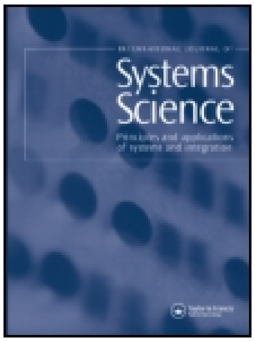

International Journal of Systems Science

ISSN: 0020-7721 (Print) 1464-5319 (Online) Journal homepage: https://www.tandfonline.com/loi/tsys20

Optimal input design for the detection of changes towards unknown hypotheses

F. Kerestecioğlu & İ. Çetin

To cite this article: F. Kerestecioğlu & İ. Çetin (2004) Optimal input design for the detection of changes towards unknown hypotheses, International Journal of Systems Science, 35:7, 435-444, DOI: 10.1080/00207720410001734219

To link to this article: https://doi.org/10.1080/00207720410001734219

Published online: 06 Oct 2011.

International Journal of Systems Science
volume 35, number 7, 15 June 2004, pages 435–444

Optimal input design for the detection of changes towards unknown hypotheses

F. Kerestecioğlu†* and İ. Çetin‡

The effects of auxiliary input signals on detecting changes in ARMAX processes via statistical tests are discussed. Two extensions to the Cumulative Sum Test are considered. The first is applicable when the direction of the change in the parameter space is known but its magnitude is unknown. The second is applicable when neither is known. The performance criteria for the design of stationary stochastic inputs are based on the asymptotic properties of the tests. It is shown that power-constrained optimal inputs have discrete spectra and a suitably chosen input can greatly improve the detection performance.

1. Introduction

Detection of abrupt changes is of crucial importance in the context of fault detection, industrial maintenance, quality control and safety of complex engineering systems as well as analysis of natural catastrophic events (earthquakes, etc.). As a result of this motivation, it has become one of the important research areas during the last two decades (e.g. Basseville and Nikiforov 1993, Kerestecioğlu 1993, Patton *et al*. 1989, 2000, and references therein).

In many practical situations, detection of the change should be performed on-line. A direct result of this is the sequential nature of the decision-making in change detection. The main objective is to detect the change as soon as possible after it has occurred. The other important point is to avoid false alarms as long as there is no change in the system or signal under change monitoring. Since one should seek a tradeoff between these two objectives, most detection mechanisms try to optimize one of these criteria while guaranteeing an acceptable specified level on the other.

Received 10 January 2001. Revised 14 May 2004. Accepted 10 June 2004.

† Department of Electronics Engineering, Kadir Has University, Cibali, Istanbul 34230, Turkey.

‡ Garanti Technology Network Services, Koçman Cad., No: 22, Güneşli, Istanbul 34555, Turkey.

* To whom correspondence should be addressed.
e-mail: kerestec@khas.edu.tr

In most cases, the decision-making part of a change monitoring system involves statistical tests since the data obtained from the monitored system are usually corrupted by noise and other disturbances that can be modelled statistically. A well-known statistical decision method used in change detection is the Cumulative Sum (CUSUM) Test (Basseville and Nikiforov 1993, Kerestecioğlu 1993). Although it was originally designed to detect a change from a known operating mode to another known one, it is possible to modify it for cases where possible operating modes after the change are unknown or partially known (Nikiforov 1980, 1986, Nikiforov and Tikhonov 1986).

The main objectives in deriving auxiliary signals for change detection purposes are inherited from the basic goals of the statistical change detection problem: namely, one is looking to improve the detection delay while keeping an acceptable level of false alarms. One of the main restrictions on such input signals can be on their magnitudes or their average power in order to ensure that they do not disturb the operating conditions of the system, which are usually maintained by other control signals. Also, they may be required to be of zero mean so that no biases are introduced to the system. Further, their spectral densities need to be constrained. Note that such inputs or perturbations are going to determine the statistical properties of the data gathered for detecting the relevant changes. From this point of view, the input design problem can be seen as a hypothesis-generation problem. That is, subject to the dynamics of the system at hand, the statistical hypotheses should be manipulated so that a desired tradeoff

International Journal of Systems Science ISSN 0020–7721 print/ISSN 1464–5319 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00207720410001734219

between the detection delay and false alarm rate is obtained.

Although the design of optimal inputs has been extensively investigated in the system identification context (e.g. Goodwin and Payne 1977, Zarrop 1979, Kalaba and Springarn 1982), there has been only a small number of works on the input design for change detection purposes (Zhang 1989, Kerestecioğlu 1993, Kerestecioğlu and Zarrop 1994). Zhang has discussed both offline designs and online algorithms to generate input signals for accelerating the detection. The design techniques introduced by Kerestecioğlu, on the other hand, were aimed not only to facilitate fast detection, but also to assure tolerable false alarm rates. In all these works, the spectrum of an optimal input signal has been shown to be discrete. Also, extensions have been presented for multi-hypothesis detection. Nevertheless, the hypotheses after the change, as well as before it, were assumed to be known. In most practical cases, the no change hypothesis describes a normal (or nominal) mode of operation. Hence, it is either known a priori or can be obtained by system identification techniques. But in many applications, it may not be possible to characterize precisely the change mode. The exact magnitude of the change may be unknown even if the changing parameters or the direction of the change in the parameter space is known. In some other cases, the hypothesis describing the after-change mode may be completely unknown. The paper aims to derive optimal off-line input signals to improve performances of modified CUSUM algorithms for detecting changes towards unknown or partially known hypotheses.

In Section 2, a brief description of CUSUM test is given and two extensions to it are presented. The first is for the case where the direction of the change in the parameter space is known, but the magnitude of the change is unknown. The other is for the case when information on the change direction is also absent. In Section 3, asymptotically optimal inputs for improving the detection performance of these modified CUSUM tests are derived. Section 4 is devoted to simulation examples that show that substantial improvements in detection performance can be obtained by a proper choice of the input signal. Some conclusions are drawn in Section 5.

