51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Quickest detection in a system with correlated noise

Hongzhong Zhang, Olympia Hadjiliadis

Abstract— **This work considers the problem of quickest detection of signals in a system of 2 sensors coupled by a negatively correlated noise, which receive continuous sequential observations from the environment. It is assumed that the signals are time invariant and with equal strength, but that their onset times may differ from sensor to sensor. The objective is the optimal detection of the first time at which any sensor in the system receives a signal. The problem is formulated as a stochastic optimization problem in which an extended Lorden's criterion is used as a measure of detection delay, with a constraint on the mean time to the first false alarm. The case in which the sensors employ their own cumulative sum (CUSUM) strategies is considered, and it is proved that the minimum of 2 CUSUMs is asymptotically optimal as the mean time to the first false alarm increases without bound. Implications of this asymptotic optimality result to the efficiency of the decentralized versus the centralized system of observations are further discussed.**

Keywords: CUSUM, correlated sensors, quickest detection

I. INTRODUCTION

We are interested in the problem of quickest detection of the first onset of a signal in a system of 2 sensors with negatively correlated noise. We consider the situation in which, the noise in one sensor is correlated with the noise in the other, and the onset of a signal (i.e., change points) can occur at different times in each of the 2 sensors; that is, the change points differ from sensor to sensor. The presence of correlations is due to the fact that, although sensors are placed typically at different locations, they are subject to the same physical environment. For example, in the case of sensors monitoring traffic in opposite directions may have negative correlations due to environmental factors such as the direction of the wind [1]. in general, negative correlations in particular can arise in the case in which sensors are hit by the signal on opposite sides. Moreover, the appearance of a signal at one location may or may not cause interference of the signal at another location, thus causing correlations. This happens when the sensors are closely spaced relative to the curvature of the field being sensed. For example, temperature sensors or humidity sensors that are in a similar geographic region will produce readings that are correlated. A constant correlation across sensors would best describe such a situation. Some of the relevant literature that includes such examples can be found in [2]–[8].

This work is a continuation of the problem considered in [9] in which the case is considered of independent observations received at each sensor. In that work, it is seen that the decentralized system of sensors in which each sensor employs its own cumulative sum (CUSUM) [10] strategy and communicates its detection through a binary asynchronous message to the Fusion center, which in turn decides at the first onset of a signal based on the first communication performs asymptotically just as well as the centralized system. In other words, the minimum of N CUSUMs is asymptotically optimal in detecting the minimum of N distinct change points in the case of independent observations as the mean time to the first false alarm increases without bound. The mean time to the first false alarm can be used as a benchmark in actual applications in which the engineer or scientist may make several runs of the system while it is in control in order to uniquely identify, the appropriate parameter that would lead to a tolerable rate of false detection. The problem of optimal detection then boils down to minimizing the detection delay subject to a tolerable rate of false alarms. Asymptotic optimality is then proven by comparing the rate of increase in detection delay to the rate of false alarms as the threshold parameter varies. In our case the detection delay is measured with respect to a generalized linear Lorden-type detection delay criterion subject to a bound on the mean time to the first false alarm. A more recent related work includes the case in which the system of sensors is coupled through the drift parameter as opposed to the noise [11], [12]. In that work it is once again seen that the minimum of N CUSUMs is also asymptotically optimal in detecting the minimum of N distinct change points with respect to a generalized Kullback-Leibler distance criterion inspired by [13]. Yet, in none of the above cases is the case of correlated noise considered even though it is very important in practical applications.

In this work we consider the special case of a system of two sensors coupled through the presence of correlated Brownian noise and in which the onset of signal can occur at distinct times, which are assumed to be unknown constants. Thus a min-max approach is taken. The problem of detecting the minimum of two change points in a Bayesian setup and a Poisson model of observations was considered in [14]. The minimum of the two distinct points signifies the first onset of a signal in such a system. So far in the literature of this type of problem (see [15]–[19]) it has been assumed that the change points are the same across sensors. Recently the case was also considered of change points that propagate

This work was supported by the NSF-CCF-MSC grant 0916452, the NSA-MSP grant 081103 and the NSF-DMS grant 1222526. Hongzhong Zhang is with the Statistics Department at Columbia University hzhang@stat.columbia.edu

Olympia Hadjiliadis is with the Department of Mathematics at Brooklyn College of the City University of New York, and with the Departments of Computer Science and Mathematics at the Graduate Center of City University of New York ohadjiliadis@brooklyn.cuny.edu

in a sensor array [20]. However, in this configuration the propagation of the change points depends on the unknown identity of the first sensor affected and considers a restricted Markovian mechanism of propagation of the change. In this paper we consider the case in which the change points can be different and do not propagate in any specific configuration. The objective is to detect the minimum (i.e., the first) of the change points.

In the next section we formulate the problem, and prove asymptotic optimality of the minimum of 2 CUSUMs as the mean time between false alarms tends to ∞, in the extended Lorden sense of [9]. We finally discuss extensions of these results to the N-sensor case and to the case of positive correlation.

II. FORMULATIONS & RESULTS

We sequentially observe the processes $\{\xi_t^{(i)}\}_{t\geq 0}$ for $i=1,2$. In order to formalize this problem we consider the measurable space $(\Omega, \mathcal{F}, \mathbb{F})$, where $\Omega = C[0,\infty]^2$ and $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ with $\mathcal{F}_t = \sigma\{(\xi_s^{(1)}, \xi_s^{(2)}); s \leq t\}$.

The processes $\{\xi_t^{(i)}\}_{t\geq 0}$ for $i=1,2$ are assumed to have the following dynamics:

$$d\xi_t^{(i)} = \begin{cases} dw_t^{(i)} & t \leq \tau_i \\ \mu\, dt + dw_t^{(i)} & t > \tau_i, \end{cases} \tag{1}$$

where $\mu > 0$ is known and $\{w_t^{(i)}\}_{t\geq 0}$ are two correlated standard Brownian motion with negative correlation $-1 < \rho < 0$. In other words, we assume that, under any measure, $E\{w_t^{(1)} w_s^{(2)}\} = \rho(s \wedge t)$. We also assume that τ_i, for $i=1,2$, are both unknown constants in the interval $[0,\infty]$.

