Sequential detection and isolation of non-orthogonal alternatives

Lionel Fillatre and Igor Nikiforov

Institut Charles Delaunay, FRE CNRS 2848,
Université de technologie de Troyes,
12 rue Marie Curie,
BP 10010, Troyes, France
E-mail: fillatre@utt.fr, nikiforov@utt.fr

Abstract. The comparison between the optimal sequential and repeated fixed size sample (FSS) strategies in the problem of abrupt change detection and isolation is discussed. The general case of non-orthogonal Gaussian hypotheses is considered. Each hypothesis is characterized by its mean vector (the change signature) and it is desirable to detect/isolate a change subject to the constraints on a pre-assigned time between false alarm and a maximum probability of false isolation. It is established that the performance of the proposed FSS algorithm is directly related to the mutual geometry between the hypotheses through the Kullback-Leibler information. This algorithm is almost as efficient as an optimal sequential one but in contrast to the sequential strategy, the FSS strategy can be easily used for monitoring in the case of variable structure systems.

Keywords. Sequential change detection/isolation; Repeated fixed size sample test; Variable structure systems.

1 Introduction and motivation

The problem of detecting and isolating abrupt changes in random signals has many important applications in system monitoring, namely fault detection and diagnosis, in quality control and automatic control. Mathematically, it is the generalization of abrupt change detection to the case of multiple ($K \geq 2$) alternative hypotheses. After the pioneering papers by A.N. Shiryaev (see details in Shiryaev (1963a,b)) the sequential change detection has been studied by many authors, see results and references in Lorden (1971); Moustakides (1986); Basseville & Nikiforov (1993); Lai (1998). Several detection/isolation criterion have been proposed in the literature : Nikiforov (1995); Malladi & Speyer (1999); Lai (2000); Nikiforov (2003); Tartakovsky (2008). In this paper, we consider the "minimax" criterion proposed in Nikiforov (2003). This criterion consists in minimizing the maximum mean delay for detection/isolation subject to the mean time before a false alarm and the maximum probability of false isolation. It seems that such a criterion is especially relevant to safety-critical applications when the system monitoring takes place in a hostile environment.

In the literature, two different kinds of algorithms are considered to solve the problem of sequential detection/isolation: the sequential algorithm and the repeated fixed size sample (FSS) one. The sequential algorithms are often theoretically optimal but, in practice, the FSS algorithms have also some advantages. First of all, usually the FSS strategy is much simpler to obtain and to process subsequent blocks of data for technical reasons (data transfer, sampling time, etc.). On the contrary, the sequential (point by point) processing is technically more sophisticated and time/resource-consuming. Second, often the monitored (typically large-scale) systems have a variable structure. This leads to an extremely complicated sequential strategy : an optimal solution to such a problem is not found. It is worth to note that the theory of sequential decision is only well-developed in the case of stationary systems (in the pre-change state). In contrast to the sequential strategy, the FSS one is easily applicable to systems with a variable structure. A typical example of fault diagnosis in the case of structure variable systems is the volume anomaly detection in an origin-destination flow's traffic over a network (for example, due to denial-of-service, viruses/worms, external routing reconfigurations, etc.). Modern networks, like Internet, are systems with a highly variable structure. Hence, it is practically impossible to consider these systems as stationary (see Fillatre et al. (2008); Casas et al. (2008)). Third, the FSS algorithms are much simpler to study and to implement than the sequential ones. For this reason practical engineers often choose the FSS algorithms for real life applications.

Nevertheless, the following problem should be solved before implementation of FSS algorithms : it is necessary to choose the best possible tuning parameters of the FSS algorithm and to compare the optimal (sequential) strategy and the FSS one in order to estimate the loss of optimality of the FSS strategy. The history of comparisons between sequential and FSS strategies in the theory of statistical hypotheses testing and signal detection is quite long, some results and references can be found in Basseville & Nikiforov (1993). The first comparison between optimal sequential and FSS strategies in the quickest change detection was performed by A.N. Shiryaev in Shiryaev (1963a), Shiryaev (1963b) and next by other authors in Pelkowitz & Schwartz (1987); Nikiforov (1997); Lai (2000).

