Sequentially Detecting Transitory Changes

George V. Moustakides
Computer Science Department
Rutgers University, New Brunswick, NJ, USA
& Department of Electrical and Computer Engineering
University of Patras, Rion, Greece.

Venugopal V. Veeravalli
Department of Electrical and Computer Engineering
University of Illinois
Urbana-Champaign, IL, USA.

***Abstract*—We are interested in the sequential detection of a change in the statistical behavior of a random process. Specifically we consider changes that are not abrupt but exhibit a transitory phase before reaching their steady-state behavior. Adopting the classical worst-case conditional detection delay proposed by Lorden as our performance measure and constraining the average false-alarm period, we derive the sequential test that optimizes, in the exact sense, the proposed criterion. The resulting optimum rule resembles the well known CUSUM rule with the corresponding test-statistic-update being not only a function of all pre- and post-change pdfs but also of the false-alarm constraint.**

***Index Terms*—Sequential detection of changes, Quickest detection, Sequential Analysis.**

I. INTRODUCTION

In the quickest change detection literature, it is generally assumed that changes are abrupt, implying that the pre- and post-change statistical behavior of the observations is stationary. Although this assumption covers most applications, there are cases where it does not hold, because changes can be gradual. Interestingly, even if the change is abrupt, it is possible, by applying transformations, to end up with a change that is transitory. For example consider observations $\{z_t\}$ satisfying $z_t = v_t + \mu \mathbb{1}_{\{t>\tau\}}$ where τ is the change-time; μ is a constant mean; $\{v_t\}$ follows the simple first-order autoregressive model $v_t = \alpha v_{t-1} + w_t$ and $\{w_t\}$ is an i.i.d. process with mean 0. If we now apply the transformation $\xi_t = z_t - \alpha z_{t-1}$ then

$$\xi_t = \begin{cases} w_t, & t = 1, \ldots, \tau \\ w_t + \mu, & t = \tau \\ w_t + (1-\alpha)\mu, & t > \tau + 1. \end{cases}$$

Consequently, we observe, that an abrupt change in the original sequence $\{z_t\}$ corresponds to a transitory change in the transformed sequence $\{\xi_t\}$, but with the latter enjoying statistical independence of its samples, a property which is always analytically desirable and the primary reason for applying the specific data transformation. Clearly the duration of the transition can become more pronounced if we increase the order of the autoregressive model describing $\{v_t\}$. As it is mentioned in [1]–[4] there are interesting applications in econometrics, statistical process control, and environmental monitoring that can be modeled with changes exhibiting a short post-change transitory phase before reaching statistical stationarity.

Let us define our problem of interest more technically. We assume that we observe, sequentially, a process $\{\xi_t\}$ with independent samples that can be described as follows:

$$\xi_t \sim \begin{cases} \mathsf{f}_\infty(\xi) & \text{for } 0 < t \leq \tau \\ \mathsf{f}_0^i(\xi) & \text{for } t = \tau + i;\ 1 \leq i \leq D \\ \mathsf{f}_0^{\text{ss}}(\xi) & \text{for } t > \tau + D. \end{cases}$$

Specifically, we assume that there exists a change-time τ such that the observations up to (and including) time τ are i.i.d. following the nominal pdf $\mathsf{f}_\infty(\xi)$, while after $\tau + D$ they are again i.i.d. following an alternative pdf $\mathsf{f}_0^{\text{ss}}(\xi)$. However, this switching in pdf takes place gradually since at time $\tau + i$, where $i = 1, \ldots, D$, the pdf $\mathsf{f}_0^i(\xi)$ is a function of the time increment i.

We would like to emphasize that the problem we consider in this work is not the same as the detection of a *transient* change [5], [6]. The latter corresponds to the case where $\mathsf{f}_0^{\text{ss}}(\xi) = \mathsf{f}_\infty(\xi)$, namely the change in statistical behavior lasts D time instances and after this time period the process returns to its *nominal* behavior. Here, after the transient period the pdf $\mathsf{f}_0^{\text{ss}}(\xi)$ becomes stationary but is *different* from the nominal $\mathsf{f}_\infty(\xi)$.