2. Extensions to the CUSUM test using a local approach

2.1. *CUSUM test*

The CUSUM test, which was originally proposed by Page (1954), is an efficient sequential method to detect changes from a known operating mode (or hypothesis, say, $\mathcal{H}_0$) to another one ($\mathcal{H}_1$). It is conducted by computing the statistics:

$$g(k) = \max[0, g(k-1) + z(k)]$$

and a change is declared as soon as $g(k)$ exceeds a predetermined positive threshold β. That means the alarm time is given as

$$n = \inf\{k : g(k) \geq \beta\}, \tag{1}$$

where $z(k)$ is the conditional log likelihood ratio of the current data $y(k)$ obtained from the process (Kerestecioğlu 1993).

We are interested in detecting a change in the dynamics of an autoregressive moving average process with exogenous input (ARMAX) given as:

$$A(q^{-1})y(k) = q^{-d}B(q^{-1})u(k) + C(q^{-1})\epsilon(k), \tag{2}$$

where $u(k)$ is an auxiliary input and q^{-1} is the backwards shift operator. It is assumed that $A(z^{-1})$ and $C(z^{-1})$ polynomials are monic and have all their zeros inside the unit circle, and $d > 0$. Further, $\epsilon(k)$ is Gaussian white noise with zero mean and variance σ^2. Note that $y(t)$ can be the output of the system under change monitoring. Nonetheless, it can also be a residual sequence generated for monitoring purposes by filtering or processing the actual outputs of the system.

In this case, the hypotheses concern the coefficients of the polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$, namely,

$$\boldsymbol{\theta} = [a_1, \ldots, a_{n_a}, b_0, \ldots, b_{n_b}, c_1, \ldots, c_{n_c}]^{\mathrm{T}}.$$

We shall denote the parameter vectors before and after the change as $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, respectively. Using (2) and the Gaussianity of the $e(k)$, it can be shown (Kerestecioğlu 1993) that the increments of the cumulative sum are computed as:

$$z(k) = \frac{1}{2\sigma^2}\left[e_0^2(k) - e_1^2(k)\right],$$

where $e_i(k)$ $(i = 0, 1)$ is the prediction error of the one-step-ahead output predictor based on the hypothesis $\mathcal{H}_i$.

It is possible to extend the cumulative sum method to the cases where the hypothesis after the change is unknown or partially known. Two such extensions have been introduced by Nikiforov (1980, 1986), which we shall briefly describe below. For detailed analyses of them, see Nikiforov (1986) and Basseville and Nikiforov (1993).

2.2. *Detecting changes of unknown magnitude*

First, we assume that the direction of the change in the parameter space is known and the magnitude is unknown. Namely, the parameters are described as $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \lambda\mathbf{d}$, where λ is the magnitude of change and

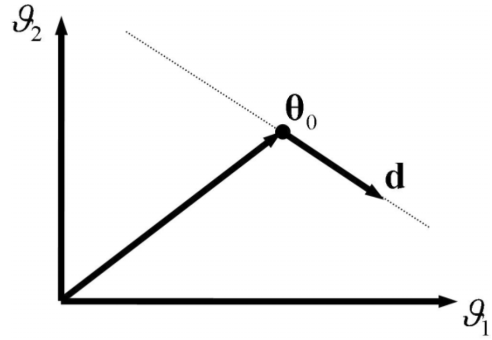

Figure 1. Change with known direction and unknown magnitude.

$$\mathbf{d} = [d_1 \quad d_2 \quad \dots \quad d_r]^{\mathrm{T}} \tag{3}$$

is the direction of change with $||\mathbf{d}|| = 1$ and $r = n_a + n_b + n_c + 1$. This case is depicted in figure 1 for a two-parameter case. The hypotheses are then given as:

$$\mathcal{H}_0: \quad \lambda \leq 0, \qquad \mathcal{H}_1: \quad \lambda > 0. \tag{4}$$

This extension of the Cumulative Sum algorithm is based on the theory of Le Cam about asymptotic expansion of the log-likelihood ratio between the hypotheses $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \lambda \mathbf{d}/\sqrt{k}$ (Ibragimov and Khasminsky 1981, Le Cam 1986). For the case of small λ, the CUSUM statistic has the form of (Nikiforov 1986):

$$g(k) = \max[0, \mathbf{d}^{\mathrm{T}} \tilde{\mathbf{z}}(\boldsymbol{\theta}_0)], \tag{5}$$

where

$$\tilde{\mathbf{z}}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_\theta(y(k), \dots, y(k - n(k) + 1)) \,|\, \mathbf{y}(k - n(k))$$

is the vector of asymptotically sufficient statistics for the observations obtained since the last resetting applied in (5) and $\mathbf{y}(k) = [y(1), \dots, y(k)]^{\mathrm{T}}$ is the observations vector. The counter $n(k)$ indicates the number of samples taken since this last resetting and is computed by formula:

$$n(k) = \begin{cases} 1 & \text{if } g(k-1) \leq 0 \\ n(k-1) + 1 & \text{if } g(k-1) > 0 \end{cases}.$$

The alarm time n is given as in (1). Note that also in this case, $g(k)$ can be written in a recursive way as:

$$g(k) = \max\big[0, g(k-1) + \mathbf{d}^{\mathrm{T}} \tilde{\mathbf{z}}(\boldsymbol{\theta}_0)\big].$$

