additional stopping variable T_r and we consider the following sequence of stopping variables :

$$N_0 = 0 < N_1 < N_2 < \cdots < N_r < \cdots,$$

where N_r is the stopping variable which is applied to $X_{N_{r-1}+1}, X_{N_{r-1}+2}, \ldots$. Let us define the sequential test (N_r, ν_r) which chooses one of the $K-1$ hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_{K-1}$.

Let us consider the statistical properties of this sequential test. From the definition of the worst mean detection/isolation delay it results immediately that

$$\begin{aligned} \overline{\tau}_l^* &= \sup_{k \geq 0} \operatorname{esssup} E_{k+1}^l(N - k | N > k, X_1, X_2, \ldots) \\ &\geq E_1^l(N - 0 | D_{r0}) = E_l(N) \end{aligned}$$

In order to apply lower bound (23) for the ASN in this case, we have to assume that in Lemma 3 $\gamma_l = 0,\ l = 1,\ldots,K-1$. The convention which interprets $0\ln\frac{0}{0}$ as zero [20] leads to the following lower bound for the ASN :

$$\overline{\tau}_l^* \geq E_l(N) \geq \max_{1\leq j\neq l\leq K-1}\left\{\frac{(1-\tilde{\gamma}_l)\ln\beta_{jl}^{-1} - \ln 2}{\rho_{lj}}\right\} \tag{37}$$

for $l = 1,\ldots,K-1$, where $\tilde{\gamma}_l = \sum_{j=1,j\neq l}^{K-1}\beta_{lj}$ and

$$\beta_{jl} = \mathbf{P}_j(\text{accepting } \mathcal{H}_l, l\neq j) = \mathbf{P}_j(\nu_r = l),\ \ j,l = 1,\ldots,K-1.$$

Since $\nu_1,\nu_2,\ldots$ are i.i.d. random variables we have immediately that :

$$E_l(R^j) = E_l(\inf\{r\geq 1 : \nu_r = j\}) = \frac{1}{\beta_{lj}}.$$

Moreover, $N_1,N_2,\ldots$ are i.i.d. too and $E_l(R^j) < \infty$. Hence, Wald's identity [22, pp.52-54; App. A.3]

$$E_l(N^{\nu=j}) = E_l\left(\inf_{r\geq 1}\{N_r|\nu_r = j\}\right) = E_l(N)E_l(R^j)$$

holds and we have

$$\beta_{lj} = \frac{E_l(N)}{E_l(N^{\nu=j})} \leq \frac{\overline{\tau}_l^*}{\gamma} \leq \frac{\overline{\tau}^*}{\gamma}. \tag{38}$$

Inserting (38) in (37), we get the following inequality :

$$\overline{\tau}_l^* \geq \max_{1\leq j\neq l\leq K-1}\left\{\frac{(1-(K-2)\frac{\overline{\tau}^*}{\gamma})\ln\frac{\gamma}{\overline{\tau}^*} - \ln 2}{\rho_{lj}}\right\}. \tag{39}$$

Note here that $\min_{x>0}(x\ln x) = e^{-1}$. Since inequality (39) holds for all values $l = 1,\ldots,K-1$, we get finally

$$\overline{\tau}^* \geq \frac{\ln\gamma - \ln\overline{\tau}^* - (K-2)e^{-1} - \ln 2}{\rho}$$

or

$$\overline{\tau}^*(1+o(1)) \geq \frac{\ln\gamma - C_2(K)}{\rho}\ \text{ as }\ \gamma\to\infty \tag{40}$$

and the proof of second part of the theorem is complete.

The "first order" lower bound (28) follows from (34), (40) and the fact that

$$\rho^* = \min\left\{\min_{1\leq l\leq K-1}(\rho_{l0}), \rho\right\}.$$

Corollary 2 *Detection/isolation algorithm (10) - (12) is asymptotically optimal in the class $\mathcal{K}_\gamma$.*

It is of interest to compare $n(\gamma)$ (28) with the infimum $n^c(\gamma)$ of the worst mean detection delay for a change detection algorithm. Let N^c be the stopping variable of a change detection algorithm. We suppose that the worst mean detection delay is

$$\overline{\tau}_c^* = \sup_{k\geq 0, 1\leq l\leq K-1} \text{esssup } E_{k+1}^l(N^c - k|N^c > k, X_1, \ldots, X_k).$$

Denote $\mathcal{K}_\gamma^c$ the class of all stopping variables satisfying $E_0(N^c) \geq \gamma^c$.

Corollary 3 *Let the following equality be true*

$$\min_{1\leq l\leq K-1} \rho_{l0} = \rho^*$$

and $(K-1)\gamma = \gamma^c$. *Then*

$$\overline{\tau}_c^*(\gamma) \sim \overline{\tau}^*(\gamma) \;\; as \;\; \gamma \to \infty. \tag{41}$$

Proof of Corollary 3: It follows immediately from Lorden's Theorem 3 [10] that

$$n_c(\gamma) = \inf_{N^c \in \mathcal{K}_\gamma^c} (\overline{\tau}_c^*) \sim \frac{\ln \gamma^c}{\min_{1\leq l\leq K-1} \rho_{l0}} \;\; \text{as} \;\; \gamma^c \to \infty.$$

Hence, it is easy to see that formula (41) is true when $\min_{1\leq l\leq K-1} \rho_{l0} = \rho^*$.

