

## DETECTION WITH THE SCAN AND THE AVERAGE LIKELIHOOD RATIO

Hock Peng Chan and Guenther Walther

*National University of Singapore and Stanford University*

*Abstract:* We investigate the performance of the scan (the maximum likelihood ratio statistic) and of the average likelihood ratio statistic in the problem of detecting a deterministic signal with unknown spatial extent in the prototypical univariate sampled data model with white Gaussian noise. Our results show that the scan statistic, a popular tool for detection problems, is optimal only for the detection of signals with the smallest spatial extent. For signals with larger spatial extent the scan is suboptimal, and the power loss can be considerable. In contrast, the average likelihood ratio statistic is optimal for the detection of signals on all scales except the smallest ones, where its performance is only slightly suboptimal. We give rigorous mathematical statements of these results as well as heuristic explanations that suggest that the essence of these findings applies to detection problems quite generally, such as the detection of clusters in models involving densities or intensities, or the detection of multivariate signals. We present a modification of the average likelihood ratio that yields optimal detection of signals with arbitrary extent and which has the additional benefit of allowing for a fast computation of the statistic. In contrast, optimal detection with the scan seems to require the use of scale-dependent critical values.

*Key words and phrases:* Average likelihood ratio statistic, fast algorithm, optimal detection, scan statistic.

### 1. Introduction and Overview of Results

We are concerned with the problem of detecting a deterministic signal with unknown spatial extent against a noisy background. This problem arises in a wide range of applications, e.g. in epidemiology and astronomy, and has received considerable attention recently due to important problems in e.g. biosurveillance. The standard statistical tool to address this problem is the scan statistic (maximum likelihood ratio statistic), that considers the maximum of local likelihood ratio statistics on certain subsets of the data. There is a large body of work on scan statistics, see e.g. the references in Glaz and Balakrishnan (1999), Glaz, Naus, and Wallenstein (2001), and Glaz, Poznyakov, and Wallenstein (2009). But there is also empirical evidence that the scan statistic is suboptimal, see e.g., Neill (2009) or Chan (2009).

Siegmund (2001) and Gangnon and Clayton (2001) propose to use the average of the likelihood ratio statistics instead of their maximum. In different contexts, various versions of the average likelihood ratio were considered by Shiryaev (1963), Burnashev and Begmatov (1990), and Dümbgen (1998). Chan (2009) and Chan and Zhang (2009) perform simulation studies for various detection problems which suggest that the average likelihood ratio statistic is superior to the scan statistic. In light of these results, it is of interest to provide a theoretical investigation of the performance of both the scan and the average likelihood ratio. Such a theoretical comparison seems to be missing in the literature and appears to be quite relevant given the widespread use of the scan statistic as a standard tool for a range of detection problems.

In the first part of this paper we show that in the prototypical univariate sampled data model with white Gaussian noise the scan statistic possesses optimal detection power only for signals with the smallest spatial extent; otherwise the scan statistic is suboptimal, and the loss of power can be considerable for signals having a large spatial extent. We also show that for average likelihood ratio (ALR) statistic these conclusions hold in reversed order: The ALR possesses optimal detection power for signals having large spatial extent, but is suboptimal for signals with small spatial extent. However, the loss of power in the latter case is so small that it is unlikely to be of concern, at least for most sample sizes considered today.

In the second part of the paper we propose a modification of the ALR that results in universal optimality and allows efficient computation. The ALR averages the likelihood ratios pertaining to  $\sim n^2$  stretches of the data, where  $n$  is the sample size, resulting in an  $O(n^2)$  algorithm. Thus the use of the ALR is computationally infeasible even for moderate sample sizes. We introduce a condensed ALR that averages only a certain subset of the likelihood ratios and we show that this condensed ALR possesses optimal detection power for signals having arbitrary spatial extent. Furthermore, this condensed ALR can be computed in almost linear time, viz. with an  $O(n \log^2 n)$  algorithm. In light of the preceding discussion, it is arguably this improvement in computation time rather than the small gain in detection power that is the main advantage of this modification. We note that typically, an approximation introduced to make a procedure computationally less intensive will on the flip side degrade its performance somewhat. It is noteworthy that in the case of the ALR, our computationally efficient modification actually leads to an improved (in fact: optimal) performance.

We give sharp theoretical results on the performance of the ALR, the scan statistic, and the newly proposed ALR in Sections 2 and 3. Since these results are asymptotic, we complement them in Section 5 with a simulation study that illustrates the results. Various modifications to the scan have been proposed in

the literature in order to improve its detection power. We describe two such modifications in Section 4, and include them in our simulation study to obtain a more informative comparison with the ALR.

As in the case of the ALR, the computation of the scan statistic requires an  $O(n^2)$  algorithm. Various efficient algorithms for computing a good approximation to the scan statistic have been introduced in Neill and Moore (2004), Arias-Castro, Donoho, and Huo (2005), Walther (2010) and Rufibach and Walther (2010). Unlike the ALR, constructing a computationally efficient approximation for the scan does not lead to universally optimal power. Rather, statistical optimality for the scan seems to require the use of size-dependent critical values. We summarize our conclusions in Section 6 and defer proofs to Section 7. The notation  $a_n \sim b_n$  means  $c < a_n/b_n < C$  for constants  $0 < c < C$ .

## 2. Comparison of the Scan and the Average Likelihood Ratio

We observe

$$Y_i = f_n\left(\frac{i}{n}\right) + Z_i, \quad i = 1, \dots, n,$$

where the  $Z_i$  are i.i.d.  $N(0, 1)$  and  $f_n(x) = \mu_n \mathbf{1}_{I_n}(x)$  with  $I_n = (j_n/n, k_n/n]$ ,  $0 \leq j_n < k_n \leq n$ . Both the amplitude  $\mu_n$  and the support  $I_n$  are unknown. The task is to decide whether a signal is present, i.e. whether  $\mu_n \neq 0$ .

The above sampled data model with Gaussian white noise serves as a prototype for many important applications. The heuristics and results we develop below suggest that our conclusions carry over, at least qualitatively, to related detection problems involving multivariate signals, non-Gaussian errors, or the detection of clusters in models involving densities or intensities, as described in Kulldorff (1997).

The likelihood ratio statistic for testing  $\mu_n = 0$  when  $I_n$  is known is computed as

$$\exp\left(\frac{(\mathbf{Y}_n(I_n))^2}{2}\right), \quad \text{where } \mathbf{Y}_n(I_n) := \frac{\sum_{i \in I_n} Y_i}{\sqrt{n|I_n|}} = \frac{\sum_{i=j_n+1}^{k_n} Y_i}{\sqrt{k_n - j_n}}.$$

Since  $I_n$  is unknown, the standard approach is to scan over all intervals  $I \in \mathcal{J}_n := \{(j/n, k/n], 0 \leq j < k \leq n\}$  for the largest likelihood ratio statistic. The resulting *scan statistic* (maximum likelihood ratio statistic) is

$$M_n := \max_{0 \leq j < k \leq n} \left| \mathbf{Y}_n\left(\left(\frac{j}{n}, \frac{k}{n}\right]\right) \right|.$$

In contrast, the *average likelihood ratio statistic* (ALR) averages the likelihood ratios over all intervals  $I \in \mathcal{J}_n$ :

$$A_n := \frac{1}{n^2} \sum_{j=0}^n \sum_{k=j+1}^n \exp\left(\frac{(\mathbf{Y}_n((j/n, k/n]))^2}{2}\right).$$

To quantify the performance of these statistics, we look for the smallest value of  $|\mu_n|$  that allows a reliable detection of the signal. As explained below, in order to achieve optimality a test must be able to asymptotically detect signals  $f_n$  with

$$|\mu_n| \sqrt{|I_n|} \geq \frac{\sqrt{2 \log 1/|I_n|} + b_n}{\sqrt{n}}, \quad \text{where } b_n \rightarrow \infty. \quad (2.1)$$

Note that for signals  $f_n$  on *small scales*,  $|I_n| \rightarrow 0$ , (2.1) is equivalent to

$$|\mu_n| \sqrt{|I_n|} \geq (\sqrt{2} + \epsilon_n) \sqrt{\frac{\log 1/|I_n|}{n}}, \quad (2.2)$$

where  $\epsilon_n$  can go to 0 but not too fast:  $\epsilon_n \sqrt{\log 1/|I_n|} \rightarrow \infty$ .

