Journal of Statistical Planning and Inference 193 (2018) 151–178

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Optimal change point detection in Gaussian processes[☆]

Hossein Keshavarz[a,*], Clayton Scott[b,a], XuanLong Nguyen[a,b]

[a] Department of Statistics, University of Michigan, United States
[b] Department of Electrical Engineering and Computer Science, University of Michigan, United States

ARTICLE INFO

Article history:
Received 10 May 2017
Received in revised form 7 September 2017
Accepted 9 September 2017
Available online 22 September 2017

Keywords:
Change-point detection
Gaussian processes
Fixed domain asymptotic analysis
Minimax optimality

ABSTRACT

We study the problem of detecting a change in the mean of one-dimensional Gaussian process data in the fixed domain regime. We propose a detection procedure based on the generalized likelihood ratio test (GLRT), and show that our method achieves asymptotically near-optimal rate in a minimax sense. The notable feature of the proposed method is that it exploits in an efficient way the data dependence captured by the Gaussian process covariance structure. When the covariance is not known, we propose the plug-in GLRT method and derive conditions under which the method remains asymptotically near-optimal. By contrast, the standard CUSUM method, which does not account for the covariance structure, is shown to be suboptimal. Our algorithms and asymptotic analysis are applicable to a number of covariance structures, including the Matern class, the powered exponential class, and others. The plug-in GLRT method is shown to perform well for maximum likelihood estimators with a dense covariance matrix.

1. Introduction

Change point detection is the problem of detecting an abrupt change or changes arising in a sequence of observed samples. A common problem of this type involves detecting shifts in the mean of a temporal or spatial process. This problem has found a variety of applications in many fields, including audio analysis (Gillet et al., 2007), EEG segmentation (Lavielle, 2005), structural health monitoring (Noh et al., 2012; Hu et al., 2007) and environment sciences (Last and Shumway, 2008; Verbesselt et al., 2010). Despite advances in the development of algorithms (Kawahara and Sugiyama, 2009; Lavielle, 2005; Liu et al., 2010; Rigaill, 2010) and asymptotic theory (Bertrand et al., 2011; Tartakovsky et al., 2006; Shao and Zhang, 2010; Levy-leduc, 2007) for a number of contexts, such studies are mainly confined to the setting of (conditionally) independently distributed data. Existing works on optimal detection of shifts in the mean in temporal data with statistically dependent observations are far less common.

Incorporating dependence structures into the modeling of random processes is a natural approach. In fact, this has been considered in detecting changes of remotely collected data (Chandola and Vatsavai, 2011; Gabriel et al., 2011). For instance, Chandola and Vatsavai (2011) proposed a Gaussian process based algorithm to identify changes in Normalized Difference Vegetation Index (NDVI) time series for a particular location in California. It is therefore of interest to study how the dependence structures of the underlying process can be accounted for, e.g., its covariance function and spectral density, in designing statistically efficient detection procedures. In this paper, we shall focus on the detection of a single change in the mean of a Gaussian process data sequence.

☆ This research is partially supported by National Science Foundation (NSF) grant ACI-1047871. Additionally, CS is partially supported by NSF grants CCF-1422157, CCF-1217880, and CCF-0953135, and LN by NSF CAREER award DMS-1351362, NSF CNS-1409303, and NSF CCF-1115769.

* Corresponding author.
E-mail addresses: hksh@umich.edu (H. Keshavarz), clayscot@umich.edu (C. Scott), xuanlong@umich.edu (X. Nguyen).

https://doi.org/10.1016/j.jspi.2017.09.003
0378-3758/

Consider a simplified setting in which we let G be a Gaussian process on a domain $\mathcal{D} \subseteq \mathbb{R}$ and $\mathcal{D}_n := \{t_k\}_{k=1}^n \subset \mathcal{D}$ denote a finite index set of sampling points. Denote the observed samples by $\boldsymbol{X} = \{X_k\}_{k=1}^n$ in which $X_k = G(t_k)$ for $k = 1, \ldots, n$. Moreover, let $t \in \mathcal{C}_{n,\alpha} \subseteq \{1, \ldots, n\}$ (the parameter α is a positive scalar which will be introduced in Section 2.1) and $b > 0$ represents the point of sudden change and the jump/shift value, respectively. Namely, there is $\mu \in \mathbb{R}$ (which will be assumed to be 0 for now) such that

$$\mathbb{E}X_k = \left(\mu - \frac{b}{2}\right) \mathbb{1}(k < t) + \left(\mu + \frac{b}{2}\right) \mathbb{1}(k \geq t), \quad k \in \{1, \ldots, n\}. \tag{1.1}$$

To design a detection procedure and analyze its performance as sample size n grows to infinity, one is confronted with two fundamentally different frameworks, the framework of *increasing domain asymptotics* and that of *fixed domain (infill) asymptotics*, cf., e.g., Rao et al. (2012). The former arises naturally in time series analysis, which is distinguished by the constraint that the distance between consecutive sampling *time* points are bounded away from zero. The simplest instance of the sampling scenario in this regime arises when the diameter of $\mathcal{D}_n$ is of order n and $\min |t_{i+1} - t_i| > \epsilon$ for some strictly positive, fixed scalar ϵ. In our notation the index set for the Gaussian process represents the sampling time points. Typically we set $\mathcal{D} = \mathbb{R}$ and $\bigcup_{n=1}^{\infty} \mathcal{D}_n = \mathbb{N}$ or $\mathbb{Z}$. There is a large literature on change point detection via the increasing domain asymptotics (Antoch et al., 1997; Horváth, 1997; Horváth and Hušková, 2012; Kokoszka and Leipus, 1998; Rencova, 2009; Yao and Davis, 1986) — which we shall return to in a moment. Fixed domain (or infill) asymptotics, one the other hand, is a more suitable setting when the index set of sampling points $\mathcal{D}$ is bounded, so that the observations get denser in $\mathcal{D}$ as n increases. Particularly for $\mathcal{D} \subset \mathbb{R}$, we have that $\min |t_{i+k} - t_i| = \mathcal{O}(k/n)$ for positive integers i, k with $i, (i+k) \in \{1, \ldots, n\}$, and it can be extended to multidimensional domains in a straightforward way. This is the case for spatially distributed data (Stein, 1999), where the domain of the index set is typically of one, two or three dimensions. This approach is also appropriate in the context of change detection for non-stationary processes (Adak, 1998; Dahlhaus, 1997; Adak, 1998; Last and Shumway, 2008). The development of detection algorithms and theory for fixed domain asymptotics are relatively rare.

To gain a quick intuition on how the different asymptotic settings can affect the detection of a change in the observed sequence $\boldsymbol{X} = \{X_k\}_{k=1}^n$, one can look into the correlation among nearby samples in the sequence. In the increasing domain regime, even for long range dependent processes the correlation among samples X_i and X_j is small when $|j - i|$ is large. By contrast, in the fixed domain regime, regardless of how large the sample size is, if $|j - i|$ is of order n^{β} for some $\beta \in (0, 1)$, the correlation among X_i and X_j is still close to one. This entails that the effective sample size is much smaller than n. As a consequence, standard techniques that work well in the increasing domain setting do not work as well in the fixed domain setting. In the latter, we shall need more effective techniques to account for the strong dependence in the observed samples.

Previous works. An early attempt to study shift in mean detection was that of Chernoff and Zacks (1964). More general settings of this problem have been studied in subsequent works, e.g., MacNeill (1974), Deshayes and Picard (1985) and Yao and Davis (1986). For instance it is assumed in Yao and Davis (1986) that the sequence of X_k's are independent Gaussian variables. They proposed a detection method based on the generalized likelihood ratio test (GLRT), also known as the *cumulative sum (CUSUM)* test, and given by

$$T_{\mathrm{CUSUM}} = \mathbb{1}\left\{\max_{t \in C_{n,\alpha}} \left\{\sqrt{\frac{t(n-t)}{n}} \left|\frac{1}{n-t} \sum_{k=t+1}^{n} X_k - \frac{1}{t} \sum_{k=1}^{t} X_k\right|\right\} \geq R_n\right\}. \tag{1.2}$$

CUSUM compares the maximum of a test statistic over $C_{n,\alpha}$ with a critical value R_n. Non-asymptotic upper bounds on the error probabilities of this simple test were obtained by the authors under the Gaussian and i.i.d. assumptions. Due to its simplicity, the CUSUM test is very popular, and has been applied to a variety of settings.

For example, subsequent works studied the behavior of the CUSUM test under weaker assumptions in the increasing domain regime (Antoch et al., 1997; Horváth, 1997; Horváth and Hušková, 2012; Rencova, 2009). We wish to mention Rencova ((Rencova, 2009), chapter 4), who studied the same test as Yao and Davis, (1986), but working with the assumption that $\boldsymbol{X}$ is a strong mixing time series. Kokoszka and Leipus (1998) also analyzed the CUSUM test, but working with a different dependent observation model with sub-squared growth of the variance of partial sums, i.e., there is $\delta \in (0, 2)$ such that for any $k < m$, $\mathrm{var}\sum_{j=k}^{m} X_j \lesssim (m - k + 1)^{\delta}$. Horváth (1997), Horváth and Hušková (2012) and Antoch et al. (1997) studied the performance of the CUSUM test for the detection of a sudden change in the mean in linear processes, i.e. $X_t = \sum_{j=0}^{\infty} w_j \epsilon_{t-j}$, in which $\{\epsilon_t\}_{t=-\infty}^{\infty}$ are i.i.d. and zero mean random variables and the weights $\left\{w_j\right\}_{j=0}^{\infty}$ satisfy some properties such as absolute or square summability. We also refer the reader to Aue and Horváth (2013) and Horváth and Rice (2014) for a comprehensive review of abrupt change detection in the increasing domain regime.

The CUSUM test may also be applied to one dimensional processes with correlated samples, after a proper standardization. For instance, Horváth and Hušková (2012) used a different normalizing factor for applying CUSUM to one dimensional Gaussian time series with long range dependence. However apart from the standardizing factor, they do not directly incorporate the correlation structures of the data in the formulation of the test statistics. Furthermore, different forms of the CUSUM test were proposed to detect abrupt changes in the sequential detection literature, see e.g., Lai (1998). At first sight, it may seem puzzling how the CUSUM test attains nearly optimal detection performance in the increasing domain even as its test statistic apparently ignore the dependence among data samples (see e.g. Antoch et al. (1997); Horváth (1997); Horváth and Hušková (2012); Rencova (2009)). As noted earlier, the covariance $\mathrm{cov}(X_s, X_t) \to 0$ as $|t - s|$ grows to infinity.

As a result, the percentage of pairs $(X_s, X_t)_{s,t=1}^n$ whose covariance is non-negligible tends to zero as $n \to \infty$. Thus, there is not much gain in accounting for the dependence structures underlying the sequence, and so the CUSUM statistic provides a good approximation of the likelihood ratio test for large n, leading to the asymptotic optimality of T_{CUSUM} in the increasing domain setting.

One may consider applying the CUSUM test to detect a change point in the fixed domain setting, but we will see in this paper that the CUSUM test has suboptimal performance. We shall consider instead a generalized likelihood ratio test and show that this achieves the asymptotically optimal rate. In comparison to the increasing domain regime, the theoretical analysis for the fixed domain setting is considerably more involved from a technical standpoint, as one needs to take into account the statistical dependence in the data sequence in a more fundamental way.