Discussion. Let us discuss the following practical interpretations of the above results:

- From Theorem 2 it follows that the Kullback-Leibler numbers ρ_{ij} play a key role in the statistical properties of the detection/isolation algorithms. The minimum ρ^* of the Kullback-Leibler "distance"[2] between the two closest hypotheses $\mathcal{H}_i$ and $\mathcal{H}_j$, $0 \leq i \neq j \leq K-1$ define the worst mean detection/isolation delay.
- Let us consider the two problems which have been mentioned in section I.
    1. The first problem is a change detection task without any isolation of the source of the change (alarm task by A.Willsky).
    2. The second problem is the joint change detection/isolation task.

[2] Strictly speaking, ρ_{ij} is not a distance in precise sense. But in some cases, for instance, in the case of a change in the mean of a Gaussian vector sequence this interpretation is precise and useful.

If we suppose that the delay for detection is the price to be paid, then the following basic question arises: Is it necessary to pay more in the case of the more complicated second problem ? If ρ (36), which is the Kullback-Leibler "distance" between the closest alternatives $\mathcal{H}_l$ and $\mathcal{H}_j$, is greater then or equal to the minimum "distance" $\min_{1\leq l\leq K-1}\rho_{l0}$ between the alternatives $\mathcal{H}_l$ and the null-hypothesis $\mathcal{H}_0$ or, equivalently, $\min_{1\leq l\leq K-1}\rho_{l0}=\rho^*$ then the answer will be "No".

V. Additive changes in linear stochastic models

In this section we introduce some linear (regression and dynamical) stochastic models with *additive changes.* We also introduce, in brief, the key concepts that are to be used for the corresponding detection/isolation problem : namely *redundancy* and *innovation* (see [3, pp.249-252] for details). After this we show how the new detection/isolation problem can be reduced to the basic problem of section II. and we discuss some new features which play a key role in linear stochastic models with additive changes.

A. Models

Basic model. We consider the following Gaussian family of distributions $\mathcal{P}=\{P_{\theta_i},\theta_i\in\Omega\subset\mathbf{R}^r\}$, where $P_{\theta_i}=N(\theta_i,\Sigma)$, and $\Sigma>0$ is a known covariance matrix. Let $(Y_n)_{n\geq 1}$ be an independent Gaussian random sequence observed sequentially :

$$\mathcal{L}(Y_n)=\left\{\begin{array}{lll} N(\theta_0,\Sigma) & \text{if} & n\leq k \\ N(\theta_l,\Sigma) & \text{if} & n\geq k+1 \end{array}\right. ,\quad k=0,1,2,\ldots, \tag{42}$$

where $\theta_0=0$ and θ_l are known constants. The change time $k+1$ and number l are unknown. We assume that the following inequality is true :

$$0<\rho_{ij}=(\theta_i-\theta_j)^T\Sigma^{-1}(\theta_i-\theta_j)<\infty,\quad 0\leq i\neq j\leq K-1. \tag{43}$$

Regression models. We consider the following regression model with additive changes :

$$Y_n=HX_n+V_n+\Upsilon_l(n,k+1), \tag{44}$$

where X_n is the unknown state, V_n is a Gaussian white noise with covariance matrix $R=\sigma^2I,\sigma^2>0$, H is a full rank matrix of size $r\times s$ with $r>s$ and $\Upsilon_l(n,k+1)$ is the l-type change occurring at time $k+1$, namely :

$$\Upsilon_l(n,k+1)=\left\{\begin{array}{lll} 0 & \text{if} & n\leq k \\ \Upsilon_l & \text{if} & n>k \end{array}\right. ,\quad l=1,\ldots,K-1.$$

The characteristic feature of model (44) is the existence of *redundancy* ($r - s > 0$) in the information contained in the observations.

Stochastic dynamical models. We consider the following linear stochastic dynamical model with additive changes :

$$A(z^{-1})Y_n = B(z^{-1})U_n + C(z^{-1})[V_n + \Upsilon_l(n, k+1)], \tag{45}$$

where U_n is the known input vector, z^{-1} is the backward shift operator, $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ are polynomial matrices in the operator z^{-1}, V_n is a Gaussian white noise with covariance matrix $\Sigma > 0$. Assume as usual that the characteristic equations

$$\det\left(\lambda^p I_r - \sum_{i=1}^{p} \lambda^{p-i} A_i\right) = 0, \ \ \det\left(\sum_{i=0}^{q} \lambda^{q-i} C_i\right) = 0$$

have zeroes outside the unit circle.

B. Algorithms

In this subsection we design the change detection/isolation algorithms for the above linear stochastic models. First, we design the algorithm for basic model (42), which is a Gaussian case of model (2). Next, we show that the algorithm for regression model (44) is based on the *residuals* of the least squares algorithm. The statistical background of this problem is multihypothesis testing with *nuisance parameters* and the minimax solution of this problem is the GLR algorithm. On the other hand this algorithm can be reduced to the *basic algorithm*. Finally, we show that the detection/isolation algorithm for dynamical model (45) is based on the *innovations* of the whitening filter. The statistical background of this algorithm is the transformation theorem [23, pp.53-59]. And again this algorithm is a particular case of the *basic algorithm*.

Basic model. Model (42) is a particular case of basic model (2), which is defined in section II. For this reason detection/isolation algorithm (10) - (12) is valid for model (42) :

$$\begin{array}{rcl}
\tilde{N} & = & \min\{\tilde{N}^1, \ldots, \tilde{N}^{K-1}\} \\
\tilde{\nu} & = & \operatorname{argmin}\{\tilde{N}^1, \ldots, \tilde{N}^{K-1}\} \\
\tilde{N}^l & = & \inf_{k \geq 1} \tilde{N}^l(k) \\
\tilde{N}^l(k) & = & \inf\left\{n \geq k : \min_{0 \leq j \neq l \leq K-1} S_k^n(l,j) \geq h\right\} \\
S_k^n(l,j) & = & \sum_{i=k}^{n} (\theta_l - \theta_j)^T \Sigma^{-1}(Y_i - \theta_j) - \dfrac{1}{2}(\theta_l - \theta_j)^T \Sigma^{-1}(\theta_l - \theta_j)
\end{array} \tag{46}$$

In the following subsections, we show how the other linear stochastic models with additive changes can be reduced to model (42).