2.3. *Detecting changes of unknown magnitude and direction*

As a second extension of CUSUM test we consider the case where both the magnitude and the direction of the change in the parameter space is unknown. The hypotheses in such a case are described as:

$$\begin{aligned} &\mathcal{H}_0: \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ &\mathcal{H}_1: \quad (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\mathrm{T}} \mathbf{F}_1(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \geq \lambda_1^2, \end{aligned} \tag{6}$$

where $\mathbf{F}_1(\boldsymbol{\theta}_0)$ is the Fisher Information Matrix for one sample and is given as:

$$\mathbf{F}_1(\boldsymbol{\theta}) = E\{\mathbf{z}(k, \boldsymbol{\theta}) \mathbf{z}^{\mathrm{T}}(k, \boldsymbol{\theta}) \,|\, \boldsymbol{\theta}\} \tag{7}$$

with

$$\mathbf{z}(k, \boldsymbol{\theta}) = \frac{\partial \ln f_\theta(y(k) \,|\, \mathbf{y}(k-1))}{\partial \boldsymbol{\theta}}. \tag{8}$$

Namely, a change needs to be declared as soon as the parameters drift outside an ellipsoid defined by the Fisher Information Matrix and the nominal parameters.

It has been shown by Nikiforov (1986) that:

$$x(k, \boldsymbol{\theta}) = \frac{1}{n(k)} \left(\sum_{i=k-n(k)+1}^{k} \mathbf{z}(k, \boldsymbol{\theta}) \right)^{\mathrm{T}} \mathbf{F}_1(\boldsymbol{\theta}) \left(\sum_{i=k-n(k)+1}^{k} \mathbf{z}(k, \boldsymbol{\theta}) \right)$$

turns out to be a sufficient statistic for this case and the decision function is obtained as:

$$g(k) = \max[0, \tilde{S}(\boldsymbol{\theta}_0)], \tag{9}$$

where

$$\tilde{S}(\boldsymbol{\theta}) = -\frac{1}{2} \lambda_1^2 n(k) + \ln G\left(\frac{r}{2}, \frac{\lambda_1^2 n(k) x(k, \boldsymbol{\theta})}{4} \right),$$

with $r = \dim(\boldsymbol{\theta}) = n_a + n_b + n_c + 1$ and

$$G(a, x) = \sum_{i=0}^{\infty} \frac{x^i}{a(a+1) \cdots (a+i-1) i!}$$

being the generalized hypergeometric function. Note that in the detection of changes in ARMAX parameters, unlike the other version of the CUSUM test mentioned above, the test statistics in (9) cannot be obtained recursively. Some methods to make the computation of $g(k)$ feasible have been introduced by Nikiforov (1986) and Nikiforov and Tikhonov (1986).

3. Input design

This section aims to investigate the effects of auxiliary inputs in detecting changes in the dynamics of ARMAX processes with the modified CUSUM algorithms mentioned above. In selecting input signals to improve the detection performance one should consider improving the average detection delay (ADD) as well as keeping the mean time between false alarms (MTBFA) at a tolerable level. These quantities and, hence, the performance of a statistical test for change detection are determined by the average run length (ARL) function of it; namely, $E\{n \mid \boldsymbol{\theta}\}$, with n as defined in (1). Note that, assuming that the test statistic is close to zero when the change occurs, for values of $\boldsymbol{\theta}$ belonging to the set describing $\mathcal{H}_1$, $E\{n \mid \boldsymbol{\theta}\}$ gives the ADD. On the other hand, $E\{n \mid \boldsymbol{\theta}_0\}$ is the mean time between false alarms.

An asymptotic relation between these two criteria of performance for independently and identically distributed observations is given by Lorden (1971) as:

$$E\{n \mid \boldsymbol{\theta}\} \sim \frac{\ln E\{n \mid \boldsymbol{\theta}_0\}}{K(\boldsymbol{\theta}, \boldsymbol{\theta}_0)} \qquad \text{when } E\{n \mid \boldsymbol{\theta}\} \to \infty, \tag{10}$$

where

$$K(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \int \ln \frac{f_{\theta}(y)}{f_{\theta_0}(y)} \, \mathrm{d}y$$

and denotes the Kullback information between the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. This result is, in fact, also shown to hold for the modified CUSUM tests, where the hypothesis after the change is partially or completely unknown, by Basseville and Nikiforov (1993). It is also extended to the correlated observations case by Lai (1998).

This suggests that to improve the test performance, the inputs should be chosen so as to maximize the Kullback information. Inputs with such a property are going to be denoted as asymptotically optimal in the sequel. Also, note that we restrict ourselves to stochastic stationary inputs, which are generated off-line, i.e. are independent of the past data gathered from the system. To have a well-posed input design problem, it is natural to assume that the input power is constrained in the sense that

$$\frac{1}{\pi} \int_0^{\pi} \mathrm{d}\xi(\omega) \leq K_u, \tag{11}$$

where $\xi(\omega)$ ($\omega \in [0, \pi]$) is the one-sided power spectral distribution of the input and K_u is the maximum allowable input power. Further, the input spectrum might be required to be limited to a predefined frequency region, say Ω. For example, constant or very low frequency inputs might not be desirable since they can introduce biases in the output.

