2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Change detection for finite dimensional Gaussian linear systems - a bound for the almost sure false alarm rate

László Gerencsér[1], Cecilia Prosdocimi[2] and Zsuzsanna Vágó[3]

***Abstract*— We consider the problem of change detection in the context of finite dimensional Gaussian linear systems. In particular a known initial system will be tested for eventual changes against a known alternative using a simplified version of the Page-Hinkley or CUSUM detector. We show that the detector is L-mixing, implying the existence of an almost sure false alarm rate. The derivation of an explicit upper bound for the latter will be outlined.**

I. Introduction

Detection of changes of statistical patterns is a fundamental problem in many applications, for a survey see [4] and [6]. In this paper we consider the problem of change detection in the context of finite dimensional Gaussian linear systems. In particular a known initial system, written in the innovation form

$$y = H(\theta)\nu, \qquad \mathbb{E}\nu\nu^T = \Lambda, \tag{1}$$

will be tested for eventual changes against a known alternative.

A basic method for detecting temporal changes is the *Cumulative Sum* (CUSUM) test or *Page-Hinkley* detector, introduced by Page [18] and analyzed later, among others, by Hinkley [13] and Lorden [15]. The Page-Hinkley detector is defined via a sequence of random variables (r.v.-s) X^θ, often called residuals in the engineering literature, such as likelihood ratios. In the case of dependent data $Y_1, ..., Y_n$, with assumed probability density function $p(Y_n|Y_{n-1}, \dots, Y_0, \theta)$, with $\theta = \theta_0$ and $\theta = \theta_1$, before and after the change-point, which is a fixed number, rather than a random variable, the residuals would be defined as

$$X_n = -\log p(Y_n|Y_{n-1}, \dots, Y_0, \theta_0) + \\ + \log p(Y_n|Y_{n-1}, \dots, Y_0, \theta_1). \tag{2}$$

As pointed out in [2], following [22], this score can also be interpreted as a relative loss in encoding y_n. Letting $S_0 = 0$ and $S_n = \sum_{k=1}^n X_k$, the CUSUM statistics or Page-Hinkley detector is defined for $n \geq 0$ as

$$g_n = S_n - \min_{0 \leq k \leq n} S_k. \tag{3}$$

[1]MTA SZTAKI, Hungarian Academy of Sciences, Budapest, Hungary. This research was partially supported by the CNR/MTA joint research project "Mathematics for telecommunications", 2010-2012.

[2]LUISS University, Rome, Italy. This research was partially supported by the CNR/MTA joint research project "Mathematics for telecommunications", 2010-2012.

[3]Péter Pázmány Catholic University, Budapest, Hungary. This research was partially supported by the TÁMOP Project of the Hungarian Government, under Grant number TMOP-4.2.1./B-11/2/KMR-2011-0002

A key performance characteristics of the Page-Hinkley detector is the almost sure *false alarm rate* obtained when monitoring a process with no change in the dynamics at all, defined as

$$\limsup_{N \longrightarrow +\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{g_n > \delta\}}. \tag{4}$$

In some special cases this is approximately the reciprocal of the so-called *Average Run Length* (ARL), the expectation of the time till the first alarm, (see Chapter 6.2 in [21] or [20]).

The trade-off between rate of false alarm (or the mean time until false alarm) and the detection delay has been converted into an optimality criterion by Lorden, see [15]. The proposed objective is to choose a stopping time so as to minimize the expected detection delay given a lower-bound on the mean time between false alarms. The optimality of the CUSUM procedure for i.i.d. data before and after the change has been established in [17] and [23], using different methods.

The applicability of the Page-Hinkley detector for dependent sequences has been discussed in a number of papers. To our best knowledge, the Page-Hinkley detector has been first rigorously studied in [14] under the very weak condition that, denoting the change point by τ,

$$\lim_N \frac{1}{N} \sum_{n=\tau}^{\tau+N} X_n = I > 0,$$

where the convergence is meant in probability, and X_n is as above. A deep theoretical analysis of the expected delay with given Average Run Length for HMM-s is provided in [7] and [8]. A general asymptotic theory for change-point detection not limited to the i.i.d case has been developed by Tartakovsky and Veeravalli, see [28], and later extended in [27]. In these works the false alarm probability is assumed to be constrained to be small in an appropriate technical sense. A well-founded heuristics for the application of the Page-Hinkley detector to ARMA systems with *unknown* dynamics before and after the change has been presented as far back as 1992 in [2].