Regression models In this case we consider regression model (44). The characteristic feature of this detection/isolation problem with respect to the above *basic problem* is the fact that the vector X is unknown. This type of statistical problems is usually called hypotheses testing problems with *nuisance parameters.* Some tutorial introduction to these problems can be found in [3, pp.141-145; 270-273]. Because $r > s$ and the matrix H has rank s, we can use the *redundancy* to solve the detection/isolation problem.

Minimax algorithm. Let us define the following hypotheses testing problem :

$$\mathcal{H}_l = \{P_\theta(Y); \theta = HX^l + \Upsilon_l, X^l\} \text{ and } \mathcal{H}_j = \{P_\theta(Y); \theta = HX^j + \Upsilon_j, X^j\}, \qquad (47)$$

where Υ_l, Υ_j are the *informative* parameters, and X^l, X^j are the *nuisance* parameters. We are interested in detecting a change from Υ_j to Υ_l, while considering X as an *unknown* parameter of model (44), but since the expectation of the distribution P_θ is a function of this unknown parameter, the design of the test is a nontrivial problem.

From Theorem 2 it results that lower bound (28) in the class $\mathcal{K}_\gamma$ is a monotone decreasing function of the minimum value of the Kullback-Leibler information ρ^*. Therefore, the design of the *minimax algorithm* consists of finding a pair of the *least favorable values* X^l and X^j for which the Kullback-Leibler information ρ_{lj} is minimum, and in computing the LR of the optimal algorithm for these values.

The expectation θ of the output Y of model (44) is

$$\theta_j = E(Y) = HX^j + \Upsilon_j \quad j = 0, \ldots, K-1,$$

where $\Upsilon_0 = 0$. Then the Kullback-Leibler information ρ_{lj} is

$$\rho_{lj} = \frac{1}{2\sigma^2}(\theta_l - \theta_j)^T(\theta_l - \theta_j).$$

Note here that ρ_{lj} is a function of the difference $x = X^l - X^j$. Therefore, we minimize $\rho_{lj}(x)$ with respect to x. The minimum is obtained for

$$x^* = (H^TH)^{-1}H^T(\Upsilon_j - \Upsilon_l)$$

and is given by

$$\rho_{lj}(x^*) = \frac{1}{2\sigma^2}(\Upsilon_j - \Upsilon_l)^T P(\Upsilon_j - \Upsilon_l), \qquad (48)$$

where $P = I - H(H^TH)^{-1}H^T$ is the projection matrix, rank $P = r - s$. Finally, we have the following formula of LR (13) for hypotheses (47) under the least favorable value x^* of the nuisance parameter

$$S_k^n(l,j) = \sum_{i=k}^{n} \frac{1}{\sigma^2}(\Upsilon_l - \Upsilon_j)^T P(Y_i - \Upsilon_j) - \frac{1}{2\sigma^2}(\Upsilon_l - \Upsilon_j)^T P(\Upsilon_l - \Upsilon_j). \quad (49)$$

Note that this LR is independent of the unknown values X^l and X^j. Therefore, let us define the following minimax algorithm :

$$\begin{array}{rcl} \overline{N} & = & \min\{\overline{N}^1, \ldots, \overline{N}^{K-1}\} \\ \overline{\nu} & = & \operatorname{argmin}\{\overline{N}^1, \ldots, \overline{N}^{K-1}\} \\ \overline{N}^l & = & \inf_{k\geq 1} \overline{N}^l(k) \\ \overline{N}^l(k) & = & \inf\left\{n \geq k : \min_{0\leq j\neq l\leq K-1} S_k^n(l,j) \geq h\right\} \\ S_k^n(l,j) & = & \sum_{i=k}^{n} \frac{1}{\sigma^2}(\Upsilon_l - \Upsilon_j)^T P(Y_i - \Upsilon_j) - \frac{1}{2\sigma^2}(\Upsilon_l - \Upsilon_j)^T P(\Upsilon_l - \Upsilon_j) \end{array} \quad (50)$$

Discussion. Let us add two comments about (49). First, it is easy to show that the minimax approach is equivalent to the GLR, which is based on the maximization of the likelihood function with respect to the unknown nuisance parameters (see also [3, p.144]). In other words

$$\begin{array}{rcl} \sum_{i=k}^{n} \ln \frac{\sup_{X_i^l} p_{\Upsilon_l}(Y_i|X_i^l)}{\sup_{X_i^j} p_{\Upsilon_j}(Y_i|X_i^j)} & = & \sum_{i=k}^{n} \frac{1}{\sigma^2}(\Upsilon_l - \Upsilon_j)^T P(Y_i - \Upsilon_j) \\ & & -\frac{1}{2\sigma^2}(\Upsilon_l - \Upsilon_j)^T P(\Upsilon_l - \Upsilon_j). \end{array}$$

Second, it is worth noting that (49) can be rewritten in the following manner

$$S_k^n(l,j) = \sum_{i=k}^{n} \frac{1}{\sigma^2}(\tilde{\Upsilon}_l - \tilde{\Upsilon}_j)^T(e_i - \tilde{\Upsilon}_j) - \frac{1}{2\sigma^2}(\tilde{\Upsilon}_l - \tilde{\Upsilon}_j)^T(\tilde{\Upsilon}_l - \tilde{\Upsilon}_j),$$

where $\tilde{\Upsilon}_l = T^T\Upsilon_l$, $\tilde{\Upsilon}_j = T^T\Upsilon_j$, $e_i = T^TY_i$, $T = (t_1, \ldots, t_{r-s})$ is a matrix of size $r \times (r-s)$, and $t_1, \ldots, t_{r-s}$ are the eigenvectors of the projection matrix P. Therefore, LR (49) is a function of the *parity vector* e_i of the analytical redundancy approach [8]. This parity vector e_i is the transformation of the measurements Y_i into a set of $r - s$ linearly independent variables by projection onto the left null space of the matrix H.

