

# Supplement to

## *Multiscale scanning with nuisance parameters*

CLAUDIA KÖNIG<sup>1</sup>, AXEL MUNK<sup>123</sup>, and FRANK WERNER<sup>45</sup>

### S.1 Implementation

As already mentioned in the introduction (Section 1), for a generic set  $\mathcal{R}_n$  all local test statistics  $T_R$  in (3) have to be computed individually. To this end, for example, the approximating rectangular system from Walther (2010) can be used to compute local averages on these. Depending on the structure of  $\mathcal{R}_n$ , other efficient computational schemes can be employed. In the following we will focus on the situation that there is a global shape  $B \subset I_n^d$  such that every  $R \in \mathcal{R}_n$  is a rescaled and shifted version of  $B$ . More precisely, for each  $R \in \mathcal{R}_n$  there exist  $t, h \in I_n^d$  with  $t_i + h_i \leq n$  for all  $1 \leq i \leq d$  such that

$$\mathbf{1}_R(x) = \mathbf{1}_B\left(\frac{x-t}{h}\right), \quad x \in I_n^d, \quad (1)$$

where division is meant component-wise and  $\mathbf{1}$  denotes the indicator function. For the set  $\mathcal{R}_n$  of all hyper-rectangles, this is for example the case with  $B = \mathbf{1}_{I_n^d}$ .

If  $R$  obeys (1), we obtain

$$\bar{Y}_R = \frac{1}{|R|} \sum_{i \in R} Y_i = \frac{1}{|R|} \sum_{i \in I_n^d} Y_i \mathbf{1}_B\left(\frac{i-t}{h}\right) = \frac{1}{|R|} \left( Y * \mathbf{1}_B\left(\frac{\cdot}{h}\right) \right)(t),$$

---

<sup>1</sup>Institute for Mathematical Stochastics, University of Göttingen, Germany

<sup>2</sup>Felix Bernstein Institute for Mathematical Statistics in the Bioscience, University of Göttingen, Germany

<sup>3</sup>DFG Cluster of Excellence "Multiscale Bioimaging - From Molecular Machines to Networks of Excitable Cells", University of Göttingen Medical Center, Germany

<sup>4</sup>Julius-Maximilians-Universität Würzburg (JMU), Institute of Mathematics, Würzburg, Germany

<sup>5</sup>Corresponding author, frank.werner@uni-wuerzburg.de

where  $*$  denotes a discrete convolution. Employing the fast Fourier transform (FFT), see Cooley and Tukey (1965), this allows to compute all  $T_R$  with a fixed scale  $h$  by means of three FFTs via

$$(\bar{Y}_R)_h = \frac{1}{h^d |B|} \text{FFT}^{-1} \left( \text{FFT}(Y) \cdot \text{FFT} \left( \mathbf{1} \left( \frac{\cdot}{h} \right) \right) \right).$$

This operation has a computational complexity almost linear in the data, i.e.  $\mathcal{O}(dn^d \log(n))$ , and hence if each  $R \in \mathcal{R}_n$  obeys (1), then  $T_n$  in (6) can be computed within  $\mathcal{O}(dNn^d \log(n))$  where  $N$  is the number of different scales. The same holds true for the evaluation of  $M_n$  in (9). A corresponding implementation in MATLAB<sup>®</sup> is available under <https://go.uniwue.de/math-ams>.

## S.2 Simulations of level and power in a homogeneous Gaussian model

At first, we present empirical quantiles of  $M_n$  in (9) for the situation considered in 4.2 and the empirical density and CDF both for the calibrated and the uncalibrated statistic, see Table S.1 and Figure S.1 below.

Table S.1: Empirical quantiles  $q_{1-\alpha}$  with different values of  $\alpha$  for the distribution of  $M_n$  in (9) in the case  $n = 128, d = 2$  and  $\mathcal{R}_n$  being the set of all rectangles in  $I_n^d$  with side-lengths between 4 and 14.

$\alpha$	0.2	0.1	0.05	0.025	0.01
$q_{1-\alpha}$	1.2906	1.4677	1.6278	1.7841	1.9768

Secondly, we consider a homogeneous Gaussian model as already discussed in the introduction (see Section 1), i.e.  $F_{\theta, \xi}$  is a normal distribution  $\mathcal{N}(\mu, \sigma^2)$ , and the variance  $\sigma^2$

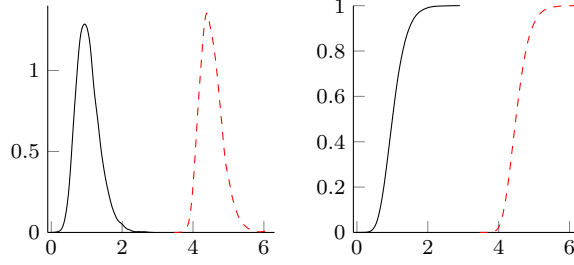


Figure S.1: Empirical density (left) and CDF (right) of  $M_n$  in (9) for  $n = 128$  in  $d = 2$  dimensions with  $\mathcal{R}_n$  being the set of all rectangles in  $I_n^d$  with even side-lengths between 4 and 14 pixels, displayed by a standard kernel density estimator, drawn from 2000 Monte Carlo runs, either calibrated (—) with penalization functions  $\tilde{\omega}_n$  and  $\omega_n$  as in (13), or uncalibrated (---).

is considered as the nuisance parameter and estimated by the sample variance

$$\hat{\sigma}_n^2 := \frac{1}{n^d - 1} \sum_{i \in I_n^d} (Y_i - \bar{Y}_{I_n^d})^2.$$

Then (16) is clearly satisfied under  $\mathbb{P}_0$  with  $s_n = n^{d/2}$ .