The Page-Hinkley detector $(g_n)_{n \geq 0}$ can be equivalently defined via a *non-linear dynamical system*, with $a_+ = \max\{0, a\}$, as follows:

$$g_n = (g_{n-1} + X_n)_+ \quad \text{with} \quad g_0 = 0. \tag{5}$$

From a system-theoretic point of view this system is not stable in any sense. On the other hand, for an i.i.d. input sequence (X_n), with $\mathrm{E}(X_n) < 0$, some stability of the output process (g_n) can be expected. The resulting non-linear stochastic system is a standard object in queuing theory (see [26] Chapter 1 and [1] Chapters 1.5 and 3.6), and in the theory of risk processes (see [19]). A number of stability properties of (g_n) have been established in [1], [16]. In [11], motivated by the problem of change detection for HMM-s, we studied the non-linear dynamical system (5) under the condition that the input X_n is L-mixing, with negative expectation. (For the definition of L-mixing see Appendix A and [9] for further details). The main technical result of the cited paper is that, under suitable technical conditions, such as boundedness, and finite mixing rate formulated in terms of an L_∞-norm the output process (g_n) of this system is L-mixing. The conditions of [11] are fairly restrictive and the results are not directly applicable to the problem of change detection of Gaussian linear systems, see (23) below.

The purpose of this paper is to adapt the arguments to the latter case, and ultimately to prove the existence of an almost sure false alarm rate and provide a reasonable upper bound for it. Such results are important as they allow one to set the detection threshold so as to make sure the false alarm rate is within a desired tolerable limit. The main technical challenge of such an extension is proving the validity of a suitable exponential inequality for partial sums of X_n, stated as Condition 1.

II. GAUSSIAN LINEAR STOCHASTIC SYSTEMS

In this section we describe the setup for the problem of change-detection for Gaussian linear stochastic systems. The system is assumed to have a known dynamics under normal operating conditions. The validity of this dynamics will be tested against a known alternative using a modified Page-Hinkley detector.

To be more specific, let us consider a family of finite-dimensional Gaussian systems in innovation form

$$y = H(\theta)\nu, \qquad \mathbb{E}\nu\nu^T = \Lambda. \tag{6}$$

Thus $y = (y_n)$ with $-\infty < n < +\infty$ is the observation process, and $\nu = (\nu_n)$ is the innovation process, assumed to be Gaussian, both with values in, say, $\mathbf{R}^m$. The square transfer function $H(\theta)$ is assumed to be rational, stable and inverse stable for a known set of θ-s, denoted by Θ. The current value of θ and Λ are assumed to be known, and will be denoted by θ_0 and Λ_0. The system parameter and the covariance matrix of the noise after an eventual change, against which our initial model will have to be tested, is assumed to be known equal to, say θ_1, and Λ_1. Both θ_0 and θ_1 belong to Θ.

To test if a change has occurred we compute a score, a kind of log-likelihood ratio, as defined in the Introduction. First we compute the likelihood of an observation sequence y_n given the infinite past of y, assuming that there is no change at all, and the system parameters and the noise covariance are θ and Λ, which may differ from θ_0 and Λ_0. Allowing infinite past rather than a finite past between 0 and $n-1$ significantly simplifies the presentation, but does not effect the validity of our arguments. The negative conditional log-likelihood of $y_n, n \geq 1$ given the infinite past of y between time $-\infty$ and $n-1$, i.e.

$$-\log p(y_n | y_{n-1}, ..., y_0, y_{-1}, ...; \theta, \Lambda)$$

can be written as

$$L_n = L_n(\theta, \Lambda) = \frac{1}{2}\bar{\nu}_n^T \Lambda^{-1/2} \bar{\nu}_n + \frac{1}{2}\log\det\Lambda, \tag{7}$$

modulo constants, which are independent of θ, Λ and n. Here $\bar{\nu}_n = \bar{\nu}_n(\theta)$ is the estimated innovation process, obtained by the inverse system

$$\bar{\nu}_n(\theta) = H(\theta)^{-1} y. \tag{8}$$

The score, or log-likelihood ratio, in the present context becomes

$$X_n = L_n(\theta_0, \Lambda_0) - L_n(\theta_1, \Lambda_1). \tag{9}$$

A fundamental property of the score, obtained by the Kullback-Leibler inequality, is that

$$\mathbb{E}_{\theta_0,\Lambda_0}\big(X_n\big) < 0, \tag{10}$$

with strict inequality, if $\theta_1 \neq \theta_0$ or $\Lambda_1 \neq \Lambda_0$, and the parametrization is such that θ is identifiable. Note that the partial sums of X_n, i.e. $S_n = \sum_0^N X_n$, to be used in the Page-Hinkley detector, can be written as

$$\begin{aligned} -\log p(Y_n, Y_{n-1}, \ldots, Y_0; \theta_0, \Lambda_0) + \\ + \log p(Y_n, Y_{n-1}, \ldots, Y_0; \theta_1, \Lambda_1). \end{aligned} \tag{11}$$

Thus we find that

$$\lim_N \frac{1}{N}\sum_{n=0}^{N} X_n = -D(P_{\theta_0,\Lambda_0} || P_{\theta_1,\Lambda_1}) \quad \text{a.s.} \tag{12}$$

where the r.h.s. denotes the divergence rate between two measures, defined over the direct product space $\times\Pi_{-\infty}^{+\infty}\,\mathbf{R}$. (Note, that $\bar{\nu}_n(\theta_0) = \nu$ is an i.i.d sequence, thus its dynamics is trivial.) This can be computed explicitly in terms of the state-space representation of $\bar{\nu}_n(\theta_1)$, see [25], and Appendix B.