2 Contribution

The contribution of the present paper is to compare optimal sequential and FSS strategies for the non-Bayesian approach by using the minimax change detection/isolation criterion (see Nikiforov (2003)) in the case of multiple ($K \geq 2$) Gaussian alternative hypotheses. It is assumed that the covariance matrix of the measurement vector is known, hence, a Gaussian hypothesis is completely defined by its mean vector. In Nikiforov (1997), a special case of orthogonal alternative hypotheses has been considered. This particular case corresponds to the detection/isolation of anomalies in K independent channels: the anomaly appears in only one channel and the random noises in different channels are independent. Next, the case of independent scalar channels has been generalized to the case of K Gaussian independent vectors with an unknown post-change mean vector in Lai (2000).

In contrast to Nikiforov (1997); Lai (2000), the general case of K non-orthogonal alternative Gaussian hypotheses is considered now. Often this general case corresponds to the linear model with K channels and n nuisance parameters. By using the theory of invariance, it can be shown that the nuisance parameters rejection implies the projection of K dimensional observations on a linear subspace of dimension $K - n$ containing the maximal invariant statistics (see Fillatre & Nikiforov (2007)). Hence, the dimension of the subspace is less than K and the previously developed theory cannot be used. This paper proposes three major extensions of the results obtained in Nikiforov (1997):

1. The relation between the dimension p of observed vector and the number of alternative hypotheses K is arbitrary.
2. The alternative hypotheses can be non-orthogonal. Obviously, some constraints of "hypotheses separability" should be respected to avoid the problem of detectability/isolability of changes.
3. The criterion of optimality used in this paper is more realistic because it considers the maximum probability of false isolation.

The rest of the paper is organized as follows. Section 3 briefly presents the change detection/isolation problem and the criterion of optimality. Section 4 is devoted to the optimal FSS algorithm. Section 5 includes the results of numerical calculation to show the efficiency of the proposed FSS test. Section 6 concludes the paper.

3 Problem statement

Let $(Y_t)_{t\geq 1}$ be an independent Gaussian random sequence observed sequentially. Firstly, it is assumed that $\mathcal{L}(Y_t) = \mathcal{N}(\boldsymbol{\theta}, \Sigma)$, where Σ is a known (positive definite) covariance matrix. By using the change of variables $g(X) = R^{-1}X$, where the symmetric matrix R is defined by $\Sigma = RR$, and the invariance properties of the gaussian family $\mathcal{N}(\theta, \Sigma)$, the change detection/isolation can be reduced to the following problem statement without loss of generality :

$$\mathcal{L}(Y_t) = \begin{cases} \mathcal{N}(\boldsymbol{\theta}_0, \sigma^2 I_p) \text{ if } t \leq t_0 \\ \mathcal{N}(\boldsymbol{\theta}_\ell, \sigma^2 I_p) \text{ if } t > t_0 \end{cases} \tag{1}$$

where the distribution of Y_t is denoted by $\mathcal{L}(Y_t)$, $\mathcal{N}(\boldsymbol{\theta}, \sigma^2 I_p)$ is the Gaussian distribution with mean $\boldsymbol{\theta}$ and covariance matrix $\sigma^2 I_p$, $Y_t \in \mathbb{R}^p$, $\boldsymbol{\theta}_\ell \in \mathbb{R}^p$, $1 \leq \ell \leq K$, $K \geq 2$, $\boldsymbol{\theta}_0^T = (0, \ldots, 0)$ and I_p is the identity matrix of order p. The vectors $\boldsymbol{\theta}_\ell$ are known and have the same norm $\|\boldsymbol{\theta}_\ell\|_2^2 = c^2$ for all ℓ. It is assumed that $\boldsymbol{\theta}_i \neq \pm\boldsymbol{\theta}_j$ for all $i \neq j$. The change detection/isolation algorithm has to compute a pair (N, ν) based on the observations Y_1, Y_2,$\cdots$, where N is the alarm time at which a ν-type change is detected/isolated and $\nu \in \{1, \cdots, K\}$ is the final decision. Let $P_\ell^{(t_0+1)}$ be the distribution of observations $Y_1, Y_2, \ldots$ when $t_0 = 0, 1, 2, \ldots$ and $\mathcal{L}(Y_t) = \mathcal{N}(\boldsymbol{\theta}_\ell, \sigma^2 I_p)$ for $t > t_0$, $\Pr_\ell^{(t_0+1)}(A)$ be the probability of the event A under $P_\ell^{(t_0+1)}$ and $E_\ell^{(t_0+1)}$ be the expectation under $P_\ell^{(t_0+1)}$. Note here that $P_0^{(\infty)} = P_0$ and $E_0(\cdot) = E_0^{(\infty)}(\cdot)$.