The parity vector sequence $(e_n)_{n\geq 1}$ can be modeled as

$$\mathcal{L}(e_n) = \left\{ \begin{array}{lll} N(0, \sigma^2 I_{r-s}) & \text{if} & n \leq k \\ N(T^T\Upsilon_l(n-k), \sigma^2 I_{r-s}) & \text{if} & n \geq k+1 \end{array} \right. , \quad k = 0, 1, 2, \ldots. \quad (51)$$

Consequently, to solve the detection/isolation problem in the case of regression model (44), we have to *transform* the observations Y_n into the parity vector e_n and then solve corresponding *basic* problem (42).

Example 1 (Radionavigation system integrity monitoring.)
Navigation systems are standard equipment for planes, boats, rockets and other mobile. On-line integrity monitoring (fault detection/isolation) is one of the main problems in the design of modern navigation systems (see examples and references in [6], [21], [13], [14], [3, pp.454-463]).

For instance, let us consider integrity monitoring of the global positioning satellite set. Simplified measurement models of this type of radionavigation systems can be described by (44). The problem is to detect and isolate a satellite clock fault which can be represented as the additional bias Υ_l in model (44). Conventional global navigation sets require measurements from 4 satellites to estimate three spatial orthogonal coordinates and a clock bias, or three orthogonal velocities and a clock bias rate ($X \in \mathbf{R}^4$). Because for 18-satellites global navigation sets, five or more satellites ($r \geq 5$) are visible 99.3% of the time, it is possible to provide integrity monitoring by using these redundant measurements [21].

Let $(Y_n)_{n\geq 1}$ be the output of model (44). Let us assume that $\Upsilon_l = (0,\ldots,0,\delta_l,0,\ldots,0)^T$. Satellite number l clock fault is represented by bias δ_l.

In this case the statistics $S_k^N(l,j)$ are defined by the following formulas :

$$
\begin{aligned}
S_k^n(l,0) &= \sum_{i=k}^{n} \ln \frac{\sup_{X_i^l} p_{\delta_l}(Y_i|X_i^l)}{\sup_{X_i^0} p_0(Y_i|X_i^0)} &=& \sum_{i=k}^{n} \left(\frac{\delta_l \varepsilon_{i,l}}{\sigma^2} - \frac{\delta_l^2 p_{ll}}{2\sigma^2} \right) \\
S_k^n(l,j) &= \sum_{i=k}^{n} \ln \frac{\sup_{X_i^l} p_{\delta_l}(Y_i|X_i^l)}{\sup_{X_i^j} p_{\delta_j}(Y_i|X_i^j)} &=& \sum_{i=k}^{n} \left(\frac{\delta_l \varepsilon_{i,l} - \delta_j \varepsilon_{i,j}}{\sigma^2} \right. \\
& & & \left. - \frac{\delta_l^2 p_{ll} + 2\delta_l \delta_j p_{lj} + \delta_j^2 p_{jj}}{2\sigma^2} \right),
\end{aligned}
$$

where $\varepsilon_{i,l}$ is the component of the LS residuals

$$\varepsilon_i = PY_i$$

and p_{lj} is the element of the matrix P.

Stochastic dynamical models In this case we consider dynamical model (45). It is obvious that the output Y_n of this model is expressible as the sum of the output of the *deterministic part* $P(z^{-1})U_n$ of model (45) and the noise process with abrupt changes $C(z^{-1})[V_n + \Upsilon_l(n,k+1)]$. Hence,

$$\tilde{Y}_n = A(z^{-1})Y_n - B(z^{-1})U_n = C(z^{-1})[V_n + \Upsilon_l(n,k+1)]. \tag{52}$$

Let us consider two Gaussian vectors $\tilde{Y}_1^n$ and X_1^n :

$$\tilde{Y}_n = C(z^{-1})X_n, \quad X_n = V_n + \Upsilon_l(n, k+1), \quad n = 1, 2, \ldots, \quad X_{n<0} = 0.$$

It is easy to show that the transformation from the X-space to the $\tilde{Y}$-space is a *diffeomorphism* (one-to-one transformation). Note J its Jacobian matrix. From transformation theorem (see [23, pp.53-59]) it results that

$$p(\tilde{Y}_1, \ldots, \tilde{Y}_n) = |\det J| p(X_1, \ldots, X_n).$$

Therefore,

$$S_k^n(l, j) = \ln \frac{p_l(\tilde{Y}_k, \ldots, \tilde{Y}_n)}{p_j(\tilde{Y}_k, \ldots, \tilde{Y}_n)} = \sum_{i=k}^{n} \ln \frac{p_l(X_i)}{p_j(X_i)}. \tag{53}$$

From formula (53) it results that this detection/isolation problem is reduced to the above *basic* problem.

C. Statistical properties of the algorithms

In this subsection, we investigate the statistical properties of the change detection/isolation algorithms. The goal of this subsection is to give some interpretations of the general results of sections III. and IV.