From a purely practical perspective, one might wonder whether there is any essential reason for seeking optimum schemes for transitory changes, especially when the corresponding duration is short. Interestingly, the answer to this question is positive. This fact can become quite apparent when we consider applications in which, during the transitory phase, observations tend to be strongly erratic before settling to their steady-state distribution (see Fig. 1). By harvesting this short but often powerful abnormal behavior it is possible to enjoy non-negligible performance gains when adopting optimum

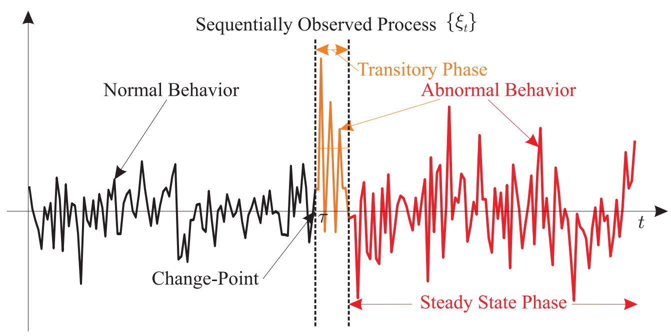

Fig. 1. Example of observations with strong transitory post-change statistical behavior.

detection strategies.

Before concluding our introduction we need to mention the work in [7], [8] that corresponds to the extremal version of our formulation with $D=\infty$ (no steady-state phase). In both articles the authors consider as post-change model a regression of the form $\xi_t = w_t + \{\mu + \delta(t-\tau)\}\mathbb{1}_{\{t>\tau\}}$ with μ, δ constants and $\{w_t\}$ i.i.d. It is clear that ξ_t, after the change, and because of the term $\delta(t-\tau)$, can never reach stationarity. Unlike [7], [8] where one can only establish first-order asymptotic optimality, here, by bounding D, we will attempt to obtain strictly optimum detection schemes.

II. Main Results

In this work we focus on the simplest case corresponding to $D=1$, namely, we consider a transitory period that involves just a *single* sample. Even though this assumption seems very restrictive the corresponding study reveals the main analytical difficulties occuring in the problem of transitory changes. We should also mention that there are significant practical applications where the assumption of $D=1$ can be satisfied. A characteristic example is that of *Line outage detection in power systems* (e.g. [9], [10]) where, after the occurrence of a line outage, the observation vector from the phasor measurement units (PMUs) exhibits a significant change in mean for a very short time-period before settling to its steady-state response. The latter resembles the nominal zero-mean pre-change behavior but with a different covariance matrix corresponding to the topology of the network after the line outage. Because of the sampling rates used to monitor the continuous-time PMU signals, we usually obtain only one sample from the transitory phase. As it has been observed in [9], [10] if we rely solely on the stationary post-change phase, we are able to detect the change, but this comes at the expense of an elevated detection delay compared to schemes that take advantage of both, the transitory and steady-state phases.

Returning to our mathematic setup, as it is well known from the sequential change detection literature, a sequential detector is simply a stopping time T which is $\{\mathscr{F}_t\}$-adapted with $\mathscr{F}_t = \sigma\{\xi_1,\ldots,\xi_t\}$. Regarding performance metrics for T, several criteria have been proposed which differ in the degree of knowledge we have about the mechanism that imposes the change (see [11] for an overview). In this work we focus on the performance measure suggested by Lorden [12]

$$J_{\rm L}(T) = \sup_{t\geq 0} \operatorname{ess\,sup} \mathsf{E}_t[T-t|T>t,\mathscr{F}_t], \tag{1}$$

where $\mathsf{E}_t[\cdot]$ denotes expectation with respect to the probability measure induced, when the change takes place at $\tau=t$. With this definition $\mathsf{E}_\infty[\cdot]$ refers to the expectation with respect to the pre-change measure, while $\mathsf{E}_0[\cdot]$ denotes that with respect to the post-change measure. We observe that in (1) we consider the worst possible history before the change and the worst possible change-time that maximize the average detection delay (conditioned on the event that there is no false-alarm before the change). In fact this criterion corresponds to complete lack of knowledge of the mechanism that generates the change [11].