An alternative score with similar property can be constructed by using the squared prediction error yielding

$$X_n = \frac{1}{2}\bar{\nu}_n^T(\theta_0)\bar{\nu}_n(\theta_0) - \frac{1}{2}\bar{\nu}_n^T(\theta_1)\bar{\nu}_n(\theta_1). \tag{13}$$

Assuming that $\nu_n = \bar{\nu}_n^T(\theta_0)$ is a second order stationary orthogonal process we have

$$\mathbb{E}_{\theta_0}\big(X_n\big) = -\varepsilon < 0, \tag{14}$$

with strict inequality, if $\theta_1 \neq \theta_0$, and the parametrization is such that θ is identifiable. For the sake of simplicity we shall restrict ourselves to the latter criterion even in the Gaussian case, although using the full likelihood is likely to be more efficient.

To compute the score X_n we have to apply the inverse filter equation (8) with $\theta = \theta_0$ (giving an identity map) and

$\theta = \theta_1$. Let

$$\eta_n = \eta_n(\theta_0, \theta_1) := (\bar{\nu}_n(\theta_0), \bar{\nu}_n(\theta_1)), \tag{15}$$

obtained concatenating the vectors $\bar{\nu}_n(\theta_0), \bar{\nu}_n(\theta_1)$. With this notation equation (13) reads

$$X_n = \frac{1}{2}\eta_n^T J \eta_n, \tag{16}$$

with $J = \mathrm{diag}(I, -I)$. The main result of this paper is the theorem below, showing that the Page-Hinkley detector in the no change case, defined in terms of X_n above is L-mixing, having finite exponential moment of some positive order. As a corollary we get a reasonable upper bound for the almost sure false alarm rate.

Let us recall the notation:

$$\mathbb{E}_{\theta_0}\big(\frac{1}{2}\bar{\nu}_n^T(\theta_0)\bar{\nu}_n(\theta_0) - \frac{1}{2}\bar{\nu}_n^T(\theta_1)\bar{\nu}_n(\theta_1)\big) = -\varepsilon. \tag{17}$$

Let $\bar{\gamma}$ be the smallest non-negative number such that

$$G(e^{i\omega})\Lambda^* G(e^{-i\omega})^T \leq \bar{\gamma} I \quad \text{for all} \quad \omega, \tag{18}$$

Finally, let x be the unique solution of the equation

$$\frac{\varepsilon}{4\bar{\gamma}} x = [-x - \ln(1-x)]. \tag{19}$$

Set $\beta^* = x/\bar{\gamma}$.

Theorem 1: Let (X_n) be defined as in equation (16). Let $(\mathcal{F}_n, \mathcal{F}_n^+)$ be the past and the future σ-field of the process (ν_n). Let the system in (1) be stable for $\theta = \theta_1$. Then the Page-Hinkley detector (g_n) defined as in (5) is L-mixing with respect to $(\mathcal{F}_n, \mathcal{F}_n^+)$. Moreover for $\beta" < \beta^*$ we have

$$\sup_n \mathbb{E}\big(\exp \beta" g_n\big) < \infty. \tag{20}$$

As a corollary to Theorem 1 we can get an upper bound for the a.s. false alarm rate. This is in fact the practically most important implication of Theorem 1.

Theorem 2: Under the conditions of Theorem 1 we have for any $\delta > 0$

$$\mathrm{FAR} = \limsup_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}_{\{g_n > \delta\}} \leq K e^{-\beta" \delta}. \tag{21}$$

The key component of the proof of Theorem 1 is the verification of Condition 1 of the next section, needed for the application of Theorem 3 of Section III. The validity of Condition 1 will be established in Theorem 7 of Section IV.

