A Detection Optimal Min–Max Test for Transient Signals

Chunming Han, Peter Willett, *Member, IEEE*, Biao Chen, and Douglas Abraham, *Member, IEEE*

***Abstract*—Page's test is optimal for detecting a permanent change in distribution, in the sense that it minimizes the worst case average delay to detection given an average distance between false alarms. When used to detect transient signals, however, it in fact becomes the generalized likelihood ratio test (GLRT). Since a GLRT is in almost all cases *ad hoc*, Page's test used as such cannot be said to be optimal in any explicit sense. The subject of this correspondence is the development of the min–max test, via the new ideas of Baygun and Hero, for the detection of a transient.**

***Index Terms*— Detection, min–max detection, Page's test. transient signals.**

I. Introduction

A transient event, or burst, can be thought of as a two-sided change: at some unknown time n_s the observations process $\{U_n\}$ switches from being governed by probability density function (pdf) f_0 to being governed by pdf f_1; and at a later time n_e there is a return to f_0.

This time between sample n_s and $n_e - 1$ represents the occurrence of a transient event, and it is desired to detect such a transient with maximum probability for a specified false-alarm rate. One technique for this detection problem is Page's test [1], [4], [6], [9]: each time the CUSUM statistic

$$\begin{aligned} Z_0 &= 0 \\ Z_n &= \max\{0, Z_{n-1} + g(U_n)\} \end{aligned} \tag{1}$$

passes a threshold h, a detection is declared. The nonlinearity g may be anything desired, but is optimally the log-likelihood ratio $g(u) = \log(f_1(u)/f_0(u))$. This is illustrated in Fig. 1, in which $n_s = 51$, $n_e = 120$, and where f_0 and f_1 denote variance-25 Gaussian densities with respective means -1 and $+1$. Note the "resetting to zero" action of Z_n for low values of n, and the general upward trend of Z_n during the transient's duration—both are as expected.

Page's (or the CUSUM) test is optimal for detecting a permanent change in distribution, in the sense that it minimizes the worst case average delay to detection given an average distance between false alarms [7], [8]. Page's test can also be used to detect transient signals [6], in which context it becomes the GLRT [4], [5]. Since a GLRT is an *ad hoc* procedure, Page's test used as such cannot be said to be optimal in any explicit sense as applied to the transient detection problem; and since it is not optimal, it is reasonable to question whether or not it can be criticized.

II. The Min–Max Test

To some extent, what we are interested in is a joint detection/estimation problem. The starting and ending times of the transient signal (in whose *detection only* we are interested) can be regarded as (nuisance) parameters. As such, the null hypothesis is simple but the alternative is composite. The min–max test [3] is defined as that which minimizes, over all the possible tests, the maximum probability of a miss (over all possible starting and ending times of the signal) subject to a false-alarm rate constraint.

Suppose each of our independent and identically distributed (i.i.d.) observations $\{U_k\}$ is distributed according to probability laws $f_1(\cdot)$ and $f_0(\cdot)$, during and outside the transient signal, respectively. Suppose also that there is at most one transient signal, that the (finite) number of observations is M, and that the duration of the shortest signal L is prior knowledge. Then the null hypothesis is

$$H_0\colon f_{H_0}(U_1, U_2, \cdots, U_M) = f_0(U_1) \cdot f_0(U_2) \cdots f_0(U_M). \tag{2}$$

The alternative hypothesis is

$$H_1 = \bigcup_{s=1}^{M-L+1} \bigcup_{l=L}^{M-s+1} H_{sl}$$

in which

$$\begin{aligned} H_{sl}\colon & f_{H_{sl}}(U_1, U_2, \cdots, U_M) && (3) \\ &= f_0(U_1) \cdot f_0(U_2) \cdots f_0(U_{s-1}) \\ &\quad \cdot [f_1(U_s) \cdots f_1(U_{s+l-1})] \cdot f_0(U_{s+l}) \cdots f_0(U_M). && (4) \end{aligned}$$

The indices s and l denote the starting sample time and the length of the transient, respectively. The likelihood ratio for a signal with a particular duration and starting time is thus

$$LR(\{U_k\}_{k=1}^{M}) = \prod_{k=s}^{s+l-1} \frac{f_1(U_k)}{f_0(U_k)}. \tag{5}$$

In this correspondence our goal is to detect whether there is a transient signal or not, with no interest in its location. As indicated earlier, the alternative is *composite*, and in such situations a generalized likelihood ratio approach (i.e., the Page procedure) is a reasonable candidate. The GLRT is not the only mode of attack, however; and further, it can be unsatisfying due to its lack of optimality properties. It is of course difficult to pose a criterion of optimality in the composite case, but one recently propounded [3], and that which we adopt, is as follows.