Overview of main results. Our focus is on the change point detection problem given samples collected from a fixed and bounded domain. In particular, the data sequence is assumed to be drawn from a Gaussian process that experiences a change in the mean. We consider two scenarios for the covariance functions: fully known or with some unknown parameters. We seek to achieve the following:

1. Given an n-sample drawn from a one dimensional Gaussian process with a known covariance structure, we propose a generalized likelihood ratio test for detecting a sudden shift in the mean. This method requires the knowledge of the dependence structure (via the covariance matrix), and will be shown to achieve asymptotically near optimal detection performance in the fixed domain setting. Our theory holds for a variety of covariance structures, such as the Matern class, powered exponential class, and several others specified in terms of the covariance kernel's spectral density. The smoothness parameter for the Gaussian process (which determines how fast the corresponding spectral density decays) plays a central role in characterizing the minimax optimal detection performance.
2. We establish an upper bound guarantee for the CUSUM detection method. This result suggests that the CUSUM is suboptimal in the fixed domain setting. The suboptimality is confirmed in our simulation study, which exhibits a wide gap between the CUSUM and GLRT. This result makes sense, in light of the minimax result described earlier.
3. Next, the Gaussian process covariance structure is assumed unknown. To address this scenario, we propose a plug-in GLRT method, and investigate its performance. Quite remarkably, we show that as long as a consistent covariance estimate is employed (the notion of consistency will be defined in Section 4), regardless of its estimation rate, the plug-in GLRT achieves asymptotically near optimal detection performance.
4. For completeness we have also derived the performance of CUSUM and GLRT based algorithms in the increasing domain regime which confirms near minimax optimality this regime. Due to space constraints, such results are included in Appendix. The interested reader is referred to the technical report (Keshavarz et al., 2017) for a more comprehensive treatment of the increasing domain setting along with technical proofs.

In addition to studying the change point detection problem for dependent data in the fixed domain regime and distinguishing this setting from the increasing domain regime, the work carried out in this paper may serve as a starting point in the study of optimal detection of discontinuities in Gaussian spatial processes on domains of higher dimensions, as initiated by Gabriel et al. (2011) and Shen et al. (2002). At a more technical level, our asymptotic analysis contains several useful proof techniques worth mentioning: they include properties of mutually orthogonal Gaussian measures, the decorrelation of samples drawn from Gaussian processes in a fixed and bounded domain, and the classical theory of minimax detection. Appendix contains several useful technical results for the analysis of Gaussian random fields and Gaussian time series that may be of independent interest.

Structure of the paper. Section 2 presents the problem of detection of a shift in the mean in a one-dimensional Gaussian process, and then introduces detection procedures for the cases that the covariance structure is known and unknown. When the covariance structure is unknown, i.e., the spectral density is given with an unknown parameter, a plug-in GLRT will be introduced. Section 3 studies sufficient conditions on shift value b and spectral density to detect the existence of shift in mean with high probability. The analysis of the plug-in GLRT and the CUSUM test in the fixed domain setting is given in Section 4 and Section 5, respectively. The minimax optimality of the proposed algorithms is established in Section 6. The empirical evaluation of these tests is carried out by a simulation study in Section 7. Section 8 contains concluding remarks and a discussion of future works. Appendix A contains the proofs of the main results and Appendix B presents and proves some auxiliary results used in Appendix A. Results on the asymptotic behavior of both CUSUM and GLRT in the increasing domain setting are stated in Appendix C.

Notation. $\wedge$ and $\vee$ stand for minimum and maximum operators and the indicator function is represented by $\mathbb{1}(\cdot)$. For any $m \in \mathbb{N}$, I_m, $\mathbf{0}_m$ and $\mathbb{1}_m$ respectively denote the m by m identity matrix, all zeros column vector of length m, and all ones column vector of length m. For two matrices of the same size M_1 and M_2, $\langle M_1, M_2\rangle := \sum_{i,j}(M_1)_{ij}(M_2)_{ij}$ denotes their usual inner product. For any symmetric matrix M, $\lambda_{\min}(M)$ represents the smallest eigenvalue of M. We will use the following matrix norms on $M \in \mathbb{R}^{m\times n}$. For any $1 \le p < \infty$, $\|M\|_{\ell_p} := \left(\sum_{i,j}|M_{ij}|^p\right)^{1/p}$ stands for the element-wise ℓ_p norm of M, while $\|M\|_{\ell_\infty} := \max_{i,j}|M_{ij}|$ represents the sup norm of M. For a function $f : \mathcal{D} \mapsto \mathbb{R}$ and $p > 0$, $\|f\|_p^p := \int_{\mathcal{D}} |f(u)|^p du$. The special case of $p = \infty$ is defined by $\|f\|_\infty := \sup_{u\in\mathcal{D}} |f(u)|$. For any $f \in \mathbb{L}^1(\mathbb{R})$, $\hat{f}$ represents its Fourier transform defined by

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t} dt, \quad \forall\, \omega \in \mathbb{R},$$

where $j^2 = -1$ denotes the imaginary unit. For two functions f and g on $\mathbb{R}$, we write $f(t) \asymp g(t)$ as $t \to t_0$, if $C_1 \le \lim_{t\to t_0} \left|\frac{f(t)}{g(t)}\right| \le C_2$ for some strictly positive bounded scalars $C_1 \le C_2$. In particular, we write $f(t) \sim g(t)$ as $t \to t_0$ to indicate the case that $C_1 = C_2 = 1$. Furthermore, for sequences a_n and b_n, we write $b_n = \Omega(a_n)$ when b_n is bounded below by a_n asymptotically, i.e. $\lim_{n\to\infty} |b_n/a_n| \ge C$ for some positive C. In case a_n and b_n are random, $b_n = o_{\mathbb{P}}(a_n)$ means that b_n/a_n converges in probability to zero as $n \to \infty$. For two sets $\Omega_1, \Omega_2 \subset \mathbb{R}^m$, $\text{dist}(\Omega_1, \Omega_2) := \inf_{\omega_i\in\Omega_i,\ i=1,2} \|\omega_1 - \omega_2\|_{\ell_2}$ stands for their mutual distance with respect to Euclidean distance. Lastly, $\Gamma(\cdot)$ denotes the gamma function.

2. Change point detection procedures

In this section we describe the formulation of the shift-in-mean detection problem associated with Gaussian process data, and then present detection procedures that account for the underlying process modeling assumptions. From here on, we assume that G is a one dimensional Gaussian process in $\mathcal{D} = [0, 1]$ with n regularly spaced samples, i.e. $\mathcal{D}_n = \{k/n\}_{k=1}^n$. Let the symmetric real functions $K : \mathbb{R} \mapsto \mathbb{R}$ and $\hat{K} : \mathbb{R} \mapsto \mathbb{R}$ respectively denote the covariance function and spectral density of G. Accordingly, the covariance matrix $\Sigma_n := \text{cov}\left(\{X_k\}_{k=1}^n\right)$ is a symmetric Toeplitz matrix given by

$$\Sigma_n = \{\text{cov}(X_r, X_s)\}_{r,s=1}^n = \left[K\left(\frac{r-s}{n}\right)\right]_{r,s=1}^n. \tag{2.1}$$

Some regularity conditions on K will be introduced in the sequel.

2.1. Detection procedure based on generalized likelihood ratio test

The problem of detecting a sudden shift in the mean of a one dimensional Gaussian process can be formulated as a composite hypothesis testing problem. Under the null hypothesis $\mathbb{H}_0$, all the random variables have zero mean, i.e. $\mathbb{E}\boldsymbol{X} = \mathbf{0}_n$. To specify the alternative hypothesis $\mathbb{H}_1$ we first introduce a few additional notations. Let $t \in \mathcal{C}_{n,\alpha}$ denote the occurrence times of the single change point. The set $\mathcal{C}_{n,\alpha} \subseteq \{1, \dots, n\}$ contains plausible occurrence time of the change. Assume there is $\alpha \in (0, 1/2)$ such that $\mathcal{C}_{n,\alpha} = \{t : t \wedge (n-t) > \alpha n\}$. The amount of shift in the mean before and after the change point is denoted by real parameter b. Thus, for a fixed change point $t \in \mathcal{C}_n$, the associated alternative hypothesis associated with t can be stated as,

$$H_{1,t} : \exists\, b \neq 0,\ \mathbb{E}\boldsymbol{X} = \frac{b}{2}\zeta_t, \tag{2.2}$$

where $\zeta_t \in \mathbb{R}^n$ is given by $\zeta_t(k) := \text{sign}(k-t)$ for any $t \in \mathcal{C}_{n,\alpha}$. We adopt the convention $\text{sign}(0) = 1$. Since t is not known a priori, the alternative hypothesis is specified by taking the union of $\mathbb{H}_{1,t}$. We arrive at the following composite hypothesis testing problem:

$$\mathbb{H}_0 : \mathbb{E}\boldsymbol{X} = \mathbf{0}_n, \quad \text{vs.} \quad \mathbb{H}_1 = \bigcup_{t\in\mathcal{C}_{n,\alpha}} \mathbb{H}_{1,t},\ \text{ i.e., } \exists\, t \in \mathcal{C}_{n,\alpha},\ b \neq 0,\ \text{ s.t. } \mathbb{E}\boldsymbol{X} = \frac{b}{2}\zeta_t. \tag{2.3}$$

Next, we propose a test statistic which is constructed by the generalized likelihood ratio (GLR). Note that the GLR is an explicit function of the joint density of samples and so the Gaussian process assumption is essential to its calculation.

Proposition 2.1. *Assuming that Σ_n is known, there exists $R_{n,\delta} > 0$ for which the GLRT is given by*

$$T_{GLRT} = \mathbb{1}\left(\max_{t\in\mathcal{C}_{n,\alpha}} \left|\frac{\zeta_t^\top(\Sigma_n)^{-1}\boldsymbol{X}}{\sqrt{\zeta_t^\top(\Sigma_n)^{-1}\zeta_t}}\right|^2 \ge R_{n,\delta}\right). \tag{2.4}$$

The threshold value $R_{n,\delta}$ depends only on n and some parameter δ determining the false alarm rate. The precise form of $R_{n,\delta}$ will be presented in subsequent sections. Setting $\mu = 0$ in (1.1) results in a substantially simplified expression of the GLR, which eases the exposition of our analysis of the computational and theoretical properties of the proposed test. The general form of the GLRT, when μ is unknown, is presented as Proposition A.1 in Appendix A.

Unlike the CUSUM test, cf. Eq. (1.2), the covariance function of G is explicitly taken into account in the GLRT. As a result, it will be shown in the sequel that the proposed detection method is optimal, while the same cannot be said for the CUSUM test, specifically in the setting of fixed domain asymptotics.

Unknown covariance. In practice, however, the covariance is not known and needs to be estimated. To address such scenarios, we propose to approximate the likelihood ratio by plugging into Eq. (2.4) a positive definite estimate of the covariance matrix, which will be denoted by $\tilde{\Sigma}_n$. This results in a *plug-in GLRT* procedure.

Definition 2.1. Let $\tilde{\Sigma}_n$ be a positive definite estimate of Σ_n. The plug-in GLRT is given by

$$\tilde{T}_{GLRT} = \mathbb{1}\left(\max_{t\in\mathcal{C}_{n,\alpha}}\left|\frac{\zeta_t^\top\left(\tilde{\Sigma}_n\right)^{-1}\boldsymbol{X}}{\sqrt{\zeta_t^\top\left(\tilde{\Sigma}_n\right)^{-1}\zeta_t}}\right|^2 \geq \tilde{R}_{n,\delta}\right), \tag{2.5}$$

for some strictly positive threshold value $\tilde{R}_{n,\delta}$.