III. The Page-Hinkley detector with L-mixing input

In this section we restate Theorem 2 of [11] under conditions that are appropriate for the application of the present paper. Thus let X^θ be a sequence of scores for a Page-Hinkley detector under the tacit assumption of no change, i.e. assume that

$$\mathbb{E}(X_n) \leq -\varepsilon < 0 \qquad \text{for all} \quad n \geq 0. \tag{22}$$

Assume that X^θ is L-mixing w.r.t. $(\mathcal{F}_n, \mathcal{F}_n^+)$. For the analysis of the Page-Hinkley detector with L-mixing, as given in [11], we needed two technical assumptions. The second condition required that

$$M_\infty(X) < +\infty \qquad \text{and} \qquad \Gamma_\infty(X) < +\infty. \tag{23}$$

(For the notations above see the Appendix). This condition will obviously not be satisfied when X_n is defined in terms of squares Gaussian random variables, as in (16). However, the proof of Theorem 3.1 of [11] reveals that we can significantly relax this condition. A careful study of the proof of the crucial Lemma 2 of the cited paper shows that what we really need is the validity of the following condition:

Condition 1: There exist $\beta^{**} > 0$ and $\alpha > 0$ such that for $0 \leq \beta' \leq \beta^{**}$ we have for all $1 \leq i \leq n$

$$\mathbb{E}\Big(\exp\big\{\beta' \sum_{k=i}^{n}(X_k - \mathbb{E}(X_k))\big\}\Big) \leq \exp\big\{\alpha\, {\beta'}^2 (n-i+1)\big\}.$$

In [11] we have proved the validity of the above condition, using the exponential inequality given as Theorem 5.1 in [10], with $\beta^{**} = \infty$, and, using the notations of that paper, with $\alpha := 4M_\infty(X)\Gamma_\infty(X)$, see the proof of Lemma 2.

A *necessary* condition for the above condition to hold is that the risk-sensitive criterion

$$J(\beta') = \limsup_n \frac{1}{n}\frac{1}{\beta'} \log \mathbb{E}\Big(\exp\big(\beta' \sum_{k=1}^{n} X_k\big)\Big)$$

is finite and has a common bound for $\beta' \leq \beta^{**}$. $J(\beta')$ in equation (24) is a well known quantity in risk sensitive control. It is known that $J(\beta') < +\infty$ for some $\beta' > 0$ if $|X_k| \leq X_k^*$ where $X_k^* = Z_k^T Z_k$ where Z_k is the stationary output of a finite-dimensional, time-invariant, stable linear Gaussian system, see Appendix F in [25].

In the next section we show that in the case when the data Y_n are generated by a finite dimensional, time-invariant, stable, linear Gaussian system, and X_n is defined as in (9), then Condition 1 holds. Furthermore potential values of β^{**} and α, will be quantified.

Define a critical exponent in terms of α as follows:

$$\beta^* := \varepsilon/\alpha, \tag{24}$$

see (12) of [11]. Then, for any $\beta' \leq \beta^*$ set

$$\lambda = \lambda(\beta') := \exp\big(\alpha(\beta')^2 - \beta'\varepsilon\big). \tag{25}$$

Note that for the critical value β^* we have $\lambda(\beta^*) = 1$, and for $\beta' < \beta^*$ we have $\lambda(\beta') < 1$. With this notations a main result of [11], stated as Theorem 2, can be restated as follows:

Theorem 3: Let X^θ be an L-mixing process w.r.t. $(\mathcal{F}_n, \mathcal{F}_n^+)$ such that (22) and Condition 1 are satisfied. Let (g_n) be the associated Page-Hinkley detector as defined in (5). Then (g_n) is L-mixing w.r.t. $(\mathcal{F}_n, \mathcal{F}_n^+)$. In addition, for any β'', β' such that $0 < \beta'' < \beta' < \beta^* \wedge \beta^{**}$, we have with $\lambda = \lambda(\beta')$

$$\mathbb{E}\big(\exp \beta'' g_n\big) \leq 1 + \Big(\frac{\beta''}{\beta' - \beta''}\Big)\frac{\lambda}{1-\lambda} =: K_{\beta'', \beta'}. \tag{26}$$

As a corollary to Theorem 3 we can get an upper bound for the a.s. false alarm rate just like in Theorem 2.

Theorem 4: Let X^θ and β^*, β^{**} be as in Theorem 3, and let (g_n) be defined as in (3). Then for any $\delta > 0$, and any

$0 < \beta'' < \beta' < \beta^* \wedge \beta^{**}$ we have

$$\limsup_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}_{\{g_n > \delta\}} \leq K_{\beta'',\beta'} \exp(-\beta''\delta) \tag{27}$$

where $K_{\beta'',\beta'}$ is defined in Theorem 3.

IV. Gaussian exponentials

The purpose of this section is to verify Condition 1 for finite dimensional Gaussian linear systems. We start by considering an R^N-valued random vector with jointly Gaussian distribution $\mathcal{N}(0, \Gamma)$, where Γ is non-singular. Let Q be a $N \times N$ symmetric, possibly indefinite matrix. Then, by elementary calculations,

$$\mathbb{E} \exp\left(\beta Z^T Q Z\right) = |I - 2\beta Q \Gamma|^{-1/2},$$

whenever

$$2\beta Q < \Gamma^{-1} \quad \text{or} \quad 2\beta\ \Gamma^{1/2} Q \Gamma^{1/2} < I. \tag{28}$$