Following [12] we can now optimize the stopping time T by solving the following constrained optimization problem

$$\inf_T J_{\rm L}(T), \quad \text{subject to: } \mathsf{E}_\infty[T] \geq \gamma > 1. \tag{2}$$

More specifically we minimize the Lorden metric for the stopping time T defined in (1), assuring, in parallel, an acceptable false-alarm level by suitably constraining the corresponding average false-alarm period.

It is clear that in order for the transitory phase to have any essential effect in our analysis we need to adopt a non-asymptotic formulation. This is because in asymptotic setups detection delays tend to infinity. Indeed we recall that, in the classical i.i.d. case, the average detection delay is of the order of $\log\gamma$ (with $\gamma\to\infty$ for the asymptotic analysis), consequently events of finite duration become insignificant.

For the classical case of abrupt changes ($D=0$), where we have a single pre- and post-change distribution $\mathsf{f}_\infty(\xi)$ and $\mathsf{f}_0(\xi)$, we know [13] that the optimum detector according to (2) is the CUSUM test defined through the CUSUM statistic

$$\begin{aligned} &R_t = \max\{R_{t-1},1\}\mathsf{L}_t = R_{t-1}\mathsf{L}_t + (1-R_{t-1})^+\mathsf{L}_t \\ &\text{where } x^+ = \max\{x,0\};\ \mathsf{L}_t = \frac{\mathsf{f}_0(\xi_t)}{\mathsf{f}_\infty(\xi_t)};\ R_0 = 0, \end{aligned} \tag{3}$$

and the corresponding CUSUM stopping time

$$T_{\rm C} = \inf\{t>0: R_t \geq \nu\}.$$

Threshold ν is selected so that the false-alarm constraint in (2) is satisfied with equality.

The recursion in (3) requires only one likelihood ratio since the post-change pdf is stationary. When, however, the change is not abrupt and $\mathsf{f}_0^{\rm tr}(\xi)$ denotes the pdf during the transitory phase while $\mathsf{f}_0^{\rm ss}(\xi)$ during the steady-state, then for each time instant t we have two possible likelihood ratios $\mathsf{L}_t^{\rm tr} = \mathsf{f}_0^{\rm tr}(\xi_t)/\mathsf{f}_\infty(\xi_t), \mathsf{L}_t^{\rm ss} = \mathsf{f}_0^{\rm ss}(\xi_t)/\mathsf{f}_\infty(\xi_t)$. To accommodate this situation we need to modify the test statistic properly so that it includes both possibilities. We therefore propose the following candidate detection scheme

$$\begin{aligned} &R_t = R_{t-1}\mathsf{L}_t^{\rm ss} + \phi(R_{t-1})\mathsf{L}_t^{\rm tr},\ R_0 = 0, \\ &T_o = \inf\{t>0: R_t \geq \nu\}. \end{aligned} \tag{4}$$

Here, $\phi(R)\geq 0$ is no longer equal to $(1-R)^+$ as in the classical CUSUM case. Actually, the main challenge in our analysis is to properly define this function so as to ensure optimality of (4) according to (2). As we discuss in the sequel, the exact form of $\phi(R)$, unlike in the classical case, depends not only on the three pdfs $\mathsf{f}_\infty(\xi), \mathsf{f}_0^{\rm tr}(\xi), \mathsf{f}_0^{\rm ss}(\xi)$, but also on the false-alarm parameter γ.

Even though we will be able to establish existence of $\phi(R)$, obtaining this function analytically, at least with the adopted analysis, seems impossible. For this reason we will develop techniques to determine it numerically and, if possible, accurately. A property that contributes towards this goal is that $\phi(R)$, as in the classical case where it is equal to $(1-R)^+$, is supported on the finite interval $[0,1]$. The latter allows for the fine sampling of the function in the interval $[0,1]$ leading to its efficient and accurate computation even with elementary numerical techniques.

