QUICKEST SEARCH OVER CORRELATED SEQUENCES WITH MODEL UNCERTAINTY

Javad Heydari

ECSE Department
Rensselaer Polytechnic Institute

Ali Tajer

ECSE Department
Rensselaer Polytechnic Institute

H. Vincent Poor

EE Department
Princeton University

ABSTRACT

An ordered set of data sequences is given where, broadly, the data sequences are categorized into normal and abnormal ones. The normal sequences consist of random variables generated according to a known distribution, while there exist uncertainties about the distributions of the abnormal sequences. Moreover, the generations of different sequences are correlated, induced by an underlying physical coupling, where a sequence being normal or abnormal depends on the status of the rest of the sequences according to a known dependency kernel. The objective is to design the quickest *sequential* and *data-adaptive* sampling procedure for identifying one abnormal sequence. This *quickest search* strategy strikes a balance between the quality and agility of the search process, as two opposing figures of merit. This paper characterizes the sampling and search strategy. Motivated by the fact that full characterization of such strategies can become computationally prohibitive, this paper also proposes *asymptotically* optimal sampling and search strategies that are computationally efficient.

Index Terms— Quickest search, correlated sequences, stopping time, model uncertainty.

1. INTRODUCTION

Advances in data acquisition and information processing have led to the generation of very large data sets in many domains, and this trend is expected to grow well into the future. This trend increases the significance of efficient searching algorithms for identifying desired or undesired features in data. Quickest search over data sets aims to perform a real-time and data-adaptive search over a set of data streams in order to identify one exhibiting a desired feature. It strikes a balance between the quality and agility of search, as two opposing figures of merit. Quickest search arises in many application domains such as detection of chemical or biological attacks, identification of free spectrum bands for opportunistic wireless transmission, and monitoring of computer networks for faults or security breaches, to name a few, [1].

Quickest search over data streams is closely related to the sequential testing problems introduced by Wald for distinguishing the distribution of one data stream [2], and further extended to address quickest detection in [3] and [4]. When the given sequence is distributed according to one of two known distributions, the optimal test is characterized by finding an optimal stopping time for the sampling procedure, which collects data until a sufficiently confident decision about the underlying distribution can be formed. The extension to multihypothesis testing and unknown distributions is studied in [5], [6] and [7].

This paper focuses on linear search in which data streams are ordered and examined sequentially. Linear search arises in applications where data streams are available in a specific order such as production line quality control [8], scanning text, audio, and video to detect a specific feature [9], and blind search by robots [10]. Quickest linear search is characterized by an information-gathering procedure in which, besides the stopping time of the sampling procedure, a switching rule determines when to discard one sequence and take samples from the next one. The problem of quickest search was first formalized in [11] as an extension of quickest detection, in which it is assumed that *multiple* data streams are available such that each one is generated according to one of the two known distributions independently of each other. The decision goal of quickest search is to identify one sequence generated according to the desired distribution in the quickest fashion. Other variations of search problems with different assumptions on settings and objectives are also studied in [12], [13] and [14].

In this paper, quickest linear search over multiple data streams is studied, and it has two major distinctions with the aforementioned studies. First, the generation of the data streams follows a certain dependency kernel. This is motivated by the fact that often in networks the observations and actions of the constituent agents are coupled, based on which the measurements collected from different agents exhibit certain correlation structures. Secondly, the data streams are, broadly, categorized into normal and abnormal ones, where only the underlying statistical behavior of the normal data streams is known, while that of the abnormal sequences is known only imperfectly. Specifically, the abnormal distribution is one of the distributions from a finite set of distributions. The number of abnormal sequences is a random variable, and the objective is to identify one of them. This problem under the setting that the distribution of the abnormal sequences is fully known is studied in [15].