The specific choice of $\tilde{\Sigma}_n$ and the accompanying theory will be given later in Section 4.

3. Detection rate of GLRT: known Σ_n

In this section we shall establish the detection rate of the GLRT in a fixed domain regime, given that Σ_n is known. We adopt the following performance measure.

Definition 3.1. For any change detection algorithm $T \in \{0, 1\}$, the conditional detection error probability (CDEP) of T, which is denoted by $\varphi_n(T)$, is defined as

$$\varphi_n(T) = \mathbb{P}(T = 1 \mid \mathbb{H}_0) + \max_{t\in\mathcal{C}_{n,\alpha}} \mathbb{P}\left(T = 0 \mid \mathbb{H}_{1,t}\right).$$

In words, φ_n is the sum of the false alarm error and the worst-case misdetection error (taken over the set of possible change point locations $\mathcal{C}_{n,\alpha}$). Clearly, CDEP hinges on the choices of $\mathcal{C}_{n,\alpha}$ — the value of φ_n increases as $\mathcal{C}_{n,\alpha}$ becomes a larger proper subset of $\{1, \ldots, n\}$. CDEP as a risk measure has been adopted for detecting abnormal clusters in a network (see, e.g., Arias-Castro et al., 2011; Butucea and Ingster, 2013). It also provides an upper bound on the Bayesian risk measure. We refer the reader to Addario-Berry et al. (2010) for a comparison of CDEP and the Bayesian risk measure.

Given a fixed $\delta \in (0, 1)$, we will present a sufficient condition expressed in terms of shift value b, sample size n, δ, and the spectral properties of the Gaussian process such that the proposed detection procedures can guarantee that CDEP is bounded from above by δ. This can be achieved by

- first, choose the critical value $R_{n,\delta}$ so that the false alarm error $\mathbb{P}(T = 1 \mid \mathbb{H}_0) < \delta/2$;
- second, the proposed sufficient condition (in terms of b, n, δ, and the parameters encoding the dependence structure of the data) guarantees that the worse-case misdetection error rate is also upper bounded by $\delta/2$.

Recall that G is a Gaussian process defined on $\mathcal{D} = [0, 1]$ whose one realization has been observed at $\mathcal{D}_n = \{k/n\}_{k=1}^n$. The covariance function and the spectral density of G are respectively denoted by K and $\hat{K}$. We study two common classes of covariance functions, one of which admits polynomially decaying spectral densities, and the other being the Gaussian covariance function.

Assumption 3.1. K is an integrable positive definite covariance function. Moreover, there exist $\nu \in (0, \infty)$ and C_K (depending on K) so the spectral density $\hat{K}$ satisfies the following condition:

$$C_K := \sup_{\omega\in\mathbb{R}}\left|\hat{K}(\omega)\left(1+\omega^2\right)^{\nu+\frac{1}{2}}\right| < \infty. \tag{3.1}$$

We shall always choose the largest possible ν that satisfies (3.1). It is simple to see that Assumption 3.1 holds if and only if $\hat{K}$ is bounded at the origin and $\hat{K}(\omega) \asymp \omega^{-(2\nu+1)}$ as ω tends to infinity. It is well-known that the tail behavior of $\hat{K}$ is closely linked to the smoothness of K at the origin (see, e.g., Section 2.8 of Stein, 1999). The following are a few examples of common covariance functions that will be studied in this paper.

(a) *Matern*: This class is widely used in geostatistics, and has a fairly simple explicit form of spectral density.

$$\hat{K}(\omega) = \frac{\sqrt{4\pi}\,\Gamma(\nu+1/2)}{\Gamma(\nu)}\sigma^2\rho^{-2\nu}\left(\frac{1}{\rho^2}+\omega^2\right)^{-(\nu+1/2)}, \tag{3.2}$$

where $\rho, \nu, \sigma \in (0, \infty)$. Regardless of the choice of ρ and σ, condition (3.1) holds for Matern spectral density with parameter ν.

(b) *Powered exponential*: Another versatile class of covariance functions is

$$K(r) = \sigma^2\exp\left(-\left|\frac{r}{\rho}\right|^\beta\right) \tag{3.3}$$

for some $\beta \in (0, 2)$ and $\rho, \sigma \in (0, \infty)$. Although the spectral density does not have a closed form in terms of simple functions, Lemma B.2 shows that $\hat{K}$ admits Assumption 3.1 with $\nu = \beta/2$.

(c) *Rational spectral densities*: Rational spectral densities form a general class admitting Assumption 3.1 . For any $\hat{K}$ in this class, there are two polynomials, Q_n and Q_d, with real coefficients, unit leading coefficients and $p := \deg(Q_d) - \deg(Q_n) \in \mathbb{N}$, such that

$$\hat{K}(\omega) = \lambda \frac{|Q_n(j\omega)|^2}{|Q_d(j\omega)|^2}. \tag{3.4}$$

Moreover, we assume that Q_d has no root on the imaginary axis and λ is a strictly positive scalar. Since $K(0) < \infty$ and $\hat{K}(\omega) \asymp \omega^{-2p}$ as $\omega \to \infty$, Assumption 3.1 holds with $\nu = p - 1/2$.

(d) *Triangular*: For $\rho, \sigma \in (0, \infty)$, the covariance function and spectral density are given by

$$K(r) = \sigma^2 \left(1 - \left|\frac{r}{\rho}\right|\right)_+, \quad \hat{K}(\omega) = \frac{\rho\sigma^2}{2}\left|\operatorname{sinc}\left(\frac{\rho\omega}{2}\right)\right|^2.$$

The triangular covariance is less favorable than the aforestated cases due to the oscillatory behavior of $\hat{K}$ (p. 31, Stein, 1999). One can easily show that this covariance fulfills Assumption 3.1 with $\nu = 1/2$.

(e) *Mixture form*: Let $K_i(\cdot; \rho_i)$, $i = 1, \ldots, m$ denote m different correlation functions satisfying Assumption 3.1 with ν_i, respectively. We construct the covariance function K by

$$K(r) = \sum_{i=1}^{m} \phi_i K_i(r; \rho_i),$$

in which $\phi_1, \ldots, \phi_m$ are strictly positive scalars. $\hat{K}$ trivially admits Assumption 3.1 with $\nu = \min_{1\le i \le m} \nu_i$.

The following theorem establishes a detection error guarantee for the GLRT, provided that the Gaussian process covariance function K is known.

Theorem 3.1. *Let $\delta \in (0, 1)$. Suppose that G is a real-valued Gaussian process defined on domain $\mathcal{D} = [0, 1]$ whose associated spectral density $\hat{K}$ admits Assumption 3.1 for some ν and C_K. G is regularly sampled at i/n, $i = 1, \ldots, n$. There exist $R_{n,\delta} > 0$ (depending only on n and δ), $n_0 := n_0(K)$ and a positive universal constant C such that if $n \ge n_0$ and*

$$|b| \ge C n^{-\nu} \sqrt{C_K \left(1 + \frac{1}{\nu}\right) \log\left(\frac{n(1-2\alpha)}{\delta}\right)}, \tag{3.5}$$

we have

$$\varphi_n(T_{GLRT}) \le \delta.$$

In the theorem statement, *universal constant* is used to refer to a fixed, finite positive scalar independent of n, δ, and all the covariance parameters. See Appendix A.2 for the proof of Theorem 3.1. The right hand side of Eq. (3.5) provides a bound on the smallest detectable shift using the GLRT algorithm associated to the CDEP risk measure. We make several comments regarding the roles of various quantities embedded in Theorem 3.1.

(a) $R_{n,\delta}$ in Theorem 3.1 can be chosen as

$$R_{n,\delta} = 1 + 2\left[\log\left(\frac{2n(1-2\alpha)}{\delta}\right) + \sqrt{\log\left(\frac{2n(1-2\alpha)}{\delta}\right)}\right]. \tag{3.6}$$

We guarantee that CDEP is less than or equal δ by controlling the false alarm and misdetection probabilities below $\delta/2$. To gain some insight into (3.6), notice that under the null hypothesis, the test statistic in Eq. (2.4) has the same distribution as the supremum of a χ_1^2 process over $\mathcal{C}_{n,\alpha}$, which is represented by $\{\Psi(t) : t \in \mathcal{C}_{n,\alpha}\}$. For controlling the false alarm probability below $\delta/2$, $R_{n,\delta}$ needs to be chosen such that

$$\mathbb{P}\left(\max_{t\in\mathcal{C}_{n,\alpha}} \Psi(t) \ge R_{n,\delta}\right) \le \frac{\delta}{2}.$$

The standard χ_1^2 tail inequality in Birgé (2001) implies that if $R_{n,\delta}$ is chosen based upon Eq. (3.6), then $\Psi(t) \le \delta/\{2n(1-2\alpha)\}$ for any $t \in \mathcal{C}_{n,\alpha}$. Thus, the union bound inequality yields

$$\mathbb{P}\left(\sup_{t\in\mathcal{C}_{n,\alpha}} \Psi(t) \ge R_{n,\delta}\right) \le \left|\mathcal{C}_{n,\alpha}\right| \max_{t\in\mathcal{C}_{n,\alpha}} \mathbb{P}\left(\Psi(t) \ge R_{n,\delta}\right) \le \frac{\delta\left|\mathcal{C}_{n,\alpha}\right|}{2n(1-2\alpha)} = \frac{\delta}{2}.$$

(b) The bound on the minimal detectable shift is proportional to $\sqrt{C_K}$, as defined in (3.1). Note that C_K is determined by both low frequency and tail behavior of the spectral density via ν. C_K is obviously linearly proportional to $\sqrt{K(0)}$ (see Eq. (3.1)), meaning that C_K also captures the notion of the standard deviation of the observations. Thus, Theorem 3.1 implicitly expresses that change detection is more challenging for Gaussian processes with larger variance.

(c) Sample size n has two opposing effects on the detection rate of the GLRT algorithm. On the one hand, $n(1-2\alpha)$ appearing in the logarithmic function, is connected to the size of alternative hypothesis which is determined by $\left|\mathcal{C}_{n,\alpha}\right| = n(1-2\alpha)$. On the other hand, the term $n^{-\nu}$ indicates the possibility of small shift detection as more observations are available.

We note that the parameter δ, the variance of observations, and sample size have similar roles in the increasing domain setting. The main difference between the two asymptotic settings is the role of the decay rate of $\hat{K}$ in the fixed domain, which is encapsulated by ν. See Appendix C for further details in the increasing domain regime. Note that ν is closely related to the smoothness of G with larger values of ν corresponding to a smoother Gaussian process in the mean squared sense (cf. Stein, 1999, Chapter 2). For smooth Gaussian processes, $G(t_0)$ can be interpolated using the observations in the vicinity of t_0 with small estimation error. This leads to a simpler shift-in-mean detection for smoother processes. More precisely, as $n \to \infty$ the lower bound on detectable b, (3.5), vanishes more rapidly for larger ν.

Remark 3.1. As an easy consequence of the theorem, we can elaborate on the asymptotic behavior (as $n \to \infty$) of the GLRT for several specific classes of spectral densities, all of which satisfy Assumption 3.1.