Basic model. Since model (42) is a special case of basic model (2), Theorems 1 and 2 are valid in the case of algorithm (46). Note, that ρ^* is given by

$$\rho^* = \min_{1 \le l \le K-1} \min_{0 \le j \ne l \le K-1} \left\{ \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j) \right\}.$$

Linear stochastic models. From the above paragraphs it follows that the change detection/isolation problems in the case of linear stochastic models (44) - (45) are reduced to the basic detection/isolation problem. It is obvious that Theorem 1 is valid for these models. Moreover, algorithm (46) is *asymptotically optimal* in the sense of Theorem 2. In the case of stochastic dynamical model (45), the proof of this fact is trivial, it is sufficient to remember that the transformation from the X-space to the $\tilde{Y}$-space is a diffeomorphism. In the case of regression model (44), we have to define the meaning of optimality.

We consider the following family of distributions $\mathcal{P} = \{P_{\theta_i, X^i}, i = 0, \ldots, K-1\}$, where θ_i are the informative parameters and X^i are the nuisance parameters. Suppose there exists a class $\mathcal{K}_\gamma$ of all sequential detection/isolation algorithms (N, ν) over this family of distributions.

Definition 2 *Let us define the* minimax *lower bound* $\overline{n}(\gamma)$ *as follows*

$$\overline{n}(\gamma) \sim \frac{\ln \gamma}{\underline{\rho^*}} \quad as \ \gamma \to \infty,$$

where

$$0 < \underline{\rho}^* = \min_{1 \le l \le K-1} \min_{0 \le j \ne l \le K-1} \inf_{X^l, X^j} \rho_{lj} < \infty.$$

We say that the detection/isolation algorithm $(\overline{N}, \overline{\nu})$ *is* asymptotically minimax *if the following condition holds :*

$$\overline{\tau}^*_{(\overline{N},\overline{\nu})} \sim \frac{\ln \gamma}{\underline{\rho}^*} \quad as \quad \gamma \to \infty.$$

Theorem 3 *Let us consider regression model (44). We assume that the following inequality is true :*

$$0 < \underline{\rho_{ij}} = \frac{1}{2\sigma^2} (\Upsilon_i - \Upsilon_j)^T P (\Upsilon_i - \Upsilon_j) < \infty, \quad 0 \le i \ne j \le K-1. \tag{54}$$

Let $(\overline{N}, \overline{\nu})$ *be detection/isolation algorithm (50). Then :*

$$\overline{\tau}^*_{(\overline{N},\overline{\nu})} \sim \frac{\ln \overline{T}}{\underline{\rho}^*} \quad as \quad \overline{T} \to \infty, \tag{55}$$

where

$$\underline{\rho}^* = \min_{1 \le l \le K-1} \min_{0 \le j \ne l \le K-1} \underline{\rho_{ij}}.$$

Proof:

As we have mentioned above, LR (49) is a function of the parity vector e, which is a Gaussian $r-s$-dimensional random variable. From (48) it follows that

$$\inf_{X^l, X^j} \rho_{ij} = \underline{\rho_{ij}} = \frac{1}{2\sigma^2} (\Upsilon_i - \Upsilon_j)^T P (\Upsilon_i - \Upsilon_j).$$

Finally, formula (55) follows at once from Theorem 1.

Corollary 4 *Detection/isolation algorithm (50) is asymptotically minimax.*

Discussion. Let us add a remark about the problem of *detectability and isolability* of changes in regression model (44). We emphasize that this problem is nontrivial in the case of the regression model.

Suppose that the vectors $\Upsilon_j \in \mathbf{R}^r$, $j = 0, \ldots, K-1$ are arbitrary chosen such that $\|\Upsilon_i - \Upsilon_j\|_2 \ge \epsilon > 0$, $0 \le i \ne j \le K-1$, where $\|X\|_2 = \sqrt{\sum_{i=1}^r x_i^2}$. From matrix theory it results immediately that

$$\inf_{\Upsilon : \|\Upsilon\|_2 \ge \epsilon > 0} \Upsilon^T P \Upsilon = 0,$$

where $\Upsilon = \Upsilon_i - \Upsilon_j$. For this reason inequality (54) is not valid for *all* arbitrary vectors Υ_i and Υ_j. Roughly speaking, it is impossible to detect and isolate *all* arbitrary changes in model (44). The fact that the norm is strictly positive is not a sufficient condition in this case. Some of these changes will be *indistinguishable* from the statistical point of view. In order to simplify the problem, let us assume *a priori* the following: i) all the vectors $\Upsilon_i - \Upsilon_j$ have $\ell \leq \operatorname{rank} P = r - s$ nonzero components only; ii) all the principal minors with order from 1 to ℓ of the matrix P are strictly positive. Then the problem is much simpler. Namely, under these constraints inequality (54) holds true for arbitrary vectors Υ_i and Υ_j. If these constraints do not apply, then we should check *a priori* inequality (54).

Example 2 (Radionavigation system integrity monitoring - contd.) *Let us pursue our discussion of Example 1. Assume that only one satellite clock can fail simultaneously. Discuss the following problem: How many visible satellites are necessary to detect and isolate this fault? Because $\Upsilon_0 = 0$, it is easy to see that the minimal number r of visible satellites for detecting this fault is equal to five (redundancy $= r - 4 = 1$). But if we wish to isolate this fault, it is necessary to have six or more visible satellites (redundancy $= r - 4 \geq 2$)!*

D. Robustness of the algorithms

The goal of this paragraph is to investigate the robustness of the above change detection/isolation algorithms with respect to design parameters. From a practical point of view, it is important to have a detection algorithm which is robust with respect to the parameters of models (42) - (45). Let us consider the two following aspects of this problem : the *unknown* dynamic profile $\grave{\theta}_l(n-k)$ of the change magnitude, and the *unknown* covariance matrix $\grave{\Sigma}$ of the observations after the change time $k+1$.