Next, write $\exp \mathbb{E}\big(\beta Z^T Q Z\big) = \exp\big(\beta \text{ tr } Q\Gamma\big)$. Then for

$$\bar{J}(\beta) = \mathbb{E} \exp\left(\beta(Z^T Q Z - \mathbb{E} Z^T Q Z)\right),$$

we can write

$$\bar{J}(\beta) = |I - 2\beta Q\Gamma|^{-1/2} \cdot \ \exp\left(-\beta \text{ tr } Q\Gamma\right). \tag{29}$$

Now, the value of the r.h.s. will not change, if we replace $Q\Gamma$ by

$$R = \Gamma^{1/2} Q \Gamma^{1/2}. \tag{30}$$

Let R be diagonalized by a similarity transformation to $\text{diag}(\rho_i)$, $i = 1, ..., N$, arranged in decreasing order. Then we can write

$$\bar{J}(\beta) = \prod_{i=1}^{N} (1 - 2\beta\rho_i)^{-1/2} \cdot \prod_{i=1}^{N} e^{-\beta\rho_i}. \tag{31}$$

Let ρ_* and ρ^* be the smallest and largest eigenvalue of R, respectively, and let

$$x_* = x_*(\beta) = 2\beta\rho_*, \qquad x^* = x^*(\beta) = 2\beta\rho^*. \tag{32}$$

Note that (28) ensures that $x^* < 1$.

Lemma 1: Let $x_* \leq x \leq x^* < 1$. Then we have, with some c depending only on x_* and x^*,

$$(1-x)^{-1/2} e^{-x/2} \leq e^{c\frac{x^2}{4}}. \tag{33}$$

More exactly we have $c = c(x_*) \vee c(x^*)$, where

$$c(x) = [-x - \ln(1-x)]/(\frac{x^2}{2}). \tag{34}$$

The proof of the lemma is based on the observation that for given x the function $c(x)$ renders the inequality (33) into an equality, and the function $c(x)$ is monotone increasing for $x \geq 0$ and monotone decreasing for $x \leq 0$. We also note the following property of $c(x)$: we have

$$c(x) \geq c(-x) \quad \text{for} \quad 0 \leq x < 1. \tag{35}$$

This follows from the convexity of $x^2 c(x)$, and the fact that its second order derivative is monotone increasing. Applying the lemma to (29) we get the following result:

Theorem 5: Assume that for $\beta^{**} = \beta$ we have $2\beta Q < \Gamma^{-1}$ and let R be as above, see (30), its eigenvalues being ρ_i. Then for $|\beta'| \leq \beta$ we have

$$\mathbb{E} \exp\left(\beta'(Z^T Q Z - \mathbb{E} Z^T Q Z)\right) \leq e^{\bar{c}(\beta)(\beta')^2 \sigma^2}, \tag{36}$$

where $\bar{c}(\beta) = c(x_*(\beta)) \vee c(x^*(\beta))$ and $\sigma^2 = \sum_{i=1}^{N} \rho_i^2$.

Apply now the result to the case when $Z = (\eta_1, ..., \eta_N)$, where (η_n) is the extended residual process defined by the linear system

$$\eta = (I, H(\theta)^{-1} H(\theta_0))\nu = (I, G)\nu. \tag{37}$$

With this notation the simplified score can be written as

$$X_n = \frac{1}{2} \eta_n^T J \eta_n,$$

with $J = \text{diag}(I, -I)$, see (16). Let $Z = (\eta_1, ..., \eta_N)$, have a covariance matrix Γ_N, and let $Q = Q_N = \frac{1}{2}\text{diag}(I, -I)$.

Let the covariance matrices of $(\nu_{0,n})$ and $(\nu_{1,n})$ with $n = 1, ..., N$, be $\Gamma_{0,N}$ and $\Gamma_{1,N}$. Then, by permuting the rows and columns we can achieve that two diagonal blocks become $\Gamma_{0,N}$ and $\Gamma_{1,N}$. Both are block Toeplitz matrices with $\Gamma_{0,N}$ being simply $\text{diag}\Lambda^*$. The eigenvalues of $\Gamma_{0,N}$ and $\Gamma_{1,N}$, in decreasing order, will be denoted by $\gamma_{0,i}$ and $\gamma_{1,i}$, respectively, with $i = 1, ..., mN$, where $m = \dim \nu$. Note that the set of $\gamma_{0,i}$-s is identical with the set of eigenvalues of Λ^* with multiplicity N.

Let γ_0^* and γ_1^* denote their maximum eigenvalues. Dependence on N will not be denoted explicitly. Obviously $\gamma_0^* = \lambda_1^*$, the maximal eigenvalue of Λ^*, independently of N. On the other hand, standard arguments show that $\gamma_1^* \leq \bar{\gamma}$, where $\bar{\gamma}$ is defined under (18).