On the space Ω, we have the following family of probability measures $\{P_{\tau_1,\tau_2}\}$, where P_{τ_1,τ_2} corresponds to the measure generated on Ω by the process $\{(\xi_t^{(1)}, \xi_t^{(2)})\}_{t\geq 0}$ when the change in the 2-tuple process occurs at time point τ_i for each $i=1,2$. Notice that the measure $P_{\infty,\infty}$ corresponds to the measure generated on Ω by 2 correlated standard Brownian motions.

Although the filtration $\mathbb{F}$ summarizes the totality of information available in this system, each sensor S_i $i=1,2$ has access to the one-dimensional process $\{\xi_t^{(i)}\}_{t\geq 0}$, each of which generates the filtration $\mathbb{G}^{(i)} = \{\mathcal{G}_t^{(i)}\}_{t\geq 0}$ with $\mathcal{G}_t^{(i)} = \sigma\{(\xi_s^{(i)}); s \leq t\}$ for $i=1,2$. On this filtration we can introduce the measures $\{P_{\tau_i}\}$ with P_∞ being one-dimensional standard Wiener measure.

Our objective is to find a stopping rule T that balances the trade-off between a small detection delay subject to a lower bound on the mean-time between false alarms and will ultimately detect $\tau_1 \wedge \tau_2$. In what follows we will use $\tilde{\tau}$ to denote $\min\{\tau_1, \tau_2\}$.

As a performance measure we consider

$$J^{(2)}(T) = \sup_{\substack{\tau_1,\tau_2 \\ \tilde{\tau}<\infty}} \operatorname{essup} \; E_{\tau_1,\tau_2}\left\{(T-\tilde{\tau})^+ \mid \mathcal{F}_{\tilde{\tau}}\right\}, \tag{2}$$

where the supremum over τ_1, τ_2 is taken over the set in which $\tilde{\tau} < \infty$. That is, we consider the worst detection delay over all possible realizations of paths of the 2-tuple of stochastic process $\{(\xi_t^{(1)}, \xi_t^{(2)})\}_{t\geq 0}$ up to $\tilde{\tau}$ and then consider the worst detection delay over all possible 2-tuples $\{\tau_1, \tau_2\}$ over a set in which at least one of them is forced to take a finite value. This is because T is a stopping rule meant to detect the minimum of the 2 change points and therefore if one of the 2 processes undergoes a regime change, any unit of time by which T delays in reacting, should be counted towards the detection delay. This gives rise to the following stochastic optimization problem:

$$\begin{gathered} \inf_{T\in\mathbb{F}_\gamma} J^{(2)}(T), \\ \text{with } \mathbb{F}_\gamma = \{\mathbb{F}\text{-stopping rule } T \;:\; E_{\infty,\infty}\{T\} \geq \gamma\}. \end{gathered} \tag{3}$$

$E_{\infty,\infty}\{T\}$ captures the mean time to the first false alarm and as such the above constraint describes the tolerance on the mean time to the first false alarm.

In the case of the presence of only one stochastic process (say, $\{\xi_t^{(1)}\}_{t\geq 0}$), the problem becomes one of detecting a one-sided change in a sequence of Brownian motion observations, whose optimality was found in [21] and [22]. The optimal solution under Lorden's criterion is the continuous time version of Page's CUSUM stopping rule, namely the first passage time of the process

$$y_t^{(1)} = \sup_{0\leq\tau_1\leq t} \frac{dP_{\tau_1}}{dP_\infty}(\mathcal{G}_t^{(1)}) \tag{4}$$

$$= \mu\xi_t^{(1)} - \frac{1}{2}\mu^2 t - \inf_{s\leq t}\left(\mu\xi_s^{(1)} - \frac{1}{2}\mu^2 s\right). \tag{5}$$

The CUSUM stopping rule is thus

$$T_{\nu^\star} = \inf\{t \geq 0 : y_t^{(1)} \geq \nu^\star\}, \tag{6}$$

where $\nu^\star$ is chosen so that $E_\infty\{T_{\nu^\star}\} =: \frac{2}{\mu^2} f(\nu^\star) = \gamma$, with $f(\nu) = e^\nu - \nu - 1$ and the corresponding optimal detection delay is given by

$$J^{(1)}(T_{\nu^\star}) = E_0\{T_{\nu^\star}\} = \frac{2}{\mu^2} f(-\nu^\star). \tag{7}$$

The fact that the worst detection delay is the same as that incurred in the case that the change point is exactly at 0 is a consequence of non-negativity and strong Markov property of the CUSUM process, from which it follows that the worst detection delay occurs when the CUSUM process is at 0 at the time of the change [23].

Returning to problem (3), it is easily seen that in seeking solutions to this problem, we can restrict our attention to stopping rules that achieve the false alarm constraint with equality [23]. Moreover, as discussed in [9] the optimal solution to (3) is achieved by equalizer rules. This means that the optimal stopping rule will be indifferent in its detection delay with respect to which of the two sensors changes dynamics. More specifically, let

$$J_{S_i}^{(2)}(T) = \sup_{\tau_i\leq\tau_j, j\neq i} \operatorname{essup} E_{\tau_1,\tau_2}\{(T-\tau_i)^+ | \mathcal{F}_{\tau_i}\}.$$

Then the optimal stopping rule should satisfy

$$J_{S_1}^{(2)} = J_{S_2}^{(2)}. \tag{8}$$

The optimality of the CUSUM stopping rule in the presence of only one observation process suggests that a CUSUM type of stopping rule might display similar optimality properties in the case of multiple observation processes. In particular, an intuitively appealing rule, when the detection of $\tilde{\tau}$ is of interest, is $T_h = T_h^1 \wedge T_h^2$, where T_h^i is the CUSUM stopping rule for the process $\{\xi_t^{(i)}\}_{t\geq 0}$ for $i = 1, 2$. That is, we use what is known as a multi-chart CUSUM stopping rule [24], which can be written as

$$T_h = \inf\left\{t \geq 0 \,:\, \max\{y_t^{(1)}, y_t^{(2)}\} \geq h\right\} \quad = \quad T_h^1 \wedge T_h^2. \tag{9}$$

where

$$y_t^{(i)} = \mu\xi_t^{(i)} - \frac{1}{2}\mu^2 t - \inf_{s\leq t}\left(\mu\xi_s^{(i)} - \frac{1}{2}\mu^2 s\right).$$

We notice that each of the T_h^i for $i = 1, 2$ are stopping times not only with respect to the large filtration $\mathbb{F}$ but also with respect to each of the smaller filtrations $\mathbb{G}^{(i)}$ and thus they can be employed by each one of the sensors S_i separately. Each of the sensors can subsequently communicate an alarm to a central fusion center once the threshold h is reached by its own cusum statistic process $y_t^{(i)}$ for $i = 1, 2$. The resulting rule, namely (9), can then be devised by the central fusion center in that it will declare a detection at the first instance either one of the two sensors communicates.