For signals  $f_n$  on *large scales*,  $\liminf_n |I_n| > 0$ , (2.1) is equivalent to

$$|\mu_n| \geq \frac{b_n}{\sqrt{n}}, \quad \text{where } b_n \rightarrow \infty. \quad (2.3)$$

It is impossible to detect signals with noticeably smaller mean: In the case of signals on small scales, a classical argument in the minimax framework (see e.g., Lepski and Tsybakov (2000), Dümbgen and Spokoiny (2001), and Dümbgen and Walther (2008)) shows that if ‘ $+\epsilon_n$ ’ is replaced by ‘ $-\epsilon_n$ ’ in (2.2), then there exists no test that can detect such  $f_n$  with nontrivial asymptotic power. Likewise, a contiguity argument, as in Dümbgen and Walther (2008), shows that in the case of large scales the condition (2.3) is necessary for any test to be consistent against  $f_n$ . On the other hand, we exhibit below a test that detects signals satisfying (2.1) with asymptotic power 1. Thus the detection threshold given by (2.1) marks a standard that is attainable but cannot be improved upon. We now examine how the scan and the ALR compare against this standard.

**Theorem 1.** *Let  $\kappa_n$  be the  $(1 - \alpha)$  quantile of the null distribution of  $M_n$ .*

1. *If  $|\mu_n| \sqrt{|I_n|} \geq (\sqrt{2} + \epsilon_n) \sqrt{(\log n)/n}$  with  $\epsilon_n \sqrt{\log n} \rightarrow \infty$ , then  $\mathbb{P}_{f_n}(M_n > \kappa_n) \rightarrow 1$ .*
2. *If  $|\mu_n| \sqrt{|I_n|} = (\sqrt{2} - \epsilon_n) \sqrt{(\log n)/n}$  with  $\epsilon_n$  as above, then  $\overline{\lim}_n \mathbb{P}_{f_n}(M_n > \kappa_n) \leq \alpha$ .*

Thus the detection threshold for the scan is  $\sqrt{2(\log n)/n}$ , irrespective of the spatial extent of the signal. Comparing to (2.2), one sees that the scan is optimal only for signals having the smallest spatial extent, i.e. for  $|I_n|$  close to  $1/n$ . As an illustration, if  $|I_n| = n^{-p}$ ,  $p \in (0, 1]$ , then detection is possible only if  $|\mu_n| \sqrt{|I_n|}$  is at least  $p^{-1/2}$  times larger than the optimal threshold. In the case of large scales, comparison with (2.3) shows that this multiplier diverges to infinity, and thus the scan suffers from a noticeably inferior performance. These results are illustrated

in the simulation study in Section 5, and explain the sometimes disappointing performance of the scan observed in the literature.

We note that an alternative way to analyze the performance of the scan is to put a prior on the unknown spatial extent of the signal, e.g. the uniform distribution on  $(0, 1)$ . It is readily seen that this analysis leads to the same conclusions as the case of large scales above, i.e. the scan is far from optimal.

The next theorem details the performance of the average likelihood ratio:

**Theorem 2.** *Let  $\tau_n$  be the  $(1 - \alpha)$  quantile of the null distribution of  $A_n$ .*

1.  *$A_n$  is optimal for detecting signals with large spatial extent:  
If  $\liminf_n |I_n| > 0$  and  $|\mu_n| = b_n/\sqrt{n}$  with  $b_n \rightarrow \infty$ , then  $\mathbb{P}_{f_n}(A_n > \tau_n) \rightarrow 1$ .*
2.  *$A_n$  is not optimal for detecting signals with small spatial extent:  
If  $|I_n| \rightarrow 0$  and  $|\mu_n|\sqrt{|I_n|} = K\sqrt{(\log 1/|I_n|)/n}$  with  $K < 2$ , then  $\overline{\lim}_n \mathbb{P}_{f_n}(A_n > \tau_n) \leq \alpha$ .*
3. *If  $K \geq 2 + \epsilon_n$ , where  $\epsilon_n\sqrt{\log 1/|I_n|} \rightarrow \infty$ , then  $\mathbb{P}_{f_n}(A_n > \tau_n) \rightarrow 1$ .*

Comparing with (2.2), one sees that on small scales the ALR requires  $|\mu_n|\sqrt{|I_n|}$  to be about  $\sqrt{2}$  times larger than the optimal threshold. This discrepancy is not very consequential: The simulations in Section 5 show that the corresponding loss of power is quite small for sample sizes up to  $n = 10,000$ , which is the largest sample size we were able to simulate due to the  $O(n^2)$  computational complexity of the ALR.

A heuristic explanation of why the scan and the ALR do not obtain optimality is as follows: There are  $n$  disjoint intervals  $I$  of length  $1/n$ . The corresponding likelihood ratio statistics  $\mathbf{Y}_n(I)$  are i.i.d.  $N(0,1)$  under the null hypothesis, thus their maximum behaves like  $\sqrt{2\log n}$ . But in the case of large intervals of length  $1/c$  (say), there are only  $c$  disjoint intervals that result in independent statistics  $\mathbf{Y}_n(I)$ . The statistics for the other intervals of length  $1/c$  are not independent of these  $\mathbf{Y}_n(I)$  since the intervals overlap. Thus the null distribution of that maximum behaves roughly like the maximum of  $c$  i.i.d.  $N(0,1)$ , which is  $O_p(1)$ . Hence the overall maximum  $M_n$  is dominated by the small intervals, with a corresponding loss of power at large intervals.

As for the ALR, if a detectable signal lives on a large interval  $I_n$ , then  $\mathbf{Y}_n(I)$  is significant provided  $I$  has a nonvanishing overlap with  $I_n$ . Since there are  $\sim n^2$  such intervals, the ALR is significant despite the divisor  $n^2$  in its definition. In the case of small intervals  $I_n$ , however, the number of intervals  $I$  that yield a sufficiently large statistic  $\mathbf{Y}_n(I)$  is so small compared to the total number of intervals ( $\sim n^2$ ) that their contribution to  $A_n$  is annihilated by the divisor  $n^2$ . More precisely: the likelihood ratio statistic is maximized at  $I = I_n$ , where its size is  $\gg |I_n|^{-1}$  (up to log terms) for signals at the detection threshold (2.1). Thus

if  $|I_n| = 1/n$ , then there are only a few significant likelihood ratios and their magnitude is about  $|I_n|^{-1} = n$ . Thus dividing by  $n^2$  will let their contribution vanish unless the size of the likelihood ratio statistics is increased to  $|I_n|^{-2} = n^2$  by doubling  $|\mu_n|^2|I_n|$  in the log likelihood ratio.

### 3. The Condensed Average Likelihood Ratio Statistic

The above heuristic suggests that an optimal version of the ALR can be constructed by averaging the likelihood ratios not over all  $\sim n^2$  intervals of  $\mathcal{J}_n$  but over a subset of  $\mathcal{J}_n$  with cardinality close to  $n$ . The general idea is that for larger intervals, there is not much lost by considering only intervals with endpoints on a coarser grid as long as the distance between such gridpoints is small compared to the length of the intervals. Then these intervals still provide a good approximation to  $\mathcal{J}_n$ , while the cardinality of this approximating set can be reduced dramatically. To implement this idea, we modify the approach in Walther (2010) and Rufibach and Walther (2010) and group intervals into  $\ell_{\max} = \lceil \log_2 n / \log n \rceil$  sets, each of which contains intervals having about the same length: the approximating set  $\mathcal{I}_{app}(\ell)$  consists of intervals that contain between  $m_\ell + 1$  and  $2m_\ell$  design points and whose endpoints are restricted to a grid consisting of every  $d_\ell$ th design point, where  $m_\ell = n2^{-\ell}$  and  $d_\ell = \lceil \sqrt{m_\ell} \ell^{4/5} / \log n \rceil$ . Our overall approximating set is then the union of these  $\mathcal{I}_{app}(\ell)$  together with all small intervals:

$$\begin{aligned} \mathcal{I}_{app} &= \bigcup_{\ell=1}^{\ell_{\max}} \mathcal{I}_{app}(\ell) \cup \mathcal{I}_{small}, \quad \text{where} \\ \mathcal{I}_{app}(\ell) &= \left\{ \left( \frac{j}{n}, \frac{k}{n} \right] \in \mathcal{J}_n : j, k \in \{id_\ell, i = 0, 1, \dots\} \text{ and } m_\ell < k - j \leq 2m_\ell \right\}, \\ \mathcal{I}_{small} &= \left\{ \left( \frac{j}{n}, \frac{k}{n} \right] \in \mathcal{J}_n : k - j \leq m_{\ell_{\max}} \right\}. \end{aligned}$$

We suppress the dependence on  $n$  for notational simplicity. Our *condensed ALR* is thus

$$A_{n,cond} := \frac{1}{\#\mathcal{I}_{app}} \sum_{I \in \mathcal{I}_{app}} \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right).$$

The above choices of  $d_\ell$  and  $m_\ell$  result in statistical and computational efficiency for the ALR and differ from the choices used in Walther (2010) and Rufibach and Walther (2010). We give an explanation for this in the proof of Theorem 3.