Other studies of quickest search problems under different settings and objectives include the scanning problem, in which a finite number of sequences are available and *only one* sequence is generated according to the desired statistical feature which makes it fundamentally different from this paper. In the scanning problems studied in [16], [17] and [18] both distributions are known, while the studies in [14], [19] and [20] consider unknown *discrete* alphabet distributions for both normal and desired distributions. In another direction, the set of available sequences contains multiple sequences with the desired distribution and the goal is to identify *all* of them. Specifically, [21] and [22] investigate anomaly detection with known distributions and [23] analyzes the setting with unknown continuous distributions for normal and outlier processes. Finally, [24] studies outlier detection with unknown discrete distributions under the setting that possibly multiple outlier sequences are available and the objective is to detect all of them. The major distinction of this paper from all the aforementioned studies is that these studies, irrespective of their discrepancies in settings and objectives, conform in the fact that different sequences are generated independently of each other.

This research was supported in part by the U. S. National Science Foundation under Grants ECCS-1455228, CNS-1456793, and ECCS-1343210.

2. PROBLEM FORMULATION

2.1. Data Model

Consider an *ordered* set of n sequences $\{\mathcal{X}^1,\dots,\mathcal{X}^n\}$ where each sequence consists of independent and identically distributed (i.i.d.) real-valued observations $\mathcal{X}^i \triangleq \{X_1^i, X_2^i,\dots\}$. Each sequence obeys the following hypothesis model:

$$\begin{array}{lll} \mathsf{H}_0: & X_j^i \sim F_0, & j=1,2,\dots \\ \mathsf{H}_1: & X_j^i \sim F_\theta, & j=1,2,\dots,\ \theta\in\{\theta_1,\dots,\theta_M\} \end{array}, \tag{1}$$

where F_0 and F_{θ_k} denote cumulative distribution functions (cdfs). The distribution F_0 captures the statistical behavior of the normal sequences and the distribution of the abnormal sequences is not known perfectly but is assumed to be one of the M possible distributions $\{F_{\theta_k}\}_{k=1}^M$. We further assume that probability density functions (pdfs) corresponding to F_0 and F_{θ_k} exist and are well-behaved and denoted by f_0 and f_{θ_k}, respectively. Induced by the underlying physical coupling, the generation of sequences follows a certain dependence structure, in which the prior probability of each sequence being abnormal is controlled by the distribution of its preceding sequence. More specifically, if T_i denotes the *true* model underlying sequence $\mathcal{X}^i$ for $i\in\{1,\dots,n-1\}$, we have the following dependency kernel:

$$\mathbb{P}(\mathsf{T}_{i+1}=\mathsf{H}_1 \mid \mathsf{T}_i=\mathsf{H}_j)=\epsilon_j\ , \quad \text{for}\ \ j\in\{0,1\}\ . \tag{2}$$

We also assume that the initial sequence $\mathcal{X}^1$ is abnormal with prior probability ϵ, i.e.,

$$\mathbb{P}(\mathsf{T}_1=\mathsf{H}_1)\ =\ \epsilon\ . \tag{3}$$

Additionally, we assume that an abnormal sequence is generated according to F_{θ_k} with prior probability μ_k for $k\in\{1,\dots,M\}$, i.e.,

$$\mathbb{P}(\theta=\theta_k)\ =\ \mu_k\ ,$$

where we have $\sum_{k=1}^M \mu_k = 1$.

2.2. Sampling Model

The objective of the search process is to identify *one* abnormal sequence. The sampling procedure examines the sequences sequentially and according to their order by taking one measurement at a time until a sufficiently confident decision can be formed. Specifically, by denoting the index of the observed sequence and its sample at time $t\in\mathbb{N}$ by s_t and Y_t, respectively, we can abstract the information accumulated sequentially by the filtration $\mathcal{F}_t \triangleq \sigma(Y_1,\dots,Y_t)$. The sampling process starts from the first sequence, i.e., $s_1=1$, and based on the information accumulated up to time t, i.e., $\mathcal{F}_t$, the sampling procedure takes one of the following actions:

A$_1$) *Detection:* stops taking more samples and declares sequence s_t to be abnormal;

A$_2$) *Observation:* due to lack of sufficient confidence to make any decision, one more sample is taken from the same sequence, i.e., $s_{t+1}=s_t$; or

A$_3$) *Exploration:* sequence s_t is discarded and the sampling procedure switches to the next sequence and takes one observation from the new sequence, i.e., $s_{t+1}=s_t+1$.