Note also that $\mathbb{E}\nu(\theta_1)\nu(\theta_1)^T \geq \mathbb{E}\nu(\theta_0)\nu(\theta_0)^T$, written in the form

$$\frac{1}{2\pi} \int G(e^{i\omega}) \Lambda^* G(e^{-i\omega})^T d\omega \geq \Lambda^* \tag{38}$$

implies $\bar{\gamma} \geq \lambda_1^*$.

Permutating the rows and columns of Q together with those of Γ_N we get a matrix of the form $\frac{1}{2}\text{diag}(I, -I)$, where I is a unit matrix of dimension Nm. We write this matrix in the form $Q^+ + Q^-$, where $Q^+ = Q_N^+ = \frac{1}{2}\text{diag}(I, 0)$, and $Q^- = Q_N^- = \frac{1}{2}\text{diag}(0, I)$. Define R_N^+ and R_N^- accordingly, following (30): first carry out the above permutation of rows and columns of each matrix, then replace the middle term by Q^+, and Q^-, respectively. Then, it is easy to see that R_N^+ is positive semi-definite, with Nm zero eigenvalues. The remaining Nm eigenvalues will be denoted by $\rho_{0,i}$, in decreasing order. Similarly, R_N^- is negative semi-definite, with Nm zero eigenvalues. The remaining eigenvalues will be denoted by eigenvalues $\rho_{1,i}$, in decreasing order. It is easy to see that

$$\rho_{0,i} = \frac{1}{2}\gamma_{0,i}, \quad \text{and} \quad \rho_{1,i} = -\frac{1}{2}\gamma_{1,i},$$

for $i = 1, ..., 2Nm$, which implies

$$\rho_{0,i} \leq \frac{1}{2}\lambda_1^* \leq \frac{1}{2}\bar{\gamma} \quad \text{and} \quad \rho_{1,i} \geq -\frac{1}{2}\gamma_1^* \geq -\frac{1}{2}\bar{\gamma}.$$

Now, obviously $R_N^+ \leq R_N \leq R_N^-$, implying that a similar inequality holds for their i-th eigenvalue, see e.g. Theorem 3, Chapter 7. [5].

It follows that

$$\rho_i \leq \rho_{0,i} \quad \text{and} \quad \rho_{mN+i} \geq \rho_{1,i} \quad \text{for} \quad i = 1, ..., mN,$$

Thus we get for the largest and smallest eigenvalue of R

$$\rho^* \leq \frac{1}{2}\lambda_1^* \leq \frac{1}{2}\bar{\gamma} \quad \text{and} \quad \rho_* \geq \frac{1}{2}\gamma_1^* \geq -\frac{1}{2}\bar{\gamma},$$

both bounds being independent of N. This yields

$$x^*(\beta) \leq \beta\lambda_1^* \leq \beta\bar{\gamma} \quad \text{and} \quad x_*(\beta) \geq -\beta\gamma_1^* \geq -\beta\bar{\gamma}.$$

To apply Theorem 5 note that

$$\sum_{i=1}^{Nm} \rho_{0,i}^2 = \frac{1}{4}\sum_{i=1}^{Nm} \gamma_{0,i}^2 = \frac{1}{4}N\sum_{i=1}^{m} \lambda_i^2 = \frac{1}{4}N\mathrm{tr}(\Lambda^*\Lambda^*).$$

Invoking Theorem 5 we finally get the following result:

Theorem 6: Assume that $\beta\lambda_1^* < 1$ and redefine

$$x^*(\beta) = \beta\lambda_1^* \quad \text{and} \quad x_*(\beta) = -\beta\bar{\gamma},$$

both independent of N. Then for $|\beta'| \leq \beta$ we have

$$\mathbb{E}\exp\big(\frac{\beta'}{2}\sum_{i=1}^{N}(\eta_i^T J\eta_i - \mathbb{E}\eta_i^T J\eta_i)\big) \leq$$

$$\leq \exp\big\{\bar{c}(\beta)\frac{\beta^2}{4}\big(N\mathrm{tr}(\Lambda^*\Lambda^*) + \sum_{i=1}^{Nm}\gamma_{1,i}^2\big)\big\},$$

where $\bar{c}(\beta) = c(x_*(\beta)) \vee c(x^*(\beta))$.

In the case of scalar processes. i.e. when $m = 1$, we can use the following result:

$$\lim_N \frac{1}{Nm}\sum_{i=1}^{Nm}\rho_{1,i}^2 = \frac{1}{2\pi}\int_0^{2\pi}|G(e^{i\omega})|^2\lambda^{*2}d\omega =: D_2(G),$$

see e.g. [12] and also[24].