The next lemma introduces a number of functions that are used to represent the average detection and false-alarm delay for T_o. In fact the two average delays will be computed not only for the initial value $R_0 = 0$ but for any value $R_0 = R$.

Lemma 1. *Fix $\phi(R) \geq 0$ to be a function with support on $[0,1]$ and let R_1 be defined from (4) but with $R_0 = R$. Define the functions $V_\infty(R), \tilde{V}_0(R)$ and $V_0(R)$ using the following equations*

$$\tilde{V}_0(R) = 1 + \mathsf{E}_0^{\rm ss}[\tilde{V}_0(R_1)\mathbb{1}_{\{R_1<\nu\}}|R_0 = R] \tag{5}$$
$$V_0(R) = 1 + \mathsf{E}_0^{\rm tr}[\tilde{V}_0(R_1)\mathbb{1}_{\{R_1<\nu\}}|R_0 = R] \tag{6}$$
$$V_\infty(R) = 1 + \mathsf{E}_\infty[V_\infty(R_1)\mathbb{1}_{\{R_1<\nu\}}|R_0 = R], \tag{7}$$

where $\mathsf{E}_0^{\rm ss}[\cdot], \mathsf{E}_0^{\rm tr}[\cdot]$ denote expectation with respect to the steady-state and transitory pdf respectively and $\mathsf{E}_\infty[\cdot]$ with respect to the nominal pdf. Then

$$\mathsf{E}_t[T_o - t|T_o > t, \mathscr{F}_t] = V_0(R_t) \tag{8}$$
$$\mathsf{E}_\infty[T_o] = V_\infty(0). \tag{9}$$

Proof: The proof is presented in the Appendix. ∎

As we mentioned, our goal is to design $\phi(R)$ in such a way that it yields the optimum version of T_o. The next theorem identifies important conditions that this function needs to satisfy.

Theorem 1. *There exists function $\phi(R) \geq 0$ supported on $[0,1]$ so that $V_0(R)$ defined through (5),(6) is constant for $R \in [0,1]$ and strictly decreasing for $R > 1$.*

Proof: This theorem is very technical and only the major steps of the proof are highlighted in the Appendix. ∎

From (8) in Lemma 1 we have that $\mathsf{E}_t[T_o - t|T_o > t, \mathscr{F}_t]$ depends on the history $\mathscr{F}_t$ solely through R_t, this suggests

$$\operatorname{ess\,sup} \mathsf{E}_t[T_o - t|T_o > t, \mathscr{F}_t] = \sup_{R_t} V_0(R_t).$$

From Theorem 1 we have that $V_0(R)$ attains its maximum for all $R \in [0,1]$ (without loss of generality we select the value $R = 0$), consequently, $\sup_{R_t} V_0(R_t) = V_0(0)$. Combining the two equalities yields $\operatorname{ess\,sup} \mathsf{E}_t[T_o - t|T_o > t, \mathscr{F}_t] = V_0(0)$. Since this is true for all t, it follows that our rule is an *equalizer* over time, and as a consequence we have

$$J_{\rm L}(T_o) = V_0(0). \tag{10}$$

We should emphasize that although not apparent from our notation, for fixed pdfs, the functions $V_\infty(R), V_0(R)$ also depend on the threshold ν.

Our candidate stopping time T_o is nearly completely defined. We only need to specify the exact value of the threshold ν and relate it to the false-alarm parameter γ. Before addressing this issue in Lemma 2, we prove an intermediate optimality result for T_o which we present with the next theorem.

Theorem 2. *Fix $\nu > 0$, and let $\phi(R)$ and T_o be the function from Theorem 1 and the corresponding stopping time from (4). Then for any stopping time T satisfying the constraint*

$$\mathsf{E}_\infty[T] \geq V_\infty(0) = \mathsf{E}_\infty[T_o]$$

we also have that

$$J_{\rm L}(T) \geq V_0(0) = J_{\rm L}(T_o).$$

Proof: Again only a sketch of the proof is presented in the Appendix. ∎

The immediate implication of Theorem 2 is that: if we fix the threshold ν, design $\phi(R)$ according to Theorem 1, and define the corresponding T_o according to (4), then T_o solves the constrained optimization problem in (2) provided we select the false-alarm parameter $\gamma = V_\infty(0)$.