Let $\mathcal{K}_{(\gamma,b)}$ be the class of all sequential detection/isolation algorithms (N, ν) which satisfy the following inequalities

$$\overline{T} \stackrel{\text{def}}{=} \mathbb{E}_0(N) \geq \gamma \quad \text{and} \quad \overline{\beta} \stackrel{\text{def}}{=} \max_{1 \leq \ell \leq K} \max_{1 \leq j \neq \ell \leq K} \sup_{t_0 \geq 0} \Pr_\ell^{(t_0+1)}(\nu = j | N > t_0) \leq b. \tag{2}$$

It is aimed to find an optimal FSS strategy within the class $\mathcal{K}_{(\gamma,b)}$ which minimizes the maximum mean delay for detection/isolation :

$$\bar{\tau} = \sup_{t_0 \geq 0, 1 \leq \ell \leq K} \mathbb{E}_\ell^{(t_0+1)}(N - t_0 \mid N > t_0). \tag{3}$$

4 Optimal FSS strategy

The repeated FSS strategy is based on the following rule : samples with fixed size m are taken, and at the end of each sample, a decision function is computed to test between the hypotheses $\mathcal{H}_0, \ldots, \mathcal{H}_K$

$$\mathcal{H}_\ell \; : \mathcal{L}(Y_t) = \mathcal{N}(\boldsymbol{\theta}_\ell, \sigma^2 I_p), \; t = (n-1)m + 1, \ldots, nm, \; \ell = 0, 1, \ldots, K \tag{4}$$

where $Y_{(n-1)m+1}, \ldots, Y_{nm}$ is the n-th sample. Sampling is stopped after the first sample of observations for which the decision $\bar{\nu}$ is taken in favor of $\mathcal{H}_{\bar{\nu}} : \{\boldsymbol{\theta} = \boldsymbol{\theta}_{\bar{\nu}}\}$ with $\bar{\nu} > 0$. The optimal solution of the multiple hypotheses testing problem (4) for a Gaussian model with non-orthogonal channels is unknown. It is proposed to use the Bayesian decision rule of level α subject to the assumption that the alternative hypotheses $\mathcal{H}_i$, $i = 1, \ldots, K$, are equally probable (see Ferguson (1967)). Hence, the decision function $\bar{\nu}$ of the Bayesian algorithm is given by :

$$\bar{\nu}\left(Y_{(n-1)m+1}, \ldots, Y_{nm}\right) = \begin{cases} 0 \text{ if } \bar{S}_{(n-1)m+1} < h, \\ \bar{d} \text{ if } \bar{S}_{(n-1)m+1} \geq h, \end{cases} \tag{5}$$

$$\bar{d}\left(Y_{(n-1)m+1}, \ldots, Y_{nm}\right) = \arg \max_{1 \leq \ell \leq K} \bar{S}_{(n-1)m+1}(\ell) \tag{6}$$

where

$$\bar{S}_{(n-1)m+1} = \max_{1 \leq \ell \leq K} \bar{S}_{(n-1)m+1}(\ell), \tag{7}$$

$$\bar{S}_{(n-1)m+1}(\ell) = \sum_{t=(n-1)m+1}^{nm} Y_t^T \boldsymbol{\theta}_\ell. \tag{8}$$