It can be shown that

$$J^{(2)}(T_h) \;=\; E_{0,\infty}\left\{T_h\right\} \;=\; E_{\infty,0}\left\{T_h\right\}.$$

In particular, we have

Lemma 1: For any $\tau_1 < \infty$ and $\tau_2 \geq \tau_1$,

$$\sup_{\tau_1,\tau_2} \operatorname{essup} E_{\tau_1,\tau_2}\{(T_h - \tau_1)^+ | \mathcal{F}_{\tau_1}\} = E_{0,\infty}\{T_h\}. \tag{10}$$

Proof: Please refer to the Appendix. ∎

What Lemma 1 states is that the worst detection delay occurs when only one of the 2 processes changes regime. An intuitive reason for this lies in the fact that the CUSUM process is a monotone function of μ, resulting in a longer on average passage time if $\mu = 0$ [25]. Thus, the worst detection delay will occur when none of the other processes changes regime, and due to the non-negativity of the CUSUM process the worst detection delay will occur when the CUSUM process of the remaining one process is at 0.

Notice that the threshold h is used for the multi-chart CUSUM stopping rule (9) in order to distinguish it from $\nu^\star$, the threshold used for the one sided CUSUM stopping rule (6).

In what follows we will demonstrate the asymptotic optimality of (9) as $\gamma \to \infty$. In view of the discussion in the previous paragraph, in order to assess the optimality properties of the multi-chart CUSUM rule (9) we will thus need to begin by evaluating $E_{0,\infty}\left\{T_h\right\}$ and $E_{\infty,\infty}\left\{T_h\right\}$ for general negative correlation $\rho \in (-1, 0]$.

In the special case that the correlation $\rho = 0$, the authors in [9] derived the asymptotic expansion of $E_{0,\infty}\{T_h\}$ and $E_{\infty,\infty}\{T_h\}$. In particular, they showed that, as $h \to \infty$,

$$E_{0,\infty}\{T_h\} \;=\; \frac{2}{\mu^2}(h - 1 + o(1)), \tag{11}$$

$$E_{\infty,\infty}\{T_h\} \;=\; \frac{1}{\mu^2}(e^h - 4 + o(1)). \tag{12}$$

In the special case that the correlation $\rho = -1$, the proposed 2-CUSUM stopping rule is identical to the 2-CUSUM used to detect a two-sided change in the drift of a Brownian motion [25], [26]. We now have that

$$E_{\infty,\infty}\{T_h\} \;=\; \frac{1}{2}E_{\infty,\infty}\{T_h^1\} = \frac{1}{\mu^2}f(-h), \tag{13}$$

where the first equality follows by the harmonic mean rule [25] and the second equality by using Itô's rule and Dynkin's formula [27] to the function $f(-y_t^{(1)})$ much along the lines of [13] or [25]. In both cases, the 2-CUSUM stopping rule are proven to exhibit asymptotic optimality under the criterion [25] as the mean time to the first false alarm γ increases without bound. In what follows we will show that the quantities in (11), (12) and (13) can be used to prove the asymptotic optimality of the 2-CUSUM stopping rule in the general case that $\rho \in (-1, 0)$. This is due to the following key result:

Proposition 1: For any $\rho \in (-1, 0)$, we have

$$E_{0,\infty}\{T_h\} \;\leq\; \frac{2}{\mu^2}(h - 1 + o(1)), \tag{14}$$

$$\frac{1}{\mu^2}f(h) \;\leq\; E_{\infty,\infty}\{T_h\} \;\leq\; \frac{1}{\mu^2}(e^h - 4 + o(1)), \tag{15}$$

as $h \to \infty$.

Proof: In observing (11), (12) and (13), it suffices to prove that the expectation of the 2-CUSUM stopping rule T_h under any of the above probability measures is bounded above by the expectation of the 2-CUSUM rule in the special case $\rho = 0$, and bounded below by the special case $\rho = -1$. Both of these claims are proved in Lemma 2 and Lemma 3 of the Appendix. ∎

As a result of Proposition 1, we can select h that solves equation $\frac{1}{\mu^2}f(h) = \gamma$, which guarantees that the 2-CUSUM stopping rule $T_h \in \mathbb{F}_\gamma$ since it satisfies $E_{\infty,\infty}\{T_h\} > \gamma$.

Corollary 1: By choosing h so that $\frac{1}{\mu^2}f(h) = \gamma$, we have $T_h \in \mathbb{F}_\gamma$, and

$$E_{0,\infty}\{T_h\} \leq \frac{2}{\mu^2}(\log\gamma + 2\log\mu - 1 + o(1)), \tag{16}$$

as $\gamma \to \infty$.

In order to demonstrate asymptotic optimality of (9), we first trivially have

$$E_{0,\infty}\{T_h\} = J^{(2)}(T_h) \geq \inf_{T\in\mathbb{F}_\gamma} J^{(2)}(T), \tag{17}$$

where h satisfies $\frac{1}{\mu}f(h) = \gamma$. On the other hand, the optimal detection delay in the 2 dimensional case is bounded below by the optimal detection delay in the 1 dimensional case. That is,

$$\inf_{T\in\mathbb{F}_\gamma} J^{(2)}(T) \geq \frac{2}{\mu^2}f(-\nu^\star). \tag{18}$$

Although this is a counterintuitive result, a moment of reflection will lead us to a better understanding of why such a relationship holds in (18). We notice that in the left hand side of (18) the infimum is taken over all $\mathbb{F}$-stopping times T that satisfy $E_{\infty,\infty}\{T\} = \gamma$, while in the right hand side of (18), which corresponds to the one dimensional case, the infimum is take over the $\mathbb{G}^{(i)}$- stopping times T which satisfy $E_{\infty}\{T\} = \gamma$. The latter is clearly a more restrictive set of stopping rules.