The results of the same simulations as described in Section 4.2 are depicted in Figure S.2. Sub-panels (a) and (b) display the signal and the corresponding data for the default setup ( $d = 2$ ,  $n = 128$ ) with an anomaly of size  $8 \times 8$  pixels and amplitude  $\mu = 0.5$  in the centre, where the variance  $\sigma^2 = 1$ . Sub-panel (c) shows the empirical level for different variances  $\sigma^2$ . We find that the level is kept quite stable over a large range of variances. Sub-panel (d) shows the empirical power for  $\sigma^2 = 1$  against different means  $\mu$ . Notably, there is nearly no power loss of the AMS method compared to the oracle caused by adaptation for an unknown  $\sigma^2$ . Sub-panel (e) is devoted to simulations in dimension  $d = 1$  with  $\sigma^2 = 1$  and different values of  $n \in \{32, 64, 96, 128\}$ . The lines depict that value of  $\mu$ , such that the AMS procedure is able to detect a present anomaly of size  $a$  marked on the  $x$ -axis with amplitude  $\mu$  with a power of  $\approx 90\%$ . We find that a larger anomaly clearly corresponds to a smaller amplitude required for detection. This is in agreement with the asymptotic separation rate in (19), and remarkably, despite the presence of nuisance

parameters and the additional difficulty that the length  $a$  is not known, the finite sample shapes of all separation lines in sub-panel (e) are of this form. Furthermore, we find that the influence of the sample size  $n$  on the separation lines is small. Sub-panel (f) shows the same situation in dimension  $d = 2$ . Due to the larger anomaly, which now contains  $a^2$  pixels, the corresponding values for  $\mu$  are chosen smaller. However, the conclusions drawn for the one-dimensional case apply here similarly. Compared with sub-panel (e) we furthermore conclude that the dimension  $d$  does not have a big influence, and again we see a good agreement with the asymptotic separation rate in (19).

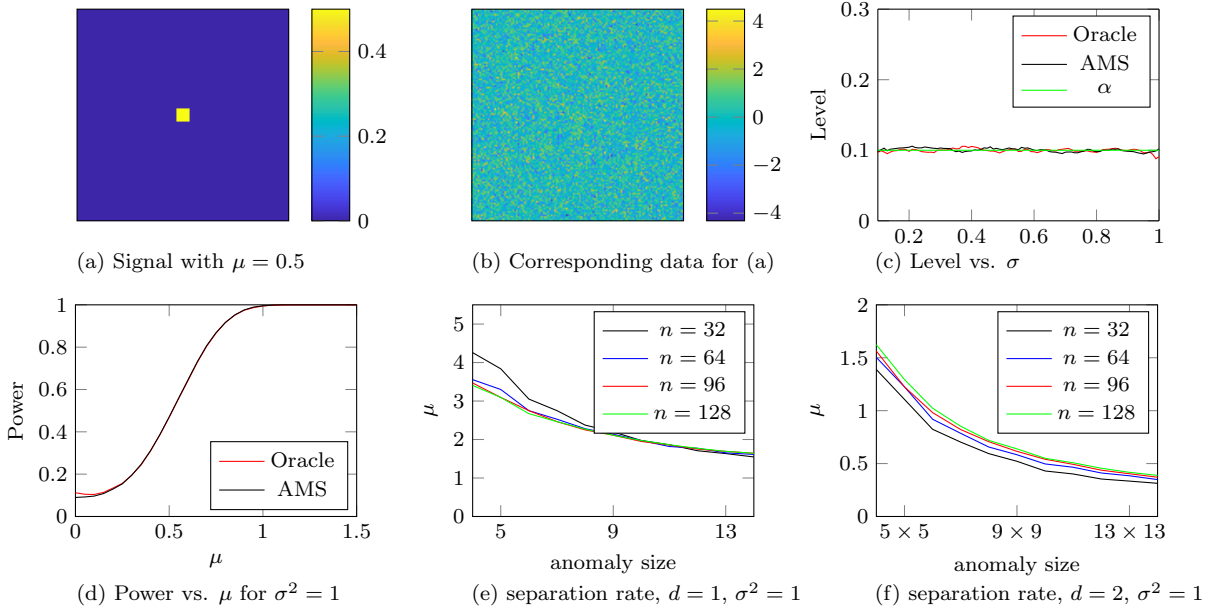


Figure S.2: Simulation results in a homogeneous Gaussian model with  $\alpha = 0.1$ . Panels (a) and (b): Signal under the alternative ( $\mu = 0.5$ ) used for this simulation and corresponding data for  $\sigma^2 = 1$ . As the signal is standardized, the standard deviation  $\sigma$  is proportional to the signal-to-noise ratio (SNR). Panel (c): Empirical level for the oracle multiscale scan ( $\sigma^2$  known) and AMS ( $\sigma^2$  unknown, i.e. considered as a nuisance parameter) for different values of  $\sigma$ , smoothed with a moving average filter of width 10. Panel (d): Empirical power versus  $\mu$  for a  $8 \times 8$  anomaly as in (a), smoothed with a moving average filter of width 10. Panels (e) and (f): Value of  $\mu$  required to obtain an empirical power of  $\beta = 0.9$  versus size of a (squared) anomaly with amplitude  $\mu$ .

## S.3 Additional data examples

Similar to the setting in Section 5.4 we investigate two further data sets.