The theorem can be simplified by taking into account the inequalities given above as follows:

Theorem 7: Assume that $\beta\bar{\gamma} < 1$ and redefine $x^*(\beta) = \beta\bar{\gamma}$. Then for $|\beta'| \leq \beta$ we have, with $\bar{c}(\beta) = c(x^*(\beta)) = c(\beta\bar{\gamma})$,

$$\mathbb{E}\exp\big(\frac{\beta'}{2}\sum_{i=1}^{N}(\eta_i^T J\eta_i - \mathbb{E}\eta_i^T J\eta_i)\big) \leq \exp(\bar{c}(\beta)\frac{\beta^2}{2}Nm\bar{\gamma}^2).$$

Using this simplified statement, the second order adjustment constant α in Condition 1 can be chosen to be

$$\alpha = \alpha(\beta) = 2c(\beta\bar{\gamma})\bar{\gamma}^2$$

Thus $\alpha(\beta)$ is an increasing function of β for $\beta > 0$, with values ranging from 0 to ∞, for $0 \leq \beta < \bar{\gamma}^{-1}$. Thus to ensure that the rightmost term in $0 < \beta'' < \beta' < \beta^* \wedge \beta^{**}$ (see Theorem 3) is maximized we have to make sure that

$$\beta^* = \varepsilon/\alpha(\beta) = \beta^{**} = \beta.$$

Thus the least strict upper bound for the exponent in Theorem 3, using the Theorem 7, is obtained by

$$\varepsilon = \beta \cdot 2c(\beta\bar{\gamma})\bar{\gamma}^2 = \frac{2}{\beta}c(\beta\bar{\gamma})\beta^2\bar{\gamma}^2 = \frac{4}{\beta}[-\beta\bar{\gamma} - \ln(1-\beta\bar{\gamma})]. \tag{39}$$

Setting $x = \beta\bar{\gamma}$ we get after rearrangement

$$\frac{\varepsilon}{4\bar{\gamma}}x = [-x - \ln(1-x)] \tag{40}$$

It is easy to see that this equation has a unique solution in x.

V. Discussion

Change-point detection in continuous time has been investigated with considerable generality in [3] under the condition of small false alarm probability. It would be interesting to establish mixing properties of the CUSUM statistics for Gaussian processes in continuous-time. In particular it would be interesting to see if a reflected Brownian motion with negative drift is L-mixing.

Appendix

A. L-mixing processes

We summarize a few definitions given in [9]. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X^θ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1: We say that X^θ is *M-bounded* if for all $1 \leq q < +\infty$

$$M_q(X) := \sup_{n\geq 0} \| X_n \|_q < +\infty.$$

We can also define $M_q(X)$ for $q = +\infty$ as

$$M_\infty(X) := \sup_{n\geq 1} \operatorname{ess\,sup} |X_n|.$$

Let $(\mathcal{F}_n)_{n\geq 1}$ be an increasing family of σ-fields and let $(\mathcal{F}_n^+)_{n\geq 1}$ be a decreasing family of σ-fields, $\mathcal{F}_n \subseteq \mathcal{F}$ and $\mathcal{F}_n^+ \subseteq \mathcal{F}$ for any n. Assume that $\mathcal{F}_n$ and $\mathcal{F}_n^+$ are independent for all n. Let $\tau > 0$ be an integer, and let for $1 \leq q < +\infty$

$$\begin{aligned}\gamma_q(\tau, X) = \gamma_q(\tau) &:= \sup_{n\geq\tau} \| X_n - \mathbb{E}(X_n|\mathcal{F}_{n-\tau}^+) \|_q, \\ \Gamma_q(X) &:= \sum_{\tau=0}^{+\infty}\gamma_q(\tau).\end{aligned}$$

We can also define

$$\begin{aligned}\gamma_\infty(\tau, X) &:= \sup_{n\geq\tau} \operatorname{ess\,sup} |X_n - \mathbb{E}(X_n|\mathcal{F}_{n-\tau}^+)|, \\ \Gamma_\infty(X) &:= \sum_{\tau=0}^{+\infty}\gamma_\infty(\tau, X).\end{aligned}$$

Definition 2: A process X^θ is *L-mixing* w.r.t. $(\mathcal{F}_n, \mathcal{F}_n^+)$ if X_n is $\mathcal{F}_n$-measurable for all $n \geq 1$, X^θ is M-bounded, and $\Gamma_q(X) < +\infty$ for all $1 \leq q < +\infty$.

A prime example of L-mixing process is the output process of a stable linear stochastic system driven by a M-bounded i.i.d. sequence.

Centered L-mixing processes satisfy the strong law of large numbers, see Corollary 1.3 in [9].