> Choose a test which minimizes the maximum probability of a miss, subject to a false-alarm rate constraint. Here the minimization is over all tests, and the maximization is over all possible alternatives.

It was shown [3], that such a "min-max" test is the weighted sum of the likelihood ratios

$$T = \sum_{s=1}^{M-L+1} \sum_{l=L}^{M-s+1} c_{sl} \prod_{k=s}^{s+l-1} \frac{f_1(U_k)}{f_0(U_k)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda \tag{6}$$

where λ, a function of $\{c_{sl}\}$, is determined by the desired false-alarm rate α

$$\Pr\{T > \lambda | H_0\} = \alpha \tag{7}$$

and $\{c_{sl}\}$ are the coefficients maximizing

$$\sum_{s,l} c_{sl} \Pr\{T < \lambda | H_{sl}\} \tag{8}$$

subject to $0 \le c_{sl} \le 1$ and $\Sigma_{s,l}\ c_{sl} = 1$. It is clear that the maximizing coefficients $\{c_{sl}\}$, called optimal coefficients, can be obtained by solving the nonlinear optimization problem (8). When possible, it is usually easier to use the "equalizer rule": more precisely, a sufficient condition for $\{c_{sl}\}$ to be the optimal coefficients is that the probability of a miss

$$\Pr\{T < \lambda | H_{sl}\}$$

Manuscript received November 21, 1996; revised May 13, 1997. This work was supported by ONR through NUWC Newport under Contracts N66604-96-C-0553 and N66604-97-M-3139. The material in this correspondence was presented in part at the IEEE ICASSP, Atlanta, GA, May 7–19, 1996.

C. Han, P. Willett, and B. Chen are with the Department of Electrical and Systems Engineering, U-157, University of Connecticut, Storrs, CT 06269-3157 USA.

D. Abraham is with the SACLANT Undersea Research Centre, CMR 426, APO AE 09613-5000.

Publisher Item Identifier S 0018-9448(98)00652-X.

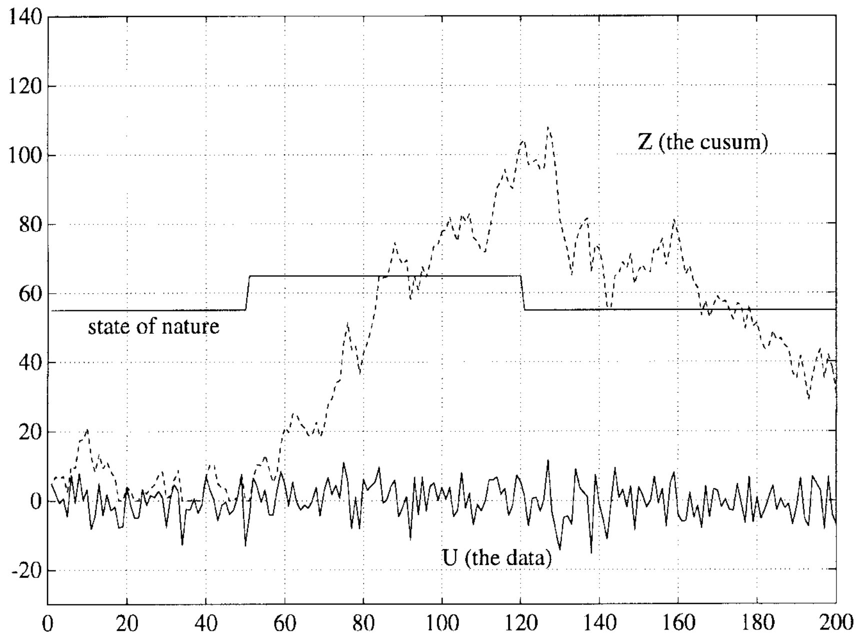

Fig. 1. An illustration of the Page's test procedure. From time samples 51–120 the data is Gaussian with variance 25 and mean $\mu = 1$; otherwise, it is Gaussian with the same variance and mean $\mu = -1$. The CUSUM statistic and the "state-of-nature" processes are also plotted. Here Page's update uses the log-likelihood ratio.