### S.3.1 Data example 1: Single sized crimson beads

As a first data example, we investigate 48nm crimson beads, i.e. carboxylate modified microspheres of 48nm diameter filled with fluorescent markers. Taking into account the microscopes effective resolution of about 60nm, we expect to see circular anomalies of size  $\sim 110\text{nm}$  in the data. Therefore we choose  $r_n$  and  $m_n$  such that the smallest box-size is 80nm, and the largest box-size is 400nm. Since each pixel has size  $20\text{nm} \times 20\text{nm}$ , this corresponds to  $r_n = 4$  and  $m_n = 20$ . The measured data for an illumination time of  $t = 5\mu\text{s}$  and  $T = 100\mu\text{s}$  are shown in Figure S.3(a) and (d). The corresponding results for AMS are shown in sub-panels (b) and (e). Surprisingly, the test detects most beads based on the  $5\mu\text{s}$  data, even though some of them are barely visible by eye there. Regarding the  $T = 100\mu\text{s}$  data as ground truth, we depict this together with the regions found in (b) in sub-panel (e), revealing that there no false positive detections. Finally, the  $T = 100\mu\text{s}$  data can be used to derive a segmentation into active and inactive regions depicted in (f).

### S.3.2 Data example 2: mixture of differently sized crimson beads

As a second data example, we now consider a mixture of 48nm and 200nm crimson beads. Consequently, the data shown in Figure S.4(a) and (d) consists of two structures of different sizes, namely around 48nm and 200nm. Due to the number of markers inside the spheres, the larger structures are significantly brighter than the small ones. We investigate  $t = 15\mu\text{s}$  and  $T = 100\mu\text{s}$ . The AMS procedure is applied with the same parameters as before (i.e. again  $m_n = 20$ , corresponding to a largest box size of 400nm), and the result is shown in sub-panels (b) and (e). It is immediately visible that it detects again nearly all of the small structures already from the measurements taken with  $15\mu\text{s}$  dwell time, which is surprising

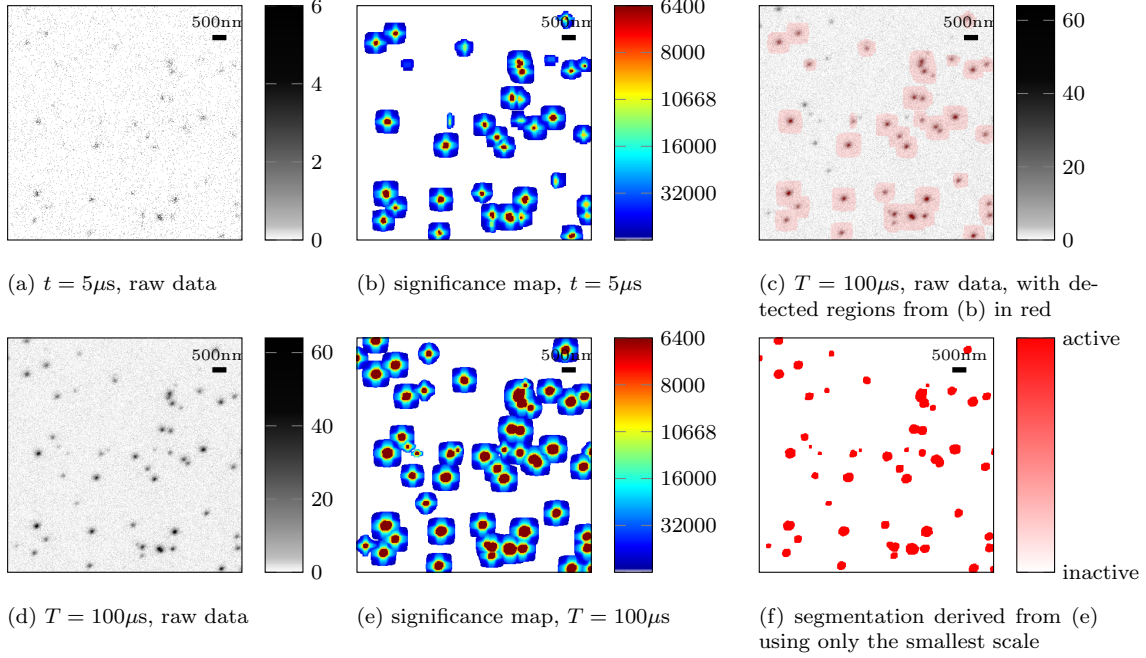


Figure S.3: Raw crimson bead data (photon counts, panels (a) and (d)) and corresponding significance maps with level  $\alpha = 0.1$  (smallest significance in  $\text{nm}^2$ , panels (b) and (d)). All significant boxes are plotted with colour indicating their size in  $\text{nm}^2$ . Due to the FWER control, with probability  $\geq 1 - \alpha = 0.9$ , all boxes marked contain anomalies. Using only the smallest scale significance, this leads to the segmentation in (f) showing active regions in red.

as some of them are not visible by eye in the data. Once again, regarding the  $T = 100\mu s$  data as ground truth, we depict this together with the regions found in (b) in sub-panel (e), revealing that there no false positive detections. Finally, the  $T = 100\mu s$  data can be used to derive a segmentation into active and inactive regions depicted in (f).

## S.4 Results for different values of $\alpha$

We also compare the resulting significance maps for different values of  $\alpha$  for the data set from Section 5.4, see Figure S.5. It turns out that the AMS methodology is remarkably stable w.r.t. the significance level  $\alpha$ .