(a) *Matern*: The smallest detectable jump is $|b| \asymp n^{-\nu}\sqrt{\log(n(1-2\alpha)/\delta)}$.
(b) *Powered exponential*: $\hat{K}$ admits Assumption 3.1 with $\nu = \beta/2$. Namely the smallest detectable b has the same order as $\sqrt{n^{-\beta}\log(n(1-2\alpha)/\delta)}$.
(c) *Rational spectral densities*: It has been discussed previously that $\hat{K}(\omega) \asymp |\omega|^{-2p}$ as $\omega \to \infty$ and $\nu = p - 1/2$. Thus $|b| = \Omega\left(n^{-(p-1/2)}\sqrt{\log(n(1-2\alpha)/\delta)}\right)$ guarantees CDEP to remain below δ.
(d) *Triangular*: Assumption 3.1 with $\nu = 1/2$ holds for $\hat{K}$, so the smallest detectable jump is of order $\sqrt{n^{-1}\log(n(1-2\alpha)/\delta)}$.
(e) *Mixture form*: The spectral density satisfies Assumption 3.1 with $\nu := \min_{1\le i\le m}\nu_i$. Thus the minimal detectable shift in mean is of order $n^{-\nu}\sqrt{\log(n(1-2\alpha)/\delta)}$.

Remark 3.2. Let us comment on the role of α in Theorem 3.1. The dependence on α in Eq. (3.5) is logarithmic, which encodes how the size of $\mathcal{C}_{n,\alpha}$ affects the detection rate. Strictly speaking, the asymptotic behavior of the smallest detectable jump remains unchanged, regardless of how small α has been chosen (even if α tends to zero). Note that we did *not* require that α is a fixed and strictly positive scalar in the theorem. For algebraic convenience, we assume that the mean of G fluctuates around $\mu = 0$ in Eq. (2.3). The fact that we assumed μ is known is the main reason behind the trifling effect of α in the detection rate of the GLRT, as we do not need to estimate μ from the data. That is why in this particular case α can even be chosen as small as $\mathcal{O}(1/n)$. The generic form of the GLRT test for unknown μ is presented in Proposition A.1.

Gaussian covariance function. The Gaussian covariance function is given by

$$K(r) = \sigma^2 \exp\left[-\frac{1}{2}\left(\frac{r}{\rho}\right)^2\right], \quad \hat{K}(\omega) = \rho\sigma^2\sqrt{2\pi}\exp\left[-\frac{(\rho\omega)^2}{2}\right]. \tag{3.7}$$

This is also a popular modeling choice of smooth Gaussian processes (Loh and Lam, 2000). Regarding this covariance, we have the following result:

Theorem 3.2. *Let G be a Gaussian process on $[0, 1]$ which is observed at i/n, $i = 1, \ldots, n$, whose covariance function is given by Eq. (3.7). Let $\delta \in (0, 1)$. There are $R_{n,\delta} > 0$, $n_0 := n_0(\rho)$, $C_0 := C_0(\rho) > 0$ and a universal constant $C > 0$, such that if $n \ge n_0$ and*

$$|b| \ge C\sqrt{\exp\left[-n\log(C_0 n)\right]\log\left(\frac{n(1-2\alpha)}{\delta}\right)}, \tag{3.8}$$

then

$$\varphi_n(T_{GLRT}) \le \delta.$$

The proof of this result is given in Appendix A.2. Because of the super-exponential decay of the Gaussian spectral density, Assumption 3.1 is actually satisfied for any $\nu > 0$. This result shows that it is possible to detect exponentially small jump size b as n increases. Theorem 3.2, along with Theorem 3.1 confirm the intuition that smoother the Gaussian process is (larger ν in Assumption 3.1), the easier it is to detect the presence of a shift in the mean.

Remark 3.3. We conclude this section noting the difference in the detection error rate guarantee in the fixed domain regime (Theorems 3.1 and 3.2) and the analogous results in the increasing domain setting (which are deferred to Appendix). The GLRT can detect the jumps of magnitude $\mathcal{O}\left(\sqrt{n^{-1}\log(n(1-2\alpha)/\delta)}\right)$ for large n (that is $\varphi_n(T_{GLRT}) \le \delta$) in the increasing domain regime, regardless of the covariance structure of G. Simply put the detection rate is not affected by the dependence structure of G. By contrast, we have seen in the earlier theorems how the detection error guarantee for the GLRT in the fixed domain setting is affected by dependence structure of G in a fundamental way.

4. Detection rate of plug-in GLRT

As we have shown in Proposition 2.1, full knowledge of Σ_n is central to computing the generalized likelihood ratio. In practice, the spectral density and covariance function of G are not known a priori, and so we take a plug-in approach, approximating the GLRT by plugging in the covariance estimate $\tilde{\Sigma}_n$ (see Definition 2.1). This section serves to investigate various ways of constructing plug-in GLRTs and assessing their detection performance. In this section we will focus only on the fixed domain setting.

We first assume that G is a Matern Gaussian process on $\mathcal{D} = [0, 1]$ with unknown parameters $\eta = (\sigma, \rho)$ in a compact space Ω, and is regularly observed on $\{k/n\}_{k=1}^n$ (see Eq. (3.2)). We use $\tilde{\eta}_m = (\tilde{\sigma}_m, \tilde{\rho}_m)$ to indicate the estimated parameters using m regularly spaced samples in $\mathcal{D}$. We also assume that $\mathcal{C}_{n,\alpha} = \{k : \ \alpha n \le k \le (1-\alpha)\, n\}$. Namely, the Gaussian process is under control for a certain number of observations. The controlled samples before the sudden change, $\boldsymbol{X}_{\mathcal{B}} := \{X_k : \ k \le \alpha n\}$, will be used to estimate η. The parameter estimation stage is typically called the *burn-in* period in the literature.

It is well-known that η is not consistently estimable in the fixed domain setting when the number of the observations in $\mathcal{D}$ grows to infinity (cf., e.g., Ying (1991) and Zhang (2004)). In particular, Zhang (2004) showed that neither σ nor ρ is consistently estimable but the quantity $\sigma\rho^{-\nu}$ can be consistently estimated using MLE. The reason behind the inconsistency is the existence of a class of mutually absolutely continuous models for G which are almost surely impossible to discern by observing one realization of G. The induced measures corresponding to two Matern Gaussian processes with parameters η and η' are absolutely continuous with respect to each other whenever $\sigma\rho^{-\nu} = \sigma'\rho'^{-\nu}$. Furthermore Zhang (2004) showed that if one fixes ρ at an arbitrary value, then the maximum likelihood estimator for $\sigma\rho^{-\nu}$ is consistent. We shall show that despite the inconsistency in estimating η, quite remarkably, the plug-in GLRT exhibits an analogous performance as the GLRT with fully known covariance function, provided that the estimate of η is consistent up to its equivalence class.

It has been noted in Kaufman and Shaby (2013) that fixing $\tilde{\rho}_m$ at large values has a trifling impact on predictive performance. Due to the complicated dependence of the Matern covariance function on ρ, estimating ρ is a computationally challenging task, particularly for large data sets. Fortunately we can accelerate the whole detection procedure without estimating ρ. In fact, our plug-in GLRT change detector $\tilde{T}_{GLRT}$ is a two stage algorithm as follows:

- *Estimation step:*
  1. Fix $\tilde{\rho}_m$ at the largest possible element in Ω. Namely, $\tilde{\rho}_m$ is a deterministic quantity given by $\tilde{\rho}_m = \max\{\rho : (\sigma, \rho) \in \Omega\}$.
  2. Estimate $\sigma\rho^{-\nu}$ given the controlled samples $\boldsymbol{X}_{\mathcal{B}}$, using any consistent procedure such as maximum likelihood (MLE) (Zhang, 2004), weighted local Whittle likelihood (Wu et al., 2013), or averaging quadratic variation (Anderes, 2010). By *consistent* we mean the condition that $\left|\sigma\rho^{-\nu} - \tilde{\sigma}_m\tilde{\rho}_m^{-\nu}\right| \xrightarrow{\mathbb{P}} 0$ as m tends to infinity.
  3. Construct the approximated covariance matrix of $\boldsymbol{X}$, as $\tilde{\Sigma}_n = \left[K\left(\frac{r-s}{n}, \tilde{\eta}_m\right)\right]_{r,s=1}^n$ (here $m := \lfloor \alpha n \rfloor$).
- *Detection step:*
  1. Apply the GLRT by plugging $\tilde{\Sigma}_n$ in place of Σ_n into (2.4), as described in Definition 2.1.

The following theorem establishes the detection performance for the plug-in GLRT.

Theorem 4.1. *Let $\delta \in (0, 1)$. Let G be Gaussian process whose associated spectral density $\hat{K}$ has Matern form with unknown parameters $(\sigma, \rho) \in \Omega$, and sampled in $\{k/n\}$, $k = 1, \ldots, n$. There are finite scalar $C, n_0 \in \mathbb{N}$, a non-negative sequence $\lim_{m\to\infty}\tau_m = 0$, and threshold level $\tilde{R}_{n,\delta} > 0$ such that for any $n \ge n_0$,*

$$\varphi_n\left(\tilde{T}_{GLRT}\right) \le \delta + 2\tau_m,$$

whenever

$$|b| \ge Cn^{-\nu}\sqrt{C_K\left(1+\frac{1}{\nu}\right)\log\left(\frac{n(1-2\alpha)}{\delta}\right)}. \tag{4.1}$$

See Appendix A.3 for the proof of Theorem 4.1.

Remark 4.1. The threshold value for the plug-in GLRT is chosen exactly the same as in Theorem 3.1;

$$\tilde{R}_{n,\delta} = \left[1 + 2\left(\log\left(\frac{2n(1-2\alpha)}{\delta}\right) + \sqrt{\log\left(\frac{2n(1-2\alpha)}{\delta}\right)}\right)\right]. \tag{4.2}$$

Since $\{\tau_k\}_{k\in\mathbb{N}}$ is a vanishing sequence and $m = \lfloor \alpha n \rfloor$ is an increasing function of n, $(\delta + 2\tau_m)$ lies in the vicinity of δ for large n. The most interesting aspect of Theorem 4.1 is that if some consistent estimate of $\sigma\rho^{-\nu}$ is available, the plug-in GLRT has asymptotically the same rate as the GLRT with fully known covariance function, regardless of how efficient the point estimate for $\sigma\rho^{-\nu}$ is. The main asymptotic cost to pay for (potentially) mis-specifying ρ and σ is that constant C appearing in Theorem 4.1 is larger than the constant C of Theorem 3.1.

We now study the performance of the plug-in GLRT when both variance and range parameters are consistently estimable. Suppose that G has a powered exponential covariance function, introduced in Eq. (3.3). Anderes (2010) proposed a consistent estimate of covariance parameters using empirical average of the quadratic variation of G. According to Theorem 5 of Anderes (2010), unlike the Matern class, both σ_0 and ρ_0 are consistently estimable when $\beta \in (0, 1/2)$. Namely, $|\rho - \tilde{\rho}_m| \vee |\sigma - \tilde{\sigma}_m| \xrightarrow{\mathbb{P}} 0$, for the method introduced in Anderes (2010). The following result, which has a similar flavor as Theorem 4.1 and can be substantiated in a similar way, determines the detection rate of the plug-in GLRT for one-dimensional powered exponential Gaussian processes.