**Theorem 3.** *The condensed ALR  $A_{n,cond}$  is optimal for detecting signals with arbitrary spatial extent. Furthermore,  $A_{n,cond}$  can be computed in  $O(n \log^2 n)$  time.*

#### 4. Modifications of the Scan That Improve Power

Here we describe two simple ways to improve the power of the scan by fixing the miscalibration across different scales described in Section 2. For the first one we adapt the penalty term introduced by Dümbgen and Spokoiny (2001) in the context of inference about a function. The idea is to subtract off the putative maximum at each scale in order to put the different scales on an equal footing: the *penalized scan* is

$$P_n := \max_{0 \leq j < k \leq n} \left( \left| \mathbf{Y}_n \left( \left( \frac{j}{n}, \frac{k}{n} \right] \right) \right| - \sqrt{2 \log \frac{en}{k-j}} \right).$$

We declare that a signal is present if  $P_n > \gamma_n(\alpha)$ , where  $\gamma_n(\alpha)$  is the  $(1 - \alpha)$  quantile of the null distribution of  $P_n$ .

A drawback of the penalized scan is that it requires the specification of the penalty term, which has to be derived for each situation at hand. The penalty term  $\sqrt{2 \log en/(k-j)}$  optimizes signal detection for all scales in the Gaussian regression setting; a different setting or a different error distribution may require a different penalty term to achieve optimal detection. The form of the penalty term depends on the tail behavior of the local test statistics, their dependence structure, and the entropy of the underlying space, see Theorem 7.1 in Dümbgen and Walther (2008). Thus these properties have to be derived on a case-by-case basis, and this derivation is typically far from straightforward.

The second way to fix the miscalibration of the scan is the *blocked scan* introduced in Walther (2010). The block method has the advantage that it is a general recipe that does not require any case-specific input. The idea is to group intervals having roughly the same length into blocks, with the  $\ell$ th block comprising all intervals that contain between  $m_\ell = n2^{-\ell}$  and  $2m_\ell$  design points. Then one assigns different critical values to different blocks such that the significance level on the  $\ell$ th block decreases as  $\sim \ell^{-2}$ .

In more detail, for  $m_\ell$  and  $\ell_{\max}$  as above, define

$$M_{n,\ell} := \max_{m_\ell < k-j \leq 2m_\ell} \left| \mathbf{Y}_n \left( \left( \frac{j}{n}, \frac{k}{n} \right] \right) \right|, \quad \ell = 1, \dots, \ell_{\max},$$

and  $M_{n,\ell_{\max}+1} := \max_{k-j \leq m_{\ell_{\max}}} |\mathbf{Y}_n((j/n, k/n])|$ . Thus the  $(\ell_{\max} + 1)$ st block comprises all small intervals that contain up to  $m_{\ell_{\max}}$  design points. The *blocked scan* declares that a signal is present if  $M_{n,\ell} > q_\ell(\tilde{\alpha}/(A + \ell)^2)$  for any  $\ell \in \{1, \dots, \ell_{\max} + 1\}$ . Here  $q_\ell(\tilde{\alpha}/(A + \ell)^2)$  is the  $(1 - \tilde{\alpha}/(A + \ell)^2)$  quantile of the null distribution of  $M_{n,\ell}$ ,  $A := 10$  (say), and  $\tilde{\alpha}$  is chosen such that the overall significance level is  $\alpha$ :

$$\mathbb{P}_0 \left( \bigcup_{\ell=1}^{\ell_{\max}+1} \left\{ M_{n,\ell} > q_\ell \left( \frac{\tilde{\alpha}}{(A + \ell)^2} \right) \right\} \right) = \alpha.$$

The critical values  $q_\ell$  and  $\tilde{\alpha}$  can be easily simulated with Monte Carlo, see Rufibach and Walther (2010). We suppress the dependence of  $q_\ell$  on  $n$  for notational simplicity.

It can be shown that both the penalized scan and the blocked scan are also optimal for detecting signals with arbitrary spatial extent, see Chan and Walther (2011).

Computationally efficient algorithms for evaluating the scan or an approximation thereof have been introduced in the literature, see e.g., Neill and Moore (2004), Arias-Castro, Donoho, and Huo (2005), Rufibach and Walther (2010) and Walther (2010). Those algorithms reduce the computational complexity for the scan from  $O(n^2)$  to almost linear time in  $n$  (apart from  $\log n$  factors), comparable to the condensed scan. Hence it is possible to modify the scan to obtain statistical optimality (via the penalized or blocked scan) and computational efficiency. But, unlike the case of the condensed ALR where the particular choice of the approximating set leads to optimal power properties, it appears that evaluating the scan on an appropriate approximating set does not lead to optimal detection by itself. Rather, it appears that optimal detection requires the use of scale-dependent critical values, and efficient computation has to be addressed separately using any of the methods cited above.

## 5. A Simulation Study

Since the results of the previous sections are asymptotic, we illustrate them in a finite sample context with a simulation study. We first consider signals  $f_n$  with fixed norm  $\|f_n\| := |\mu_n| \sqrt{|I_n|}$  but varying spatial extent. Table 1 gives the power of the scan, the ALR, the condensed ALR, the penalized scan, and the blocked scan for a sample size of  $n = 10,000$ . The results are visualized in the left plot in Figure 1. One sees that the overall performance of the scan is inferior to that of the other four methods, whose performances are quite similar. In particular, the power of the scan is not increasing with the spatial extent of the signal as opposed to the other four methods. As a consequence, the scan is competitive only for signals on the smallest scales.

Table 1 also shows an improvement in power of the condensed ALR vis-a-vis the ALR on small scales, illustrating Theorems 2 and 3. However, this improvement is modest, at least for the sample size under consideration, and thus the main advantage of the condensed scan is arguably the dramatic reduction in computation time to  $O(n \log^2 n)$  versus  $O(n^2)$  for the ALR. We were able to accurately simulate critical values for the condensed ALR with a sample size of 1 million in a matter of hours, whereas this computation would take hundreds of days for the ALR.



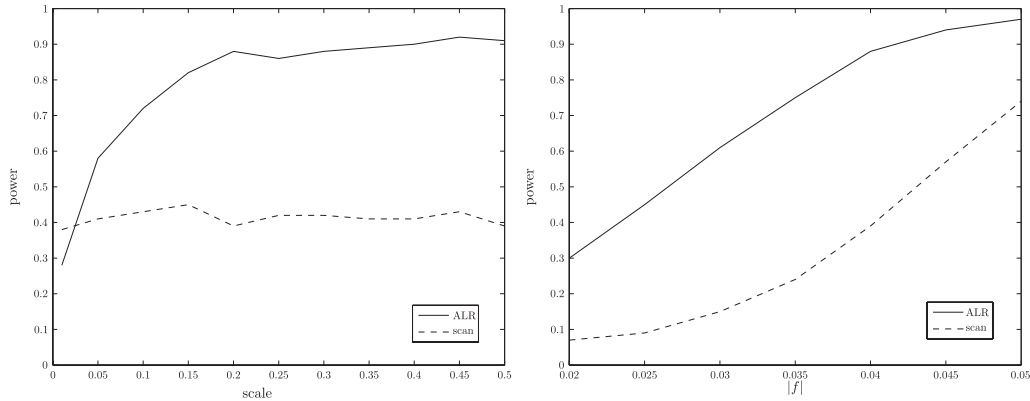


Figure 1. Left: Power of the scan and the ALR for detecting signals  $f_n$  with fixed norm  $\|f_n\|=0.04$  but varying spatial extent  $|I_n|$ ,  $n = 10,000$ . Right: Power of the scan and the ALR for detecting signals  $f_n$  with varying norms  $\|f_n\|$  and random spatial extent. The power curves for the condensed ALR, the penalized scan, and the blocked scan are similar to those of the ALR and are not plotted, see Tables 1 and 2 for the numerical results.