is constant over all the alternative hypotheses $\{H_{sl}\}$. However, sometimes the equalization over all the alternative hypotheses $\{H_{sl}\}$ is impossible (the case with our problem), because the probability of a miss under some alternative hypotheses, called dominating hypotheses in [2], is always larger than the probability of a miss under the other alternative hypotheses no matter what $\{c_{sl}\}$ are. In this case, equalization of only the dominating probabilities of a miss becomes a sufficient condition for constructing the min–max test, provided that the weights for the nondominating hypotheses are set to zero [2]. This results in the theorem below. We begin with the following lemma.

Lemma 1: With $f_H(\vec{U})$ and $f_K(\vec{U})$ being probability density functions of vector observation $\vec{U}$ under hypotheses H and K, respectively,

$$\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})} < \tau | H\right\} \geq \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})} < \tau | K\right\} \tag{9}$$

is true for all thresholds τ. Further, if f_H and f_K are not "essentially the same" and τ is nontrivial, meaning that at least one of

$$\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})} < \tau | H\right\} \quad \text{and} \quad \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})} < \tau | K\right\} \tag{10}$$

is in the open interval $(0, 1)$, then the inequality in (9) is strict.

While the lemma itself is intuitive, its proof is included in the appendix for completeness.

Theorem 1: Suppose the prior knowledge is that there can be at most one transient signal and the duration of the shortest possible signal is L. As the number of i.i.d. observations M approaches infinity, the min–max test can be written as

$$\sum_{i=1}^{M-L+1} \prod_{j=i}^{i+L-1} \frac{f_1(U_j)}{f_0(U_j)} \underset{H_1}{\overset{H_0}{\lessgtr}} \tau \tag{11}$$

where $f_1(U_j)$ and $f_0(U_j)$ are the pdf's of an individual independent observation under the signal-present and -absent hypotheses, respectively. The threshold τ is determined by the false-alarm rate constraint.

That is, the min–max optimal test for detection of a transient of length at least L can be thought of as the Neyman–Pearson optimal test for the alternative

$$f_{H_1}(U_1, U_2, \cdots, U_M) = \frac{1}{M-L+1} \sum_{s=1}^{M-L+1} f_{H_{sL}}(U_1, U_2, \cdots, U_M) \tag{12}$$

meaning that any transient is *exactly* of length L, and among these all are equally likely.

Proof of Theorem 1: As the number of observations approaches infinity, the distribution of the likelihood ratio (5) for the hypotheses with the same length of signal becomes permutation invariant. The result is that the coefficients for the same length of signal in test (6) are independent of the starting point s. Or, in short, $c_{sl} = c_l$. For easy reference, we rewrite test (6) below

$$T = \sum_{s=1}^{M-L+1} \sum_{l=L}^{M-s+1} c_{sl} \prod_{k=s}^{s+l-1} \frac{f_1(U_k)}{f_0(U_k)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda. \tag{13}$$

Constraint $\Sigma_{s,l}\, c_{sl} = 1$ implies that there exists at least one l_1 such that c_{l_1} is positive. Now for arbitrary s_1 and $L \leq l_1 \leq M - s_1 + 1$, consider the two hypotheses $H_{s_1 l_1}$ and $H_{s_1(l_1+1)}$. The difference between these two hypotheses is that the distributions for $U_{s_1+l_1}$ are different, while, for all other $k \neq s_1 + l_1$, U_k's have the same distribution. Specifically, under $H_{s_1 l_1}$, we have

$$U_1, U_2, \cdots, U_{s_1-1}, U_{s_1+l_1}, U_{s_1+l_1+1}, \cdots, U_M \sim f_0$$
$$U_{s_1}, \cdots, U_{s_1+l_1-1} \sim f_1$$

while under $H_{s_1(l_1+1)}$, we have

$$U_1, U_2, \cdots, U_{s_1-1}, U_{s_1+l_1+1}, \cdots, U_M \sim f_0$$
$$U_{s_1}, \cdots, U_{s_1+l_1} \sim f_1.$$