Theorem 4.2. *Let $\delta \in (0, 1)$. Let G be a Gaussian process with powered exponential covariance function with unknown parameters (σ, ρ) and known $\beta \in (0, 1/2)$, sampled in $\{k/n\}, \ k = 1, \ldots, n$. Given a consistent estimate of (σ, ρ) (e.g., the method in Anderes (2010)), there are finite scalars $n_0 \in \mathbb{N}$ and C (which depends on the covariance parameters β, σ and ρ), and a non-negative sequence $\lim_{m\to\infty} \tau_m = 0$, such that for any $n \geq n_0$,*

$$\varphi_n\left(\tilde{T}_{GLRT}\right) \leq \delta + 2\tau_m,$$

whenever

$$|b| \geq C\sqrt{n^{-\beta} \log\left(\frac{n\,(1-2\alpha)}{\delta}\right)},$$

and

$$\tilde{R}_{n,\delta} = \left[1 + 2\left(\log\left(\frac{2n\,(1-2\alpha)}{\delta}\right) + \sqrt{\log\left(\frac{2n\,(1-2\alpha)}{\delta}\right)}\right)\right].$$

Theorem 4.2 states that given a consistent estimate of $\eta = (\sigma_0, \rho_0)$, the plug-in GLRT procedure has the same asymptotic behavior as the GLRT with fully known parameters (see part (b) of Remark 3.1 for the detection rate of the GLRT with known σ_0 and ρ_0).

4.1. Detection rate for correct and misspecified mixture covariance models

Theorems 4.1 and 4.2 establish the asymptotic detection rate of the plug-in GLRT under the assumption of having a consistent parametric estimation of the covariance function. However such models, particularly with a few number of unknown parameters cannot fully capture the complexity of dependence structure and are usually prone to misspecification error in numerous practical situations. Estimating the covariance function within a broader class (such as mixture of different parametric forms) is a common effective way of reducing the misspecification error. However, dealing with misspecified covariance models is imperative in the fixed domain asymptotic regime, as it is infeasible to identify all the unknown covariance parameters. This section studies the detection rate of the plug-in GLRT when the covariance function can be formulated as a mixture of Matern functions with known smoothness parameters, i.e.,

$$\operatorname{cov}\Big(G(s), G(t)\Big) = \sum_{i=1}^{p} \sigma_i^2 K\left(s - t; \nu_i, \rho_i\right) \tag{4.3}$$

in which $\nu_1 < \cdots < \nu_p$ and $K(\cdot, \nu)$ is a standard Matern correlation function with smoothness paramagnet ν. G can alternatively be viewed as sum of p independent Matern Gaussian processes with known smoothness parameters. Notice that the spectral density of such process is given by

$$\hat{K}(\omega) = \sum_{i=1}^{p} \sigma_i^2 \hat{K}\left(\omega, \rho_i, \nu_i.\right)$$

The notion of equivalence of two probability measures on the sample paths of a Gaussian process is of extreme importance in the fixed domain asymptotic analysis. In summary, it is not possible to discern the unknown parameters of two Gaussian processes inducing equivalent probability measures on the Hilbert space $\mathbb{L}^2(\mathcal{D})$. Therefore, before going further we rigorously introduce the concept of equivalent Gaussian processes. Consider two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on real line. Let G be a zero mean stationary Gaussian processes on $\mathbb{R}$ with the spectral density $f_i, \ i = 1, 2$ under $\mathbb{P}_i, \ i = 1, 2$, respectively. Then $\mathbb{P}_1$ and $\mathbb{P}_2$ are equivalent on the sample paths of $G(s), \ s \in \mathcal{D}$ for any bounded set $\mathcal{D} \subset \mathbb{R}$, if

- There exists $\alpha > 1$ such that $f_1(\omega)\,|\omega|^{\alpha}$ is bounded away from zero and infinity as $\omega \to \infty$.
- There exists $c > 0$ such that $\int_{|\omega|>c} \left(\frac{f_2(\omega)}{f_1(\omega)} - 1\right)^2 d\omega < \infty$.

We refer the reader to Chapter 4 of Stein (1999) for further details. The above criteria establishes sufficient conditions under which a centered Gaussian process with covariance function (4.3) is indistinguishable from a zero-mean one-dimensional Matern process. In the following we state such a result without proof. We omit the proof because of its simplicity.

Corollary 4.1. *Let $\mathbb{P}_1$ be a probability measure under which G is a stationary Matern Gaussian process with spectral density parameters (σ, ρ, ν_1). Furthermore, consider a probability measure $\mathbb{P}_2$ under which the spectral density of G is given by Eq. (4.3). Then $\mathbb{P}_1$ and $\mathbb{P}_2$ are equivalent measures if $\sigma_1\rho_1^{-\nu_1} = \sigma\rho^{-2\nu_1}$ and $\nu_i > \nu_1 + 1/4$ for any $i = 2, \ldots, p$.*

According to Corollary 4.1, it is only feasible to consistently estimate $\sigma_1\rho_1^{-\nu}$ for a Gaussian process with mixture of Matern covariance functions, if $\nu_i > \nu_1 + 1/4$ for any $i = 2, \ldots, p$. In other words, it is not possible to detect the existence of sufficiently smoother components than the least smooth term in the covariance function. However we do not pay a high price for mis-specifying the form of covariance function in this setting. Roughly speaking, given a consistent estimate of $\sigma_1\rho_1^{-\nu_1}$ (using the aforementioned approach in Section 4), the plug-in GLRT has the same detection rate as in Theorem 4.1. The following result refines our argument.

Theorem 4.3. *Let $\delta \in (0, 1)$ and let G be Gaussian process whose covariance function is given by Eq. (4.3), with known smoothness parameters $\nu_1, \ldots, \nu_p$. Suppose that $\nu_i > \nu_1 + 1/4$ for any $i = 2, \ldots, p$. Furthermore suppose that G is sampled in $\{k/n\}$, $k = 1, \ldots, n$. There are finite scalar C, $n_0 \in \mathbb{N}$, a non-negative sequence $\lim_{m\to\infty}\tau_m = 0$, and threshold level $\tilde{R}_{n,\delta} > 0$ such that for any $n \geq n_0$,*

$$\varphi_n\left(\tilde{T}_{GLRT}\right) \leq \delta + 2\tau_m,$$

whenever

$$|b| \geq Cn^{-\nu_1}\sqrt{C_K\left(1 + \frac{1}{\nu_1}\right)\log\left(\frac{n(1-2\alpha)}{\delta}\right)}.$$

We again choose the critical value of the plug-in GLRT test based upon Remark 4.2. We refer the reader to Appendix A.3 for the proof of Theorem 4.3. Next we consider the case that a Gaussian process with mixture covariance model is discernible from a Matern process. For brevity suppose that $p = 2$ and $\nu_1 < \nu_2 < \nu_1 + 1/4$ in Eq. (4.3). Now, both quantities $\sigma_i\rho_i^{-2\nu_i}$, $i = 1, 2$ are consistently estimable using the quadratic variation technique proposed by Anderes (2010). Thus the covariance model (4.3) does not lead to a misspecified formulation, which means that the plug-in GLRT has the same detection rate as the GLRT with fully known covariance function.

Theorem 4.4. *Let $\delta \in (0, 1)$ and let G be Gaussian process whose covariance function is given by Eq. (4.3), with $p = 2$ and known smoothness parameters $\nu_1, \ldots, \nu_p$. Suppose that $\nu_1 < \nu_2 >< \nu_1 + 1/4$. Furthermore assume that G is sampled in $\{k/n\}$, $k = 1, \ldots, n$. Given a consistent estimate of $\sigma_i\rho_i^{-2\nu_i}$, $i = 1, 2$, there are finite scalar C, $n_0 \in \mathbb{N}$, a non-negative sequence $\lim_{m\to\infty}\tau_m = 0$, and threshold level $\tilde{R}_{n,\delta} > 0$ such that for any $n \geq n_0$,*

$$\varphi_n\left(\tilde{T}_{GLRT}\right) \leq \delta + 2\tau_m,$$

whenever

$$|b| \geq Cn^{-\nu_1}\sqrt{C_K\left(1 + \frac{1}{\nu_1}\right)\log\left(\frac{n(1-2\alpha)}{\delta}\right)}.$$

The proof of Theorem 4.4 is omitted to avoid repetition, as it is similar to that of Theorems 4.1–4.3.

Remark 4.2. It is worthwhile to mention that the results in Sections 3 and 4 are restricted to the equally spaced sampling regime. Such assumption can significantly reduce the cumbersome algebra of evaluating the mean and variance of the plug-in GLR test statistic under the alternative hypothesis. Furthermore, it also simplifies the procedure of obtaining closed form expressions for the constants appearing at Theorems 4.1–4.4. However based upon our proof techniques, we believe that the asymptotic behavior of proposed tests remains untouched (apart from the constants) for evenly distributed sampling locations in $\mathcal{D} = [0, 1]$. Namely, G is observed at $\mathcal{D}_n = \{t_i\}_{i=1}^n$ such that $n(t_i - t_{i-1})$, $i = 2, \ldots, n$ are bounded from 0 and ∞, as $n \to \infty$.

5. Detection rate of CUSUM

In this section we revisit the classical CUSUM test and obtain its detection rate in the fixed domain setting. This result should be contrasted with our earlier theorems on the performance of the proposed exact and plug-in GLRTs, and highlights the need for accounting for the dependence structures underlying the data. Theorem 5.1 introduces sufficient conditions on $|b|$ under which CUSUM can distinguish null and alternative hypotheses with high probability.

Theorem 5.1. *Suppose that $\|K\|_1 < \infty$ and $\left\|\hat{K}'\right\|_\infty < \infty$. Moreover let $\delta \in (0, 1)$, $\alpha \in (0, 1/2)$ and $\mathcal{C}_{n,\alpha} = [\alpha n, (1-\alpha) n] \cap \mathbb{N}$. There are $R_{n,\delta} > 0$, and $n_0 := n_0(\delta, \alpha)$ such that if $n \geq n_0$ and*

$$|b| \geq 4\sqrt{\frac{\log\left(\frac{2n(1-2\alpha)}{\delta}\right)}{\alpha(1-\alpha)}}, \tag{5.1}$$

then,

$$\varphi_n(T_{\mathrm{CUSUM}}) \leq \delta.$$

The proof of this theorem is deferred to Appendix A.4. The risk of fixed domain-CUSUM has been controlled from above under mild conditions on K, which holds true for all the examples of covariance functions considered in this paper. Due to the following inequality, K satisfies the assumptions in Theorem 5.1 if $a(r) := rK(r)$ is absolutely integrable:

$$\left\|\hat{K}'\right\|_\infty = \sup_{\omega\in\mathbb{R}}\left|\int_{-\infty}^{\infty} a(r)\, e^{-j\omega r} dr\right| \leq \int_{-\infty}^{\infty} |rK(r)|\, dr.$$

The main feature of the above theorem is the sufficient condition that the jump size increases (at the order of $\log n$ at least) in order to have an upper bound guarantee on the detection error. Although we do not have a definitive proof that this sufficient condition is also necessary, the theorem suggests that the CUSUM test is *inconsistent* in the fixed domain setting: the detection error may not vanish as data sample size increases, when the jump size is a constant. This statement is in fact verified by a careful simulation study. By contrast, we have shown earlier that using the GLRT, we can guarantee vanishing detection error as long as the jump size is either constant or (better yet) bounded from below by a suitable vanishing term.