More mathematically, for any stopping rule $T \in \mathbb{F}_{\gamma}$, we have

$$J^{(2)}(T) \geq J_1^{(2)}(T), \tag{19}$$

where

$$J_1^{(2)}(T) = \sup_{\tau_1 < \infty} \operatorname{essup} E_{\tau_1,\infty}\{(T-\tau_1)^+ | \mathcal{F}_{\tau_1}\}, \tag{20}$$

which, although is a performance measure in 2 dimensions, it is essentially a performance measure in 1. In particular, it can be shown that the one-dimensional CUSUM stopping rule T_1 of (6) with threshold parameter ν^* is the optimal solution to the problem of

$$\inf_{T \in \mathbb{F}_{\gamma}} J_1^{(2)}(T), \quad \text{with } \mathbb{F}_{\gamma} = \{\mathbb{F}\text{-stopping rule } T \ : \ E_{\infty,\infty}\{T\} \geq \gamma\}. \tag{21}$$

We notice that (21) is a two-dimensional problem whose optimal solution coincides with the optimal solution to the one-dimensional problem

$$\inf_{T \in \mathbb{G}_{\gamma}} J^{(1)}(T), \quad \text{with } \mathbb{G}_{\gamma} = \{\mathbb{G}\text{-stopping rule } T \ : \ E_{\infty}\{T\} \geq \gamma\}, \tag{22}$$

whose optimal solution is known to be the one-dimensional CUSUM stopping rule of (6).

The steps required to prove that the one-dimensional CUSUM T_1 of (6) with threshold parameter ν^* (which we denote by T_{ν^*} in the sequel) chosen so that $\frac{2}{\mu^2} f(\nu^*) = \gamma$ is the optimal solution to (21) basically replicate the proof of optimality of the CUSUM in [13] for the choice $\alpha_t = \mu$ and can be summarized as follows:

1) Step 1: Every $\mathbb{F}$ stopping time S satisfies the lower bound

$$J_1^{(2)}(S) \geq \frac{E_{\infty,\infty}\{e^{y^{(1)}_{S_\nu}} f(-y^{(1)}_{S_\nu})\}}{E_{\infty,\infty}\{e^{y^{(1)}_{S_\nu}}\}}, \tag{23}$$

where $S_\nu = S \wedge T_\nu$ for any $\nu > 0$.

2) Step 2: The function $\Psi_S(\nu) := E_{\infty,\infty}\{f(y^{(1)}_{S_\nu})\}$ is continuous.

3) Step 3: Any $\mathbb{F}$-stopping time S that satisfies the false alarm constraint of (21) with equality satisfies

$$J_1^{(2)}(S) \quad \geq \quad \frac{2}{\mu^2} f(-\nu*). \tag{24}$$

Thus the optimal detection delay $\inf_{T \in \mathbb{F}_{\gamma}} J^{(2)}(T)$ is bounded below by the optimal detection delay $\frac{2}{\mu^2} f(\nu^\star)$. By combining (17) and (18) we have

$$E_{0,\infty}\{T_h\} \geq \inf_{T \in \mathbb{F}_{\gamma}} J^{(2)}(T_h) \geq \frac{2}{\mu^2} f(-\nu^\star). \tag{25}$$

We will demonstrate that the difference between the upper and the lower bounds is bounded by a constant as $\gamma \to \infty$, with thresholds h and $\nu^\star$ chosen so that

$$\frac{1}{\mu^2} f(h) \quad = \quad \gamma, \tag{26}$$

$$\frac{2}{\mu^2} f(\nu^\star) \quad = \quad \gamma. \tag{27}$$

In view of Corollary 1, we have

Theorem 1: For any $\rho \in (-1, 0)$, the difference between the detection delay of T_h, $J^{(2)}(T_h)$, and the optimal detection delay $\inf_{T \in \mathbb{F}_{\gamma}} J^{(2)}(T)$, is bounded above by $\frac{2}{\mu^2} \log 2$, as $\gamma \to \infty$.

Proof: It is easily seen from (27) that

$$\nu^\star = \log \gamma + 2 \log \mu - \log 2.$$

Thus the lower bound of (25) is

$$\frac{2}{\mu^2} f(-\nu^\star) = \frac{2}{\mu^2} (\log \gamma + 2 \log \mu - \log 2 - 1 + o(1)).$$

Using (16) and (25) we have

$$\begin{aligned} 0 \leq & J^{(2)}(T_h) - \inf_{T \in \mathbb{F}_{\gamma}} J^{(2)}(T) \\ \leq & E_{0,\infty}\{T_h\} - \frac{2}{\mu^2} f(-\nu^\star) \\ \leq & \frac{2}{\mu^2} \log 2. \end{aligned}$$

This completes the proof. ∎

The consequence of Theorem 1 is the asymptotic optimality of (9). We notice that this asymptotic optimality holds for any negative correlation between the observation at different sensors.

III. CONCLUSIONS AND FUTURE WORKS

In this paper we have demonstrated the asymptotic optimality of the minimum of 2 CUSUMs for detecting the minimum of 2 different change points in a system of 2 sensors, coupled by correlated noise, which receive sequential observations from the environment.

In order to generalize the argument to N sensors, one may consider a covariance matrix with non-positive off-diagonal entries. Then it is appealing to use the minimum of N CUSUM stopping rules, the so called N-CUSUM stopping rule to detect the first instance of a signal. In this case, we will still be able to bound the expected delay of this stopping rule by the expected delay of the N-CUSUM stopping rule if the sensors are independent. However, we do not necessarily have a realistic lower bound for the mean time to the first false alarm. This is because, the simple N dimensional analogue of the covariance structure in the 2 dimensional case, $\Gamma := (\gamma_{i,j})_{i,j=1}^N$ with $\gamma_{i,j} = \mathbf{1}_{\{i=j\}} - \mathbf{1}_{\{i \neq j\}}$ is in fact not non-negative definite, and thus cannot be a covariance matrix.

The case $0 < \rho < 1$ presents difficulties in that we will need to investigate the expectation of the 2-CUSUM stopping rule under $P_{0,\infty}$ in the case $\rho = 1$. And the analysis becomes more complicated due to the asymmetry of the drifts of the CUSUM processes $\{y_t^{(i)}\}_{t\geq 0}$, $i = 1, 2$ under $P_{0,\infty}$. However, we believe it may be tractable and is open for future work.