In order to formalize the sampling procedure we define τ as the stopping time of the procedure, which is the instance at which detection action (A$_1$) is performed. To characterize dynamic switching between observation and exploration actions we define the binary function $\psi:\{1,\dots,\tau-1\}\rightarrow\{0,1\}$ such that at time $t\in\{1,\dots,\tau-1\}$ if the decision is in favor of performing observation (A$_2$) we set $\psi(t)=0$, while $\psi(t)=1$ indicates a decision in favor of exploration (A$_3$). Hence, $\forall t\in\{1,\dots,\tau-1\}$

$$\psi(t)=\begin{cases} 0 & \text{action A}_2 \\ 1 & \text{action A}_3 \end{cases}. \tag{4}$$

A sampling strategy is completely characterized by the set $\Phi \triangleq \{\tau,\psi(1),\dots,\psi(\tau-1)\}$.

2.3. Problem Formulation

The optimal search procedure can be found by determining the sampling strategy Φ. Two natural performance measures for evaluating the efficiency of the sampling procedure are the quality of the final decision, which is captured by the frequency of the erroneous decisions $\mathsf{P_e}(\Phi)\triangleq\mathbb{P}(\mathsf{T}_{s_\tau}=\mathsf{H}_0)$, and the *average* delay in reaching a decision, i.e., $\mathsf{AD}(\Phi)\triangleq\mathbb{E}\{\tau\}$. There exists an inherent tension between these two measures as improving one penalizes the other one. By integrating these two figures of merit into one cost function, the optimal sampling strategy can be obtained as the solution to

$$\inf_{\Phi}\ \ \mathsf{P_e}(\Phi)+c\cdot\mathsf{AD}(\Phi)\ , \tag{5}$$

where $c>0$ is a constant that controls the balance between quality and agility of the search process.

3. QUICKEST SEARCH ALGORITHM

We consider an infinite-horizon setting for the search process, in which the process does not have to be terminated prior to a hard deadline. In order to formalize the search process, we define π_t as the posterior probability that the sequence observed at time t is abnormal, i.e., $\pi_t\triangleq\mathbb{P}(\mathsf{T}_{s_t}=\mathsf{H}_1\mid\mathcal{F}_t)$. It can be readily verified that

$$\pi_1=\frac{\epsilon\sum_{k=1}^M\mu_k f_{\theta_k}(Y_1)}{\epsilon\sum_{k=1}^M\mu_k f_{\theta_k}(Y_1)+(1-\epsilon)f_0(Y_1)}\ , \tag{6}$$

and we have the following temporal evolution for π_t:

$$\begin{aligned}\pi_{t+1}=&\frac{\pi_t\sum_{k=1}^M\mu_k^t f_{\theta_k}(Y_{t+1})}{\pi_t\sum_{k=1}^M\mu_k^t f_{\theta_k}(Y_{t+1})+(1-\pi_t)f_0(Y_{t+1})}\mathbb{1}_{(\psi(t)=0)}\\ &+\frac{\bar{\pi}_t\sum_{k=1}^M\tilde{\mu}_k^t f_{\theta_k}(Y_{t+1})}{\bar{\pi}_t\sum_{k=1}^M\tilde{\mu}_k^t f_{\theta_k}(Y_{t+1})+(1-\bar{\pi}_t)f_0(Y_{t+1})}\mathbb{1}_{(\psi(t)=1)}\ ,\end{aligned} \tag{7}$$

where $\mathbb{1}_{(\cdot)}$ denotes the indicator function, $\mu_k^0\triangleq\mu_k$, and we have defined