In order to completely establish our original optimality claim about T_o we need to show that for any value of the parameter $\gamma > 1$ there always exists a threshold $\nu > 0$ that can meet the false-alarm constraint with equality, namely $V_\infty(0) = \gamma$. This is the aim of the following lemma.

Lemma 2. *For fixed pre- and post-change pdfs and for $\phi(R)$ selected from Theorem 1, the corresponding function $V_\infty(0)$ is increasing with respect to the threshold ν satisfying: $\lim_{\nu\to 0} V_\infty(0) = 1$ and $\lim_{\nu\to\infty} V_\infty(0) = \infty$. Furthermore, if the two likelihood ratios $\mathsf{L}_1^{\rm tr}, \mathsf{L}_1^{\rm ss}$ contain no atoms under $\mathsf{f}_\infty(\xi)$ then $V_\infty(0)$ is continuous in ν.*

Proof: The proof of this lemma is also very technical and its details are omitted. A brief sketch can be found in the Appendix. ∎

An immediate consequence of Lemma 2 is the following corollary that assures exact optimality of the proposed detector.

Corollary. *The stopping time T_o introduced in (4) with $\phi(R)$ defined from Theorem 1 and the threshold ν selected to satisfy the false-alarm constraint with equality, that is, $V_\infty(0) = \gamma$, is optimum according to (2).*

With the available results the proof is straightforward: Lemma 2, using continuity arguments, guarantees existence of ν such that $V_\infty(0) = \gamma$, while Theorem 2 establishes optimality of T_o according to Lorden's min-max constrained optimization depicted in (2).

III. Numerical Computations

In our analysis in the previous section we emphasized that the most crucial issue in this problem is the computation of the function $\phi(R)$ which can only be obtained numerically. To develop a numerical method we could follow the iterative logic employed in the Appendix (in "Proof of Theorem 1") for the existence of this function. Unfortunately the convergence speed of the resulting scheme tends to be very low. This is mainly due to the convergence of the solution of the integral equation in (11), which is particularly slow when we employ large threshold values ν.

If $\phi^{(k)}(R)$ is the estimate of $\phi(R)$ during the kth iteration, we can increase the convergence speed considerably by finding the solution to the integral equation in (5) directly and not iteratively as in (11). This can be achieved by sampling the range of R and then transforming the integral equation into a system of linear equations that can be solved numerically using classical (non-iterative) solvers thus generating the sampled version of $\tilde{V}_0(R)$. This estimate can then be used in (12) to solve the first order differential equation using simple finite differences. The latter provides the update of the estimate of $\phi(R)$ for the next iteration. As we mentioned, this idea has a positive impact on convergence, speeding up computations considerably.

To illustrate the proposed numerical technique we apply it to a Gaussian process $\{\xi_t\}$ that has mean equal to 0 before the change, while after the change the mean equals 1 for just one sample (transitory phase) and stabilizes to the value 2 for all subsequent samples. All three Gaussian pdfs have variance equal to 1. Clearly this is not a case where one expects drastic differences between $\phi(R)$ and the classical version $(1-R)^+$. But even with this simple example one can test if our analysis is of any practical importance by examining whether the optimum $\phi(R)$ can diverge from the classical version.

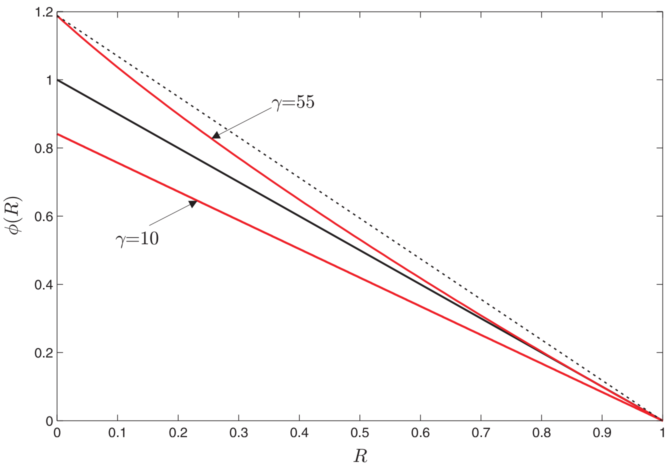

Fig. 2. In red the optimum function $\phi(R)$ for the Gaussian case with mean μ switching from 0 to 1 (transitory value) and then to 2 (steady-state value), for two different values of the false-alarm parameter $\gamma = 10, 55$. In solid black the classical version $(1-R)^+$.