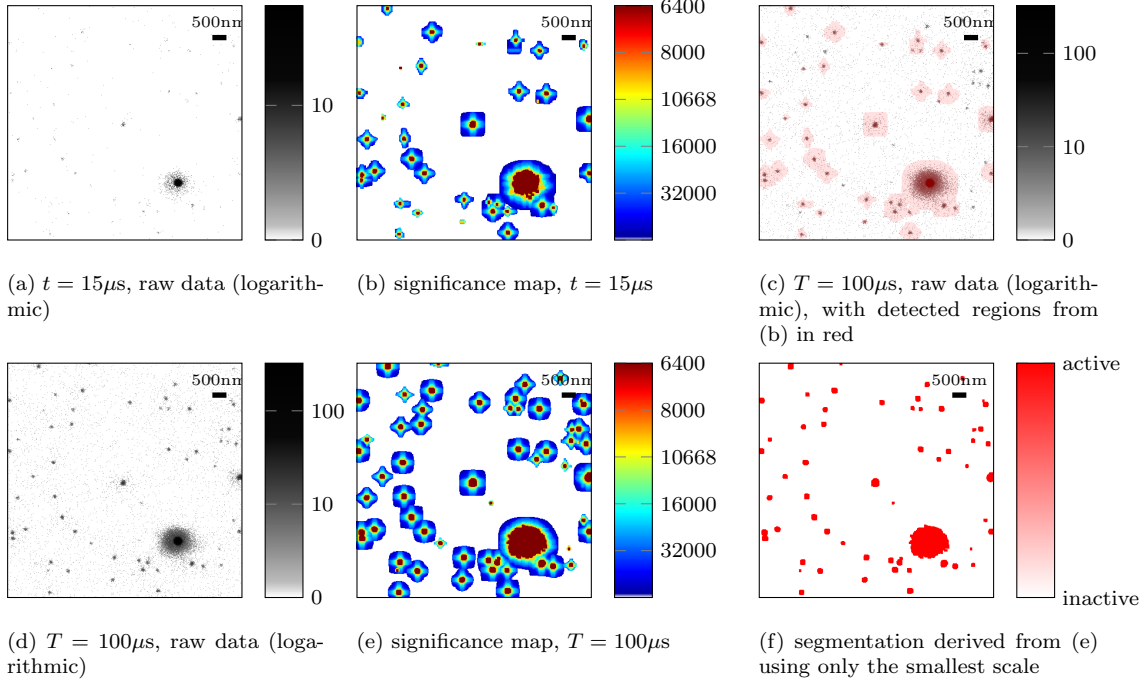


Figure S.4: Raw crimson bead data (photon counts, logarithmic scale, panels (a) and (d)) and corresponding significance maps with level  $\alpha = 0.1$  (smallest significance in  $\text{nm}^2$ , panels (b) and (e)). All significant boxes are plotted with colour indicating their size in  $\text{nm}^2$ . Due to the FWER control, with probability  $\geq 1 - \alpha = 0.9$ , all boxes marked contain anomalies. Using only the smallest scale significance, this leads to the segmentation in (f) showing active regions in red.

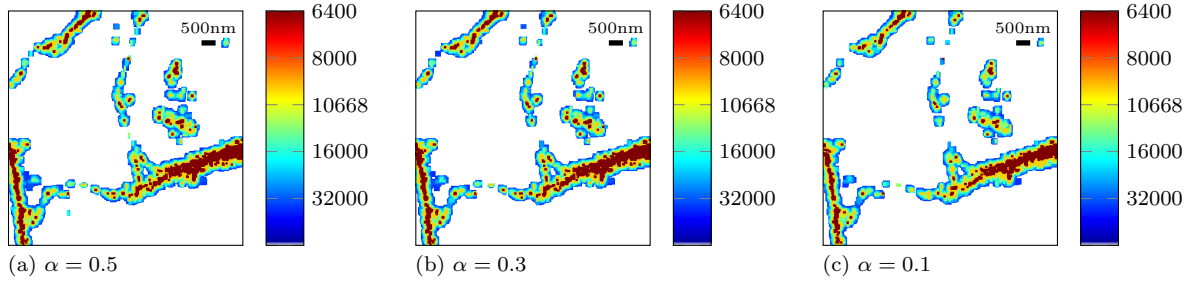


Figure S.5: Significance maps from the raw tubulin data in Figure 4(a) for different values of  $\alpha$ .

## S.5 Proofs

We start with some helpful abbreviations. For the whole section, let  $X = (X_i)_{i \in I_n^d}$  be a field of i.i.d. standard Gaussians,  $X_i \sim \mathcal{N}(0, 1)$ . As we rely heavily on centred and

standardized partial sums, we introduce for  $R \in \mathcal{R}_n$  the quantity

$$Y_R(\theta, \xi) := |R|^{-1} \sum_{i \in R} \left( \frac{Y_i - m(\theta, \xi)}{\sqrt{v(\theta, \xi)}} \right) = \frac{\bar{Y}_R - m(\theta, \xi)}{\sqrt{v(\theta, \xi)}}, \quad (\theta, \xi) \in \Theta \times \Xi$$

with the mean and variance functions  $m$  and  $v$  from Assumption 1. Also recall our abbreviation

$$\bar{X}_R := \frac{1}{|R|} \sum_{i \in R} X_i.$$

### S.5.1 Preparations

Next we recall a series of helpful results taken from König et al. (2020).

**Lemma 1** (Coupling). *Let Assumptions 1(a) and 2 hold true, and let  $Y = (Y_i)_{i \in I_n^d}$  be an array of i.i.d. random variables  $Y_i \sim F_{\theta, \xi}$  with  $\theta \in U, \xi \in V$  and let  $(r_n)_{n \in \mathbb{N}} \subset \mathbb{N}$  be a sequence. Then, on the same probability space, there exists an array  $X = (X_i)_{i \in I_n^d}$  of i.i.d. standard Gaussians  $X_i \sim \mathcal{N}(0, 1)$  such that*

$$\mathbb{P}_{0,n}^* \left[ \left| \max_{R \in \mathcal{R}_n(r_n, n^d)} |R|^{\frac{1}{2}} |Y_R(\theta, \xi)| - \max_{R \in \mathcal{R}_n(r_n, n^d)} |R|^{\frac{1}{2}} |\bar{X}_R| \right| > \delta \right] \leq C \delta^{-3} \left( \frac{\log^{10} n}{r_n} \right)^{1/2}$$

for all  $\delta > 0$  with some universal constant  $C > 0$ .