The special cases $\rho = -1$ and $\rho = 1$ are degenerate and reduce the problem to the one-dimensional change point problem with two-sided and one-sided alternatives respectively. Notice that, in our setting, the change-points τ_1 and τ_2 can be distinct, which will make P_{τ_1,τ_2} not absolutely continuous with respect to the nominal measure $P_{\infty,\infty}$. As a result, the CUSUM stopping rule considered in this paper cannot be used to treat these two special cases.

REFERENCES

[1] R. Cuchiara, M. Piccardi, P. Mello, Image analysis and rule-based reasoning for a traffic monitoring system, *IEEE Intelligent Transportation Systems Society*, Vol. 1, No.2, 2000, pp. 119-130.

[2] M. Basseville, M. Abdelghani and A. Benveniste, Subspace-based fault detection algorithms for vibration monitoring, *Automatica*, Vol. 36, No.1, 2000, pp. 101- 109.

[3] M. Basseville, A. Benveniste, M. Goursat and L. Mevel, Subspace-based algorithms for structural identification, damage detection, and sensor data fusion, *Journal of Applied Signal Processing, Special Issue on Advances in Subspace-Based Techniques for Signal Processing and Communications*, Vol. 2007, No. 1, 2007, pp. 200-213.

[4] M. Basseville, L. Mevel and M. Goursat, Statistical model-based damage detection and localization, subspace-based residuals and damage-to-noise sensitivity ratios, *Journal of Sound and Vibration*, Vol. 275, No. 3-5, 2004, pp. 769-794.

[5] D. J. Ewins, *Modal Testing: Theory, Practice and Applications*, Research Studies Press, Letchworth, Hertfordshire UK, 2nd edition, 2000.

[6] W. Heylen, S. Lammens, and P. Sas, *Modal Analysis Theory and Testing*, Department of Mechanical Engineering, Katholieke Universiteit Leuven, Leuven, Belgium, 1995.

[7] J. N. Juang, *Applied System Identification*, Prentice Hall, Englewood Cliffs, NJ, USA, 1994.

[8] B. Peeters and G. De Roeck, Reference-based stochastic subspace identification for output-only modal analysis, *Mechanical Systems and Signal Processing*, Vol. 13, No. 6, 1999, pp. 855-877.

[9] O. Hadjiliadis, H. Zhang and H.V. Poor, One shot schemes for decentralized quickest change detection, *IEEE Transactions of Information Theory*, Vol. 55, No. 7, pp. 3346-3359, 2009.

[10] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, Cambridge UK, 2008.

[11] O. Hadjiliadis, T. Schäfer, and H. V. Poor, Quickest detection in coupled systems, *Proceedings of the 48th IEEE Conference on Decision and Control*, Shanghai, China, December 16-18, 2009.

[12] O. Hadjiliadis, T. Schäfer, H. Zhang and H. V. Poor, Quickest detection in coupled systems, 2012, under review by the SIAM Journal on Control and Optimization.

[13] G. V. Moustakides, Optimality of the CUSUM procedure in continuous time, *Annals of Statistics*, Vol. 32, No. 1, 2004, pp. 302-315.

[14] E. Bayraktar and H. V. Poor, Quickest detection of a minimum of two Poisson disorder times, *SIAM Journal on Control and Optimization*, Vol. 46, No. 1, pp. 308-331, 2007.

[15] G. V. Moustakides, Decentralized CUSUM change detection, *Proceedings of the 9th International Conference on Information Fusion (ICIF)*, Florence, Italy, 2006, pp. 1 - 6.

[16] A. G. Tartakovsky and H. Kim, Performance of certain decentralized distributed change detection procedures, *9th International Conference on Information Fusion (ICIF)*, Florence, Italy, July 10-13 2006, pp. 1-8.

[17] A. G. Tartakovsky and V. V. Veeravalli, Quickest change detection in distributed sensor systems, *Proceedings of the 6th Conference on Information Fusion*, Cairns, Australia, July 8-11 2003.

[18] A. G. Tartakovsky and V. V. Veeravalli, Change-point detection in multichannel and distributed systems with applications, in *Applications of Sequential Methodologies*, pp. 331-363, (N. Mukhopadhay, S.Datta and S. Chattopadhay, Eds), Marcel Dekker, New York, 2004.

[19] A. G. Tartakovsky and V. V. Veeravalli, Asymptotically optimum quickest change detection in distributed sensor systems, *Sequential Analysis*, Vol.27, No. 4, 2008, pp. 441-475.

[20] V. Raghavan and V. V. Veeravalli, Quickest detection of a change process across a sensor array, *Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, July 1-3 2008.

[21] M. Beibel, A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure, *Annals of Statistics*, Vol. 24, No. 2, pp. 1804-1812, 1996.

[22] A. N. Shiryaev, Minimax optimality of the method of cumulative sums (CUSUM) in the continuous case, *Russian Mathematical Surveys*, Vol. 51, No. 4, pp. 750-751, 1996.

[23] G. V. Moustakides, Optimal stopping times for detecting changes in distributions, *Annals of Statistics*, Vol. 14, No. 4, pp. 1379-1387, 1986.

[24] A. G. Tartakovsky, Asymptotic performance of a multichart CUSUM test under false alarm probability constraint, *Proceedings of the 44th IEEE Conference on Decision and Control*, Seville, Spain, December 12 - 15 2005, pp. 320 - 325.

[25] O. Hadjiliadis and G. V. Moustakides, Optimal and asymptotically optimal CUSUM rules for change point detection in the Brownian motion model with multiple alternatives, *Theory of Probability and its Applications*, Vol. 50, No. 1, pp. 131-144, 2006.

[26] O. Hadjiliadis, G. Hernandez del-Valle and I. Stamos, A comparison of 2-CUSUM stopping time rules for quickest detection of two-sided alternatives in a Brownian motion model, *Sequentially Analysis*, Vol 28, No. 1, pp. 92-114, 2009.

[27] B. Oksendal, *Stochastic Differential Equations*, Springer-Verlag, New York, 2002.

[28] T. W. Anderson, A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size, *Annals of Mathematical Statistics* Vol. 31, 165-197, 1960.

IV. APPENDIX

In the Appendix we give proofs of the lemmas used in the main text.