Fig. 2 depicts the resulting optimum function $\phi(R)$ for values of the false-alarm parameter $\gamma = 10, 55$. We can see that in the first case, $\phi(R)$ is very close to a linear function. When, however, we employ the larger value for γ, the resulting optimum $\phi(R)$ is clearly nonlinear as can be noted by comparing it against the dashed black straight line. Therefore one can expect that in examples where the transitory behavior is more extreme, the nonlinear form of $\phi(R)$ will be more pronounced making it very different from the classical case depicted in solid black.

In the future (extended) version of our work we will target the analysis of a model that is inspired by the power system line outage detection problem described in Section II. This involves a nominal Gaussian pdf with mean 0 and variance 1, a transitory Gaussian pdf with mean $\mu \gg 1$ and variance 1 and, finally, a steady-state Gaussian pdf with mean 0 and variance $\sigma^2 > 1$. We anticipate that in this example the usage of the optimum detector will result in significant performance gains compared to alternative detection structures that ignore the transitory phase.

ACKNOWLEDGEMENT

This work was supported by the US National Science Foundation under Grant CIF 1513373, through Rutgers University, and by Grant CCF 1514245, through the University of Illinois at Urbana-Champaign.

APPENDIX

Proof of Lemma 1. The functions $\tilde{V}_0(R), V_\infty(R)$ are defined through the integral equations (5),(7). Existence of the solutions is guaranteed through classical integral equation theory and can be established using iterative solvers. In fact we can demonstrate that the sequence of functions that is generated is increasing for every fixed R and also bounded from above, therefore the pointwise (in R) limit is assured. Continuity can also be established by assuming that the two likelihood ratios exhibit no atoms with respect to $\mathsf{f}_0^{\rm tr}(\xi), \mathsf{f}_0^{\rm ss}(\xi)$ and $\mathsf{f}_\infty(\xi)$. The steps required are standard.

Once we establish existence of $\tilde{V}_0(R)$ the existence of $V_0(R)$ is also guaranteed since it involves a simple expectation of $\tilde{V}_0(R)$, as we can verify from (6)

To show (8), due to stationarity, we can limit ourselves to $t = 0$ and assume that $R_0 = R < \nu$. We can then write

$$\mathsf{E}_0[T_o] = \mathsf{E}_0\left[\sum_{s=0}^{\infty} \mathbb{1}_{\{T_o>s\}} \big| \mathscr{F}_0\right] \\ = 1 + \mathsf{E}_0^{\rm tr}\left[\mathsf{E}_0^{\rm ss}\left[\sum_{s=1}^{\infty} \mathbb{1}_{\{T_o>s\}} \big| \mathscr{F}_1\right] \big| \mathscr{F}_0\right].$$

It is well known that the stationary part involving identically distributed samples is related to the function $\tilde{V}_0(R)$ defined in (5). More precisely

$$\mathsf{E}_0^{\rm ss}\left[\sum_{s=1}^{\infty} \mathbb{1}_{\{T_o>s\}} \big| \mathscr{F}_1\right] = \tilde{V}_0(R_1)\mathbb{1}_{\{T_o>1\}}.$$

When this expression is substituted in the previous equation we obtain

$$\mathsf{E}_0[T_o] = 1 + \mathsf{E}_0^{\rm tr}[\tilde{V}_0(R_1)\mathbb{1}_{\{T_o>1\}} | \mathscr{F}_0] \\ = 1 + \mathsf{E}_0^{\rm tr}[\tilde{V}_0(R_1)\mathbb{1}_{\{R_1<\nu\}} | \mathscr{F}_0] = V_0(R),$$

which proves (8). Equation (9) involves only identically distributed samples, and it is therefore known to be valid. ■