*Proof.* As the distributions  $F_{\theta, \xi}$  have uniformly sub-exponential tails, this follows directly from Theorem 4.3 in König et al. (2020) in combination with a symmetrization argument as employed in the proof of Theorem 2.5 there.  $\square$

Note, that in terms of  $\mathcal{O}$  notation, Lemma 1 yields

$$\max_{R \in \mathcal{R}_n(r_n, n^d)} |R|^{\frac{1}{2}} |Y_R(\theta, \xi)| - \max_{R \in \mathcal{R}_n(r_n, n^d)} |R|^{\frac{1}{2}} |\bar{X}_R| = \mathcal{O}_{\mathbb{P}_{0,n}^*}^* \left( \left( \frac{\log^{10}(n)}{r_n} \right)^{\frac{1}{6}} \right), \quad (2)$$

uniformly in  $(\theta, \xi) \in U \times V$ .



**Lemma 2** (Taylor expansion, see Lemma 5.1 in König et al. (2020)). *Let  $Y = (Y_i)_{i \in I_n^d}$  be an array of i.i.d. random variables  $Y_i \sim F_{\theta, \xi}$  and let  $(r_n)_{n \in \mathbb{N}} \subset \mathbb{N}$  be a sequence satisfying (17). If Assumptions 1(a) and 2 hold true, then*

$$\max_{R \in \mathcal{R}_n(r_n, n^d)} \left| T_R(Y, \theta, \xi) - |R|^{\frac{1}{2}} Y_R(\theta, \xi) \right| = \mathcal{O}_{\mathbb{P}_{0,n}^*} \left( \left( \frac{\log^3(n)}{r_n} \right)^{1/4} \right)$$

*uniformly for  $\theta \in U, \xi \in V$ .*

Note, that the previous results remain true if we replace the upper bound  $n^d$  for the scale size by  $m_n$  with any sequence  $(m_n)_{n \in \mathbb{N}} \subset \mathbb{N}$  satisfying  $r_n \leq m_n$  for all  $n \in \mathbb{N}$ .

**Lemma 3** (Uniform remainder). *Let Assumptions 1(a) and 2 hold true, and let  $Y = (Y_i)_{i \in I_n^d}$  be an array of i.i.d. random variables  $Y_i \sim F_{\theta, \xi}$  with fixed  $\theta \in U, \xi \in V$ . If  $(r_n)_{n \in \mathbb{N}} \subset \mathbb{N}$  is a sequence satisfying (17), then we obtain*

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} |R| |Y_R(\theta_0, \xi)|^3 = \mathcal{O}_{\mathbb{P}_{0,n}^*} \left( \left( \frac{\log^3(n)}{r_n} \right)^{\frac{1}{2}} \right)$$

*uniformly for  $\theta \in U, \xi \in V$ .*

*Proof.* Note, that due to  $\alpha, \tilde{\alpha} > 0$ , (17) implies especially that  $\gamma \geq 4$  and hence  $\log^4(n) = o(r_n)$ .

It is well known (see e.g. Kabluchko (2011)), that for a standard Gaussian array  $X = (X_i)_{i \in I_n^d}$ ,  $X_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ , one has

$$\mathbb{E} \left[ \max_{R \in \mathcal{R}_n(r_n, m_n)} |R|^{\frac{1}{2}} |\bar{X}_R| \right] \leq C \sqrt{\log(\#\mathcal{R}_n(r_n, m_n))} \quad (3)$$

with some constant  $C > 0$ . Due to Assumption 2 this implies that

$$\frac{1}{\sqrt{\log(n)}} \max_{R \in \mathcal{R}_n(r_n, m_n)} |R|^{\frac{1}{2}} |\bar{X}_R| = \mathcal{O}_{\mathbb{P}}(1).$$

Together with (2) this yields

$$\frac{1}{\sqrt{\log(n)}} \max_{R \in \mathcal{R}_n(r_n, m_n)} |R|^{\frac{1}{2}} |Y_R(\theta, \xi)| = \mathcal{O}_{\mathbb{P}_{0,n}^*}(1)$$

uniformly for  $(\theta, \xi) \in U \times V$ , where we used that  $\log^4(n) = o(r_n)$ . Consequently we find

$$\sqrt{\frac{r_n}{\log^3(n)}} \max_{R \in \mathcal{R}_n(r_n, m_n)} |R| |Y_R(\theta_0, \xi)|^3 \leq \sqrt{\frac{1}{\log^3(n)}} \max_{R \in \mathcal{R}_n(r_n, m_n)} |R|^{\frac{3}{2}} |Y_R(\theta_0, \xi)|^3 = \mathcal{O}_{\mathbb{P}_{0,n}^*}$$

uniformly for  $(\theta, \xi) \in U \times V$ , which proves the claim.  $\square$

## S.5.2 An adjusted Taylor expansion

Now we are in position to derive some preparations for the proof of Theorem 1. Therefore we first replace the likelihood ratio statistic  $T_R$  by its Taylor expansion according to Assumption 1(e):

**Theorem 1.** *Let  $(r_n)_{n \in \mathbb{N}} \subset \mathbb{N}$  and  $(m_n)_{n \in \mathbb{N}} \subset \mathbb{N}$  be sequences tending to  $\infty$  and consider the local statistic  $T_R$  from (3). Suppose that  $Y$  is as in (1) and let Assumptions 1 and 2 be fulfilled. If the estimators  $\hat{\theta}_n$  and  $\hat{\xi}_n$  satisfy (16), then*

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R(Y, \hat{\theta}_n, \hat{\xi}_n) - |R|^{1/2} |Y_R(\theta_0, \xi)| \right| = \mathcal{O}_{\mathbb{P}_{0,n}^*} \left( \left( \frac{\log^7(n)}{r_n} \right)^{\frac{1}{4}} + \sqrt{\frac{m_n}{s_n}} \right). \quad (4)$$

*Proof.* Let  $U \times V$  be the neighbourhood of  $(\theta_0, \xi)$  as in Assumption 1. By (16), the probability that  $\hat{\theta}_n \in U$  and  $\hat{\xi}_n \in V$  tends to 1, as  $n \rightarrow \infty$ . Hence, it suffices to prove the result conditional on this event. Consequently we can assume  $\hat{\theta}_n \in U$  and  $\hat{\xi}_n \in V$  in the following.