Therefore, the decision rule $(\overline{N}, \bar{\nu})$ of the FSS change detection/isolation algorithm is given by:

$$\overline{N} \stackrel{\text{def}}{=} \inf_{n \geq 1} \left\{ n\, m : \bar{S}_{(n-1)m+1} \geq h \right\}, \tag{9}$$

$$\bar{\nu} \stackrel{\text{def}}{=} \bar{d}(Y_{\overline{N}-m+1}, \ldots, Y_{\overline{N}}). \tag{10}$$

When the vectors $\boldsymbol{\theta}_\ell$ are not orthogonal, the optimal FSS algorithm performance depends on the mutual "geometry" of the hypotheses. So, let us introduce some notations to formulate these facts. Let $\delta_{i,j} = \frac{1}{2} \|\mathbf{e}_i - \mathbf{e}_j\|_2^2$ be the distance between two "unit alternatives", where $\mathbf{e}_i = \boldsymbol{\theta}_i / c$ and $\mathbf{e}_i \neq \pm \mathbf{e}_j$ for all $1 \leq i \neq j \leq K$. The real numbers $\{\delta_{i,j}\}_{1 \leq i \neq j \leq K}$ describe the mutual "geometry" of the hypotheses. Let $\bar{\delta}_{\text{d}} = \min_{1 \leq j \leq K} \delta_{0,j} = 1/2$, $\bar{\delta}_{\text{i}} = \min_{1 \leq i \neq j \leq K} \delta_{i,j}$ (clearly, $0 < \bar{\delta}_{\text{i}} < 2$) and $\omega^2 = c^2/\sigma^2$. Finally, let us define $\rho_{\text{d}} = \omega^2\, \bar{\delta}_{\text{d}}$ and $\rho_{\text{i}} = \omega^2\, \bar{\delta}_{\text{i}}$.

The following theorem gives the minimum achievable mean detection/isolation delay of the optimal FSS change detection/isolation algorithm (9) - (10) and the associated optimal parameters m^* and h^*.

Theorem 1. *Let us consider model (1). Let $(\overline{N}, \bar{\nu})$ be the FSS change detection/isolation algorithm (9) - (10). Then, the minimum mean detection/isolation delay within the class $\mathcal{K}_{(\gamma,b)}$ and the optimal tuning parameters h^*, m^* are given by :*

$$\bar{\tau}^* \lesssim \frac{4 \ln \gamma}{\omega^2} = \frac{2 \ln \gamma}{\rho_{\mathrm{d}}}, \tag{11}$$
$$h^* \sim 2 \ln \gamma,$$
$$m^* \sim \frac{2 \ln \gamma}{\omega^2} = \frac{\ln \gamma}{\rho_{\mathrm{d}}}$$
$$\text{subject to } \min \left\{ \bar{\delta}_{\mathrm{i}}^2; \frac{\bar{\delta}_{\mathrm{i}}}{2} \right\} \ln \gamma \gtrsim^+ \ln b^{-1} \;\; \text{as} \;\; b^{-1} \to +\infty \tag{12}$$

where $x \gtrsim^+ y$ is equivalent to $x \geq y(1 + |o(1)|)$ when $y \to +\infty$.

By comparing the mean detection/isolation delay from Theorem 1 with the lower bounds given in Nikiforov (2003), it is easy to see that the FSS strategy is almost as efficient as an optimal one in the case of general non-orthogonal Gaussian hypotheses.

Let us continue our discussion of Theorem 1. The FSS algorithm defined by equation (5) - (10) has two tuning parameters : m and h. By using only these parameters, it is impossible to respect any given combination of the minimum mean time before a false alarm γ and the maximum probability of false isolation b which define the class $\mathcal{K}_{(\gamma,b)}$ and, simultaneously, to get the best possible maximum mean delay for detection/isolation $\bar{\tau}$. To minimize the maximum mean delay for detection/isolation it is assumed that an additional constraint is imposed on the relation between γ and b by the parameter $\bar{\delta}_{\mathrm{i}} > 0$, namely $\min\{\bar{\delta}_{\mathrm{i}}^2; \frac{\bar{\delta}_{\mathrm{i}}}{2}\} \ln \gamma \gtrsim^+ \ln b^{-1}$ as $b^{-1} \to +\infty$. This condition means that the prescribed level of false isolations must be fixed by taking into account the minimum Kullback-Leibler distance $\bar{\delta}_{\mathrm{i}}$ between the alternative hypotheses. If this distance $\bar{\delta}_{\mathrm{i}}$ is very small, it is natural to expect a high probability of false isolation.