Table 1. Power in percent for detecting signals  $f_n$  with fixed norm  $\|f_n\|=0.04$  but varying spatial extent  $|I_n|$ ,  $n = 10,000$ .

	scale	0.01	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
scan		38	41	43	45	39	42	42	41	41	43	39
ALR		28	58	72	82	88	86	88	89	90	92	91
condensed ALR		36	61	72	80	87	85	87	88	90	91	91
penalized scan		37	61	72	80	85	84	85	86	87	90	89
blocked scan		41	59	69	77	82	80	82	82	84	87	86

Table 2 shows how the power varies as function of  $\|f_n\|$ , see the right plot in Figure 1 for a visual representation. The spatial extent of the signal was chosen uniformly in  $[0, 1]$  in each of the 2,000 Monte Carlo simulations. The power curves of the last four methods are again quite similar, and superior to that of the scan. One sees that the scan requires a signal with almost twice the norm to achieve the power of the four other methods. According to the results in the previous sections, this discrepancy increases with the sample size.

All power values in Tables 1 and 2 are with respect to a 5% significance level. The corresponding critical values were simulated with 10,000 Monte Carlo samples, and the power was simulated with 2,000 Monte Carlo samples. The location of the signal was chosen at random in each of these simulations to avoid confounding the results with the approximation scheme of the condensed ALR.

Table 2. Power in percent for detecting signals  $f_n$  with varying norms  $\|f_n\|$  and random spatial extent,  $n = 10,000$ .

$\ f_n\  \times 100$	2	2.5	3	3.5	4	4.5	5
scan	7	9	15	24	39	57	74
ALR	30	45	61	75	88	94	97
condensed ALR	30	44	60	75	87	94	97
penalized scan	26	40	57	74	85	93	97
blocked scan	24	35	51	69	82	92	96

## 6. Conclusion

The scan is optimal only for detecting signals on the smallest scales. The ALR has a superior overall performance and is optimal for detecting signals on all scales except on the smallest ones, but the loss of power there appears to be modest. Moreover, by averaging the likelihood ratios over a particular subset of intervals rather than over all intervals, the resulting condensed ALR is simultaneously optimal for all scales and also allows for efficient computation. In contrast, improved versions of the scan, such as the penalized scan and the blocked scan, appear to require the use of scale-dependent critical values, and thus it appears that statistical optimality and computational efficiency have to be addressed separately for the scan.

The results of this paper are developed in the Gaussian white noise model since it is known that the conceptual results in that model are applicable and relevant for a wide range of related problems. We note that the concrete implementation of the results derived in the Gaussian white noise model requires additional work that depends on the concrete problem at hand. For example, in the univariate regression setting, Rohde (2008) employs local signed rank tests to transform non-Gaussian data into statistics with sub-gaussian tails, and Cai, Jeng, and Li (2011) employ a local median transformation for the same purpose. To see why such an additional step is required, note that the null distribution of both the scan and the ALR, as well as the form of the penalty term for the penalized scan, depend sensitively on the tails of the error distribution, and hence on the assumption of Gaussianity. The above papers show rigorously that the Gaussian white noise model is applicable after a local signed rank or a local median transformation, assuming only e.g. symmetry of the error distribution. These arguments are technically sophisticated and thus the transformation step constitutes a piece of methodological work by itself. It is thus helpful to separate the conceptual issues involving the scan and the ALR from the particular implementation and to present an unencumbered exposition in the Gaussian white noise model. In particular, the heuristics developed in this model give guidance how the corresponding problems might be addressed in related detection problems,

such as the detection of clusters in models involving densities and intensities, as well as the important case of detecting multivariate signals. For example, in a regression type setting with irregularly spaced or multivariate data, the proportion of observations falling into the set  $I$  would take the place of the size  $|I|$  used above. Section 3 shows how to construct an approximating set that, on the one hand, results in computational efficiency and, on the other, provides the right ‘weighting’ of the various sizes of intervals in the condensed ALR to achieve statistical optimality. This idea can presumably be mimicked in a multivariate situation, where the construction would then depend on the entropy of the class of scanning windows. The implementation of this idea in the multivariate context is an interesting problem for future research. Another open problem is a theoretical result on the precision with which the scan and the ALR allow one to localize a signal once it is detected.

**7. Proofs**

Note that  $f_n(x) = \mu_n \mathbf{1}_{I_n}(x)$  implies for any interval  $I \in \mathcal{J}_n$ :

$$\mathbf{Y}_n(I) = \mathbf{Z}_n(I) + \text{sign}(\mu_n) \sqrt{n} |\mu_n| \frac{|I \cap I_n|}{\sqrt{|I|}}. \tag{7.1}$$

We use the following consequence of a result of Dümbgen and Spokoiny (2001).

**Lemma 1.** *Let  $\ell \in (0, 1)$  and  $J \in \mathcal{J}_n$ , where  $J$  does not depend on  $\mathbf{Z}_n$ . Then*

$$\max_{I \in \mathcal{J}_n: I \subset J, |I| \geq \ell} |\mathbf{Z}_n(I)| \stackrel{d}{\leq} L + \sqrt{2 \log \frac{e|J|}{\ell}}$$

for a universal random variable  $L$  which is finite almost surely.

**Proof of Lemma 1.** Writing  $W$  for Brownian motion and  $j, k$  for integer indices:

$$\begin{aligned} & \max_{\substack{I \in \mathcal{J}_n: I \subset J \\ |I| \geq \ell}} \left( |\mathbf{Z}_n(I)| - \sqrt{2 \log \frac{e|J|}{|I|}} \right) \\ & \stackrel{d}{=} \max_{\substack{0 \leq j < k \leq n|J| \\ k-j \geq n\ell}} \left( \frac{|W(k) - W(j)|}{\sqrt{k-j}} - \sqrt{2 \log \frac{e|J|n}{k-j}} \right) \\ & \leq \sup_{\substack{s, t \in \mathbf{R}: \\ 0 \leq s < t \leq n|J|}} \left( \frac{|W(t) - W(s)|}{\sqrt{t-s}} - \sqrt{2 \log \frac{e|J|n}{t-s}} \right) \\ & \stackrel{d}{=} \sup_{0 \leq s < t \leq 1} \left( \frac{|W(t) - W(s)|}{\sqrt{t-s}} - \sqrt{2 \log \frac{e}{t-s}} \right) =: L \end{aligned} \tag{7.2}$$

by Brownian scaling. Thus the random variable  $L$  defined above is universally applicable for all  $n, \ell$ , and  $J$ . Importantly,  $L$  is finite almost surely, see Sec. 6.1 in Dümbgen and Spokoiny (2001).

**Proof of Theorem 1.** As for part 1, (7.1) implies  $M_n \geq |\mathbf{Y}_n(I_n)| \geq -|\mathbf{Z}_n(I_n)| + \sqrt{2 \log n} + \epsilon_n \sqrt{\log n}$ . Since  $\mathbf{Z}_n(I_n) \sim N(0, 1)$  and  $\epsilon_n \sqrt{\log n} \rightarrow \infty$ , the claim follows from  $\kappa_n = \sqrt{2 \log n} + O(1)$ , see (7.3).