The statistic (13) is complicated in form; but it is simply the weighted sum of products of $f_1(U_k)/f_0(U_k)$. Some of these products contain the term $f_1(U_{s_1+l_1})/f_0(U_{s_1+l_1})$ and some do not. We hence split T in (13) according to

$$T = \frac{f_1(U_{s_1+l_1})}{f_0(U_{s_1+l_1})} T_1 + T_2 \tag{14}$$

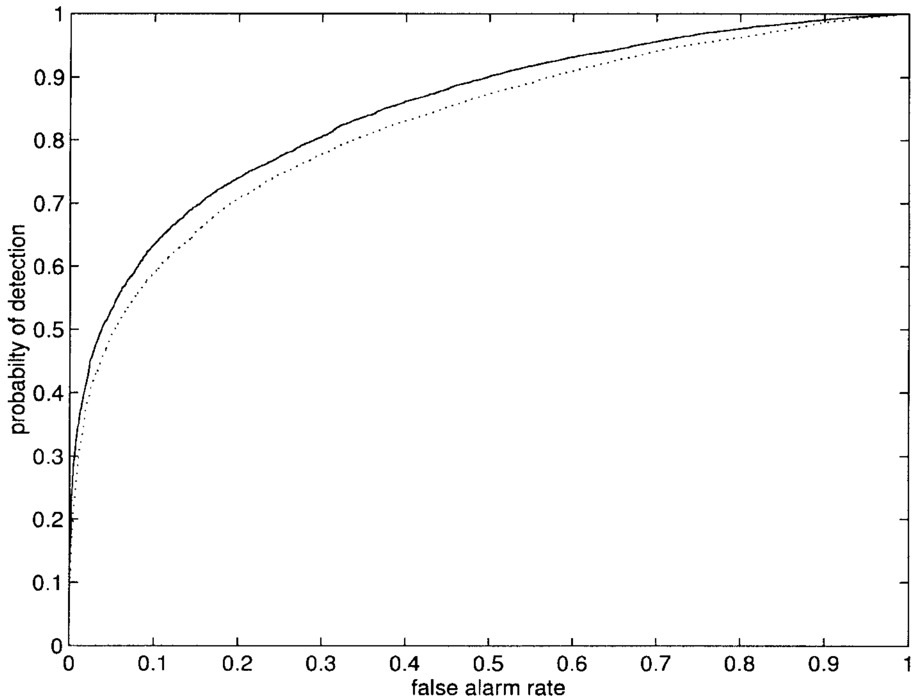

Fig. 2. Relative performance of Page's procedure (dotted) and the min–max test (solid) when the latter's design minimum-transient length is $L = 4$. The actual transient here is of length $l = 4$; the situation is Gaussian with mean shift $\mu = 0.9$.

where neither T_1 nor T_2 involve $U_{s_1+l_1}$. We have the following

$$\begin{aligned}\Pr(T<\tau) &= \Pr\left(\frac{f_1(U_{s_1+l_1})}{f_0(U_{s_1+l_1})}T_1+T_2<\tau\right)\\ &= \Pr\left(\frac{f_1(U_{s_1+l_1})}{f_0(U_{s_1+l_1})}<\frac{\tau-T_2}{T_1}\right). \end{aligned} \tag{15}$$

Note that under both $H_{s_1l_1}$ and $H_{s_1(l_1+1)}$, the distributions for U_k with $k \neq s_1 + l_1$ are identical to each other. Therefore, we can apply Lemma 1 to (15) after conditioning on and averaging over T_1 and T_2 (i.e., all U_k with $k \neq s_1 + l_1$). For nontrivial τ (which results in a nontrivial $\tau' = (\tau - T_2)/T_1$, as T_2 and T_1 are just sums of likelihood ratios and equal to zero or infinity with zero probability) we get

$$\Pr\{T<\lambda|H_{s_1l_1}\} > \Pr\{T<\lambda|H_{s_1(l_1+1)}\} \tag{16}$$

is true for any nontrivial λ. This implies that the probability of a miss under hypothesis $H_{s_1(l_1+1)}$ is always smaller than that under $H_{s_1l_1}$ regardless of the coefficients $\{c_{sl}\}$. The hypotheses with $l = L$ dominate the probabilities of a miss. Hence the optimal $\{c_{sl}\}$ is achieved by equalization over hypotheses $\{H_{sL}\}$, with $\{c_{sl}\}$ set to zero [2] $\forall l > L$. Then test (6) becomes (11) with $\lambda = c_L\tau$. □