3.1. *Detecting changes of unknown magnitude*

The Kullback information between the hypotheses in (4) can be shown to be (Basseville and Nikiforov 1993):

$$K(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0) \approx \frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\mathrm{T}} \mathbf{F}_1(\boldsymbol{\theta}_0)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0) \tag{12}$$

for the cases where the difference between the parameter vectors describing the hypotheses before and after the change is small. Further, since $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 + \lambda \mathbf{d}$, it follows that:

$$K(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0) = \frac{\lambda^2}{2} \mathbf{d}^{\mathrm{T}} \mathbf{F}_1(\boldsymbol{\theta}_0) \mathbf{d}. \tag{13}$$

This suggests that asymptotically optimal auxiliary input signals should be chosen so as to maximize $\mathbf{d}^{\mathrm{T}} \mathbf{F}_1(\boldsymbol{\theta}_0) \mathbf{d}$. As shown by Nikiforov (1986), $\mathbf{d}^{\mathrm{T}} \mathbf{F}_1(\boldsymbol{\theta}_0) \mathbf{d}$ also determines the slope of the ARL function at $\lambda = 0$. To gain more insight on the choice of (13) as the cost function for input optimization, let us consider the ideal values of ARL for this modified CUSUM test which are depicted in figure 2 for a scalar-parameter case. For the hypotheses in (4), the ideal values for the ARL function for $\lambda \leq 0$ and $\lambda > 0$ are infinity and unity, respectively. In other words, ideally speaking, the change is to be detected as soon as it occurs and false alarms should be avoided forever. Therefore, better discrimination between the hypotheses is achieved as the magnitude of the slope of the ARL curve at $\lambda = 0$ is increased.

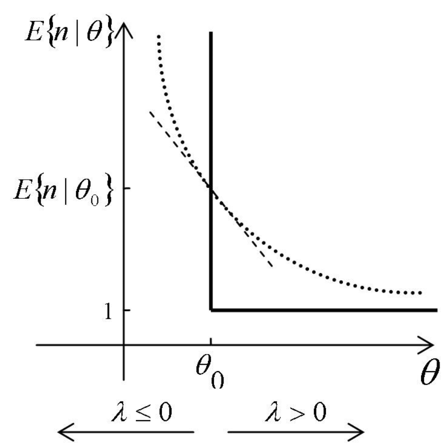

Figure 2. Real (dotted) and ideal (solid) ARL functions for a scalar parameter.

Optimal inputs in the above sense are given in the following theorem.

Theorem 1: *Asymptotically optimal power-constrained offline stationary input signals for the CUSUM test defined by* (5) *and* (1) *consist of a single frequency and are given as*:

$$\begin{aligned} u(k) &= \operatorname{sign}(\varphi)\sqrt{K_u}\cos(\omega^* k) \qquad \text{if } \omega^* = 0 \text{ or } \pi \\ u(k) &= \sqrt{2K_u}\cos(\omega^* k + \varphi) \qquad \text{if } \omega^* \in (0, \pi), \end{aligned} \tag{14}$$

where

$$\omega^* = \arg\max_{\omega\in\Omega} \left| \frac{A(e^{j\omega})D_B(e^{j\omega}) - B(e^{j\omega})D_A(e^{j\omega})}{A(e^{j\omega})C(e^{j\omega})} \right|^2_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

with

$$\begin{aligned} D_A(q^{-1}) &= d_1 q^{-1} + d_2 q^{-2} + \cdots + d_{n_a} q^{-n_a} \\ D_B(q^{-1}) &= d_{n_a+1} + d_{n_a+2} q^{-1} + \cdots + d_{n_a+n_b+1} q^{-n_b} \end{aligned}$$

and φ *is uniformly distributed in* $[-\pi, \pi]$.

Proof: To optimize the Kullback information, first note that the conditional distribution of a single observation obtained from the process (2) can be written as:

$$f_\theta(y(k) \mid \boldsymbol{\phi}(k-1)) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}\epsilon^2(k)\right\}, \tag{15}$$

where $\boldsymbol{\phi}(k-1) = [y(k-1), \ldots, u(k-d), \ldots]^{\mathrm{T}}$ is the vector containing all the data available at time $k-1$.

From (15) and (8), it follows that

$$\mathbf{z}(k, \boldsymbol{\theta}) = -\frac{\epsilon(k)}{\sigma^2} \cdot \frac{\partial\epsilon(k)}{\partial\boldsymbol{\theta}}. \tag{16}$$

The partial (sensitivity) derivatives of $\epsilon(k)$ with respect to the parameters can be obtained from (2) as

$$\begin{aligned} \frac{\partial\epsilon(k)}{\partial a_i} &= \frac{B(q^{-1})}{A(q^{-1})C(q^{-1})} u(k-d-i) \\ &\quad + \frac{1}{A(q^{-1})}\epsilon(k-i) \qquad i = 1, \ldots, n_a \end{aligned} \tag{17}$$

$$\frac{\partial\epsilon(k)}{\partial b_i} = -\frac{1}{C(q^{-1})} u(k-d-i) \qquad i = 0, \ldots, n_b \tag{18}$$

$$\frac{\partial\epsilon(k)}{\partial c_i} = -\frac{1}{C(q^{-1})}\epsilon(k-i) \qquad i = 1, \ldots, n_c. \tag{19}$$

The partial derivatives in (17–19) can be substituted into (16), to rewrite $\mathbf{z}(k, \boldsymbol{\theta})$ as

$$\mathbf{z}(k, \boldsymbol{\theta}) = \frac{\epsilon(k)}{\sigma^2}\big(\mathbf{p}_u + \mathbf{p}_\epsilon)\big), \tag{20}$$

where

$$\mathbf{p}_u = \begin{bmatrix} -\dfrac{B(q^{-1})}{A(q^{-1})C(q^{-1})}\mathbf{u}_{1,n_a} \\ \dfrac{1}{C(q^{-1})}\mathbf{u}_{0,n_b} \\ \mathbf{0}_{n_c} \end{bmatrix} \tag{21}$$

and

$$\mathbf{p}_\epsilon = \begin{bmatrix} -\dfrac{1}{A(q^{-1})}\mathbf{e}_{1,n_a} \\ \mathbf{0}_{n_b+1} \\ \dfrac{1}{C(q^{-1})}\mathbf{e}_{1,n_c} \end{bmatrix},$$

where $\mathbf{u}_{i,j}$ and $\mathbf{e}_{i,j}$ are the vectors composed of relevant recent samples of the input and the innovations, respectively. That is,

$$\begin{aligned} \mathbf{u}_{i,j} &= [u(k-d-i), \ldots, u(k-d-j)]^{\mathrm{T}}, \\ \mathbf{e}_{i,j} &= [\epsilon(k-i), \ldots, \epsilon(k-j)]^{\mathrm{T}} \end{aligned}$$

and $\mathbf{0}_i$ denotes an i-dimensional zero vector.