Remark 5.1. Let us give a qualitative argument for the inconsistency of the CUSUM test in the fixed domain setting. Suppose that b tends to zero as $n \to \infty$. Define

$$U_t := \sqrt{\frac{t(n-t)}{n}}\left[\frac{1}{n-t}\sum_{k=t+1}^{n} X_k - \frac{1}{t}\sum_{k=1}^{t} X_k\right].$$

The expected value of U_t is zero, under the null hypothesis and for any t. Regardless of the existence of a shift in the mean, the standard deviation of U_t remains the same. A careful look at the proof of Theorem 5.1 reveals that the smallest value of the standard deviation of U_t over $t \in \mathcal{C}_{n,\alpha}$ is order $\sqrt{n}$. Moreover, if there is a shift in the mean occurring at the change point $\tilde{t} \in \mathcal{C}_{n,\alpha}$, then the expected value of $U_{\tilde{t}}$ is given by $b\sqrt{\tilde{t}\left(n-\tilde{t}\right)/n} = \mathcal{O}\left(b\sqrt{n}\right)$ (Recall that $\alpha n \leq \tilde{t} \leq (1-\alpha) n$). Generally speaking, as the mean of U_t under the null hypothesis, denoted by $\mathbb{E}(U_t \mid \mathbb{H}_0)$, is zero for any $t \in \mathcal{C}_{n,\alpha}$, the CUSUM test cannot distinguish between the null and the alternative (even for large sample size), since

$$\left|\frac{\mathbb{E}(U_{\tilde{t}} \mid \mathbb{H}_1)}{\sqrt{\mathrm{var}(U_{\tilde{t}})}}\right| = \mathcal{O}\left(\frac{b\sqrt{n}}{\sqrt{n}}\right) = \mathcal{O}(b) \to 0, \quad \text{as } n \nearrow \infty.$$

Here, $\mathbb{E}(U_{\tilde{t}} \mid \mathbb{H}_1)$ represents the expected value of U_t under the alternative. This suggests that regardless of the sample size, the CUSUM test cannot detect the existence of a small shift in the mean in the fixed domain setting.

Remark 5.2. To a prudent reader, CUSUM can be modified to a consistent detection scheme in the infill regime, if $\mathrm{var}(U_{\tilde{t}})$ can be consistently estimated under $\mathbb{H}_1$ (see Perron and Vogelsang (1992) for further details). We rigorously exclude the possibility of such phenomenon for a Matern Gaussian process with known smoothness parameter and unknown standard deviation σ and range parameter ρ. For brevity define $\Psi_n := \mathrm{var}\left(U_{\tilde{t}}\right)/(nc)$, in which $c := \sigma\rho^{-\nu}$. From Remark 5.1, we know that Ψ_n remains bounded as $n \to \infty$. Recall that in this case no estimation algorithm can consistently estimate both the pair (σ, ρ), and it is only feasible to estimate c. So $\mathrm{var}(U_{\tilde{t}})$ can be estimated if there exists a consistent estimate for Ψ_n. Furthermore, assume that $\tilde{t}/n$ is fully known a priori and tends to some $\beta \in (0, 1)$, as $n \to \infty$. Namely, we assume that the normalized location of change point is either fully known or can be consistently estimated. Then straightforward algebra leads to

$$\Psi_n = \tilde{t}\left(n-\tilde{t}\right)\int_{\mathbb{R}}\left|\frac{1}{n-\tilde{t}}\sum_{\tilde{t}}^{n}\frac{e^{-jk\omega/n}}{n} - \frac{1}{\tilde{t}}\sum_{k=1}^{\tilde{t}}\frac{e^{-jk\omega/n}}{n}\right|^2\left(\omega^2 + \frac{1}{\rho^2}\right)^{-\nu-\frac{1}{2}} d\omega.$$

For large n, $\left|\frac{1}{n-\tilde{t}}\sum_{\tilde{t}}^{n}\frac{e^{-jk\omega/n}}{n}-\frac{1}{\tilde{t}}\sum_{k=1}^{\tilde{t}}\frac{e^{-jk\omega/n}}{n}\right|$ can be well approximated by an integral. Thus one can show that

$$\Psi_n \to \int_{\mathbb{R}} \frac{f(\omega)-\beta f(\beta\omega)-(1-\beta)f((1-\beta)\omega)}{\left(\omega^2+\frac{1}{\rho^2}\right)^{\nu+\frac{1}{2}}}d\omega,$$

in which $f(x)=4sin^2(x/2)/x^2$. In other words, even for a fully known change point location, the limiting behavior of Ψ_n is not estimable, since it directly depends on a non-identifiable parameter (ρ).

Remark 5.3. The threshold value of the CUSUM test in Theorem 5.1 is given by

$$R_{n,\delta}=\sqrt{n\left(1+2\log\left(\frac{2n(1-2\alpha)}{\delta}\right)+2\sqrt{\log\left(\frac{2n(1-2\alpha)}{\delta}\right)}\right)}.$$

This threshold has a different form of dependence on n than that of the threshold of the GLRT in Eq. (3.6), since, unlike the GLRT, the CUSUM test does not reduce the correlation among the samples. In order to remove the gap between the threshold of GLRT and CUSUM in the fixed domain setting, we further normalize U_t by considering $U_n^\star=U_n/\sqrt{n}$. Equivalently, CUSUM test in this regime can be written as

$$T_{\mathrm{CUSUM}}=\mathbb{1}\left(\max_{t\in\mathcal{C}_{n,\alpha}}\left|U_n^\star\right|^2>R_{n,\delta}^\star:=\frac{R_{n,\delta}^2}{n}\right).$$

Here, $R_{n,\delta}^\star$ is exactly the same as the critical value of the GLRT.

6. Minimax lower bound on detection rate

In this section, we establish minimax lower bounds on the detectable jump in the mean of G in the fixed domain regime. Theorem 6.1 shows that the obtained rate for the plug-in GLRT (Theorem 4.1) is asymptotically near-optimal in a minimax sense. This result is applicable for rational spectral densities. First, let us formalize the notion of asymptotic (near)-optimality.

Definition 6.1. Given n samples, let $T\in\{0,1\}$ be a shift-in-mean detection algorithm whose CDEP is denoted by $\varphi_n(T)$. T is said to be asymptotically near-optimal in a minimax sense if, for any $\delta\in(0,2)$, there are sequences $\{h_n\}_{n=1}^{\infty}$ dependent on n, δ and the spectral density, such that

1. As $n\to\infty$, T can detect the existence of any abrupt change with the CDEP guarantee $\varphi_n(T)\le\delta$, provided that the jump size b satisfies $h_n\log n=o(b)$.
2. There is a large enough n_0 (depending on the model parameters) such that if $n\ge n_0$ and $|b|\le h_n$, then there is no algorithm whose CDEP is strictly less than δ.

Recall the fixed domain regime in which G is a Gaussian process defined on $[0,1]$ and is observed at $\{i/n\}_{i=1}^{n}$. We formally introduce a suitable class of spectral densities that we consider in this section. While somewhat more restrictive than Assumption 3.1, it still provides a sufficiently rich class of commonly used spectral densities.

Assumption 6.1. There are constants $p\in\mathbb{N}$ and $\beta\in(1/2,\infty)$ such that

1. $\lim_{\omega\to\infty}\hat{K}(\omega)|\omega|^{2p}$ exists and $C_K':=\lim_{\omega\to\infty}\hat{K}(\omega)|\omega|^{2p}\in(0,\infty)$.
2. $\limsup_{\omega\to\infty}\left|\left(\frac{\hat{K}(\omega)|\omega|^{2p}}{C_K'}-1\right)\omega^\beta\right|<\infty$.

Generally speaking, Assumption 6.1 contains the class of spectral densities $\hat{K}(\omega)$ for which there is some $p\in\mathbb{N}$ such that $\hat{K}(\omega)\asymp|\omega|^{-2p}$ as ω tends to infinity. Note that the second condition in Assumption 6.1 is of theoretical purposes and does not have a simple qualitative interpretation. It can be observed that Assumption 6.1 excludes any $\hat{K}(\omega)$ satisfying Assumption 3.1 with $(\nu+1/2)\notin\mathbb{N}$. For instance, Assumption 6.1 does not hold for Matern covariance functions with $(\nu+1/2)\notin\mathbb{N}$.

Remark 6.1. Here, we name a salient class of spectral densities satisfying Assumption 6.1.

- Simple calculations show that any rational spectral density $\hat{K}$ (See (3.4)) admits Assumption 6.1 with $C_K'=\lambda,\beta=1$ and $p=\deg(Q_d)-\deg(Q_n)\in\mathbb{N}$. Moreover, $\hat{K}$ satisfies Assumption 3.1 with $\nu=p-1/2$. Indeed the Matern covariance function with $p:=(\nu+1/2)\in\mathbb{N}$ has a rational spectral density. These particular instances of Matern covariance, which are commonly used in machine learning and geostatistics, are of the form $K(r)=Q(|r|)e^{-d|r|}$, where $Q(\cdot)$ is a polynomial of degree $p-1$.

Theorem 6.1. *Let $\delta \in (0, 2)$ and assume that Assumption 6.1 holds for K. Consider the change point detection problem (2.3) in which $\operatorname{cov}(\boldsymbol{X}) = \left[K\left(\frac{r-s}{n}\right)\right]_{r,s=1}^{n}$. There are positive scalars $\bar{C}_K$ and $n_0 := n_0(K)$ such that if $n \geq n_0$ and*

$$|b| \leq \bar{C}_K n^{-p+1/2} \sqrt{\log\left(\frac{1}{\delta(2-\delta)}\right)}, \tag{6.1}$$

then for any test T,

$$\varphi_n(T) \geq \delta.$$

See Appendix A.5 for the proof of Theorem 6.1.

Remark 6.2. Comparing the detection rate of the GLRT (see Theorem 3.1) and plug-in GLRT (see Theorem 4.1), with the rate established in Eq. (6.1) entails the asymptotically near optimality of the GLRT with known covariance structure and the plug-in GLRT for the class of spectral densities considered in Remark 6.1. Strictly speaking, under the fixed domain setting, there is a gap of order $\sqrt{\log n}$ between (6.1) and the detection rate of the GLRT based algorithms. Although we do not have a proof to establish the asymptotic near optimality of the GLRT and plug-in GLRT for the broader class of spectral densities admitting Assumption 3.1, our conjecture is that Theorem 6.1 can be extended to this broader class.

7. Simulation study

To illustrate the performance of the proposed shift-in-mean detection algorithms, we conduct a set of controlled simulation studies for verifying the results in Sections 3–5. We also present other simulation studies to assess the performance of CUSUM and GLRT in the increasing domain regime. Our goals are two-fold:

(a) comparing the performance of the GLRT based algorithms with the standard CUSUM test in the two asymptotic frameworks.
(b) assessing the sensitivity of algorithm (2.5) to the parameters of the covariance function.

In all the numerical studies in this section we fix $n = 500$ and $\alpha = 0.1$.

The area under the receiver operating characteristic (ROC) curve, which will be referred as AUC, is a standard way for assessing the performance of a test. The ROC curve plots the power against the false alarm probability. Since the ROC curve is confined in the unit square, the AUC ranges in [0,1]. The ROC curve of a test based on pure random guessing is the diagonal line between origin and $(1, 1)$ and so the AUC of any realistic test is at least 0.5.