Proof of Lemma 1:

Proof: It suffices to show that for any $\tau_1 \leq \tau_2$ such that $\tau_1 < \infty$,

$$E_{\tau_1,\tau_2}\{(T_h - \tau_1)^+ | \mathcal{F}_{\tau_1}\} \leq E_{0,\infty}\{T_h\}, \ P_{0,\infty}\text{-a.s.} \tag{28}$$

To show this, we apply Itô's lemma to processes $\{g^-(y^{(1)}_{t\wedge\tau_2}, y^{(2)}_{t\wedge\tau_2})\}_{t\geq\tau_1}$ and $\{g^+(y^{(1)}_t, y^{(2)}_t)\}_{t\geq\tau_2}$ to obtain

$$\begin{aligned}
&E_{\tau_1,\tau_2}\{(T_h - \tau_1)^+ | \mathcal{F}_{\tau_1}\} \times \frac{\mu^2}{2} \\
=&E_{\tau_1,\infty}\{\mathbf{1}_{\{T_h\geq\tau_2\}}[g^-(y^{(1)}_{\tau_1}, y^{(2)}_{\tau_1}) - g^-(y^{(1)}_{\tau_2}, y^{(2)}_{\tau_2})] | \mathcal{F}_{\tau_1}\} \\
&\times \mathbf{1}_{\{T_h\geq\tau_1\}} + E_{\tau_1,\infty}\{\mathbf{1}_{\{T_h\geq\tau_2\}} g^+(y^{(1)}_{\tau_2}, y^{(2)}_{\tau_2}) | \mathcal{F}_{\tau_1}\}\mathbf{1}_{\{T_h\geq\tau_1\}} \\
&+ E_{\tau_1,\infty}\{\mathbf{1}_{\{T_h<\tau_2\}} g^-(y^{(1)}_{\tau_1}, y^{(2)}_{\tau_1}) | \mathcal{F}_{\tau_1}\}\mathbf{1}_{\{T_h\geq\tau_1\}},
\end{aligned}$$

where $g^{\pm}$ are solutions to the following partial differential equations on $[0, h]^2$:

$$\frac{\partial^2 g}{\partial x^2} + 2\rho\frac{\partial^2 g}{\partial x \partial y} + \frac{\partial^2 g}{\partial y^2} + \frac{\partial g}{\partial x} \pm \frac{\partial g}{\partial y} = -1,$$

with boundary condition

$$g|_{x=h} = g|_{y=h} = \frac{\partial g}{\partial x}\Big|_{x=0} = \frac{\partial g}{\partial y}\Big|_{y=0} = 0. \tag{29}$$

Using Lemma 4 we have that $g^-(x, y) \leq g^-(0, 0)$, and

$$\frac{\partial g^-}{\partial y} \leq 0.$$

Now notice that for $\Delta := g^- - g^+$ satisfies the boundary condition (29) and

$$\frac{\partial^2 \Delta}{\partial x^2} + 2\rho \frac{\partial^2 \Delta}{\partial x \partial y} + \frac{\partial^2 \Delta}{\partial y^2} + \frac{\partial \Delta}{\partial x} + \frac{\partial \Delta}{\partial y} = 2\frac{\partial g^-}{\partial y} \leq 0,$$

which implies that $\{\Delta(y_t^{(1)}, y_t^{(2)})\}$ is a super-martingale under $P_{0,0}$. Hence,

$$g^-(x,y) - g^+(x,y) = h(x,y) \geq E_{0,0}\{\Delta(y_{T_h}^{(1)}, y_{T_h}^{(2)})\} = 0.$$

As a result, we obtain

$$\frac{\mu^2}{2} E_{\tau_1,\tau_2}\{(T_h - \tau_1)^+ | \mathcal{F}_{\tau_1}\} \leq g^-(0,0)\mathbf{1}_{\{T_h \geq \tau_2\}},\ P_{\tau_1,\infty}\text{-a.s.}$$

On the other hand, a similar argument using Itô's lemma's lemma yields

$$\frac{\mu^2}{2} E_{0,\infty}\{T_h\} = g^-(0,0).$$

Therefore, (28) holds and this completes the proof. ■

We now proceed to prove the results used in the proof of Proposition 1.

Lemma 2: For any constant correlation $\rho \in (-1, 0]$, consider two correlated reflected Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$:

(30) $$dX_t = \mu_X dt + dw_t^{(1)} + dL_t^X,$$

(31) $$dY_t = \mu_Y dt + dw_t^{(2)} + dL_t^Y,$$

where μ_X, μ_Y are constants, $dw_t^{(1)} \wedge dw_t^{(2)} = \rho dt$, and

$$L_t^X = -\min\{0, \inf_{s \leq t}(X_0 + \mu_X t + w_t^{(1)})\},$$

$$L_t^Y = -\min\{0, \inf_{s \leq t}(Y_0 + \mu_Y t + w_t^{(1)})\}.$$

Let us define for a $h > 0$ that

(32) $$T_h^\rho := \inf\{t \geq 0\ ;\ \max(X_t, Y_t) \geq h\}.$$

Then we have

(33) $$E\{T_h^\rho\} \leq E\{T_h^0\}.$$

Proof: We begin with the generator of the 2-dimensional process $\{(X_t, Y_t)\}_{t \geq 0}$:

$$\mathcal{L}_\rho := \frac{1}{2}\frac{\partial^2}{\partial x^2} + \rho \frac{\partial^2}{\partial x \partial y} + \frac{1}{2}\frac{\partial^2}{\partial y^2} + \mu_X \frac{\partial}{\partial x} + \mu_Y \frac{\partial}{\partial y},$$

where the domain of $\mathcal{L}_\rho$ is

$$\{C^2 \text{ function } g(x,y)\ :\ \frac{\partial g}{\partial x}|_{x=0} = \frac{\partial g}{\partial y}|_{y=0} = 0\}.$$

It is easily seen using Itô's lemma that for the solution to the partial differential equation

$$(\mathcal{L}_\rho g_\rho)(x,y) = -1,\ \forall (x,y) \in [0,h]^2 =: \mathcal{D},$$

with boundary condition (29), we have a martingale.

$$M_t := g_\rho(X_t, Y_t) - g_\rho(x,y) + t.$$

And thus, optional sampling theorem implies that

$$E\{T_h^\rho | X_0 = x, Y_0 = y\} = g_\rho(x,y).$$

To finish the proof, we will show that

$$g_0(x,y) \geq g_\rho(x,y),\ \forall (x,y) \in \mathcal{D}.$$

But then it suffices to prove that $\{\Delta_0(X_t, Y_t)\}_{t \geq 0}$ with $\Delta_0 := g_\rho - g_0$ is a sub-martingale since that will imply

$$g_\rho(x,y) - g_0(x,y) = \Delta_0(x,y) \leq E\{\Delta_0(X_{T_h^\rho}, Y_{T_h^\rho})\} = 0.$$