Therefore, from (7) and (20) it follows that

$$\begin{aligned} \mathbf{d}^{\mathrm{T}}\mathbf{F}_1(\boldsymbol{\theta})\mathbf{d} &= \mathbf{d}^{\mathrm{T}} E\left\{\frac{\epsilon(k)}{\sigma^2}(\mathbf{p}_u + \mathbf{p}_\epsilon)(\mathbf{p}_u + \mathbf{p}_\epsilon)^{\mathrm{T}} \;\Big|\; \boldsymbol{\theta}\right\}\mathbf{d} \\ &= \frac{1}{\sigma^2}\mathbf{d}^{\mathrm{T}} E\left\{(\mathbf{p}_u + \mathbf{p}_\epsilon)(\mathbf{p}_u + \mathbf{p}_\epsilon)^{\mathrm{T}} \;\Big|\; \boldsymbol{\theta}\right\}\mathbf{d}. \end{aligned}$$

Since the auxiliary stochastic input $u(k)$ and $\epsilon(l)$ are statistically independent, so are $\mathbf{p}_u$ and $\mathbf{p}_\epsilon$. Therefore, we have

$$E\big\{(\mathbf{p}_u + \mathbf{p}_\epsilon)(\mathbf{p}_u + \mathbf{p}_\epsilon)^{\mathrm{T}} \mid \boldsymbol{\theta}\big\} = E\big\{\mathbf{p}_u\mathbf{p}_u^{\mathrm{T}} \mid \boldsymbol{\theta}\big\} + E\big\{\mathbf{p}_\epsilon\mathbf{p}_\epsilon^{\mathrm{T}} \mid \boldsymbol{\theta}\big\}.$$

Hence, the input signal should be chosen so as to maximize

$$J = E\big\{\mathbf{d}^{\mathrm{T}}\mathbf{p}_u\mathbf{p}_u\mathbf{d}^{\mathrm{T}} \mid \boldsymbol{\theta} = \boldsymbol{\theta}_0\big\}. \tag{22}$$

In view of (21) and (3), (22) can be written as

$$J = E\left\{\left[\left(\frac{D_B(q^{-1})}{C(q^{-1})} - \frac{B(q^{-1})D_A(q^{-1})}{A(q^{-1})C(q^{-1})}\right)u(k)\right]^2 \;\Big|\; \boldsymbol{\theta} = \boldsymbol{\theta}_0\right\}. \tag{23}$$

The cost function in (23) can be expressed in frequency domain as

$$J = \frac{1}{\pi}\int_0^{\pi} |R(e^{j\omega})|^2 \,\mathrm{d}\xi(\omega), \tag{24}$$

where

$$R(e^{j\omega}) = \left.\frac{A(e^{j\omega})D_B(e^{j\omega}) - B(e^{j\omega})D_A(e^{j\omega})}{A(e^{j\omega})C(e^{j\omega})}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Under the power constraint given in (11), one concludes from (24) that all input power should be concentrated at the frequency

$$\omega^* = \arg\max_{\omega\in\Omega} |R(e^{j\omega})|^2.$$

Hence, the optimal stationary inputs can be generated as in (14). □

It is interesting to note that (23) does not contain any information about the changes in the coefficients of $C(q^{-1})$. Therefore, the test performance cannot be affected by the input signal if the change is expected only in the $C(q^{-1})$ polynomial.

3.2. *Detecting changes of unknown magnitude and direction*

When the change direction is not specified, the Kullback information cannot be represented as a multiple of the change magnitude, i.e. as in (13), any more. In view of (10) and (12), an asymptotically optimal input should maximize a suitable scalar function of $F_1(\boldsymbol{\theta}_0)$. Different choices are possible for such a scalar function, such as the determinant, trace, largest eigenvalue, etc. Note that all such optimizations aim to contract the ellipsoid describing the possible parameters after the change, and hence are expected to improve the detection performance. In particular, optimizing $\det F_1(\boldsymbol{\theta}_0)$ would minimize the volume of the ellipsoid defining $\mathcal{H}_1$ hypothesis, if an a priori Gaussian distribution is assumed for the parameter vector (Cramer 1946). On the other hand, maximizing the largest eigenvalue of the Information Matrix would minimize the length along its largest axis. We shall adopt the largest-determinant criterion, that is, maximize $\det F_1(\boldsymbol{\theta}_0)$. In fact, optimization of the Fisher Information Matrix in the context of parameter estimation has been treated in detail by Goodwin and Payne (1977). The analysis below will follow their work closely.