By the triangle inequality and Lemma 2, we find

$$\begin{aligned}
& \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R(Y, \hat{\theta}_n, \hat{\xi}_n) - |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| \right| \\
& \leq \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R(Y, \theta_0, \xi) - |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| \right| \\
& \quad + \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R(Y, \hat{\theta}_n, \hat{\xi}_n) - T_R(Y, \theta_0, \xi) \right| \\
& \leq \sqrt{\max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y, \hat{\theta}_n, \hat{\xi}_n) - T_R^2(Y, \theta_0, \xi) \right|} + \mathcal{O}_{\mathbb{P}_{0,n}^*}^* \left( \left( \frac{\log^3(n)}{r_n} \right)^{\frac{1}{4}} \right),
\end{aligned}$$

where we exploited Lemma 2 and  $|a - b| \leq \sqrt{|a^2 - b^2|}$ . By the triangle inequality and Assumption 1(e) we find furthermore that

$$\begin{aligned}
& \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y, \hat{\theta}, \hat{\xi}) - T_R^2(Y, \theta_0, \xi) \right| \\
& \leq \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y, \hat{\theta}, \hat{\xi}) - |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n) \right|^2 \right| + \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| T_R^2(Y, \theta_0, \xi) - |R| \left| Y_R(\theta_0, \xi) \right|^2 \right| \\
& \quad + \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n) \right|^2 - |R| \left| Y_R(\theta_0, \xi) \right|^2 \right| \\
& \leq C_T \max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n) \right|^3 + C_T \max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| Y_R(\theta_0, \xi) \right|^3 \\
& \quad + \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n) \right|^2 - |R| \left| Y_R(\theta_0, \xi) \right|^2 \right| \\
& \leq 2C_T \max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| Y_R(\theta_0, \xi) \right|^3 + \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n) \right|^2 - |R| \left| Y_R(\theta_0, \xi) \right|^2 \right| \\
& \quad + C_T \max_{R \in \mathcal{R}_n(r_n, m_n)} \left| |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n) \right|^3 - |R| \left| Y_R(\theta_0, \xi) \right|^3 \right|,
\end{aligned}$$

where we used  $||a|^3 - |b|^3| = ||a^3| - |b^3|| \leq |a^3 - b^3|$  for  $a, b \in \mathbb{R}$ . In view of Assumptions 1(a) and 2, the first term in the last display can be bounded using Lemma 3. The second and third term will be handled using the mean value theorem. Therefore, note that all derivatives of

$$(\theta, \xi) \mapsto Y_R(\theta, \xi)^k, \quad k \in \{2, 3\}$$

are a.s. bounded due to Assumption 1(b)–(d) and the fact that  $\bar{Y}_R$  is a.s. bounded in view of Lemma 3. This together with (16) implies that

$$\max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| Y_R(\hat{\theta}, \hat{\xi})^2 - Y_R(\theta_0, \xi)^2 \right| \leq m_n C \left( \|\hat{\theta}_n - \theta_0\|_2 + \|\hat{\xi}_n - \xi\|_2 \right) = \mathcal{O}_{\mathbb{P}_{0,n}^*} \left( \frac{m_n}{s_n} \right),$$

where  $C$  is some generic constant  $C > 0$ . As  $\max_{R \in \mathcal{R}_n(r_n, m_n)} |R| \left| Y_R(\hat{\theta}_n, \hat{\xi}_n)^3 - Y_R(\theta_0, \xi)^3 \right|$  can be treated similarly, this yields the claim.  $\square$

### S.5.3 Proof of Theorem 1

Now we are ready to prove the Gaussian approximation result from Theorem 1.

*Proof of Theorem 1.* For notional simplicity, we will throughout this proof drop the index of  $\tilde{\omega}_n$  and  $\omega_n$ . Furthermore we abbreviate  $\tilde{\mathcal{R}}_n := \mathcal{R}_n(r_n, m_n)$ . First of all, we estimate

$$\begin{aligned} & \left| T_n(Y, \tilde{\mathcal{R}}_n, \hat{\theta}, \hat{\xi}) - M_n(\tilde{\mathcal{R}}_n) \right| \\ &= \left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( T_R(Y, \hat{\theta}_n, \hat{\xi}_n) - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \right| \\ &\leq \left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( T_R(Y, \hat{\theta}_n, \hat{\xi}_n) - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \omega(|R|) \right) \right| \\ &\quad + \left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \right| \\ &\leq \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \max_{R \in \tilde{\mathcal{R}}_n} \left| T_R(Y, \hat{\theta}_n, \hat{\xi}_n) - |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| \right| \\ &\quad + \left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \right|, \end{aligned}$$

where we used  $|||x||_\infty - ||y||_\infty| \leq \|x - y\|_\infty$ . The first term can be controlled using As-

sumption 3 and Theorem 1 by

$$\begin{aligned} \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \max_{R \in \tilde{\mathcal{R}}_n} \left| T_R \left( Y, \hat{\theta}_n, \hat{\xi}_n \right) - |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| \right| \\ = \mathcal{O}_{\mathbb{P}_{0,n}^*} \left( \frac{\log^{\tilde{\alpha}}(n)}{r_n} \left( \left( \frac{\log^3(n)}{r_n} \right)^{\frac{1}{4}} + \sqrt{\frac{m_n}{s_n}} \right) \right). \end{aligned}$$

So it remains to estimate the second term, i.e. for a suitable sequence  $(c_n)_{n \in \mathbb{N}}$  we want to show that

$$\begin{aligned} \lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}_0 \left[ \left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| \right. \right. \right. \\ \left. \left. \left. - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \right| > c_n M \right] = 0 \quad (5) \end{aligned}$$