The subsequent figures in this section exhibit empirical AUC versus b. For a fixed value of b, covariance function K and a detection algorithm T, we apply the following method to compute the AUC of T:

1. Set $T_1 = 500$ and $T_2 = 50$.
2. For $k = 1$ to T_2 repeat independently
   (a) For $\ell = 1$ to T_1 repeat independently
      i. Choose $p \in \{0, 1\}$ with equal probability which denotes null or alternative hypotheses. Thus, approximately $T_1/2 = 250$ experiments correspond to both null and alternative.
      ii. If $p = 0$, generate zero mean $\boldsymbol{X} \in \mathbb{R}^n$ according to covariance function K. That is, $\boldsymbol{X}$ are sampled from a Gaussian process with no abrupt shift in mean. Otherwise, choose $t \in [\alpha n, (1-\alpha)n] = \{50, 51, \ldots, 450\}$ uniformly at random (recall that t represents the location of the mean shift) and generate $\boldsymbol{X} \in \mathbb{R}^n$ according to $\mathbb{H}_{1,t}$.
      iii. Compute T score.
   (b) Numerically obtain the ROC curve of T based upon T_1 experiments in part *i*.
   (c) Given the ROC curve, compute AUC_k using trapezoidal integration method.
3. Compute the average AUC by $\overline{AUC} = \frac{1}{T_2}\sum_{k=1}^{T_2} AUC_k$.

The first simulation study aims to compare CUSUM and GLRT based algorithms in the fixed domain regime and assess the role of smoothness and other parameters of K in the performance of the GLRT. For this experiment G is a Gaussian process in [0, 1] which is observed at regularly spaced samples, $\mathcal{D}_n = \{k/n\}_{k=1}^{n}$, i.e., $X_k = G(k/n)$, $k = 1, \ldots, n$. The covariance function of G is assumed to have Matern form with parameters (σ_0, ρ_0, ν). Strictly speaking,

$$\operatorname{cov}(X_i, X_l) = \sigma_0^2 K_\nu\left(\left|\frac{i-l}{n\rho_0}\right|\right), \quad i, l = 1, \ldots, n,$$
$$K_\nu(x) = \frac{\sqrt{4\pi}\,\Gamma(\nu+1/2)}{\Gamma(\nu)} \int_{-\infty}^{\infty} e^{-j\omega x}\left(1+\omega^2\right)^{-(\nu+1/2)} d\omega, \quad \forall x \geq 0.$$

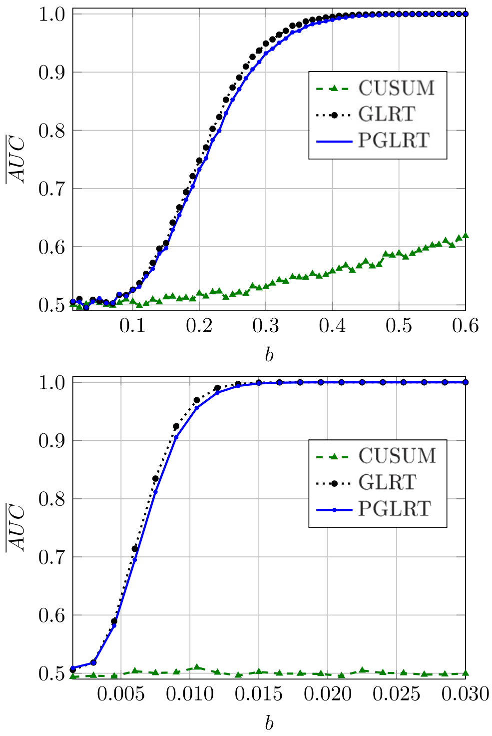

Fig. 1. The above figures assess the performance of different detection algorithms when G is a one dimensional Matern Gaussian process, with parameters (ν, σ_0, ρ_0), and regularly sampled in $[0, 1]$. From left to right then from top to bottom, $(\nu, \sigma_0, \rho_0) = (0.5, 1, 0.5)$, $(1, 1, 0.5)$, $(1.5, 1, 0.5)$. In each panel the horizontal axis displays the jump value b and the three curves (dashed black, solid blue and green) respectively exhibit the AUC of the GLRT with known covariance structure, plug in GLRT (PGLRT) using full MLE and CUSUM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We consider $\nu = 0.5, 1$, and 1.5. We also set $\sigma_0 = 1$ and $\rho_0 = 1/2$. As customary in the literature, we assume that ν is known and so ν will not be estimated. For conducting the plug-in GLRT procedure, both parameters (σ_0, ρ_0) are estimated using the MLE. Due to the low dimensionality of the unknown parameters, the most effective way to estimate (σ_0, ρ_0) is to apply a brute force grid search over a pre-specified set $\mathcal{P}$. Here, we choose $\mathcal{P} = \{0.2, 0.4, \dots, 2\} \times \{1/4, 1/3.9, \dots, 1/0.1\}$. The final results are exhibited in Fig. 1. We observe the following:

- GLRT and plug-in GLRT have a significantly better detection performance than CUSUM. This performance improvement is more pronounced for smoother covariance functions (larger ν). In particular, the CUSUM test is completely impractical for detection of a small change when $\nu = 1$ or 1.5.
- In each panel of Fig. 1, the GLRT has a slightly larger AUC than that of the plug-in GLRT suggesting a small gap between the smallest detectable jump of these two scenarios. Although the two GLRTs have the same rate, this gap is likely accounted for by the differing constants in Eqs. (3.5) and (4.1). In short, having full knowledge of the covariance parameters slightly improves the detection performance and so our proposed algorithm is robust to the estimation error of the unknown parameters of K.
- Comparing the range of b in each panel of Fig. 1 discloses that more rapid decay of the spectral density can decrease the smallest detectable jump. This observation substantiates the role of ν in the theory established in Sections 3 and 4.

Next, we compare the performance of the GLRT with known parameters and the CUSUM in the increasing domain setting. We have concisely discussed that the two methods have analogous asymptotic rates (see Appendix C.2 for further details). In the left panel of Fig. 2, we choose an exponentially decaying covariance function

$$\operatorname{cov}(X_i, X_l) = \sigma_0^2 \exp\left(-\frac{|i-l|}{\rho_0}\right), \qquad i, l = 1, \dots, n,$$

in which $\sigma_0 = 1$ and $\rho_0 = 2$. That is Σ_n has exponentially decaying off-diagonal entries. However, in the right panel, the chosen covariance function has a polynomially decaying tail given by

$$\operatorname{cov}(X_i, X_l) = \sigma_0^2 \left(1 + \frac{|i-l|}{\rho_0}\right)^{-(1+\lambda)},$$

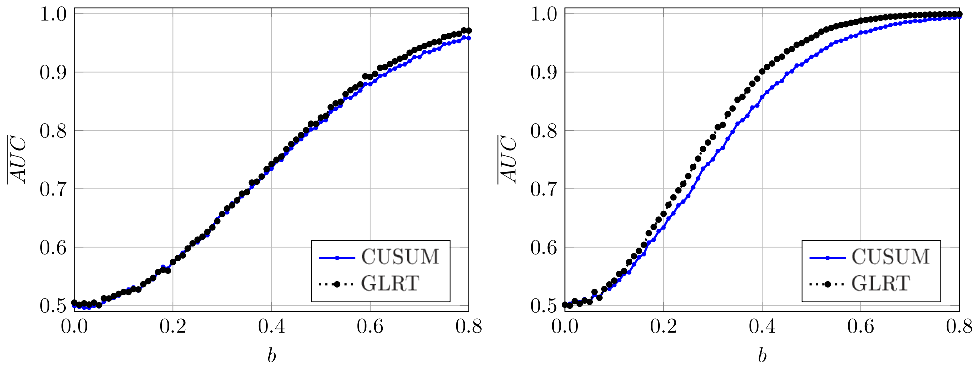

Fig. 2. The above figure assesses the performance of increasing domain detection algorithms. In each panel the horizontal axis displays the jump value b and the two curves (dashed black and solid blue) respectively exhibit the AUC of the GLRT with known covariance structure and CUSUM. In the right panel, we choose $\operatorname{cov}(X_i, X_l) = \sigma_0^2(1 + |i - l| / \rho_0)^{-(1+\lambda)}$ in which $(\sigma_0, \rho_0) = (1, 2)$ and $\lambda = 0.5$. For the left panel, the covariance function is given by $\operatorname{cov}(X_i, X_l) = \sigma_0^2 \exp(-|i - l| / \rho_0)$ where $(\sigma_0, \rho_0) = (1, 2)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with $\sigma_0 = 1$, $\rho_0 = 2$ and $\lambda = 0.5$. In this case, Σ_n has heavier off-diagonal terms. It is evident from Fig. 2 that the GLRT exhibits a slightly better performance than the CUSUM, and the gap between the two AUC curves is more visible in the case of polynomially decaying covariance function. Thus, we still recommend the use of GLRT in the presence of strong correlation among samples in applications described by the increasing domain regime.

8. Discussion

As indicated in the Introduction, a systematic treatment of the problem of detection of shift-in-mean of a Gaussian process in the fixed domain regime has remained unexplored. This work seeks to fill this gap, and to highlight the importance of accounting the dependence structure of observed samples. Moreover, the basic model (1.1) that we studied is only one of several plausible scenarios which can be subject to further investigation. We note that the probabilistic model of G can be extended in some possible ways:

(a) Here, we deal with a single abrupt change in $\mathbb{E}G$. However, we believe that our methods and techniques provide a good starting point to study the performance of the GLRT and plug-in GLRT for the case where there may be multiple shifts in the mean.

(b) Recently, Ivanoff and Merzbach (2010) studied the problem of *change-set* detection in two dimensional Poisson processes. Specifically, G is a Poisson process in $\mathbb{R}^2$ and there are scalars $\mu_0 \neq \mu_1$ and $\Omega \subset \mathbb{R}^2$ such that the intensity of G can be formulated by

$$\mathbb{E}G(s) = \mu_0 \mathbb{1}_{s\in\Omega} + \mu_1 \mathbb{1}_{s\notin\Omega}, \quad s \in \mathbb{R}^2. \tag{8.1}$$

The objective is to detect Ω as well as possible based on samples of G. However, a systematic study of change-set detection methods for multi-dimensional Gaussian processes remains unavailable. The fixed domain setting is clearly the natural way to study asymptotic behavior of detection procedures for these spatial processes. We hope this paper may provide valuable insights into this more challenging problem. We expect that aside from the sample size and smoothness of the covariance function, geometric properties of change-sets will have a crucial role in the design and analysis of detection algorithms.

Appendices

Appendix A contains the proofs of the main results in Sections 2–6. Appendix B states and proves the technical results required in the proofs of the main results. In addition, Appendix C establishes the increasing domain asymptotic behavior of CUSUM and GLRT detection procedures.