Let us rewrite the ARMAX process (2) as

$$y(k) = G_u(q^{-1})u(k-d) + G_\epsilon(q^{-1})\epsilon(k),$$

where

$$G_u(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})} \qquad G_\epsilon(q^{-1}) = \frac{C(q^{-1})}{A(q^{-1})}.$$

Note that

$$\begin{aligned}\frac{\partial\epsilon(k)}{\partial\boldsymbol{\theta}} = &-G_\epsilon^{-1}(q^{-1})\frac{\partial G_\epsilon(q^{-1})}{\partial\boldsymbol{\theta}}\epsilon(k)\\ &- G_\epsilon^{-1}(q^{-1})\frac{\partial G_u(q^{-1})}{\partial\boldsymbol{\theta}}u(k-d).\end{aligned} \tag{25}$$

Therefore, since the innovations and the off-line input are statistically independent, using (7), (16), (25) and the stationarity of $u(k)$, the Information Matrix can be written as

$$\mathbf{F}_1(\boldsymbol{\theta}) = \mathbf{F}_u(\boldsymbol{\theta}) + \mathbf{F}_\epsilon(\boldsymbol{\theta}), \tag{26}$$

where

$$\mathbf{F}_u(\boldsymbol{\theta}) = \frac{1}{\sigma^2}E\left\{\left(G_\epsilon^{-1}(q^{-1})\frac{\partial G_u(q^{-1})}{\partial\boldsymbol{\theta}}u(k)\right) \times\left(G_\epsilon^{-1}(q^{-1})\frac{\partial G_u(q^{-1})}{\partial\boldsymbol{\theta}}u(k)\right)^{\mathrm{T}}\middle|\,\boldsymbol{\theta}\right\}$$

$$\mathbf{F}_\epsilon(\boldsymbol{\theta}) = \frac{1}{\sigma^2}E\left\{\left(G_\epsilon^{-1}(q^{-1})\frac{\partial G_\epsilon(q^{-1})}{\partial\boldsymbol{\theta}}\epsilon(k)\right) \times\left(G_\epsilon^{-1}(q^{-1})\frac{\partial G_\epsilon(q^{-1})}{\partial\boldsymbol{\theta}}\epsilon(k)\right)^{\mathrm{T}}\middle|\,\boldsymbol{\theta}\right\}.$$

Hence, the problem of finding asymptotically optimal inputs under a power constraint can be cast in the frequency domain as

$$\begin{aligned}&\text{maximize} \quad \det\left[\frac{1}{\pi}\int_0^{\pi}\tilde{\mathbf{F}}\,\mathrm{d}\xi(\omega) + \mathbf{F}_\epsilon\right]\\ &\text{subject to} \quad \frac{1}{\pi}\int_0^{\pi}\mathrm{d}\xi(\omega) \le K_u,\end{aligned} \tag{27}$$

where

$$\tilde{\mathbf{F}}(\omega) = \left.\mathrm{Re}\left\{\frac{1}{\sigma^2}\left|G_\epsilon^{-1}(e^{j\omega})\right|^2\frac{\partial G_u(e^{j\omega})}{\partial\boldsymbol{\theta}}\left(\frac{\partial G_u(e^{-j\omega})}{\partial\boldsymbol{\theta}}\right)^{\mathrm{T}}\right\}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Following Goodwin and Payne, it is straightforward to show that

$$\tilde{\mathbf{F}}(\omega) = \frac{1}{\sigma^2}\sum_{k=1}^{n_a+n_b+1}\frac{\cos((k-1)\omega)}{|C(e^{j\omega})|^2|A(e^{j\omega})|^2}\mathbf{T}\boldsymbol{\Omega}_k\mathbf{T}^{\mathrm{T}}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad (28)$$

where $\boldsymbol{\Omega}_k$ is a matrix with the (i,j)-th element $\delta_{|i-j|-k+1}$, δ being the Kronecker delta, and

$$\mathbf{T} = \begin{bmatrix} 1 & a_1 & \cdots & a_{n_a} & 0 & \cdots & 0 \\ 0 & 1 & a_1 & \cdots & a_{n_a} & & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & 0 \\ 0 & & & 1 & a_1 & \cdots & a_{n_a} \\ 0 & -b_0 & \cdots & \cdots & -b_{n_b} & & 0 \\ \vdots & & \ddots & & & \ddots & \vdots \\ 0 & \cdots & 0 & -b_0 & \cdots & \cdots & -b_{n_b} \end{bmatrix}.$$

From (26–28), it follows that $\det \mathbf{F}_1(\boldsymbol{\theta}_0)$ is an (n_a+n_b+1)-dimensional variety and the following theorem follows by applying Caratheodory's theorem (Rockafellar 1970).

Theorem 2 (Goodwin and Payne 1977): *An optimal power-constrained input maximizing* $\det \mathbf{F}_1(\boldsymbol{\theta}_0)$ *exists comprising not more than* (n_a+n_b+1) *frequencies.*

Note that the proof given by Goodwin and Payne assumes that $n_a = n_b + 1$; nevertheless, a generalization to the case where one has arbitrary degrees for $A(q^{-1})$ and $B(q^{-1})$ is straightforward.

The above theorem is quite useful in determining an optimal spectrum for the input signal. It reduces the search for the optimal spectrum to a search over n_a+n_b+1 frequencies and, in view of the power constraint in (27), n_a+n_b magnitudes corresponding to these frequencies. In other words, an unconstrained search has to be done in a $(2(n_a+n_b)+1)$-dimensional space. With the optimal frequencies (ω_i, $i=1,\ldots,n_a+n_b+1$) and the optimal powers at these frequencies (p_i, $i=1,\ldots,n_a+n_b+1$) at hand, the optimal input is generated as

$$u(k) = \sum_{i=1}^{n_a+n_b+1} u_i(k),$$

where

$$u_i(k) = \begin{cases} \sqrt{p_i}\ \mathrm{sign}(\varphi_i)\cos(\omega_i^* k) & \text{for } \omega_i^* = 0 \text{ or } \pi \\ \sqrt{2p_i}\ \cos(\omega_i^* k + \varphi_i) & \text{for } \omega_i^* \in (0,\pi) \end{cases}$$

and φ_i's are random variables uniformly distributed in $[-\pi,\pi]$.

4. Simulation examples

This section presents two examples to demonstrate the effect of suitably chosen inputs on the detection performance of modified CUSUM tests. Monte Carlo simulations have been used to estimate the ADD and MTBFA by taking the means obtained from 500 runs for each case. To estimate ADD the data are generated according to the $\mathcal{H}_0$ hypothesis up to $k=50$, which is the instant when the change occurs. Both the maximum allowable input power and the noise variance are taken as unity.