The parameter $\bar{\delta}_{\mathrm{i}}$ defines a solid angle around each vector $\mathbf{e}_i$, i.e. an "indifference zone", forbidden for the other vectors $\mathbf{e}_\ell$, $\ell \neq i$. This additional constraint permits us to avoid an unfavorable situation when there are two (or more) close hypotheses which severely penalize the FSS algorithm. At first glance this constraint seems to be too restrictive for the number of potential alternatives K but in reality, if p tends to infinity, it is not too restrictive. The following geometric interpretation illustrates the impact of $\bar{\delta}_{\mathrm{i}}$. Let us defined (by $\bar{\delta}_{\mathrm{i}}$) a spherical cap on the unit p-sphere with apex angle $0 < \alpha < \pi/2$. According to Weisstein (1999), if the dimension p tends to infinity then the number $\overline{K}$ of spherical caps, necessary to recover the unit sphere, tends to infinity with an exponential speed. Hence, an upper bound $\overline{K}$ for the number K of potential alternative hypotheses goes to infinity with an exponential speed for any $0 < \alpha < \pi/2$ (the case $\alpha = \pi/2$ corresponds to orthogonal vectors $\mathbf{e}_\ell$). This geometric interpretation shows that the above mentioned constraint imposed on the relation between γ and b is not too severe.

5 Numerical results

The optimization of the FSS tuning parameters m and h is reduced to the following minimization problem with constraint :

$$(\hat{m}, \hat{h}) = \arg \min_{m,h} \; \bar{\tau}(m,h) \quad \text{subject to} \quad \overline{T}(m,h) = \gamma. \tag{13}$$

As it has been mentioned before, a simultaneous minimisation of the mean detection delay under two constraints γ and b is impossible. For this reason the following method has been adopted for this paper : in the optimisation problem given by equation (13), the active constraint is γ. The probability of false isolation $\overline{\beta} = \overline{\beta}(\hat{h}, \hat{m})$ is calculated as a function of $(\hat{m}, \hat{h})$ obtained from (13).

Let us now compare the asymptotic equation for the mean detection/isolation delay with the results of non-asymptotic numerical optimization (13) of the FSS algorithm and with the numerical results obtained in Nikiforov (1997). This comparison is presented in Figure 1 (left), resp. (right), for the following set of parameters : $\omega^2 = c^2 = 1$, $\sigma = 1$, $p = 25$ and $K = 1$, resp. $K = 10$. In the first case,