For the proof of part 2, set  $b_n := \epsilon_n \sqrt{\log n} \rightarrow \infty$  and consider first the collection of intervals  $\mathcal{J}_{n,1} := \left\{ I \in \mathcal{J}_n : I \cap I_n \neq \emptyset \text{ and } |I_n|/b_n \leq |I| \leq b_n |I_n| \right\}$ . So  $I \in \mathcal{J}_{n,1}$  implies  $I \subset \left( (j_n - \lfloor b_n n |I_n| \rfloor)/n, (k_n + \lfloor b_n n |I_n| \rfloor)/n \right) \cap (0, 1]$  and thus Lemma 1 gives

$$\max_{I \in \mathcal{J}_{n,1}} |\mathbf{Z}_n(I)| \stackrel{d}{\leq} L + \sqrt{2 \log \frac{e(1+2b_n)|I_n|}{|I_n|/b_n}} \leq L + 2\sqrt{\log(3b_n)}.$$

Together with (7.1) and  $|I \cap I_n| \leq \sqrt{|I||I_n|}$ , this yields

$$\begin{aligned} \mathbb{P}_{f_n} \left( \max_{I \in \mathcal{J}_{n,1}} |\mathbf{Y}_n(I)| > \kappa_n \right) &\leq \mathbb{P} \left( \max_{I \in \mathcal{J}_{n,1}} |\mathbf{Z}_n(I)| + \sqrt{2 \log n} - b_n > \kappa_n \right) \\ &\leq \mathbb{P} \left( L > \kappa_n - \sqrt{2 \log n} + b_n - 2\sqrt{\log(3b_n)} \right) \\ &\rightarrow 0 \quad \text{since } \kappa_n = \sqrt{2 \log n} + O(1) \text{ by (7.3)}. \end{aligned}$$

Next, only if  $b_n \leq \log^3 n$  do we need to consider  $\mathcal{J}_{n,2} := \{I \in \mathcal{J}_n : I \cap I_n \neq \emptyset \text{ and either } |I_n|/\log^3 n \leq |I| \leq |I_n|/b_n \text{ or } b_n |I_n| < |I| \leq |I_n| \log^3 n\}$ . Similarly as above,  $I \in \mathcal{J}_{n,2}$  implies that  $I$  is contained in an interval  $J \in \mathcal{J}_n$  with  $|J| \leq (1 + 2 \log^3 n)|I_n|$ . Thus Lemma 1 yields

$$\max_{I \in \mathcal{J}_{n,2}} |\mathbf{Z}_n(I)| \stackrel{d}{\leq} L + \sqrt{2 \log \frac{e(1+2 \log^3 n)|I_n|}{|I_n|/\log^3 n}} \leq L + 4\sqrt{\log \log n}.$$

One readily checks that  $I \in \mathcal{J}_{n,2}$  implies  $|I \cap I_n| \leq \sqrt{|I||I_n|/b_n}$ . Thus (7.1) gives

$$\begin{aligned} \mathbb{P}_{f_n} \left( \max_{I \in \mathcal{J}_{n,2}} |\mathbf{Y}_n(I)| > \kappa_n \right) &\leq \mathbb{P} \left( \max_{I \in \mathcal{J}_{n,2}} |\mathbf{Z}_n(I)| + \frac{\sqrt{2 \log n} - b_n}{\sqrt{b_n}} > \kappa_n \right) \\ &\leq \mathbb{P} \left( L > \kappa_n - \frac{\sqrt{2 \log n} - b_n}{\sqrt{b_n}} - 4\sqrt{\log \log n} \right) \\ &\rightarrow 0 \quad \text{since } \kappa_n = \sqrt{2 \log n} + O(1). \end{aligned}$$

Finally, consider  $\mathcal{J}_{n,3} := \left\{ I \in \mathcal{J}_n : I \cap I_n = \emptyset \text{ or } |I| \leq |I_n|/\log^3 n \text{ or } |I| > |I_n| \log^3 n \right\}$ . Since  $I \in \mathcal{J}_{n,3}$  implies  $|I \cap I_n| \leq \sqrt{|I||I_n|/\log^3 n}$ , we get by

(7.1),

$$\mathbb{P}_{f_n} \left( \max_{I \in \mathcal{J}_{n,3}} |\mathbf{Y}_n(I)| > \kappa_n \right) \leq \mathbb{P} \left( \max_{I \in \mathcal{J}_n} |\mathbf{Z}_n(I)| > \kappa_n - \frac{\sqrt{2 \log n} - b_n}{\sqrt{\log^3 n}} \right) \rightarrow \alpha,$$

where the convergence follows from the following.

A sequence  $\{c_n\}$  satisfies  $\lim_n \mathbb{P} \left( \max_{I \in \mathcal{J}_n} |\mathbf{Z}_n(I)| > c_n \right) = \alpha$  if and only if

$$c_n = \sqrt{2 \log n} + (2 \log n)^{-1/2} \left( \frac{1}{2} \log \log n + C(\alpha) \right) + o \left( (\log n)^{-1/2} \right) \quad (7.3)$$

for a certain constant  $C(\alpha)$ . This follows from Theorem 1.3 in Kabluchko (2008) or, with some work, from the earlier Theorem 1 in Siegmund and Venkatraman (1995).

Since  $\mathcal{J}_n = \mathcal{J}_{n,1} \cup \mathcal{J}_{n,2} \cup \mathcal{J}_{n,3}$  the theorem is proved.

**Proof of Theorem 2.** We begin by showing that in the null case of no signal,  $\mathbf{Y}_n = \mathbf{Z}_n$ , we have

$$A_n = O_p(1). \quad (7.4)$$

Note that  $A_n$  is an average of correlated random variables that do not possess a finite first moment. In light of the converse to the strong law it is thus not at all obvious that (7.4) holds. For a proof let  $m > 0$  and define the event  $\mathcal{B}_{m,n} := \{ |\mathbf{Z}_n((j/n, k/n))| \leq C((k-j)/n) + m \text{ for all } 0 \leq j < k \leq n \}$ , where  $C(\delta) := \sqrt{2 \log e/\delta}$ . Then Markov's inequality gives, for  $\lambda > 0$ ,

$$\begin{aligned} \mathbb{P}_0(A_n > \lambda) &\leq \frac{1}{\lambda n^2} \sum_{j=0}^{n-1} \sum_{k=j+1}^n \mathbb{E} \left[ \exp \left( \frac{(\mathbf{Z}_n((j/n, k/n)))^2}{2} \right) 1(\mathcal{B}_{m,n}) \right] + \mathbb{P}(\mathcal{B}_{m,n}^c) \\ &\leq \frac{1}{\lambda n^2} \sum_{j=0}^{n-1} \sum_{k=j+1}^n \int_{-C(\frac{k-j}{n})-m}^{C(\frac{k-j}{n})+m} e^{z^2/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz + \mathbb{P}(\mathcal{B}_{m,n}^c) \\ &\leq \frac{1}{\lambda n^2} \sum_{j=0}^{n-1} \sum_{k=j+1}^n \left( \sqrt{2 \log \frac{en}{k-j}} + m \right) + \mathbb{P}(\mathcal{B}_{m,n}^c) \\ &\leq \frac{1}{\lambda} \left( \frac{1}{n} \sum_{j=0}^{n-1} \int_0^1 \sqrt{2 \log \frac{e}{u}} du + m \right) + \mathbb{P}(\mathcal{B}_{m,n}^c) \\ &\leq \frac{4+m}{\lambda} + \mathbb{P}(L > m) \end{aligned}$$

by (7.2). This sum can be made arbitrarily small by choosing  $m$  and  $\lambda$  appropriately, proving (7.4).

To prove parts 1 and 3 together we consider  $f_n$  with arbitrary spatial extent and  $|\mu_n|\sqrt{|I_n|} \geq (\sqrt{4\log 1/|I_n|} + b_n)/\sqrt{n}$  with  $b_n \rightarrow \infty$ . Set  $\epsilon_n := \min(1, b_n(\log e/|I_n|)^{-1/2})$  and  $\mathcal{J}(I_n) := \{I \in \mathcal{J}_n : I \subset I_n \text{ and } |I| \geq |I_n|(1-\epsilon_n/2)\}$ . Then  $\#\mathcal{J}(I_n) \geq \lceil \epsilon_n n |I_n|/4 \rceil^2$  since each of the  $\lceil \epsilon_n n |I_n|/4 \rceil$  smallest (largest) design points in  $cl(I_n)$  may serve as a left (right) endpoint for some  $I \in \mathcal{J}(I_n)$ . Lemma 1 gives

$$\max_{I \in \mathcal{J}(I_n)} |\mathbf{Z}_n(I)| \stackrel{d}{\leq} L + \sqrt{2 \log \frac{e}{1-\epsilon_n/2}} \leq L + 2.$$