$$\bar{\pi}_t\triangleq(\epsilon_1-\epsilon_0)\pi_t+\epsilon_0\ , \tag{8}$$

$$\mu_k^t\triangleq\mathbb{P}(\theta=\theta_k|\mathsf{T}_{s_t}=\mathsf{H}_1,\mathcal{F}_t) \tag{9}$$

$$=\frac{\mu_k^{t-1}f_{\theta_k}(Y_t)}{\sum_{m=1}^M\mu_m^{t-1}f_{\theta_m}(Y_t)}\mathbb{1}_{(\psi(t-1)=0)}+\frac{\tilde{\mu}_k^{t-1}f_{\theta_k}(Y_t)}{\sum_{m=1}^M\tilde{\mu}_m^{t-1}f_{\theta_m}(Y_t)}\mathbb{1}_{(\psi(t-1)=1)}\ , \tag{10}$$

$$\tilde{\mu}_k^t\triangleq\frac{\epsilon_1\pi_t\mu_k^t+\epsilon_0(1-\pi_t)\eta_k^t}{\epsilon_1\pi_t+\epsilon_0(1-\pi_t)}\ , \tag{11}$$

$$\text{and}\ ,\quad \eta_k^t\triangleq\frac{(1-\epsilon_0)(1-\pi_t)\eta_k^{t-1}+(1-\epsilon_1)\pi_t\mu_k^{t-1}}{(1-\epsilon_0)(1-\pi_t)+(1-\epsilon_1)\pi_t}\ . \tag{12}$$

Next, in order to identify which of the three actions *detection, observation,* and *exploration* should be taken at each time t, we define three cost functions associated with these three actions. The optimal action at time t is determined by the one that yields the minimal cost-to-go. Specifically, given the filtration $\mathcal{F}_t$ and the structure of the cost function in (5), the minimal cost-to-go at time t, denoted by $\tilde{G}_t(\mathcal{F}_t)$, is related to the costs associated with actions $\{\mathrm{A}_i\}_{i=1}^3$ according to

$$\tilde{G}_t(\mathcal{F}_t) = \min\left\{1 - \pi_t \ , \ c + \min_{i=0,1} \tilde{J}_{t;i}(\mathcal{F}_t)\right\} , \qquad (13)$$

where $(1 - \pi_t)$ is the cost of stopping without taking further measurements and detecting the sequence s_t as an abnormal sequence, and $(c + \tilde{J}_{t;i}(\mathcal{F}_t))$ denotes the cost associated with the observation and exploration actions for $i = 0$ and $i = 1$, respectively. These cost functions, in turn, are related to $\tilde{G}_t(\mathcal{F}_t)$ through

$$\tilde{J}_{t;i}(\mathcal{F}_t) \triangleq \mathbb{E}\Big\{\tilde{G}_t(\mathcal{F}_{t+1}) \,|\, \mathcal{F}_t, \psi(t) = i\Big\} , \quad \text{for } i \in \{0, 1\}. \quad (14)$$

From the recursive connections among π_t and $\{\mu_k^t\}_{k=1}^M$, it can be established that they form a sufficient statistic for the cost functions $\tilde{G}_t(\mathcal{F}_t)$ and $\{\tilde{J}_{t;i}(\mathcal{F}_t)\}_{i=0}^1$, as formalized in the next lemma.

Lemma 1 *The cost functions $\tilde{G}_t(\mathcal{F}_t)$ and $\{\tilde{J}_{t;i}(\mathcal{F}_t)\}_{i=0}^1$ depend on $\mathcal{F}_t$ only through $(\pi_t, \mu_1^t, \dots, \mu_{M-1}^t)$, and can be cast as functions of $(\pi_t, \mu_1^t, \dots, \mu_{M-1}^t)$, denoted by $G_t(\pi_t, \mu_1^t, \dots, \mu_{M-1}^t)$ and $\{J_{t;i}(\pi_t, \mu_1^t, \dots, \mu_{M-1}^t)\}_{i=0}^1$.*

Based on Lemma 1 and by taking into account that we are considering the infinite-horizon setting, it can be verified that the structures of the cost functions are independent of time t, and we denote them by $G(\pi_t, \mu_1^t, \dots, \mu_{M-1}^t)$ and $J_i(\pi_t, \mu_1^t, \dots, \mu_{M-1}^t)$, respectively. Next, we characterize some properties of these cost functions, which are instrumental in designing the optimal sampling strategy.