This is more subtle, as the coupling in Lemma 1 only yields a result for the difference of the maxima, and not for the maximum of the differences (as Theorem 1 does). We therefore borrow a slicing technique from Proksch et al. (2018), dividing the set of scales into families on which  $\tilde{\omega}$  and  $\omega$  are almost constant, i.e. into

$$\mathcal{R}_{n,j} := \{R \in \mathcal{R}_n \mid \exp(\epsilon_j) \leq |R| < \exp(\epsilon_{j+1})\}.$$

Here and in what follows we choose

$$\epsilon_1 := \log(r_n), \quad \epsilon_j = \epsilon_1 + \frac{j-1}{J} \log\left(\frac{m_n}{r_n}\right), \quad j = 2, \dots, J.$$

With this definition, for any choice  $J \in \mathbb{N}_{\geq 2}$  we obtain

$$\tilde{\mathcal{R}}_n = \mathcal{R}_n(r_n, m_n) = \{R \in \mathbb{R}_n \mid r_n \leq |R| \leq m_n\} = \bigcup_{j \in J} \mathcal{R}_{n,j}.$$

The parameter  $J \in \mathbb{N}$  will be defined later. On  $\mathcal{R}_{n,j}$  we can approximate  $\tilde{\omega}(|R|)$  and

$\omega(|R|)$  by

$$\tilde{\omega}_{j,n} := \tilde{\omega}(\exp(\epsilon_j)) \quad \text{and} \quad \omega_{j,n} := \omega(\exp(\epsilon_j)),$$

respectively. Then it holds

$$\omega_{j+1,n} \leq \omega(|R|) \leq \omega_{j,n} \quad \text{and} \quad \tilde{\omega}_{j+1,n} \leq \tilde{\omega}(|R|) \leq \tilde{\omega}_{j,n}$$

for all  $R \in \mathcal{R}_{n,j}$ . Now we compute

$$\begin{aligned} & \max_{R \in \mathcal{R}_{n,j}} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \omega(|R|) \right) - \max_{R \in \mathcal{R}_{n,j}} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \\ & \leq \left( \tilde{\omega}_{j,n} \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \tilde{\omega}_{j+1,n} \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |\bar{X}_R| \right) + (\tilde{\omega}_{j,n} \omega_{j,n} - \tilde{\omega}_{j+1,n} \omega_{j+1,n}) \\ & = \tilde{\omega}_{j,n} \left( \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |\bar{X}_R| \right) + (\tilde{\omega}_{j,n} - \tilde{\omega}_{j+1,n}) \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |\bar{X}_R| \\ & \quad + (\tilde{\omega}_{j,n} \omega_{j,n} - \tilde{\omega}_{j+1,n} \omega_{j+1,n}), \end{aligned}$$

and a similar relation holds true if the roles of  $\bar{X}_R$  and  $Y_R(\theta_0, \xi)$  are interchanged. To bound the differences involving the  $\tilde{\omega}_{j,n}$  terms, we exploit the rough bound

$$\tilde{\omega}_{j,n} = \tilde{\omega}(\exp(\epsilon_j)) \leq C_\omega \left( \log \left( \frac{n^d}{\exp(\epsilon_j)} \right) \right)^{\tilde{\alpha}} \leq C_\omega (\log(n^d))^{\tilde{\alpha}} =: A_n \quad (6)$$

for all  $1 \leq j \leq J$ , and a similar analogue for  $\omega_{j,n}$ . From the mean value theorem we get

$$\begin{aligned}
\tilde{\omega}_{j,n} - \tilde{\omega}_{j+1,n} &= \tilde{\omega}(\exp(\epsilon_{j+1})) - \tilde{\omega}(\exp(\epsilon_j)) \\
&\leq |\exp(\epsilon_j) - \exp(\epsilon_{j+1})| \max_{\zeta \in (\epsilon_j, \epsilon_{j+1})} \tilde{\omega}'(\exp(\zeta)) \\
&\leq C_\omega (\log(n^d))^{\max(\tilde{\beta}, 0)} (\exp(\epsilon_{j+1} - \epsilon_j) - 1) \\
&= C_\omega (\log(n^d))^{\max(\tilde{\beta}, 0)} \left( \exp\left(\frac{1}{J} \log\left(\frac{m_n}{r_n}\right)\right) - 1 \right) \\
&= C_\omega (\log(n^d))^{\max(\tilde{\beta}, 0)} \left( \left(\frac{m_n}{r_n}\right)^{\frac{1}{J}} - 1 \right) =: B_n
\end{aligned}$$

Similar statements hold for  $\omega_{j,n}$  and  $\omega_{j,n} - \omega_{j+1,n}$  and hence using (6) we obtain

$$\begin{aligned}
|\tilde{\omega}_{j,n}\omega_{j,n} - \tilde{\omega}_{j+1,n}\omega_{j+1,n}| &= |\tilde{\omega}_{j,n}(\omega_{j,n} - \omega_{j+1,n}) - \omega_{j+1,n}(\tilde{\omega}_{j,n} - \tilde{\omega}_{j+1,n})| \\
&\leq C_\omega^2 \left( \left(\frac{m_n}{r_n}\right)^{\frac{1}{J}} - 1 \right) \left( (\log(n^d))^{\tilde{\alpha} + \max\{\beta, 0\}} + (\log(n^d))^{\alpha + \max\{\tilde{\beta}, 0\}} \right) =: C_n
\end{aligned}$$