It is interesting to note that if the prior knowledge is that the minimum length of a transient signal is $L = 1$, then as the number of observations M goes to infinity, the min–max test asymptotically approaches the following simple form:

$$\sum_{i=1}^{M} \frac{f_1(U_i)}{f_0(U_i)} \underset{H_1}{\overset{H_0}{\lessgtr}} \tau.$$

Example Suppose the minimum length of a possible transient signal $L = 4$ and the number of observations is much larger: $M = 500$. This is a case where Theorem 1 can be used as an approximation because $M \gg L$. The resulting test is

$$\sum_{i=1}^{497}\prod_{j=i}^{i+3} \frac{f_1(U_j)}{f_0(U_j)} \underset{H_1}{\overset{H_0}{\lessgtr}} \tau. \tag{17}$$

In this example we use the Gaussian shift-in-mean transient

$$f_0(U_j) = \frac{1}{\sqrt{2\pi}} e^{-((U_j+\mu)^2/2)}$$

and

$$f_1(U_j) = \frac{1}{\sqrt{2\pi}} e^{-((U_j-\mu)^2/2)}.$$

Fig. 2 shows the simulation results of the receiver's operating characteristic (ROC) of the min–max test versus that of Page's test for the "least favorable" transient signal of length $l = L = 4$, and with mean shift $\mu = 0.9$.

In this case, the min–max test should (and does) perform better than any other test, including Page's test. In Fig. 3, in which the *actual* transient signal is of length $l = 10$, Page's procedure outperforms the min–max $(L = 4)$ test; this is the price paid for robustness.

III. Summary

In this correspondence, we proposed a min–max test for detecting a transient signal. We have shown that when the number of observations approaches infinity we can explicitly determine the min–max test without doing the actual optimization. The min–max test is compared with Page's test. As must be so, the former is better when "ground-truth" is worst case; that is, when the actual transient signal length is its minimum (design) value. In more favorable situations (longer transients), Page's procedure can be superior.

Appendix
Proof of Lemma 1

$$\begin{aligned}&\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|H\right\} - \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|K\right\}\\ &\quad= \int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} f_H(\vec{U})\,d\vec{U}\\ &\quad\quad- \int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} f_K(\vec{U})\,d\vec{U}\end{aligned}$$