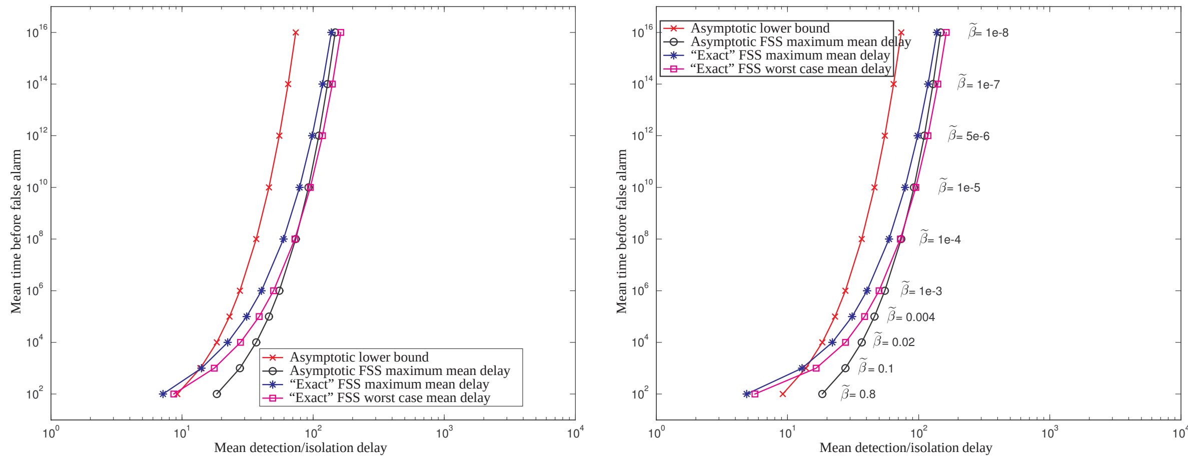

Fig. 1. Comparison between the results of numerical optimization of the FSS algorithm and the asymptotic equations in the case of $K = 1$ (left) and $K = 10$ (right) : asymptotic lower bound (cross marks); asymptotic FSS maximum mean delay (circles); "exact" FSS maximum mean delay (stars); "exact" FSS worst case mean delay (squares).

$K = 1$, the alternative hypothesis is given by $\boldsymbol{\theta}_1^T = (c, 0, \ldots, 0)$ and in the second case, $K = 10$, the alternatives hypotheses $\boldsymbol{\theta}_\ell^T = (0, \ldots, 0, c, 0, \ldots, 0)$ are orthogonal : i.e. the only non-zero element is ℓ-th, $1 \leq \ell \leq 10$. The maximum mean delay $\bar{\tau}(\hat{m}(\gamma), \hat{h}(\gamma))$ obtained by numerical optimization (13), as a function of γ, is called the "exact" FSS maximum mean delay. The expression of the "exact" FSS maximum mean delay is omitted due to its complexity. The "exact" FSS worst case mean delay (with $\operatorname{ess\,sup}$) is obtained by numerical optimization in Nikiforov (1997). This curve is only applicable when the alternatives hypotheses are orthogonal (see Figure 1). Next, the asymptotic lower bound presented in Nikiforov (2003) and the optimal FSS asymptotic maximum mean delay $\bar{\tau}^*$ given by equation (11) as functions of γ are also shown in Figures 1. In the case of $K = 10$ alternative hypotheses, the conservative bound $\widetilde{\beta}$ for the maximum probability of false isolation (not given in the paper due to the lack of place) is also shown in Figure 1 (right). This figure confirms the following : the "exact" FSS curves are close to the asymptotic one and the obtained results are relevant to the results previously published in Nikiforov (1997). Naturally, the "exact" FSS worst case mean delay obtained in Nikiforov (1997) corresponds to a more pessimistic criterion (with $\operatorname{ess\,sup}$), for this reason the curve of the "exact" FSS maximum mean delay (13) is shifted left from the curve of the "exact" FSS worst case mean delay given in Nikiforov (1997).

Let us now consider the non-orthogonal alternative hypotheses. Let us assume that the SNR is $\omega^2 = 1$, $\sigma = 1$, $p = 25$ and $K = 10$. The vectors $\boldsymbol{\theta}_\ell$ are the same as in the previous case except $\boldsymbol{\theta}_2^T = (0.4472, 0.8944, 0, \ldots, 0)$ and $\boldsymbol{\theta}_4^T = (0, 0, 0.2873, 0.9568, 0, \ldots, 0)$ which leads to $\bar{\delta}_{\mathrm{i}} = 0.5528$. The results are presented in Figure 2. This figure shows the following : first, the "exact" FSS curve is close to the asymptotic one and, second, the conservative bound for the maximum probability of false alarm remains relatively important (even for large values of γ) due to the impact of non-orthogonal alternatives.

6 Conclusion

This paper studies the asymptotic performances of the FSS fault detection/isolation strategy. It is shown that the FSS strategy is almost optimal in the case of general non-orthogonal Gaussian hypotheses. The performance of the proposed algorithm is directly related to the mutual geometry (in term of Kullback-Leibler information) between the hypotheses.