4.1. *Example 1*

The process is assumed to be operating under the normal mode as

$$\begin{aligned}(1 - 0.4q^{-1} + 0.6q^{-2} + 0.3q^{-3})y(k) \\ = (1+0.9q^{-1})u(k) + (1+0.2q^{-1} - 0.15q^{-2})\epsilon(k).\end{aligned}$$

The change direction is specified by

$$\mathbf{d} = [0.318 \quad 0.106 \quad -0.423 \quad 0 \quad -0.318 \quad 0.741 \quad 0.243]^{\mathrm{T}},$$

and the change magnitude is unknown. The system is simulated after the change as

$$\begin{aligned}(1 - 0.1q^{-1} + 0.7q^{-2} - 0.1q^{-3})y(k) \\ = (1+0.6q^{-1})u(k) + (1+0.9q^{-1} + 0.08q^{-2})\epsilon(k),\end{aligned}$$

which correspond to a change with $\lambda = 0.945$. The optimal input frequency can be found by a search over the $|R(e^{j\omega})|^2$, which is plotted in figure 3, as $\omega^* = 1.155$. Therefore, the optimal input is chosen as

$$u(k) = \sqrt{2}\cos(1.155k + \varphi).$$

Also note that the worst possible single frequency in this case turns out to be $\omega = 2.388$, which gives the minimum of $|R(e^{j\omega})|^2$.

Table 1 presents the estimates of ADD and MTBFA, which have been generated by selecting the test threshold as $\beta = 200$. It is seen that any input can improve the detection delay as compared with no-input case. Nevertheless, the price paid for this improvement is a degradation in the MTBFA. So, the choice of the input signal should be so that the best ADD versus MTBFA tradeoff is achieved.

To facilitate a fair comparison among different types of input, simulations are repeated with thresholds chosen separately for each type of input so as to obtain similar MTBFAs. Different input schemes can then be compared for their detection performance. From the results shown in table 2, the optimal input, which

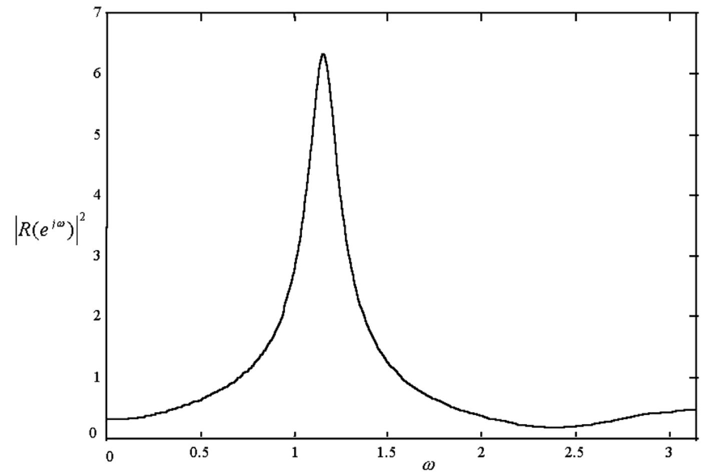

Figure 3. Function $|R(e^{j\omega})|^2$ in example 1.

Table 1. Estimates of ADD and MTBFA in example 1

	ADD	MTBFA
No input	171	26.3×10^3
White input	84	9.9×10^3
Optimal input	13	4.1×10^3

Table 2. Estimates of ADD and MTBFA for different test thresholds in example 1

	β	ADD	MTBFA
No input	200	171	26.3×10^3
Worst input	200	166	25.5×10^3
White input	275	118	25.5×10^3
Optimal input	350	47	27.2×10^3

delivered the fastest ADD and yet the longest average false alarm time, achieved a far better tradeoff in the above sense as compared with other types of inputs. A white noise input also gives some improvement in ADD, but not as much as the optimal one. Note that if the input is generated with the frequency, which minimizes $|R(e^{j\omega})|^2$, there is only a marginal reduction in the ADD, even much less than that obtained by the white input. This fact emphasizes the relevance of a proper choice for the input frequency.

On the other hand, figure 4 depicts the behaviour of ARL around $\lambda = 0$. The roll-off of the ARL curve for the optimal-input case is steeper than that corresponding to the no-input case. This means that the optimal input achieves a better discrimination between $\mathcal{H}_0$ and $\mathcal{H}_1$, and, hence, improves the detection performance.

4.2. *Example 2*

To demonstrate the effect of auxiliary inputs on the performance of the modified CUSUM test defined by (9) and (1), let us consider the following normal operating mode for an ARMA process, where the output is corrupted by white noise

$$y(k) = \frac{0.8}{1 - 0.3q^{-1}}\, u(k-1) + \epsilon(k)$$

and the dynamics after the change is unknown. A change is to be declared as soon as possible after the system dynamics switches to $\mathcal{H}_1$ specified by (6) with $\lambda_1 = 0.4$. In simulation, the process has been changed to

$$y(k) = \frac{0.5}{1 - 0.7q^{-1}}\, u(k-1) + \epsilon(k).$$

By Theorem 2, a power-constrained optimal input can be generated using at most two frequencies. In fact, a numerical search over two frequencies and the input power at these frequencies yield the result that the optimal input consists of only one frequency in this case. An optimal input turns out to be

$$u(k) = \sqrt{2}\cos(0.685k + \varphi), \tag{29}$$

where φ is a uniformly-distributed random phase.

A comparison of the test performances of the input in (29) and a white-noise input can be made in view of table 3, which is obtained by using a test threshold of $\beta = 6$. It is seen that the optimal offline input greatly improves both the ADD and the MTBFA.