Lemma 2 *Functions $G(\cdot)$ and $J_i(\cdot)$ are non-negative, concave functions of π over $[0, 1]$, and they are bounded above by 1.*

Based on the Bayesian cost function defined in (5), the sampling process terminates when the cost-to-go associated with *detection* falls below those associated with *exploration* and *observation*, and the optimal stopping time is the first instance at which such a relationship holds. Based on Lemmas 1 and 2, the optimal stopping rule and time are characterized in the following theorem.

Theorem 1 *The optimal stopping time of the search process is*

$$\tau \triangleq \inf \left\{ t \,:\, \pi_t \geq \pi_U\big(\mu_1^t, \dots, \mu_{M-1}^t\big)\right\} , \qquad (15)$$

where $\pi_U\big(\mu_1^t, \dots, \mu_{M-1}^t\big)$ is the solution of

$$1-\pi_U\big(\mu_1^t, \dots, \mu_{M-1}^t\big) = c + \min_{i=0,1} J_i(\pi_U\big(\mu_1^t, \dots, \mu_{M-1}^t\big), \mu_1^t, \dots, \mu_{M-1}^t) . \qquad (16)$$

This theorem indicates that the optimal stopping rule can be simplified to comparing the posterior probability π_t with a fixed threshold at time t, which can be found as the unique solution of (16). This threshold is time-dependent and varies over time. However, by leveraging the fact that the optimal decision rule has a thresholding structure, we devise an alternative decision rule that can be computed more efficiently and is asymptotically optimal.

Next, we characterize the optimal decision rules prior to the stopping time in order to dynamically decide between *exploration* and *observation* actions. Based on the cost-to-go functions defined earlier, at each time t the optimal action is the one with the smaller associated cost, which leads to the following rule:

$$\psi(t) = \begin{cases} 0 & J_0(\pi_t, \mu_1^t, \dots, \mu_{k-1}^t) \leq J_1(\pi_t, \mu_1^t, \dots, \mu_{k-1}^t) \\ 1 & J_0(\pi_t, \mu_1^t, \dots, \mu_{k-1}^t) > J_1(\pi_t, \mu_1^t, \dots, \mu_{k-1}^t) \end{cases} . \qquad (17)$$

Computing the optimal stopping time and switching rules in (15) and (17) involves extensive dynamic programming and can become computationally prohibitive. As a remedy, we next propose alternative decision rules that can incur significantly less computational complexity and enjoy asymptotic optimality when c (the cost per sample) tends to zero. Such asymptotic optimality, equivalently, can be also considered optimal when the rate of erroneous detection decisions tends to zero. Specifically, we define the sampling strategy $S(\pi_L^*, \pi_U^1, \pi_U^2, \dots, \pi_U^M)$, where we have defined

$$\pi_L^* \triangleq \frac{\epsilon_0}{1 - \epsilon_1 + \epsilon_0} \ , \quad \text{and} \quad \pi_U^k \triangleq \pi_U\big(\boldsymbol{e}_k\big) \ , \qquad (18)$$

with $\boldsymbol{e}_k \in \mathbb{R}^{M-1}$ being the vector that is zero everywhere except in its kth element which is 1. This sampling strategy discards sequence s_t and switches to sequence $s_t + 1$ when $\pi_t < \pi_L^*$, i.e.,

$$\psi(t) = \begin{cases} 1 & \text{If } \pi_t < \pi_L^* \\ 0 & \text{If } \pi_t \geq \pi_L^* \end{cases} , \qquad (19)$$

and it terminates the sampling procedure at time t if $\pi_t \geq \pi_U^*$, where we have defined

$$\pi_U^* \triangleq \pi_U^k \ , \quad \text{when } k = \operatorname*{arg\,max}_{m\in\{1,\dots,M\}} \ \mu_m^t \ . \qquad (20)$$