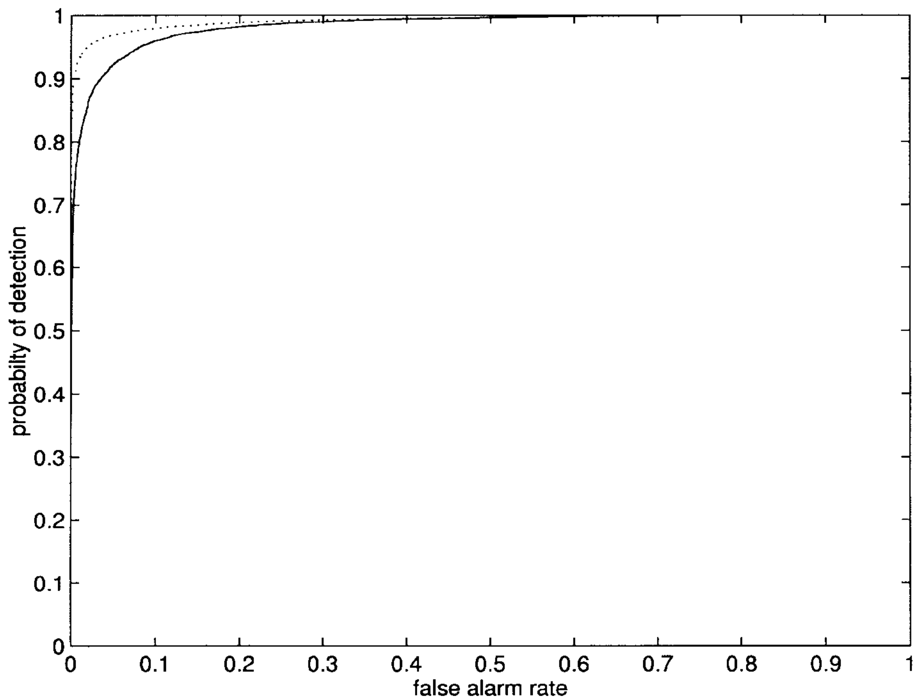

Fig. 3. Relative performance of Page's procedure (dotted) and the min–max test (solid) when the latter's design minimum-transient length is $L = 4$. The actual transient here is of length $l = 10$; the situation is Gaussian with mean shift $\mu = 0.9$.

$$\begin{aligned}&\ge \int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} f_H(\vec{U})\, d\vec{U}\\ &\quad - \int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} \tau f_H(\vec{U})\, d\vec{U}\\ &= (1-\tau)\int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} f_H(\vec{U})\, d\vec{U}. \qquad (18)\end{aligned}$$

This means

$$\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|H\right\} - \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|K\right\} \ge 0, \qquad \text{if } \tau \le 1.$$

On the other hand,

$$\begin{aligned}&\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|H\right\} - \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|K\right\}\\ &= \int_{(f_K(\vec{U})/f_H(\vec{U}))\ge\tau} f_K(\vec{U})\, d\vec{U}\\ &\quad - \int_{(f_K(\vec{U})/f_H(\vec{U}))\ge\tau} f_H(\vec{U})\, d\vec{U}\\ &\ge \int_{(f_K(\vec{U})/f_H(\vec{U}))\ge\tau} \tau f_H(\vec{U})\, d\vec{U}\\ &\quad - \int_{(f_K(\vec{U})/f_H(\vec{U}))\ge\tau} f_H(\vec{U})\, d\vec{U}\\ &= (\tau-1)\int_{(f_K(\vec{U})/f_H(\vec{U}))\ge\tau} f_H(\vec{U})\, d\vec{U}. \qquad (19)\end{aligned}$$

This means

$$\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|H\right\} - \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|K\right\} \ge 0, \qquad \text{if } \tau > 1.$$

As such, inequality (9) is proved for all τ. With inequality (9) proved, condition (10) becomes equivalent to

$$\int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} f_H(\vec{U})\, d\vec{U} > 0$$

and

$$\int_{(f_K(\vec{U})/f_H(\vec{U}))<\tau} f_H(\vec{U})\, d\vec{U} > 0.$$

If f_H and f_K are not essentially the same, (18) and (19) become strict inequalities, thus inequality (9) becomes strict for $\tau \ne 1$. If $\tau = 1$

$$\Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|H\right\} - \Pr\left\{\frac{f_K(\vec{U})}{f_H(\vec{U})}<\tau|K\right\} = \int_{f_K(\vec{U})<f_H(\vec{U})} [f_H(\vec{U}) - f_K(\vec{U})]\, d\vec{U} > 0.$$

This completes the proof that (10) is the sufficient condition for inequality (9) to be strict.

Acknowledgment

The authors wish to thank an anonymous reviewer for his diligence.

References

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[2] B. Baygun, "Optimal strategies and tradeoffs for joint detection and estimation," Ph.D. dissertation, Dept. EECS, Univ. Michigan, Ann Arbor, MI, 1992.
[3] B. Baygun and A. Hero, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Inform. Theory*, vol. 41, pp. 688–703, May 1995.
[4] B. Broder and S. Schwartz, "Quickest detection procedures and transient signal detection," ONR Tech. Rep. 21, Nov. 1990.
[5] C. Han, P. Willett, and D. Abraham, "Some bounds and approximations for the probability of detection of a transient signal," Univ. Conn., Tech. Rep. TR-94-4, 1994.
[6] ——, "On the probability of detection of a transient," in *Proc. IEEE ICASSP-95* (Detroit, MI, May 1995), pp. 2000–2003.
[7] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, 1971.
[8] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, pp. 1379–1387. 1986.
[9] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.