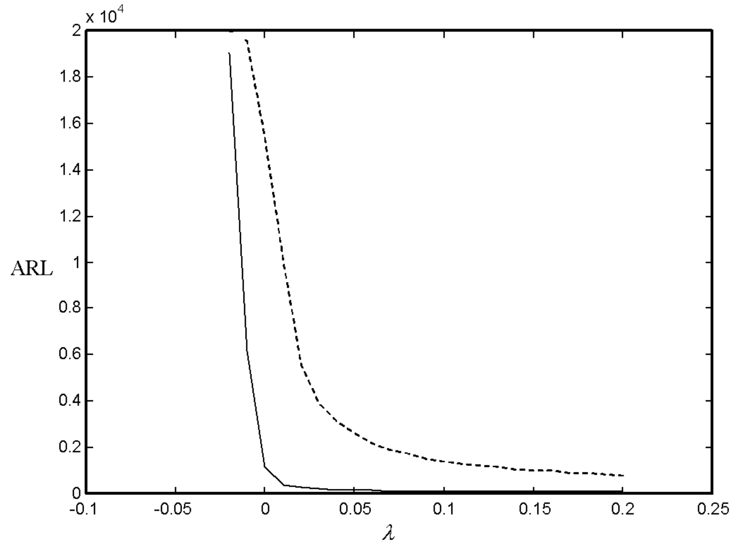

Figure 4. Estimated ARL curves with optimal off-line input (solid) and no input (dashed) in example 1.

Table 3. Estimates of ADD and MTBFA in example 2

	ADD	MTBFA
White input	113	7.6×10^3
Optimal input	63	19.4×10^3

5. Conclusions

We have derived optimal off-line inputs to improve performances of extensions to the CUSUM algorithm for cases where the after-change hypotheses are not specified completely in detecting changes in the parameters of a ARMAX process. It is shown that asymptotically optimal inputs have discrete spectra. If the change direction is known, a single-frequency input will be sufficient. For a more general case where the direction of change in the parameter space is also unknown, the inputs are obtained by optimizing the determinant (or other suitable scalar functions) of the Fisher Information Matrix corresponding to the parameters before the change. In this case, the number of frequencies needed is determined by the number of poles and zeros of the input–output transfer function of the ARMAX process.

For both types of extensions of the CUSUM test, it is possible to obtain significant improvement in the detection delay and/or false alarm rate, if the input is wisely chosen. It is interesting to note that the simulations suggest that the classical tradeoff in statistical change detection (i.e. the one between ADD and MTBFA) is also valid for the first CUSUM extension. On the other hand, as far as the second extension is concerned, in particular changes the inputs can improve both ADD and MTBFA.

We should also note that the input in the known-change-direction case cannot be effective if the change is on the $C(q^{-1})$ polynomial only. Nor does the $C(q^{-1})$ polynomial have any effect on the number of frequencies for the second extension of the CUSUM test.

Acknowledgements

Work was supported by grants EEEAG-DS-6 and 94A0209 of the Technical and Scientific Research Council of Turkey and the Research Fund of Boğaziçi University, respectively.

References

Basseville, M., and Nikiforov, I. V., 1993, *Detection of Abrupt Changes: Theory and Application* (Englewood Cliffs: Prentice-Hall).
Cramer, H., 1946, *Mathematical Methods of Statistics* (Princeton: Princeton University Press).
Goodwin, G. C., and Payne, R. L., 1977, *Dynamic System Identification: Experiment Design and Data Analysis* (New York: Academic Press).
Ibragimov, I. A., and Khasminsky, R. Z., 1981, *Statistical Estimation—Asymptotic Theory* (New York: Springer).
Kalaba, R., and Springarn, K., 1982, *Control, Identification and Input Optimization* (New York: Plenum).
Kerestecioğlu, F., 1993, *Change Detection and Input Design in Dynamical Systems* (Somerset: Research Studies Press).
Kerestecioğlu, F., and Zarrop, M. B., 1994, Input design for detection of abrupt changes in dynamical systems. *International Journal of Control*, **59,** 1063–1084.
Lai, T. Z., 1998, Information bounds and quick detection of parameter changes in stochastic systems, *IEEE Transactions on Information Theory*, **IT-44,** 2917–2929.
Le Cam, L., 1986, *Asymptotic Methods in Statistical Decision Theory* (New York: Springer).

Lorden, G., 1971, Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, **42,** 1897–1908.

Nikiforov, I. V., 1980, Modification and analysis of the cumulative sum procedure. *Automation and Remote Control*, **41,** 74–80.

Nikiforov, I. V., 1986, Sequential detection of changes in stochastic systems. In A. Benveniste and M. Basseville (eds), *Detection of Abrupt Changes in Signals and Dynamical Systems* (Berlin: Springer), pp. 216–258.

Nikiforov, I. V., and Tikhonov, I. N., 1986. Application of change detection theory to seismic signal processing. In A. Benveniste and M. Basseville (eds), *Detection of Abrupt Changes in Signals and Dynamical Systems* (Berlin: Springer), pp. 355–373.

Page, E. S., 1954, Continuous inspection schemes. *Biometrika*, **41,** 100–115.

Patton, R., Frank, P., and Clarke R., 1989, *Fault Diagnosis in Dynamic Systems* (Hemel Hempstead: Prentice-Hall).

Patton, R., Frank, P., and Clarke, R., 2000, *Issues of Fault Diagnosis in Dynamic Systems* (London: Springer).

Rockafellar, R., 1970, *Convex Analysis* (Princeton: Princeton University Press).

Zarrop, M. B., 1979, *Optimal Experiment Design for Dynamic System Identification* (Berlin: Springer).

Zhang, X. J., 1989, *Auxiliary Signal Design in Fault Detection and Diagnosis* (Berlin: Springer).