According to this stopping rule, for the stopping time we have

$$\tau \triangleq \inf \ \big\{t \,:\, \pi_t \geq \pi_U^k \ , \ k = \operatorname*{arg\,max}_{m\in\{1,\dots,M\}} \ \mu_m^t\big\} \ . \qquad (21)$$

The following theorem proves that this sampling procedure is asymptotically optimal.

Theorem 2 *As P_e approaches zero and for every abnormal distribution, the sampling strategy $S(\pi_L^*, \pi_U^1, \pi_U^2, \dots, \pi_U^M)$ is first-order asymptotically optimal, i.e.,*

$$\begin{aligned} \inf_{\Phi} \ \mathbb{E}^{\theta_k}\{\tau\} &= \mathbb{E}^{\theta_k}\{\tau\}\big(1 + o(1)\big) \\ &= \frac{|\log \mathsf{P}_\mathsf{e}|}{D(f_{\theta_k} \| f_0)}\big(1 + o(1)\big) \ , \end{aligned} \qquad (22)$$

where $\mathbb{E}^{\theta_k}\{\cdot\}$ denotes expected value under distribution F_{θ_k}.

Even though the threshold π_L^* is not necessarily the optimal solution, it exhibits an important property. By recalling the definition of $\bar{\pi}_t$ in (8) we observe that the condition $\pi_t < \pi_L^*$ is equivalent to $\bar{\pi}_t > \pi_t$ and, conversely, the condition $\pi_t > \pi_L^*$ is equivalent to $\bar{\pi}_t < \pi_t$. Hence, the switching rule in (19) is equivalent to

$$\psi(t) \triangleq \begin{cases} 1 & \text{If } \pi_t < \bar{\pi}_t \\ 0 & \text{If } \pi_t \geq \bar{\pi}_t \end{cases} . \qquad (23)$$

Moreover, by recalling that π_t is the *posterior* probability that sequence s_t is an abnormal sequence and $\bar{\pi}_t$ is the *prior* probability that sequence $s_t + 1$ is an abnormal sequence, the switching rule is equivalent to following the decision that yields the best instantaneous chance of a correct decision. Hence, besides being asymptotically optimal, it also exhibits desirable non-asymptotic performance. Selecting this threshold ensures that the expected time before visiting the last sequence remains finite as $\mathsf{P_e}$ tends to zero, which is a fundamental part of proving optimality. Similarly, the stopping rule given in (21), while not being non-asymptotically optimal, exhibits a desirable property. According to Theorem 1, the optimal stopping rule reduces to dynamically comparing the posterior probability with a threshold, which generally varies over time. However, under the setting that c (or equivalently $\mathsf{P_e}$) tends to zero, the last sequence under observation, i.e., s_τ, is an anomalous sequence with high probability. By taking more samples from this sequence, the value of μ_k^t for the true θ_k tends to 1 and other values diminish to zero. This observation motivates finding the maximum value of μ_k^t for $k \in \{1, \ldots, M\}$ and selecting π_U^k as the upper threshold, which controls the error rate and forces it to zero as c tends to zero.

4. CONCLUSION

We have analyzed the problem of quickest sequential search for an anomaly among a group of correlated data streams. We have assumed a setting in which the normal data streams have a known distribution, but the distribution of the abnormal stream is known imperfectly. Also, induced by some underlying physical coupling, the state of the sequences are correlated according to a known dependency kernel. A sequential sampling strategy for identifying one abnormal sequence has been designed for the setting in which the abnormal distribution takes one of a finite number of possible forms. We have shown that an asymptotically optimal decision rule reduces to comparing the posterior probability of the sequence under observation being abnormal with specific thresholds. If it falls below the lower threshold, which is a function of the correlation structure, then the sequence under scrutiny should be discarded and the following sequence should be observed next, while when it exceeds a dynamically changing upper threshold, the sampling process terminates and the last sequence observed is declared as one with an abnormal distribution.