Proof of Theorem 1. Existence of $\phi(R)$ will be established by proposing an iterative computation of this function. Specifically we will compute a sequence of functions $\phi^{(k)}(R)$, $\tilde{V}_0^{(k)}(R)$, $V_0^{(k)}(R)$ that converge to the desired ones. We start by setting $\phi^{(0)}(R) = (1-R)^+$ and $\tilde{V}_0^{(0)}(R) = 1$. Consider now that we are in the kth iteration having available $\phi^{(k-1)}(R)$ and $\tilde{V}_0^{(k-1)}(R)$. Then, according to (5) in Lemma 1, we can apply the following update

$$\tilde{V}_0^{(k)}(R) = 1 + \mathsf{E}_0^{\rm ss}[\tilde{V}_0^{(k-1)}(R_1)\mathbb{1}_{\{R_1<\nu\}}] \tag{11}$$

with $R_1 = R\mathsf{L}_1^{\rm ss} + \phi^{(k-1)}(R)\mathsf{L}_1^{\rm tr}$ (i.e. we replace R_0 with R in (4)).

We can now use $\tilde{V}_0^{(k)}(R)$ in (6) to *update* the estimate of $\phi(R)$ and compute $\phi^{(k)}(R)$. We recall that $\phi(R)$ is supported on $[0,1]$ and it must be such that the resulting $V_0(R)$ function is *constant* in the same interval. We will therefore impose the same condition on $V_0^{(k)}(R)$ by properly designing $\phi^{(k)}(R)$. This suggests that if we differentiate $V_0^{(k)}(R)$ with respect to R we must obtain 0 for $R \in [0,1]$. This means that for $R \in [0,1]$

$$
\begin{aligned}
0 = \frac{dV_0^{(k)}(R)}{dR} &= \frac{d}{dR}\mathsf{E}_0^{\rm tr}[\tilde{V}_0^{(k)}(R_1)\mathbb{1}_{\{R_1<\nu\}}] \\
&= \mathsf{E}_0^{\rm tr}\left[\partial_{R_1}\tilde{V}_0^{(k)}(R_1)\left\{\mathsf{L}_1^{\rm ss} + \frac{d\phi^{(k)}(R)}{dR}\mathsf{L}_1^{\rm tr}\right\}\mathbb{1}_{\{R_1<\nu\}}\right];
\end{aligned}
$$

from which we obtain the following first order differential equation

$$
\frac{d\phi^{(k)}(R)}{dR} = -\frac{\mathsf{E}_0^{\rm tr}[\partial_{R_1}\tilde{V}_0^{(k)}(R_1)\mathsf{L}_1^{\rm ss}\mathbb{1}_{\{R_1<\nu\}}]}{\mathsf{E}_0^{\rm tr}[\partial_{R_1}\tilde{V}_0^{(k)}(R_1)\mathsf{L}_1^{\rm tr}\mathbb{1}_{\{R_1<\nu\}}]}; \quad \phi^{(k)}(1) = 0, \tag{12}
$$

that needs to be solved starting from $R = 1$ with $\phi^{(k)}(1) = 0$ and going backwards towards $R = 0$. The boundary condition at $R = 1$ is required to secure continuity for $\phi(R)$ at this point (recall that the support of $\phi(R)$ is $[0,1]$ suggesting that $\phi(R) = 0$ for $R > 1$).