To this end, we notice that

$$(\mathcal{L}_\rho \Delta_0)(x,y) = -\rho \frac{\partial^2 g_0}{\partial x \partial y}.$$

Using Lemma 4 we have that $\frac{\partial^2 g_0}{\partial x \partial y} \geq 0$. Therefor,e $(\mathcal{L}_\rho \Delta_0) \geq 0$ and $\{\Delta_0(X_t, Y_t)\}_{t \geq 0}$ is a sub-martingale. This completes the proof. ■

Lemma 3: Under the assumption of Lemma 2 and $\mu_X, \mu_Y < 0$, we have

(34) $$E\{T_h^\rho\} \geq E\{T_h^{-1}\}.$$

Proof: To simplify notations, let us begin by defining $P^{x,y}(\cdot) := P(\cdot | X_0 = x, Y_0 = y)$. Then we define

(35) $$g_{-1}(x,y) = E^{x,y}\{T_h^{-1}\},\ \forall (x,y) \in \mathcal{D}.$$

We will bound g_ρ by g_{-1} from below. To this end, let us define

$$\Delta_-(x,y) := g_\rho(x,y) - g_{-1}(x,y).$$

We will show that the process $\{\Delta_-(X_t, Y_t)\}_{t \geq 0}$ is a super-martingale, so that

$$\Delta_-(x,y) \geq E^{x,y}\{\Delta_-(X_{T_h^{-1}}, Y_{T_h^{-1}})\} = 0,\ \forall (x,y) \in \mathcal{D}.$$

The formula for g_{-1} can be derived. Let us denote by

$$\mathcal{D}_- := \{(x,y) \in \mathcal{D}\ :\ x + y \leq k\},\ \mathcal{D}_+ := \mathcal{D} - \mathcal{D}_-.$$

Using Lemma 3.1 of Hadjiliadis et. al. [26], we have that for $(x,y) \in \mathcal{D}_-$,

$$E^{x,y}\{\theta_1\} = E^{x,y}\{T_h^{-1}\} + E^{0,h}\{\theta_1\}P^{x,y}(\theta_1 > \theta_2),$$
$$E^{x,y}\{\theta_2\} = E^{x,y}\{T_h^{-1}\} + E^{h,0}\{\theta_2\}P^{x,y}(\theta_1 < \theta_2),$$

where

$$\theta_1 = \inf\{t \geq 0\ :\ X_t \geq h\},$$
$$\theta_2 = \inf\{t \geq 0\ :\ Y_t \geq h\}.$$

It is easily seen using Itô's lemma that

$$E^{x,y}\{\theta_1\} = \frac{f(-2\mu_X h) - f(-2\mu_X x)}{2\mu_X^2} =: U(x),$$

$$E^{x,y}\{\theta_2\} = \frac{f(-2\mu_Y h) - f(-2\mu_Y y)}{2\mu_X^2} =: V(y),$$

where $f(\nu) = \mathrm{e}^\nu - \nu - 1$. As a result, we have

$$g_{-1}(x,y) = \frac{V(0)U(x) + U(0)V(y) - U(0)V(0)}{U(0) + V(0)},$$

for any $(x,y) \in \mathcal{D}_-$. For any $(x,y) \in \mathcal{D}_+$, let us define

$$\varrho = \inf\{t \geq 0\ :\ X_t = k \text{ or } Y_t = k \text{ or } X_t + Y_t \leq k\}.$$

By strong Markov property, we have that for $x+y>k$,

$$E^{x,y}\{T_h^{-1}\} = E^{x,y}\{\varrho\}+ \\ + \int_0^k P^{x,y}(X_\varrho \in du, X_\varrho + Y_\varrho = k) g_{-1}(u, k-u).$$

The distribution of ρ and X_ρ can be obtained from Anderson [28]. In particular, by letting $\gamma_1 = k-y$, $\gamma_2 = x-k$, $\delta_1 = -\mu_Y$ and $\delta_2 = -\mu_X$ in Theorem 5.1 of [28], we have

$$E^{x,y}\{\varrho\} = \int_0^{\sigma_{x,y}} t\left(\frac{dP_1(t)}{dt} + \frac{dP_2(t)}{dt}\right)dt+ \\ + \sigma_{x,y}[1-P_1(\sigma_{x,y}) - P_2(\sigma_{x,y})],$$

where $\sigma_{x,y} = \frac{x+y-k}{-\mu_X-\mu_Y}$. Similarly, using the same set of parameters as the ones of Theorem 4.2 of [28], we have

$$P^{x,y}(X_\rho \in du, X_\rho + Y_\rho = k) \\ = [1-P_1(\sigma_{x,y}, u-x-\mu_X\sigma_{x,y}) - P_2(\sigma_{x,y}, u-x-\mu_X\sigma_{x,y})] \\ \times \frac{1}{\sqrt{2\pi\sigma_{x,y}}} \mathrm{e}^{-\frac{(u-x-\mu_X\sigma_{x,y})^2}{2\sigma_{x,y}}} du.$$

The exact formulas for densities and probabilities P_1, P_2 used above are complicated, but it can be easily seen that $g_{-1}|_{\mathcal{D}_+}$ is a smooth function.

Let us denote by $F = g_{-1}|_{\mathcal{D}_-}$ and $G = g_{-1}|_{\mathcal{D}_+}$. Then

$$(\mathcal{L}_{-1}F)|_{\mathcal{D}_-} \equiv -1, \ (\mathcal{L}_{-1}G)|_{\mathcal{D}_+} \equiv -1.$$

Moreover, we can trivially write,

$$g_{-1}(x,y) \quad = \quad F(x,y) + [G(x,y) - F(x,y)]\mathbf{1}_{\{x+y>k\}}.$$

Due to exponential decaying of P_1, P_2 as $x+y \searrow k$ and $\sigma_{x,y} \searrow 0$, we have that

$$\frac{\partial G}{\partial x}\Big|_{x+y=k} = \frac{\mu_Y}{\mu_X+\mu_X}\frac{\partial}{\partial x} g_{-1}(x, k-x), \\ \frac{\partial G}{\partial y}\Big|_{x+y=k} = -\frac{\mu_X}{\mu_X+\mu_Y}\frac{\partial}{\partial x} g_{-1}(x, k-x).$$