Together with  $\frac{|I \cap I_n|}{\sqrt{|I||I_n|}} = \sqrt{\frac{|I|}{|I_n|}} \geq \sqrt{1-\frac{\epsilon_n}{2}}$  and (7.1) we get

$$\begin{aligned} \min_{I \in \mathcal{J}(I_n)} |\mathbf{Y}_n(I)| &\stackrel{d}{\geq} \left( \sqrt{4 \log \frac{1}{|I_n|}} + b_n \right) \sqrt{1-\frac{\epsilon_n}{2}} - L - 2 \\ &\geq \sqrt{4 \log \frac{1}{|I_n|}} + \frac{b_n}{9} - L - 2, \end{aligned}$$

since  $(x+y)\sqrt{1-\min(1,y)/2} \geq x+y/9$  for  $x \in [0,2], y > 0$ . Thus, writing  $R_n := (b_n/9) - L - 2$ ,

$$\begin{aligned} A_n &\geq \frac{\#\mathcal{J}(I_n)}{n^2} \min_{I \in \mathcal{J}(I_n)} \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) \\ &\stackrel{d}{\geq} \frac{\epsilon_n^2 |I_n|^2}{4^2} \exp\left\{2 \log \frac{1}{|I_n|} + R_n \left(\frac{R_n}{2} + \sqrt{4 \log \frac{1}{|I_n|}}\right)\right\} \\ &= \frac{\epsilon_n^2}{4^2} \exp\left\{R_n \left(\frac{R_n}{2} + \sqrt{4 \log \frac{1}{|I_n|}}\right)\right\} \\ &\geq \frac{\epsilon_n^2}{4^2} \left(\log \frac{e}{|I_n|}\right) \exp\left(\frac{R_n^2}{2}\right) 1(R_n \geq 1) \\ &\stackrel{a.s.}{\rightarrow} \infty \quad \text{since } R_n \stackrel{a.s.}{\rightarrow} \infty \text{ and } \epsilon_n \sqrt{\log \frac{e}{|I_n|}} \geq 1 \text{ eventually.} \end{aligned}$$

The claim follows since  $\tau_n = O(1)$  by (7.4).

For the proof of part 2 we partition  $\mathcal{J}_n$  into  $\mathcal{J}_{n,1} := \{I \in \mathcal{J}_n : I \cap I_n \neq \emptyset \text{ and } |I_n|/(\log^4 1/|I_n|) \leq |I| \leq |I_n| \log^4 1/|I_n|\}$  and  $\mathcal{J}_{n,2} := \mathcal{J}_n \cap \mathcal{J}_{n,1}^c$ . We show for  $\mathcal{J} = \mathcal{J}_{n,1}, \mathcal{J}_{n,2}$  that

$$\frac{1}{n^2} \sum_{I \in \mathcal{J}} \left| \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) - \exp\left(\frac{(\mathbf{Z}_n(I))^2}{2}\right) \right| \xrightarrow{P} 0. \quad (7.5)$$

Since it can be shown that in the null case  $f_n \equiv 0$ , the ALR  $A_n$  converges weakly to a continuous limit, the claim of part 2 follows from (7.5).

To prove (7.5) for  $\mathcal{J} = \mathcal{J}_{n,1}$  we follow the proof of part 2 of Theorem 1 (set  $b_n := \log^4 1/|I_n|$  there) and conclude  $\max_{I \in \mathcal{J}_{n,1}} |\mathbf{Z}_n(I)| \leq L + 2\sqrt{\log(3 \log^4 1/|I_n|)}$ . Hence for a fixed constant  $\lambda$  which is specified below we obtain

$$\mathbb{P}\left(\mathcal{A}_n := \left\{ \max_{I \in \mathcal{J}_{n,1}} |\mathbf{Z}_n(I)| \leq \lambda \sqrt{\log \frac{1}{|I_n|}} \right\}\right) \rightarrow 1.$$

That proof also shows that every  $I \in \mathcal{J}_{n,1}$  is contained in a certain interval of length  $(1 + 2 \log^4 1/|I_n|)|I_n|$ , thus  $\#\mathcal{J}_{n,1} \leq ((1 + 2 \log^4 1/|I_n|)n|I_n|)^2$ . On the event  $\mathcal{A}_n$  we have, by (7.1),

$$\begin{aligned} & \frac{1}{n^2} \sum_{I \in \mathcal{J}_{n,1}} \left| \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) - \exp\left(\frac{(\mathbf{Z}_n(I))^2}{2}\right) \right| \\ & \leq \frac{\#\mathcal{J}_{n,1}}{n^2} 2 \exp\left(\max_{I \in \mathcal{J}_{n,1}} \frac{(\mathbf{Z}_n(I))^2}{2} + \frac{n|\mu_n|^2|I_n|}{2} + \max_{I \in \mathcal{J}_{n,1}} |\mathbf{Z}_n(I)|\sqrt{n|I_n|}|\mu_n|\right) \\ & \leq 18\left(\log 1/|I_n|\right)^8 |I_n|^2 \exp\left(\frac{\lambda^2 \log 1/|I_n|}{2} + \frac{K^2}{2} \log \frac{1}{|I_n|} + \lambda K \log \frac{1}{|I_n|}\right) \\ & = 18\left(\log \frac{1}{|I_n|}\right)^8 |I_n|^{2-\frac{K^2}{2}-\frac{\lambda^2}{2}-\lambda K}. \end{aligned}$$

Since  $K^2/2 < 2$  we can choose  $\lambda = \lambda(K) > 0$  such that the above expression goes to 0 as  $|I_n| \rightarrow 0$ , proving (7.5) for  $\mathcal{J} = \mathcal{J}_{n,1}$ .

To prove (7.5) for  $\mathcal{J} = \mathcal{J}_{n,2}$ , we proceed similarly as in the proof of (7.4) and employ the event  $\mathcal{B}_{m,n}$  defined there. Then for  $\lambda, m > 0$ , Markov's inequality gives

$$\begin{aligned} & \mathbb{P}_{f_n}\left(\frac{1}{n^2} \sum_{I \in \mathcal{J}_{n,2}} \left| \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) - \exp\left(\frac{(\mathbf{Z}_n(I))^2}{2}\right) \right| > \lambda\right) \\ & \leq \frac{1}{\lambda n^2} \sum_{I \in \mathcal{J}_{n,2}} \mathbb{E}_{f_n}\left(\left| \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) - \exp\left(\frac{(\mathbf{Z}_n(I))^2}{2}\right) \right| 1(\mathcal{B}_{m,n})\right) + \mathbb{P}(\mathcal{B}_{m,n}^c). \end{aligned}$$

Since (7.2) gives  $\mathbb{P}(\mathcal{B}_{m,n}^c) \leq \mathbb{P}(L > m) \rightarrow 0$  as  $m \rightarrow \infty$ , it is enough to show that for any fixed  $m$  the above expectation converges to 0 as  $n \rightarrow \infty$ , uniformly in  $I \in \mathcal{J}_{n,2}$ . Using  $\mathbf{Z}_n(I) \sim N(0,1)$  and writing  $\delta_I := \sqrt{n}|\mu_n|(|I \cap I_n|/\sqrt{|I|})$  and  $C(|I|) := \sqrt{2 \log e/|I|}$ , we get with (7.1),

$$\mathbb{E}_{f_n}\left(\left| \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) - \exp\left(\frac{(\mathbf{Z}_n(I))^2}{2}\right) \right| 1(\mathcal{B}_{m,n})\right)$$

$$\begin{aligned} &\leq \frac{1}{\sqrt{2\pi}} \int_{-C(|I|)-m}^{C(|I|)+m} \left| \exp\left(z\delta_I + \frac{\delta_I^2}{2}\right) - 1 \right| dz \\ &\leq \sqrt{\frac{2}{\pi}} \exp\left(\delta_I(C(|I|) + m) + \frac{\delta_I^2}{2}\right) \left(\delta_I(C(|I|) + m)^2 + \frac{\delta_I^2(C(|I|) + m)}{2}\right) \end{aligned} \quad (7.6)$$

by bounding the function  $z \mapsto \exp(z\delta + \delta^2/2) - 1$  above and below by the Mean Value Theorem. Next we show that, as  $n \rightarrow \infty$ ,

$$\delta_I \rightarrow 0 \quad \text{and} \quad \delta_I C^2(|I|) \rightarrow 0 \quad \text{uniformly in } I \in \mathcal{I}_{n,2}. \quad (7.7)$$

This conclusion then also holds for the expression in (7.6), and (7.5) follows.