Let us consider change detection/isolation algorithm (46). From the above paragraphs it follows that this algorithm is asymptotically optimal in the case of Gaussian basic model (42). Suppose now, that observations $Y_{k+1}, Y_{k+2}, \ldots$ are generated by *another* Gaussian distribution. In other words, the model is :

$$\mathcal{L}(Y_n) = \begin{cases} N(\theta_0, \Sigma) & \text{if} \quad n \leq k \\ N(\grave{\theta}_l(n-k), \grave{\Sigma}) & \text{if} \quad n \geq k+1 \end{cases}, \quad k = 0, 1, 2, \ldots, \tag{56}$$

where $\theta_0 = 0$. The profile $\grave{\theta}_l(n-k)$ and the covariance matrix $0 < \grave{\Sigma} < \infty$ are unknown *a priori*. We assume that the following condition holds :

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \grave{\theta}_l(n) = \grave{\theta}_l, \quad l = 1, \ldots, K-1. \tag{57}$$

Let $\grave{E}^l_{k+1}(\ )$ denote the expectation under $N(\grave{\theta}_l(n-k), \grave{\Sigma})$ and $\grave{E}_l(\ ) = \grave{E}^l_1(\ )$.

Let us define the following *robust* worst mean detection/isolation delay :

$$\overline{\tau}^*_r = \sup_{k\geq 0, 1\leq l \leq K-1} \operatorname{esssup} \grave{E}^l_{k+1}(\tilde{N} - k | \tilde{N} > k, Y_1, \ldots, Y_k)$$

The goal of this subsection is to show that $\overline{\tau}_r^*$ is expressible by *the same asymptotic formula* as in the case of true model (42) (see formula (19) in Corollary 1).

Theorem 4 *Let us consider model (56) - (57). We assume that the following condition holds :*

$$0 < \grave{\rho} = \min_{1 \leq l \leq K-1} \min_{0 \leq j \neq l \leq K-1} \left\{ (\theta_l - \theta_j)^T \Sigma^{-1} (\grave{\theta}_l - \theta_j) - \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j) \right\}.$$

Let $(\tilde{N}, \tilde{\nu})$ be detection/isolation algorithm (46). Then :

$$\overline{\tau}_r^* \leq \max_{1 \leq j \leq K-1} \grave{E}_j \left[\tilde{N}^j(1) \right] \sim \frac{h}{\grave{\rho}} \quad as \;\; h \to \infty. \tag{58}$$

Proof of Theorem 4:

The first part of formula (58) follows from Lemma 2 and Corollary 1. It is sufficient to show that

$$\grave{E}_l \left[\tilde{N}^l(1) \right] \sim \frac{h}{\min_{0 \leq j \neq l \leq K-1} \left\{ (\theta_l - \theta_j)^T \Sigma^{-1} (\grave{\theta}_l - \theta_j) - \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j) \right\}},$$

as $h \to \infty$, to prove the second part of formula (58) :

$$\max_{1 \leq j \leq K-1} \grave{E}_j \left[\tilde{N}^j(1) \right] \sim \frac{h}{\grave{\rho}} \quad \text{as} \;\; h \to \infty. \tag{59}$$

It results from Berk's Theorem 3.1 (see [5]) that the mean delay for detection satisfies

$$\lim_{n \to \infty} \frac{\grave{E}_l \left[\tilde{N}^l(1) \right]}{h} = \frac{1}{\varrho} \tag{60}$$

provided that

$$\frac{\min_{0 \leq j \neq l \leq K-1} S_1^n(l, j)}{n} \overset{w.p.1}{\to} \varrho \in (0, +\infty] \tag{61}$$

and also, for some $\tilde{\varrho} \in (0, \varrho)$, the "large deviation" probability

$$p_n = \grave{\mathbf{P}}_l \left[\frac{\min_{0 \leq j \neq l \leq K-1} S_1^n(l, j)}{n} < \tilde{\varrho} \right]$$

satisfies the two following conditions

$$\lim_{n \to \infty} n p_n = 0 \tag{62}$$

$$\sum_{n=1}^{\infty} p_n < \infty \tag{63}$$

Let us show that (61) is true and ϱ is defined by

$$\varrho = \min_{0 \leq j \neq l \leq K-1} \left\{ (\theta_l - \theta_j)^T \Sigma^{-1} (\grave{\theta}_l - \theta_j) - \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j) \right\}.$$

The left hand side of (61) can be rewritten as

$$f_n = \min_{0 \leq j \neq l \leq K-1} \left\{ \frac{1}{n} S_1^n(l,j) \right\} = \min_{0 \leq j \neq l \leq K-1} \left\{ \overline{z}_n(l,j) + \overline{d}_n(l,j) \right\},$$

where $\mathcal{L}[\overline{z}_n(l,j)] = N(0, \frac{\sigma_z^2}{n})$ and

$$\overline{d}_n(l,j) = (\theta_l - \theta_j)^T \Sigma^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \grave{\theta}_l(n) - \theta_j \right) - \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j).$$