One can show that this iteration converges confirming the existence of the desired $\phi(R)$ and of the corresponding $V_0(R)$ enjoying the required properties. ∎

Proof of Theorem 2. Let $\phi(R)$ and ν be the function and the threshold needed by T_o to satisfy the false-alarm constraint with equality and the requirements of Theorem 1. Using ideas similar to the ones introduced in [13], we can prove optimality of T_o in three major steps. The first consists in showing that for any stopping time T with finite $\mathsf{E}_\infty[T]$ (actually it is sufficient to limit ourselves to such stopping times) we have

$$
J_{\rm L}(T) \geq \frac{\mathsf{E}_\infty\left[\sum_{t=0}^{T-1} R_t\right]}{\mathsf{E}_\infty\left[\sum_{t=0}^{T-1}\phi(R_t)\right]} = \tilde{J}_{\rm L}(T).
$$

The second step, which is simple due to the special form of $\phi(R)$ and $V_0(R)$ (namely that $\phi(R)$ has support on those values of R where $V_0(R)$ attains its maximum), consists in showing that for T_o it is true that $J_{\rm L}(T_o) = \tilde{J}_{\rm L}(T_o)$. Finally the third step is concerned with the minimization of $\tilde{J}_{\rm L}(T)$ instead of the original $J_{\rm L}(T)$. In particular we can show that among all stopping times that satisfy the false-alarm constraint, the expression $\tilde{J}_{\rm L}(T)$ is minimized by T_o.

With these facts at hand we can now establish optimality of T_o with respect to the *original* measure. Indeed note that

$$
J_{\rm L}(T_o) \geq \inf_T J_{\rm L}(T) \geq \inf_T \tilde{J}_{\rm L}(T) = \tilde{J}_{\rm L}(T_o) = J_{\rm L}(T_o)
$$

from which we can conclude that $\inf_T J_{\rm L}(T) = J_{\rm L}(T_o)$ and prove optimality of T_o. ∎

Proof of Lemma 2. For fixed pre- and post-change pdfs the function $\phi(R)$ depends on ν. Using the iterative solution introduced in the proof of Theorem 2, we can show that in each iteration the corresponding function $\phi^{(k)}(R)$ has the mentioned properties, which are retained in the limit. ∎

References

[1] T. J. Harris and W. H. Ross. Statistical process control procedures for correlated observations. *Canadian Journal of Chemical Engineering*, vol. 69, pp. 48–57, Feb. 1991.

[2] S. Knoth and M. Frisén. Minimax optimality of CUSUM for an autoregressive model. *Statistica Neerlandica*, vol. 66, no. 4, pp. 357–379, 2012.

[3] Y. Okhrin and W. Schmid. *Surveillance of Univariate and Multivariate Linear Time Series, in Financial Surveillance (ed. M. Frisén).* John Wiley & Sons, Ltd, 2007.

[4] E. Yashchin. Performance of CUSUM control schemes for serially correlated observations. *Technometrics*, vol. 35, no. 1, pp. 37–52, Feb. 1993.

[5] B. K. Guépié, L. Fillatre and I. Nikiforov. Sequential detection of transient changes. *Seq. Anal.*, vol. 31, pp. 528–547, 2012.

[6] E. Ebrahimzadeh and A. Tchamkerten. Sequential detection of transient changes in stochastic systems under a sampling constraint. Proceedings of the IEEE International Symposium on Information Theory, pp. 156–160, 2015.

[7] A. M. Krieger, M Pollak, B Yakir, Detecting a change in regression: first-order optimality. *Ann. Stat.*, vol. 27, no. 6, pp. 1896–1913, 1999.

[8] A. M. Krieger, M Pollak, B Yakir, Surveillance of a simple linear regression. *J. Amer. Stat. Assoc.*, vol. 98, no. 462, pp. 456–469, June 2003.

[9] X. He, M.-O. Pun, and C.-C. J. Kuo. Quickest detection of unknown power quality events for smart grids. *Proc. of Annual Summit of Asia-Pacific Signal and Information Processing Association*, pp. 335(1–4), 2012.

[10] Y. C. Chen, T. Banerjee, A. D. Domínguez-García, and V. V. Veeravalli. Quickest line outage detection and identification. *IEEE Tran. Power Syst.*, vol. 31, no. 1, pp. 749–758, Jan. 2016.

[11] G. V. Moustakides. Sequential change detection revisited. *Ann. Stat.*, vol. 36, no. 2, pp. 787–807, April 2008.

[12] G. Lorden. Procedures for reacting to a change in distribution. *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[13] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *Ann. Stat.*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.