To prove (7.7), note that  $\delta_I = K \sqrt{\log 1/|I_n|} |I \cap I_n| / \sqrt{|I||I_n|}$ . If  $I \cap I_n = \emptyset$ , then  $\delta_I = 0$ . If  $|I| < |I_n|/\log^4 1/|I_n|$ , then the bound  $|I \cap I_n| \leq |I|$  yields  $\delta_I \leq K(\log 1/|I_n|)^{-3/2}$ , while the monotonicity of the function  $x \mapsto \sqrt{x} \log e/x$  for  $x \in (0, e^{-1})$  gives

$$\begin{aligned} \delta_I C^2(|I|) &\leq 2K \sqrt{\log \frac{1}{|I_n|}} \sqrt{\frac{|I|}{|I_n|}} \log \frac{e}{|I|} \\ &\leq 2K \left(\log \frac{1}{|I_n|}\right)^{-3/2} \log \frac{e \log^4 1/|I_n|}{|I_n|} \\ &\leq 6K \left(\log \frac{1}{|I_n|}\right)^{-1/2} \end{aligned}$$

if  $n$  is large enough so that  $|I_n|/\log^4 1/|I_n| \leq e^{-1}$ .

If  $|I| > |I_n| \log^4 1/|I_n|$ , then the bound  $|I \cap I_n| \leq |I_n|$  yields again  $\delta_I \leq K(\log 1/|I_n|)^{-3/2}$ , while

$$\delta_I C^2(|I|) \leq 2K \left(\log \frac{1}{|I_n|}\right)^{-3/2} \log \frac{e}{|I|} \leq 2K \left(\log \frac{1}{|I_n|}\right)^{-1/2}.$$

**Proof of Theorem 3.** Before proceeding to the proof, we sketch an explanation for the choice of the grid spacing  $d_\ell$ . For given  $I_n$ , let  $\ell$  be such that the intervals in  $\mathcal{I}_{app}(\ell)$  have length about  $|I_n|$ . Thus  $m_\ell \sim n|I_n|$ , i.e.  $\ell \approx \log_2 e/|I_n|$ . An interval  $I$  results in a significant likelihood ratio provided its endpoints lie in a  $(\epsilon_n/2)|I_n|$  neighborhood of the endpoints of  $I_n$ , where  $\epsilon_n := \min(1, b_n(\log e/|I_n|)^{-1/2})$ . Thus the number of significant intervals in  $\mathcal{I}_{app}(\ell)$  is  $\sim (\epsilon_n|I_n|n/d_\ell)^2$ . Under (2.1) the size of the corresponding likelihood ratios is  $\gg (\log_2 e/|I_n|)^p/|I_n|$  for arbitrary  $p > 0$ . Thus optimality of  $A_{n,cond}$  obtains if the number of significant intervals in  $\mathcal{I}_{app}(\ell)$  is at least  $\#\mathcal{I}_{app}|I_n|/(\log_2 e/|I_n|)^p$  for some fixed  $p > 0$ , and even if the number is smaller by some factor  $\epsilon_n^k$ , since  $\epsilon_n \sqrt{\log_2 e/|I_n|} \geq 1$ . Solving this inequality for  $d_\ell$  yields  $d_\ell \leq \sqrt{n^2|I_n|\ell^p/\#\mathcal{I}_{app}}$ . Requiring  $\#\mathcal{I}_{app} \sim n(\log n)^q$  for



some  $q > 0$  for computational efficiency suggests the choice  $d_\ell \sim \sqrt{m_\ell \ell^p (\log n)^{-q}}$ . Computing  $\#\mathcal{I}_{app}$  shows that this choice of  $d_\ell$  is indeed consistent with  $\#\mathcal{I}_{app} \sim n(\log n)^q$  provided  $p > 1$ . Further, it will be seen that  $A_{n,cond} = O_p(1)$  under the null hypothesis requires  $\sum_\ell \ell^{1/2-p} < \infty$ , i.e.  $p > 3/2$ . Finally, optimal detection for very small intervals  $I_n$  requires that their endpoints are approximated exactly, i.e. it is necessary to have  $d_\ell = 1$  for large  $\ell$ . Thus we need an appropriate combination of a small  $p > 3/2$  and a large  $q$ . Since the required large  $q$  results in a noticeably worse computation time, we prefer to stick to a  $O(n \log^2 n)$  algorithm by setting  $p = 8/5$ ,  $q = 2$ , and by explicitly considering all small intervals containing up to  $\log n$  design points in lieu of choosing a larger  $q$ .

We now prove the theorem, starting with the claim about the computational complexity. Since  $\mathcal{I}_{app}(\ell)$  allows only every  $d_\ell$ th design point as a potential endpoint for an interval, there are at most  $\lceil n/d_\ell \rceil$  left endpoints. For each left endpoint there are at most  $\lceil 2m_\ell/d_\ell \rceil$  right endpoints since each interval contains not more than  $2m_\ell$  design points. Thus

$$\#\mathcal{I}_{app}(\ell) \leq \left\lceil \frac{n}{d_\ell} \right\rceil \left\lceil \frac{2m_\ell}{d_\ell} \right\rceil \leq 3 \frac{n \log^2 n}{\ell^{8/5}} \quad \text{and} \tag{7.8}$$

$$\#\mathcal{I}_{small} \leq n \log n \quad \text{since } m_{\ell_{\max}} \leq \log n. \quad \text{Hence} \tag{7.9}$$

$$\#\mathcal{I}_{app} \leq \sum_{\ell=1}^{\ell_{\max}} 3 \frac{n \log^2 n}{\ell^{8/5}} + n \log n \leq 9n \log^2 n. \tag{7.10}$$

$\exp((\mathbf{Y}_n(I))^2/2)$  can be evaluated in a constant number of steps after an initial one-time computation of the cumulative sum vector of  $(Y_i, 1 \leq i \leq n)$ . Since that computation has complexity  $O(n)$ , the overall computational complexity of computing  $A_{n,cond}$  is dominated by the cardinality of  $\mathcal{I}_{app}$  and hence is  $O(n \log^2 n)$ .

Next we show that for  $f_n \equiv 0$ :

$$A_{n,cond} = O_p(1) \quad \text{as } n \rightarrow \infty. \tag{7.11}$$

Proceeding as in the proof of (7.4) it is enough to show that

$$\left( \frac{1}{\#\mathcal{I}_{app}} \right) \sum_{I \in \mathcal{I}_{app}} \sqrt{2 \log \left( \frac{1}{|I|} \right)} = O(1). \tag{7.12}$$

Since  $I \in \mathcal{I}_{app}(\ell)$  implies  $|I| > m_\ell/n = 2^{-\ell}$ , we obtain with (7.8) and (7.9),

$$\begin{aligned} \sum_{I \in \mathcal{I}_{app}} \sqrt{2 \log \frac{1}{|I|}} &\leq \sum_{\ell=1}^{\ell_{\max}} \#\mathcal{I}_{app}(\ell) \sqrt{2\ell} + n(\log n) \sqrt{2 \log n} \\ &\leq 5n \log^2 n \sum_{\ell=1}^{\infty} \frac{\ell^{1/2}}{\ell^{8/5}} + n(\log n)^2 \leq 56n \log^2 n. \end{aligned}$$

On the other hand,  $\#\mathcal{I}_{app} \geq \#\mathcal{I}_{app}(2) \geq (1/10)n(\log n)^2$  by considerations similar to those establishing (7.8). (7.12) follows.

To establish optimality of  $A_{n,cond}$ , we proceed as in the proof of Theorem 2 and consider  $f_n$  with arbitrary spatial extent and  $|\mu_n|\sqrt{|I_n|} \geq (\sqrt{2\log 1/|I_n|} + b_n)/\sqrt{n}$ , where  $b_n \rightarrow \infty$ . As before we define  $\epsilon_n := \min(1, b_n(\log e/|I_n|)^{-1/2})$  and  $\mathcal{J}(I_n) := \{I \in \mathcal{J}_n : I \subset I_n \text{ and } |I| \geq |I_n|(1 - \epsilon_n/2)\}$ . Intervals  $I \in \mathcal{J}(I_n)$  contribute significant LR statistics to  $A_{n,cond}$ . Since we now require that the endpoints of these intervals fall on a  $d_\ell$ -grid, there are now many fewer of these intervals. But this is more than compensated by the small cardinality of  $\mathcal{I}_{app}$  appearing in the divisor of  $A_{n,cond}$ . This fact allows us to detect  $f_n$  with a norm that is smaller than in Theorem 2. In more detail, take the integer  $\ell$  so  $m_\ell < n|I_n|(1 - \epsilon_n/4) \leq 2m_\ell$ .