This implies

$$\begin{aligned}
&\left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \right| \\
&\leq A_n \max_{1 \leq j \leq J} \left| \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \max_{R \in \tilde{\mathcal{R}}_n} |R|^{\frac{1}{2}} |\bar{X}_R| \right| + B_n \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |\bar{X}_R| + C_n,
\end{aligned}$$

and hence

$$\begin{aligned}
&\mathbb{P}_{0,n}^* \left[ \left| \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \omega(|R|) \right) - \max_{R \in \tilde{\mathcal{R}}_n} \tilde{\omega}(|R|) \left( |R|^{\frac{1}{2}} |\bar{X}_R| - \omega(|R|) \right) \right| > c_n M \right] \\
&\leq \mathbb{P}_{0,n}^* \left[ A_n \max_{1 \leq j \leq J} \left| \max_{R \in \mathcal{R}_{n,j}} |R|^{\frac{1}{2}} |Y_R(\theta_0, \xi)| - \max_{R \in \tilde{\mathcal{R}}_{n,j}} |R|^{\frac{1}{2}} |\bar{X}_R| \right| > \frac{c_n M}{3} \right] \\
&\quad + \mathbb{P}_{0,n}^* \left[ B_n \max_{R \in \mathcal{R}_n} |R|^{\frac{1}{2}} |\bar{X}_R| > \frac{c_n M}{3} \right] + \mathbb{P}_{0,n}^* \left[ C_n > \frac{c_n M}{3} \right] \\
&=: \text{I} + \text{II} + \text{III}.
\end{aligned}$$

To bound I, we use Lemma 1 and the union bound to obtain

$$\text{I} \leq C |J| \frac{3^3 A_n^3}{c_n^3 M^3} \left( \frac{\log^{10}(n)}{r_n} \right)^{\frac{1}{2}}.$$

Here and in what follows,  $C > 0$  is some generic constant, the value of which can change from line to line. For II, we exploit Markov's inequality as well as (3) and find in view of Assumption 2, that

$$\text{II} \leq \frac{\mathbb{E} \left[ \max_{R \in \mathcal{R}_n} |R|^{\frac{1}{2}} |\bar{X}_R| \right]}{c_n M 3^{-1} B_n^{-1}} \leq C \frac{B_n \sqrt{\log(n)}}{c_n M}$$

with some constant  $C > 0$ . This shows that (5) is satisfied as soon as

$$(A) \quad |J| A_n^3 c_n^{-3} \log^5(n) r_n^{-\frac{1}{2}} = \mathcal{O}(1),$$

$$(B) \quad B_n c_n^{-1} \sqrt{\log(n)} = \mathcal{O}(1), \text{ and}$$

$$(B) \quad C_n = o(c_n)$$

hold true. To obtain this, we set

$$J := \lfloor \frac{\log^{\nu+1}(n)}{c_n} \rfloor$$

with some parameter  $\nu > 0$  to be determined. This implies

$$|J| \frac{A_n^3}{c_n^3} \left( \frac{\log^{10}(n)}{r_n} \right)^{\frac{1}{2}} \leq C \frac{\log^{6+3\tilde{\alpha}+\nu}}{c_n^4 \sqrt{r_n}},$$

which proves (B) for

$$c_n := \left( \frac{\log^{12+6\tilde{\alpha}+2\nu}(n)}{r_n} \right)^{\frac{1}{8}}. \tag{7}$$



Next we use  $m_n/r_n \leq n^d$  and use  $\exp(x) - 1 \leq 2x$  for  $x \in [0, 1]$  to find

$$\left(\frac{m_n}{r_n}\right)^{\frac{1}{J}} \leq n^{\frac{d}{J}} = \exp\left(\frac{dc_n}{\log^\nu(n)}\right) \leq 1 + C \frac{c_n}{\log^\nu(n)}$$

for any sufficiently large  $n$  supposed that  $c_n = o(\log^\nu(n))$ . Thus it holds

$$B_n \leq C \frac{c_n}{\log^{\nu - \max\{\tilde{\beta}, 0\}}(n)},$$

and hence (A) is satisfied if  $\nu \geq \max\{\tilde{\beta}, 0\} + \frac{1}{2}$ . It can readily be seen that (C) is now satisfied as soon as we define

$$\nu := \max\left\{\frac{1}{2}, \alpha, \tilde{\alpha}\right\} + \max\left\{\beta, \tilde{\beta}, 0\right\}.$$

Note that this also ensures the required condition  $c_n = o(\log^\nu(n))$  under (17). Hence (5) is satisfied with  $c_n$  as in (7), and thus we have

$$\begin{aligned} & \left| T_n(Y, \tilde{\mathcal{R}}_n, \hat{\theta}, \hat{\xi}) - M_n(\tilde{\mathcal{R}}_n) \right| \\ &= \mathcal{O}_{\mathbb{P}_{0,n}^*} \left( \frac{\log^{\tilde{\alpha}}(n)}{r_n} \left( \left( \frac{\log^3(n)}{r_n} \right)^{\frac{1}{4}} + \sqrt{\frac{m_n}{s_n}} \right) + \left( \frac{\log^{12+6\tilde{\alpha}+2\nu}(n)}{r_n} \right)^{\frac{1}{8}} \right), \end{aligned}$$

which together with  $\frac{\log^{\tilde{\alpha}+3/44}(n)}{r_n^{5/4}} = \mathcal{O}\left(\frac{\log^{\gamma/8}(n)}{r_n^{1/8}}\right)$  proves the claim.  $\square$

## References

Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19:297–301.

Kabluchko, Z. (2011). Extremes of the standardized Gaussian noise. *Stochastic Process*.

*Appl.*, 121(3):515–533.

König, C., Munk, A., and Werner, F. (2020). Multidimensional multiscale scanning in exponential families: limit theory and statistical consequences. *Ann. Statist.*, 48(2):655–678.

Proksch, K., Werner, F., and Munk, A. (2018). Multiscale scanning in inverse problems. *Ann. Statist.*, 46(6B):3569–3602.

Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.*, 38(2):1010–1033.