It results from the strong law of large numbers that $\overline{z}_n(l,j) \overset{w.p.1}{\to} 0$. Hence, the continuity theorem [4, Ch.1, §5] and the fact that

$$\lim_{n \to \infty} \overline{d}_n(l,j) = (\theta_l - \theta_j)^T \Sigma^{-1} (\grave{\theta}_l - \theta_j) - \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j)$$

lead to the following formula

$$f_n \overset{w.p.1}{\to} \min_{0 \leq j \neq l \leq K-1} \left\{ (\theta_l - \theta_j)^T \Sigma^{-1} (\grave{\theta}_l - \theta_j) - \frac{1}{2} (\theta_l - \theta_j)^T \Sigma^{-1} (\theta_l - \theta_j) \right\}.$$

Let us estimate the "large deviation" probability p_n and prove (62) - (63). First, find the following upper bound for p_n :

$$\begin{aligned}
p_n &= \grave{\mathbf{P}}_l \left\{ \frac{1}{n} \min_{0 \leq j \neq l \leq K-1} S_1^n(l,j) < \tilde{\varrho} \right\} \\
&= \grave{\mathbf{P}}_l \left\{ \min_{0 \leq j \neq l \leq K-1} \left[\overline{z}_n(l,j) + \overline{d}_n(l,j) \right] < \tilde{\varrho} \right\} \\
&\leq \grave{\mathbf{P}}_l \left\{ \min_{0 \leq j \neq l \leq K-1} \left[\overline{z}_n(l,j) + \min_{0 \leq j \neq l \leq K-1} \overline{d}_n(l,j) \right] < \tilde{\varrho} \right\} \\
&= \grave{\mathbf{P}}_l \left\{ \min_{0 \leq j \neq l \leq K-1} \left[\overline{z}_n(l,j) \right] < c \min_{0 \leq j \neq l \leq K-1} \overline{d}(l,j) - \min_{0 \leq j \neq l \leq K-1} \overline{d}_n(l,j) \right\},
\end{aligned}$$

where $c \in (0,1)$.

It is obvious that for all $c \in (0,1)$ there exists $N(c)$ such that for all $n > N(c)$ the following inequality holds

$$C(N) = (c-1) \min_{0 \leq j \neq l \leq K-1} \overline{d}(l,j) + \left| \min_{0 \leq j \neq l \leq K-1} \overline{d}(l,j) - \min_{0 \leq j \neq l \leq K-1} \overline{d}_n(l,j) \right| < 0.$$

Hence, we have for all $n > N(c)$

$$p_n \leq \grave{\mathbf{P}}_l \left\{ \min_{0\leq j\neq l\leq K-1} [\overline{z}_n(l,j)] < C(N) \right\} = 1 - \grave{\mathbf{P}}_l \left\{ \min_{0\leq j\neq l\leq K-1} [\overline{z}_n(l,j)] \geq C(N) \right\}.$$

Let us consider the Gaussian random vector $Z_n \in \mathbf{R}^{K-1}$ such that

$$Z_n = (\sqrt{n}\overline{z}_n(l,0), \ldots, \sqrt{n}\overline{z}_n(l,K-1))^T, \;\; \mathcal{L}(Z_n) = N(0, \Sigma_Z).$$

It is known that the family $X \sim N(0, I)$ remains invariant[3] under the transformation $gX = RX$, where $\Sigma_Z = RR^T$. Therefore,

$$\Phi_{0,I}(A) = \Phi_{\overline{g}(0,I)}(gA),$$

where

$$\begin{array}{rcl} \overline{g}(\theta, I) & = & (R\theta, \Sigma_z) \\ A & = & \{X : X^T X \leq \lambda^2\} \\ gA & = & \{Z : Z^T \Sigma_Z^{-1} Z \leq \lambda^2\}, \;\; Z = RX \\ \Phi_{\theta,\Sigma}(A) & = & \displaystyle\int_A \varphi_{\theta,\Sigma}(X)dX \\ \varphi_{\theta,\Sigma}(X) & = & (2\pi)^{-\frac{K-1}{2}} (\det\Sigma)^{-\frac{1}{2}} \exp\left\{ -\dfrac{1}{2}(X-\theta)^T\Sigma^{-1}(X-\theta) \right\}. \end{array}$$

Define the following ellipsoid

$$Z^T \Sigma_Z^{-1} Z = \lambda^2,$$

where $\lambda^2 = C^2(N) \min_i \sigma_{ii}^{-1}$ and σ_{ii}^{-1} are the diagonal elements of the matrix Σ_Z^{-1}. It is easy to see that

$$p_n \leq 1 - \grave{\mathbf{P}}_l \left\{ \min_{0\leq j\neq l\leq K-1} [\overline{z}_n(l,j)] \geq C(N) \right\} < 1 - \Phi_{0,n^{-1}\Sigma_Z}(gA) = 1 - \Phi_{0,n^{-1}I}(A)$$

and, finally,

$$p_n \leq 1 - \Phi_{0,n^{-1}I}(A) < \overline{p}_n = 1 - \{1 - 2\phi(-\overline{\lambda}\sqrt{n})\}^{K-1},$$

where $\overline{\lambda} = \frac{\lambda}{\sqrt{K-1}} > 0$. From this and the asymptotic formula

$$\phi(-x) \sim \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(1 - \frac{1}{x^2} + \frac{3}{x^4} + \ldots\right), \;\; \phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

[3] A parametric family of distributions $\mathcal{P} = \{P_\theta\}$ remains invariant *under a group of transformation* $\mathcal{G}$ if $\forall g \in \mathcal{G}$ and $\forall \theta$, $\exists \theta_g$ such that: $\mathbf{P}_\theta(Y \in A) = \mathbf{P}_{\theta_g}(Y \in gA)$, where $\theta_g = \overline{g}\theta$.

we deduce that

$$\lim_{n\to\infty} n\overline{p}_n = 0.$$

Moreover, straightforward computations show that

$$\sum_{n=1}^{\infty} \overline{p}_n < \infty,$$

and the proof of (62) - (63) is complete.