Bibliography

M. Basseville & I. V. Nikiforov (1993). *Detection of abrupt changes : theory and application*. Prentice Hall.

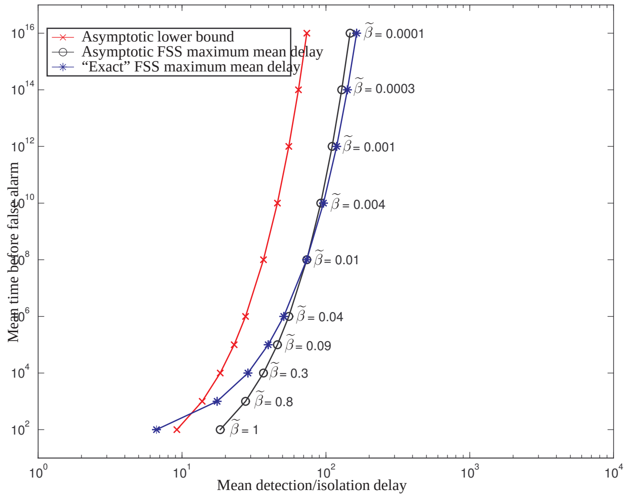

Fig. 2. Comparison between the results of numerical optimization of the FSS algorithm and the asymptotic equations in the case of $K = 10$ non-orthogonal alternative hypotheses : asymptotic lower bound (cross marks); asymptotic FSS maximum mean delay (circles); "exact" FSS maximum mean delay (stars).

P. Casas, et al. (2008). 'Robust and Reactive Traffic Engineering for Dynamic Traffic Demands'. In *Next Generation Internet Networks*, pp. 69–76.

T. Ferguson (1967). *Mathematical Statistics : A Decision Theoretic Approach*. Academic Press.

L. Fillatre & I. Nikiforov (2007). 'Non-bayesian detection and detectability of anomalies from a few noisy tomographic projections'. *IEEE Trans. Signal Process.* **55**(2):401–413.

L. Fillatre, et al. (2008). 'Sequential Non-Bayesian Network Traffic Flows Anomaly Detection and Isolation'. In *International Workshop on Applied Probability (IWAP)*, pp. 1–6.

T. Lai (1998). 'Information bounds and quick detection of parameter changes in stochastic systems'. *IEEE Trans. Inf. Theory* **44**(7):2917–2929.

T. Lai (2000). 'Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems'. *IEEE Trans. Inf. Theory* **46**(2):595–608.

G. Lorden (1971). 'Procedures for reacting to a change in distribution'. *Annals Math. Statistics* **42**:1897–1908.

D. Malladi & J. Speyer (1999). 'A generalized Shiryayev sequential probability ratio test for change detection and isolation'. *IEEE Trans. Autom. Control* **44**(8):1522–1534.

G. Moustakides (1986). 'Optimal procedures for detecting changes in distributions'. *Annals Statistics* **14**:1379–1387.

I. Nikiforov (1995). 'A generalized change detection problem'. *IEEE Trans. Inf. Theory* **41**(1):171–187.

I. Nikiforov (1997). 'Two strategies in the problem of change detection and isolation'. *IEEE Trans. Inf. Theory* **43**(2):770–776.

I. Nikiforov (2003). 'A lower bound for the detection/isolation delay in a class of sequential tests'. *IEEE Trans. Inf. Theory* **49**(11):3037–3046.

L. Pelkowitz & S. Schwartz (1987). 'Asymptotically optimum sample size for quickest detection'. *EEE Trans. Aeros. and Elect. Systems* **AES-23**:263–272.

A. Shiryaev (1963a). 'On detection of disorder in a manufacturing process II'. *Theory Prob. Appl.* **8**:402–413.

A. Shiryaev (1963b). 'On optimum methods in quickest detection problems'. *Theory Prob. Appl.* **8**:22–46.

A. G. Tartakovsky (2008). 'Multidecision Quickest Change-Point Detection: Previous Achievements and Open Problems'. *Sequential Analysis* **27**(2):201–231.

E. Weisstein (1999). *CRC Concise Encyclopedia of Mathematics*. Chapman & Hall.