B. Divergence rate

We report below a result on the divergence rate between two Gaussian white processes. The result is stated as Theorem E.5. in [25]

Theorem 8: Let $y_1, y_2 : \Omega \times \mathbb{Z} \to \mathbb{R}^p$ be two stationary Gaussian processes. Assume they have a mean function that is zero. Let $\hat{W}_1, \hat{W}_2$ be the spectral densities of y_1, y_2 respectively, and

$$\begin{aligned} \hat{W}_1 &= G_1^\top(z^{-1})G_1(z) \\ \hat{W}_2 &= G_2^\top(z^{-1})G_2(z) \\ S &= G_1(z)G_2(z)^{-1} \end{aligned}$$

Let y be the output of a system with transfer function S. Assume that the transfer function S admits a realization as a finite-dimension linear system with a minimal realization parameterized by

$$\begin{aligned} x(n+1) &= Ax(n) + Bv(n) \\ y(n) &= Cx(n) + Dv(n), \end{aligned}$$

where $sp(A) \subset \mathbb{C}^-$. Assume that there exists a matrix Q such that the following relations are satisfied

$$\begin{aligned} &Q = Q^\top \geq 0, \\ &N = D^\top D + B^\top QB > 0, \\ &M = B^\top QA + D^\top C, \\ &Q = A^\top QA + C^\top C - M^\top N^{-1} M, \\ &sp\big(A - BN^{-1}(B^\top QA + D^\top C)\big) \subset \mathbb{C}^-. \end{aligned} \tag{41}$$

and that this relation admits a unique solution. Let R be the unique solution of the Lyapunov equation

$$R = A^\top RA + C^\top C.$$

Then the divergence rate between the measures induced by the two processes is given by

$$\begin{aligned} &D(P_1||P_2) = \\ &\quad -\frac{1}{2}\ln\det N + \frac{1}{2}\mathrm{tr}(B^\top RB + D^\top D - \mathbb{I}). \end{aligned}$$

REFERENCES

[1] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.

[2] J. Baikovicius and L. Gerencsér. Change point detection in a stochastic complexity framework. In *Proc. 29th IEEE CDC*, volume 6, pages 3554–3555, 1990.

[3] M. Baron and A.G. Tartakovsky. Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Analysis*, 25(3):257296, 2006.

[4] M. Basseville and I.V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, 1993.

[5] R. Bellman. *Introduction to matrix analysis*. MCGraw-Hill, 1960.

[6] B.E. Brodsky and B.S. Darkhovsky. *Nonparametric Methods in Change-Point Problems*. Kluwer Academic Publishers, 1993.

[7] C.D. Fuh. SPRT and CUSUM in hidden Markov models. *Annals of Statistics*, 31(3):942–977, 2003.

[8] C.D. Fuh. Asymptotic operating characteristics of an optimal change point detection in hidden Markov models. *Annals of Statistics*, 32(5):2305–2339, 2004.

[9] L. Gerencsér. On a class of mixing processes. *Stochastics*, 26:165–191, 1989.

[10] L. Gerencsér. Almost sure exponential stability of random linear differential equations. *Stochastics and Stochastics Reports*, 36:91–107, 1991.

[11] L. Gerencsér and C. Prosdocimi. Input-output properties of the Page-Hinkley detector. *Systems and Control Letter*, 60(7):486–491, 2011.

[12] R. Gray. *Toeplitz and Circulant Matrices: A Review*. Published as a paperback book by Now Publishers, 2006. Revised version.

[13] D.V. Hinkley. Inference about the change-point from cumulative sum tests. *Biometrika*, 58(3):509–523, 1971.

[14] T.L. Lai. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Transactions on Information Theory*, 44:2917–2929, 1998.

[15] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.

[16] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

[17] G.V. Moustakides. Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14(4):1379–1387.

[18] E.S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[19] H.H. Panjer and G.E. Willmot. *Insurance Risk Models*. Society of Actuaries, 1992.

[20] M. Pollak and A.G. Tartakovsky. Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Theory of Probability and its Applications Probability and Applications*, 53(3):430–442, 2009.

[21] H.V. Poor and O. Hadjiliadis. *Quickest Detection*. Cambridge University Press, 2009.

[22] J. Rissanen. *Stochastic complexity in statistical inquiry*. Word Scientific, Singapore, 1989.

[23] Y. Ritov. Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics*, 18(3):1464–1469.

[24] T.O. Silva. On the asymptotic eigenvalue distribution of block-Toeplitz matrices related to the generalized orthonormal basis functions model. *Automatica*, 35:1653–1661.

[25] A.A. Stoorvogel and J.H. van Schuppen. System identification with information theoretic criteria. In S. Bittanti and G. Picci, editors, *Identification, adaptation, learning*, pages 289–338. Springer-Verlag, Berlin, 1996.

[26] L. Takacs. *Introduction to the Theory of Queues*. Oxford University Press, New York, 1962.

[27] A.G. Tartakovsky. Asymptotic optimality in Bayesian changepoint detection problems under global false alarm probability constraint. *Theory Probab. Appl.*, 53(3):443–466, 2009.

[28] A.G. Tartakovsky and V.V. Veeravalli. General asymptotic Bayesian theory of quickest change detection. *Theory Probab. Appl.*, 49(3):458–497, 2005.