Thus we have proved that (61) - (63) hold true. From this we then have (60) and, finally, we get (59). The proof of Theorem 4 is complete.

Corollary 5 *Let us assume that* $\grave{\theta}_l = \theta_l,\ l = 1, \ldots, K-1$. *Then*

$$\overline{\tau}^*(h) \sim \overline{\tau}_r^*(h) \ \ \textit{as} \ \ h \to \infty.$$

VI. Discussion

A new statistical approach to the change diagnosis problem is proposed. This approach consists of the joint detection and isolation of abrupt changes in a stochastic system.

Our main results are the following :

1. We introduced a minimax criterion of optimality (6) - (7) for this detection/isolation problem.
2. A new statistical change detection/isolation algorithm has been designed. This algorithm is expressible by equations (10) - (12).
3. We investigated the statistical properties of this algorithm. The result is stated in Theorem 1. We proved a lower bound for the worst mean detection/isolation delay in a certain class of sequential change detection/isolation algorithms. This result is given by Theorem 2. From Theorems 1 and 2 it follows that the proposed algorithm is asymptotically optimal in this class.
4. As we demonstrated in the previous sections the general results can be applied to some classical linear stochastic models with additive abrupt changes. The nontrivial problem of detectability and isolability, which arises in the case of regression model with redundancy (44), has been adressed.
5. It has been proved that the detection/isolation algorithm is *robust* with respect to the parameters of linear stochastic models with additive changes. Particularly, the algorithm is robust with respect to the *unknown* dynamic profile of the change magnitude if this profile converges to a known constant.

Let us add a concluding remark. It is obvious that the proposed scheme (10) - (12) is not a *resursive* algorithm. Hence, the problem of interest is to find *another* appropriate recursive computational scheme in order to reduce the amount of numerical operations, which should be performed for every new observation, without losing optimality.

Acknowledgments - The author is grateful to A.Benveniste and M.Basseville for their many constructive comments on the paper.

References

[1] P. Armitage, "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis," *J. Roy. Statist. Soc. B.*, vol.12, pp.137-144, 1950.

[2] R.K. Bansal and P. Papantoni-Kazakos, "An algorithm for detecting a change in a stochastic process," *IEEE Trans. Information Theory*, vol.IT-32, pp.227-235, March 1986.

[3] M. Basseville and I. Nikiforov, *Detection of abrupt changes. Theory and applications.* Prentice Hall, Information and System Sciences Series, 1993.

[4] A.A. Borovkov, *Theory of Mathematical Statistics - Estimation and Hypotheses Testing.* Moscow: Naouka, 1984, (in Russian).

[5] R.H. Berk, " Some asymptotic aspects of sequential analysis," *Annals Statistics*, vol.1, pp. 1126-1138, 1973.

[6] T.-T. Chien and M.B. Adams, "A sequential failure detection technique and its application," *IEEE Trans. Automatic Control,* vol.AC-21, pp.750-757, October 1976.

[7] R.H. Farrell, "Limit theorems for stopped random walks," *Annals Math. Statistics,* vol.35, 1332-1343, 1964.

[8] P.M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge based redundancy - A survey and new results," *Automatica,* vol.26, no.3, pp. 459-474, 1990.

[9] B.K.Ghosh and P.K.Sen, Ed. *Handbook of sequential analysis.* New York: Marcel Dekker, Inc.,1991.

[10] G. Lorden, "Procedures for reacting to a change in distribution," *Annals Math. Statistics,* vol.42, pp. 1897-1908, 1971.

[11] G. Lorden, "Open-ended tests for Koopman-Darmois families," *Annals Statistics,* vol.1, pp.633-643, 1973.

[12] G. Moustakides, "Optimal procedures for detecting changes in distributions," *Annals Statistics,* vol.14, pp.1379-1387, 1986.

[13] I. Nikiforov, V.Varavva, and V.Kireichikov, "GNSS integrity monitoring based on statistical change detection algorithms," in *Proc. TOOLDIAG'93,* Toulose, France, vol.2, pp.477-486, April 1993.

[14] I. Nikiforov, V.Varavva, and V.Kireichikov, "Application of statistical fault detection algorithms to navigation system monitoring," *Automatica,* vol.29, n.5, 1993.

[15] E.S.Page, "Continuous inspection schemes," *Biometrika,* vol.41, pp. 100-115, 1954.

[16] M. Pollak, "Optimal detection of a change in distribution," *Annals Statistics,* vol.13, pp.206-227, 1985.

[17] Y. Ritov, "Decision theoretic optimality of the CUSUM procedure," *Annals Statistics,* vol.18, pp.1464-1469, 1990.

[18] A.N. Shiryaev, "The problem of the most rapid detection of a disturbance in a stationary process,"*Soviet Math. Dokl.,* no.2, pp. 795-799, 1961.

[19] A.N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probability and Applications,* vol.8, pp.22-46, 1963.

[20] G. Simons, "Lower bounds for average sample number of sequential multihypothesis tests," *Annals Math. Statistics,* vol.38, pp.1343-1364, 1967.

[21] M.A. Sturza, "Navigation system integrity monitoring using redundant measurements," *Navigation*, vol.35, pp.483-501, Winter 1988-89.

[22] A. Wald, *Sequential analysis.* New York: John Wiley, 1947.

[23] S.S. Wilks, *Mathematical Statistics.* New York: John Wiley, 1963.

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

Unité de recherche INRIA Lorraine, Technôpole de Nancy-Brabois, Campus scientifique,
615 rue de Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

ISSN 0249-6399