Therefore, we have for $(x,y) \in \partial\mathcal{D}_- \cap \partial\mathcal{D}_+$,

$$\begin{aligned} H(x,y) \quad &:= \quad \frac{1}{2}\frac{\partial^2 G}{\partial x^2} - \frac{\partial^2 G}{\partial x \partial y} + \frac{1}{2}\frac{\partial^2 G}{\partial y^2} - \frac{1}{2}\frac{\partial^2 F}{\partial x^2} - \frac{1}{2}\frac{\partial^2 F}{\partial y^2} \\ &= \quad -1 + \frac{V(0)\mathrm{e}^{-2\mu_X x} + U(0)\mathrm{e}^{-2\mu_Y y}}{U(0)+V(0)} \geq 0 \end{aligned}$$

Moreover, using exponential decay of P_1, P_2 we can show that

$$\frac{\partial^2 G}{\partial x \partial y}\Big|_{x+y=k} \quad = \quad -\frac{\mu_X \mu_Y}{(\mu_X+\mu_Y)^2}\frac{\partial^2}{\partial x^2} g_{-1}(x, k-x) \geq 0.$$

Using the argument in Lemma 4, we obtain that

$$\frac{\partial^2 G}{\partial x \partial y}\Big|_{x=h}, \frac{\partial^2 G}{\partial x \partial y}\Big|_{y=h}, \quad \geq \quad 0.$$

Because $\frac{\partial^2 G}{\partial x \partial y}$ solves parabolic equation

$$(\mathcal{L}_{-1}g)(x,y) = 0, \ \forall (x,y) \in \mathcal{D}^+,$$

and it has non-negative boundary values, we conclude that

$$\frac{\partial^2 G}{\partial x \partial y} \geq 0, \ \forall (x,y) \in \mathcal{D}^+,$$

by the maximum principle of parabolic equation. Thus, for process $\{(X_t, Y_t)\}_{t\geq 0}$ with correlation function $\rho > -1$, we have

$$\begin{aligned} & dg_{-1}(X_t, Y_t) \\ =& (\mathcal{L}_\rho F)(X_t, Y_t)\mathbf{1}_{\{X_t+Y_t<k\}}dt + (\mathcal{L}_\rho G)(X_t, Y_t)\mathbf{1}_{\{X_t+Y_t>k\}} \\ &+ \left(\frac{1}{2}\frac{\partial^2 G}{\partial x^2} + \rho\frac{\partial^2 G}{\partial x \partial y} + \frac{1}{2}\frac{\partial^2 G}{\partial y^2} - \frac{1}{2}\frac{\partial^2 F}{\partial x^2} - \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\right) dL_t^{X+Y-k} \\ &+ \frac{\partial g_{-1}}{\partial x^-}(X_t, Y_t)dw_t^{(1)} + \frac{\partial g_{-1}}{\partial y^-}(X_t, Y_t)dw_t^{(2)} \\ =& \left(-\mathbf{1}_{\{X_t+Y_t\neq k\}} + \mathbf{1}_{\{X_t+Y_t>k\}}[1+\rho(X_t, Y_t)]\frac{\partial^2 G}{\partial x \partial y}\right)dt \\ &+ \left(H(X_t, Y_t) + [1+\rho(X_t, Y_t)]\frac{\partial^2 G}{\partial x \partial y}(X_t, Y_t)\right) dL_t^{X+Y-k} \\ &+ \frac{\partial g_{-1}}{\partial x^-}(X_t, Y_t)dw_t^{(1)} + \frac{\partial g_{-1}}{\partial y^-}(X_t, Y_t)dw_t^{(2)}, \end{aligned}$$

where $\{L_t^{X+Y-k}\}_{t\geq 0}$ is the local time of $\{X_t + Y_t - k\}_{t\geq 0}$ at zero. It follows that the process $\{\Delta_{-1}(X_t, Y_t)\}_{t\geq 0}$ is a super-martingale. This completes the proof. ■

Lemma 4: For any constant $\rho \in (-1, 1)$ and the solution of the partial differential equation,

$$\mathcal{L}_\rho g = -1, \ \forall (x,y) \in \mathcal{D}, \\ g|_{x=h} = g|_{y=h} = \frac{\partial g}{\partial x}\Big|_{x=0} = \frac{\partial g}{\partial y}\Big|_{y=0} = 0.$$

We have $\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \leq 0$ and $\frac{\partial^2 g}{\partial x \partial y} \geq 0$, for all $(x,y) \in \mathcal{D}$.

Proof: By Feynmann-Kac theorem we can express g as expectation

$$g(x,y) = E\{T_\rho^{1,x} \wedge T_\rho^{2,y}\}, \text{ where}$$

$$\begin{aligned} T_\rho^{1,x} \quad &= \quad \inf\{t \geq 0 \ : \ X_t^x \geq h\}, \ dX_t^x = dX_t, X_0^x = x, \\ T_\rho^{2,y} \quad &= \quad \inf\{t \geq 0 \ : \ Y_t^y \geq h\}, \ dY_t^y = dY_t, Y_0^y = y. \end{aligned}$$

For $x' \in (x, k)$, by strong Markov property we have

$$T_\rho^{1,x} > T_\rho^{2,x'}, \ P\text{-a.s.}$$

Thus we have

$$g(x,y) = E\{T_\rho^{1,x} \wedge T_\rho^{2,y}\} \geq E\{T_\rho^{1,x'} \wedge T_\rho^{2,y}\} = g(x', y).$$

This proves the monotonicity of g.

Moreover, it is easily seen that the solution g is a smooth function on $\mathcal{D}$. Thus its derivative $u := \frac{\partial^2 g}{\partial x \partial y}$ satisfies partial differential equation

$$(\mathcal{L}_\rho u)(x,y) = 0, \ \forall (x,y) \in \mathcal{D}.$$

However, regularity of u on $\overline{\mathcal{D}}$ and monotonicity of g implies that $(u|_{x=k}) \geq 0$ and $(u|_{y=k}) \geq 0$, and Neumann boundary condition implies $(u|_{x=0}) = (u|_{y=0}) = 0$. By the maximum principle of elliptic equations, we conclude that

$$\frac{\partial^2 g}{\partial x \partial y} = u(x,y) \geq 0, \ \forall (x,y) \in \mathcal{D}.$$

This completes the proof. ■