**Lemma 2.**

$$\begin{aligned} \text{If } \ell \leq \ell_{\max}, \text{ then } \frac{\#\left(\mathcal{J}(I_n) \cap \mathcal{I}_{app}(\ell)\right)}{\#\mathcal{I}_{app}} &\geq \frac{\epsilon_n^2 |I_n|}{9^3 \left(\log_2 e/|I_n|\right)^{8/5}}. \\ \text{If } \ell > \ell_{\max}, \text{ then } \frac{\#\left(\mathcal{J}(I_n) \cap \mathcal{I}_{small}\right)}{\#\mathcal{I}_{app}} &\geq \frac{\epsilon_n^2 |I_n|}{32^2 \left(\log_2 e/|I_n|\right)^2}. \end{aligned}$$

As in the proof of Theorem 2 we find  $\min_{I \in \mathcal{J}(I_n)} |\mathbf{Y}_n(I)| \stackrel{d}{\geq} \sqrt{2\log 1/|I_n|} + R_n$ , where  $R_n := b_n/9 - L - 2$ . Thus in the case  $\ell \leq \ell_{\max}$ , Lemma 2 gives

$$\begin{aligned} A_{n,cond} &\geq \frac{\#\left(\mathcal{J}(I_n) \cap \mathcal{I}_{app}(\ell)\right)}{\#\mathcal{I}_{app}} \min_{I \in \mathcal{J}(I_n)} \exp\left(\frac{(\mathbf{Y}_n(I))^2}{2}\right) \\ &\stackrel{d}{\geq} \frac{\epsilon_n^2 |I_n|}{9^3 \left(\log_2 \frac{e}{|I_n|}\right)^{8/5}} \exp\left\{\log \frac{1}{|I_n|} + R_n \left(\frac{R_n}{2} + \sqrt{2\log \frac{1}{|I_n|}}\right)\right\} \\ &= \frac{\epsilon_n^2}{9^3 \left(\log_2 \frac{e}{|I_n|}\right)^{8/5}} \exp\left\{R_n \left(\frac{R_n}{2} + \sqrt{2\log \frac{1}{|I_n|}}\right)\right\} \\ &\geq C \epsilon_n^2 \left(\log \frac{e}{|I_n|}\right) \exp\left(\frac{R_n^2}{2}\right) 1(R_n \geq 1) \text{ for some universal } C > 0 \\ &\stackrel{a.s.}{\rightarrow} \infty \quad \text{since } R_n \stackrel{a.s.}{\rightarrow} \infty \text{ and } \epsilon_n \sqrt{\log \frac{e}{|I_n|}} \geq 1 \text{ eventually.} \end{aligned}$$

In the case  $\ell > \ell_{\max}$ , the same conclusion obtains by using  $\mathcal{I}_{small}$  in place of  $\mathcal{I}_{app}(\ell)$ . The claim then follows since the critical value of  $A_{n,cond}$  stays bounded, by (7.11).

Thus the crucial difference with Theorem 2 is the stronger inequality provided by Lemma 2. The corresponding inequality for Theorem 2 is  $\#\mathcal{J}(I_n)/n^2 \geq \epsilon_n^2 |I_n|^2/4^2$ , and the extra term  $|I_n|$  causes the loss of efficiency in the case when the spatial extent  $|I_n|$  is small.

It remains to prove Lemma 2. Elementary considerations show that one can find sets  $\mathcal{I}_{n,left}$  and  $\mathcal{I}_{n,right}$ , each consisting of  $p := \lceil \epsilon_n n |I_n|/8 \rceil$  consecutive integers, such that  $(j, k) \in \mathcal{I}_{n,left} \times \mathcal{I}_{n,right}$  implies  $(j/n, k/n) \in \mathcal{J}(I_n)$ , and also  $m_\ell < k - j \leq 2m_\ell$  if  $\ell \leq \ell_{\max}$ , resp.  $k - j \leq m_{\ell_{\max}}$  if  $\ell > \ell_{\max}$ . Thus, in the latter case, we immediately obtain  $\#(\mathcal{J}(I_n) \cap \mathcal{I}_{small}) \geq p^2$ , and the claim of the Lemma obtains with (7.10),  $n|I_n| \geq 1$ , and  $\log n \leq (4/3) \log_2 e/|I_n|$ , which follows from  $\ell > \ell_{\max}$ . In the case  $\ell \leq \ell_{\max}$ , only a subset of the  $p^2$  intervals  $\{(j/n, k/n) : (j, k) \in \mathcal{I}_{n,left} \times \mathcal{I}_{n,right}\}$  belongs to  $\mathcal{J}(I_n) \cap \mathcal{I}_{app}(\ell)$ , namely those for which both  $j$  and  $k$  lie on the  $d_\ell$ -grid. The number of such indices  $j$  is at least  $\lfloor p/d_\ell \rfloor \geq \lceil \epsilon_n \sqrt{n|I_n|} (\log n)/(9 \ell^{4/5}) \rceil$  for  $n$  large enough, where this inequality follows from the fact that  $\ell \leq \ell_{\max}$  implies  $n|I_n| > m_\ell \geq (\log n)/2$  and  $d_\ell \geq (\log n)^{3/10}$ ; further,  $\epsilon_n \sqrt{\log n} \rightarrow \infty$  by the definition of  $\epsilon_n$ . The same bound obtains for the number of indices  $k$  that lie on the  $d_\ell$ -grid. The claim of the Lemma then follows with (7.10) and  $\ell \leq \log_2 e/|I_n|$ , which is a consequence of the definition of  $\ell$ .

## Acknowledgement

The first author was supported by NUS grant R-155-000-090-112. The second author was supported by NSF grant DMS-1007722 and NIH grant AI077395.

## References

- Arias-Castro, E., Donoho, D. L. and Huo, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory* **51**, 2402-2425.
- Burnashev, M. V. and Begmatov, I. A. (1990). On a problem of detecting a signal that leads to stable distributions. *Theory Probab. Appl.* **35**, 556-560.
- Cai, T., Jeng, X. J. and Li, H. (2011). Robust detection and identification of sparse segments in ultra-high dimensional data analysis. Manuscript.
- Chan, H. P. (2009). Detection of spatial clustering with average likelihood ratio test statistics. *Ann. Statist.* **37**, 3985-4010.
- Chan, H. P. and Walther, G. (2011). Detection with the scan and the average likelihood ratio. Manuscript. January 2011.
- Chan, H. P. and Zhang, N. R. (2009). Local average likelihood ratio test statistics with applications in genomics and change-point detection. Manuscript.
- Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Ann. Statist.* **26**, 288-314.
- Dümbgen, L. and Spokoiny, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29**, 124-152.

- Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36**, 1758-1785.
- Gangnon, R. E. and Clayton, M. K. (2001). The weighted average likelihood ratio test for spatial disease clustering. *Statist. Medicine* **20**, 2977-2987.
- Glaz, J. and Balakrishnan, N. (ed.) (1999). *Scan statistics and Applications*. Birkhäuser, Boston
- Glaz, J., Naus, J. and Wallenstein, S. (2001). *Scan Statistics*. Springer, New York
- Glaz, J., Poznyakov, V., and Wallenstein, S. (ed.) (2009). *Scan Statistics: Methods and Applications*. Birkhauser, Boston
- Kabluchko, Z. (2008). Extreme-value analysis of standardized Gaussian increments. Manuscript.
- Kulldorff, M. (1997). A spatial scan statistic. *Commun. Statist. - Theory Meth.* **26**, 1481-1496.
- Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields* **117**, 17-48.
- Neill, D. (2009). An empirical comparison of spatial scan statistics for outbreak detection. *Internat. J. Health Geographics* **8**, 1-16.
- Neill, D. and Moore, A. (2004). A fast multi-resolution method for detection of significant spatial disease clusters. *Adv. Neur. Info. Proc. Sys.* **10**, 651-658.
- Rohde, A. (2008). Adaptive goodness-of-fit tests based on signed ranks. *Ann. Statist.* **36**, 1346-1374.
- Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *J. Comput. Graph. Statist.* **19**, 175-190.
- Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8**, 22-46.
- Siegmund, D. (2001). Is peak height sufficient? *Genetic Epidemiology* **20**, 403-408.
- Siegmund, D. and Venkatraman, E. S. (1995). Using the generalized likelihood ratio statistic for sequential detection of a change-point. *Ann. Statist.* **23**, 255-271.
- Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38**, 1010-1033.

Department of Statistics and Applied Probability, National University of Singapore, Singapore 119260, Republic of Singapore.

E-mail: stachp@nus.edu.sg

Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford University, Stanford, CA 94305-4065, USA.

E-mail: walther@stat.stanford.edu

(Received July 2011; accepted November 2011)