5. REFERENCES

[1] A. Tajer, V. V. Veeravalli, and H. V. Poor, "Outlying sequence detection in large data sets: A data-driven approach," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 44–56, Sep. 2014.

[2] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, Jun. 1945.

[3] A. N. Shiryaev, "On the theory of decision functions and control by observation from incomplete data," in *Proc. Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague, Czech Republic, Jun. 1962, pp. 657–687.

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

[5] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1994–2007, Nov. 1994.

[6] D. R. Cox, "Sequential tests for composite hypotheses," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, pp. 290–299, Apr. 1952.

[7] N. Chand, "Sequential tests of composite hypotheses," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 394–397, 1974.

[8] B. Torcianti and J. Vass, "Non-contact measurement for mechanical fault detection in production line," in *Proc. IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, Cracow, Poland, Sep. 2007, pp. 297–301.

[9] K. Kashino, T. Kurozumi, and H. Murase, "A quick search method for audio and video signals based on histogram pruning," *IEEE Transactions on Multimedia*, vol. 5, no. 3, pp. 348–357, Sep. 2003.

[10] S. P. Fekete, R. Klein, and A. Nüchter, "Online searching with an autonomous robot," *Computational Geometry*, vol. 34, no. 2, pp. 102 – 115, May 2006.

[11] L. Lai, H. V. Poor, Y. Xin, and G. Georgiadis, "Quickest search over multiple sequences," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5375–5386, Aug. 2011.

[12] A. Tajer and H. V. Poor, "Quick search for rare events," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4462–4481, Jul. 2013.

[13] M. L. Malloy and R. D. Nowak, "Sequential testing for sparse recovery," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7862–7873, Dec. 2014.

[14] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal sequential outlier hypothesis testing," in *Proc. IEEE International Symposium on Information Theory*, Honolulu, HI, Jun.-Jul. 2014, pp. 3205–3209.

[15] J. Heydari and A. Tajer, "Quickest linear search over correlated sequences," in *Proc. International Symposium on Information Theory*, Hong Kong, China, Jun. 2015, pp. 765–769.

[16] E. C. Posner and H. Rumsey Jr., "Continuous sequential decision in the presence of a finite number of hypotheses," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 248–255, Apr. 1966.

[17] K. Zigangirov, "On a problem in optimal scanning," *Theory of Probability and Its Applications*, vol. 11, no. 2, pp. 294–298, Jan. 1966.

[18] V. Dragalin, "A simple and effective scanning rule for a multichannel system," *Metrika*, vol. 43, no. 1, pp. 165–182, Dec. 1996.

[19] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal outlier hypothesis testing," in *Proc. IEEE International Symposium on Information Theory*, Istanbul, Turkey, Jul. 2013, pp. 2666–2670.

[20] S. Nitinawarat and V. V. Veeravalli, "Universal scheme for optimal search and stop," *arXiv:1412.4870v1*, 2014.

[21] K. Cohen, Q. Zhao, and A. Swami, "Optimal index policies for anomaly localization in resource-constrained cyber systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4224–4236, Aug. 2014.

[22] K. Cohen and Q. Zhao, "Asymptotically optimal anomaly detection via sequential testing," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2929–2941, Jun. 2015.

[23] S. Zou, Y. Liang, H. V. Poor, and X. Shi, "Nonparametric detection of anomalous data via kernel mean embedding," *arXiv:1405.2294v1*, 2014.

[24] Y. Li, S. Nitinawarat, and V. V. Veeravalli, "Universal outlier hypothesis testing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4066–4082, Jul. 2014.