Appendix A. Proofs

A.1. Proofs for Section 2

Proof of Proposition 2.1. In the following $\mathfrak{L}$ stands for the generalized negative log-likelihood ratio.

$$2\mathfrak{L} = \boldsymbol{X}^\top (\Sigma_n)^{-1} \boldsymbol{X} - \min_{t\in\mathcal{C}_{n,\alpha}} \min_{b\neq 0} \left[\left(\boldsymbol{X} - \frac{b}{2}\zeta_t \right)^\top (\Sigma_n)^{-1} \left(\boldsymbol{X} - \frac{b}{2}\zeta_t \right) \right]. \tag{A.1}$$

Note that the objective function in (A.1) is quadratic in terms of b. The explicit form of $2\mathfrak{L}$ can be obtained with a bit of algebraic derivations. The algebra has been skipped to save space; we arrive at

$$2\mathfrak{L} = \max_{t\in\mathcal{C}_{n,\alpha}} \max_{b\neq 0} \left(-\frac{\zeta_t^\top (\Sigma_n)^{-1}\zeta_t}{4} b^2 + b\zeta_t^\top (\Sigma_n)^{-1}\boldsymbol{X} \right) = \max_{t\in\mathcal{C}_{n,\alpha}} \left| \frac{\zeta_t^\top (\Sigma_n)^{-1}\boldsymbol{X}}{\sqrt{\zeta_t^\top (\Sigma_n)^{-1}\zeta_t}} \right|^2 .$$

So, there is a threshold value, $R_{n,\delta} > 0$, for which the GLRT is given by (2.4). □

The following result expressing the form of GLRT in the generic case of unknown μ can be proved in an analogous way as Proposition 2.1.

Proposition A.1. *There is $R_{n,\delta} > 0$ for which the GLRT is given by*

$$T_{GLRT} = \mathbb{1}\left(\max_{t\in\mathcal{C}_{n,\alpha}} \left| \frac{\langle \boldsymbol{Y}, \zeta_t - B_1(t)\mathbb{1}_n\rangle}{\sqrt{B_2(t)}} \right|^2 \geq R_{n,\delta} \right), \tag{A.2}$$

where $\boldsymbol{Y} = (\Sigma_n)^{-1}\boldsymbol{X}$ and

$$B_1(t) = \frac{\zeta_t^\top (\Sigma_n)^{-1}\mathbb{1}_n}{\mathbb{1}_n^\top (\Sigma_n)^{-1}\mathbb{1}_n}, \quad B_2(t) = \zeta_t^\top (\Sigma_n)^{-1}\zeta_t - \frac{\left(\zeta_t^\top (\Sigma_n)^{-1}\mathbb{1}_n\right)^2}{\mathbb{1}_n^\top (\Sigma_n)^{-1}\mathbb{1}_n}.$$

A.2. Proofs for Section 3

Proof of Theorem 3.1. Let $p = \lceil \nu + 1/2 \rceil$, $P = \{1, \ldots, p\}$ and $\theta_n = \exp(-1/n)$. Construct a banded triangular matrix $A_n \in \mathbb{R}^{n\times n}$ by the following procedure.

$$A_n[k, k-j] = \binom{p}{j} (-\theta_n)^j, \; j \in \{0, \ldots, p\}, \; k \in \{p+1, \ldots, n\},$$
$$(A_n)_{P,P} = n^{-2\nu} I_p .$$

It is relatively simple to verify that A_n is invertible. In addition, for brevity let $Z_t = \frac{\zeta_t^\top (\Sigma_n)^{-1}\boldsymbol{X}}{\sqrt{\zeta_t^\top (\Sigma_n)^{-1}\zeta_t}}$ for any $t \in \mathcal{C}_{n,\alpha}$, in which ζ_t has been defined in (2.2). Lastly, define $U_{n,t} := A_n\zeta_t \in \mathbb{R}^n$, $W := A_n\boldsymbol{X}$ and $D_n := \mathrm{cov}(W)$.

Easy calculations show that under the null hypothesis $\{Z_t\}_{t\in\mathcal{C}_{n,\alpha}}$ is a set of standard Gaussian random variables and so, by Lemma B.1, we have $\mathbb{P}\left(\max_{t\in\mathcal{C}_{n,\alpha}} Z_t^2 \geq R_{n,\delta}\right) \leq \delta/2$. That is, the false alarm probability is less than $\delta/2$. Moreover if the alternative hypothesis $\mathbb{H}_{1,\tilde{t}}$ (for some $\tilde{t} \in \mathcal{C}_{n,\alpha}$) holds then $\left\{Z_t^2\right\}_{t\in\mathcal{C}_{n,\alpha}}$ are non-central χ_1^2 random variables and the non-centrality parameter of $Z_{\tilde{t}}^2$ is given by

$$\mathbb{E}\left(Z_{\tilde{t}} \mid \mathbb{H}_{1,\tilde{t}}\right) = \frac{|b|}{2}\sqrt{\zeta_{\tilde{t}}^\top (\Sigma_n)^{-1}\zeta_{\tilde{t}}}.$$

Applying Lemma B.1 ($\sigma_0 = \sigma_k = 1$ for any k) demonstrates that $\varphi_n(T_2) \leq \delta$, whenever

$$|b|\sqrt{\zeta_{\tilde{t}}^\top (\Sigma_n)^{-1}\zeta_{\tilde{t}}} \geq |b| \min_{t\in\mathcal{C}_{n,\alpha}} \sqrt{\zeta_t^\top (\Sigma_n)^{-1}\zeta_t} \geq 8\sqrt{\log\left(\frac{4n}{\delta}\right)}. \tag{A.3}$$

Thus, in order to get a sufficient condition on detectable b, it suffices to find a tight uniform lower bound on $\zeta_t^\top (\Sigma_n)^{-1}\zeta_t$ for $t \in \mathcal{C}_{n,\alpha}$.

The identity $\Sigma_n^{-1} = A_n^\top (D_n)^{-1} A_n$ can be shown using the linearity of covariance operator and non-singularity of A. Choose $t \in \mathcal{C}_{n,\alpha}$ in an arbitrary way. As a result of this alternative representation of Σ_n^{-1}, we have $\zeta_t^\top (\Sigma_n)^{-1}\zeta_t = U_{n,t}^\top (D_n)^{-1} U_{n,t}$. Applying *Kantorovich* inequality (cf. Appendix B) and the triangle inequality yields

$$\zeta_t^\top (\Sigma_n)^{-1}\zeta_t = U_{n,t}^\top (D_n)^{-1} U_{n,t} \geq \frac{\left\|U_{n,t}\right\|_{\ell_2}^4}{U_{n,t}^\top D_n U_{n,t}} \geq \left[\frac{\left\|U_{n,t}\right\|_{\ell_2}^2}{\left\|U_{n,t}\right\|_{\ell_1}}\right]^2 \frac{1}{\|D_n\|_{\ell_\infty}}. \tag{A.4}$$

Now, we show that $\frac{\|U_{n,t}\|_{\ell_2}^2}{\|U_{n,t}\|_{\ell_1}} \geq \frac{1}{3}$, for large enough n. Indeed, after some algebra, we can get

$$\begin{aligned}
\left\|U_{n,t}\right\|_{\ell_2}^2 &\geq \sum_{k=t+1}^{t+p} U_{n,t}^2(k) = \sum_{k=1}^{p}\left[-(1-\theta_n)^p + 2\sum_{j=0}^{k-1}\binom{p}{j}(-\theta_n)^j\right]^2 \\
&\overset{(a)}{\geq} 2\sum_{k=1}^{p}\left[\sum_{j=0}^{k-1}\binom{p}{j}(-1)^j\right]^2 = 2\sum_{k=1}^{p}\left[\binom{p-1}{k-1}(-1)^{k-1}\right]^2 = 2\binom{2(p-1)}{p-1} \geq 2^p,
\end{aligned} \tag{A.5}$$

where inequality (a) follows from the fact that for large enough n, θ_n is arbitrarily close to 1. To get an upper bound on $\left\|U_{n,t}\right\|_{\ell_1}$,

$$\begin{aligned}
\left\|U_{n,t}\right\|_{\ell_1} &= \sum_{k=1}^{p}\left|U_{n,t}(k)\right| + \sum_{k=p+1}^{t}\left|U_{n,t}(k)\right| + \sum_{k=t+p+1}^{n}\left|U_{n,t}(k)\right| + \sum_{k=t+1}^{t+p}\left|U_{n,t}(k)\right| \\
&= \sum_{k=1}^{p} n^{-2\nu} + \sum_{k=p+1}^{t}(1-\theta_n)^p + \sum_{k=t+p+1}^{n}(1-\theta_n)^p + \sum_{k=1}^{p}\left|-(1-\theta_n)^p + 2\sum_{j=0}^{k-1}\binom{p}{j}(-\theta_n)^j\right| \\
&\leq pn^{-2\nu} + n^{1-p} + 2\sum_{k=1}^{p}\left|\sum_{j=0}^{k-1}\binom{p}{j}(-\theta_n)^j\right| \overset{(b)}{\leq} 2 + 2\sum_{k=1}^{p}\left|\sum_{j=0}^{k-1}\binom{p}{j}(-\theta_n)^j\right| \\
&\leq 2 + 4\sum_{k=1}^{p}\left|\sum_{j=0}^{k-1}\binom{p}{j}(-1)^j\right| = 2 + 4\sum_{k=1}^{p}\left|\binom{p-1}{k-1}(-1)^{k-1}\right| = 2 + 2^{p+1} \leq 3\,2^p.
\end{aligned} \tag{A.6}$$

Note that inequality (b) is valid when $pn^{-2\nu} + n^{1-p} \leq 2$, which obviously holds for sufficiently large $n = \mathcal{O}(1)$. The remaining inequalities and identities in (A.6) can be easily verified via basic properties of the binomial coefficients. Combining (A.5) and (A.6) yields the desired goal. Now, inequality (A.4) can be rewritten as

$$\zeta_t^\top (\Sigma_n)^{-1}\zeta_t \geq \frac{1}{9\left\|D_n\right\|_{\ell_\infty}} = \left[9 \max_{1\leq k\leq n} \operatorname{var}(W_k)\right]^{-1}. \tag{A.7}$$

In the final phase of the proof, we achieve a tight upper bound on $\max_{1\leq k\leq n}\operatorname{var}(W_k)$. It is obvious from the formulation of A_n and the stationarity of $\boldsymbol{X} - \mathbb{E}\boldsymbol{X}$ that $\max_{1\leq k\leq n}\operatorname{var}(W_k) = n^{-2\nu} \vee \operatorname{var}\left(W_{p+1}\right)$. So, the goal is reduced to give an upper bound on the variance of W_{p+1}.

$$\begin{aligned}
\operatorname{var}\left(W_{p+1}\right) &= \operatorname{var}\left(\sum_{r=0}^{p}\binom{p}{r}(-\theta_n)^r X_{p+1-r}\right) \overset{(c)}{=} \frac{1}{2\pi}\int_{\mathbb{R}} \hat{K}(\omega)\left|\sum_{r=0}^{p}\binom{p}{r}(-\theta_n)^r \exp\left(\frac{-jr\omega}{n}\right)\right|^2 d\omega \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{K}(\omega)\left|\sum_{r=0}^{p}\binom{p}{r}\left(-\exp\left(\frac{-(1+j\omega)}{n}\right)\right)^r\right|^2 d\omega = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{K}(\omega)\left|1 - e^{\frac{-(1+j\omega)}{n}}\right|^{2p} d\omega \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{K}(\omega)\left[1+\theta_n^2 - 2\theta_n\cos(\omega/n)\right]^p d\omega \overset{(d)}{\leq} \frac{C_K}{2\pi}\int_{\mathbb{R}} \frac{\left[1+\theta_n^2 - 2\theta_n\cos\left(\frac{\omega}{n}\right)\right]^p}{\left(1+\omega^2\right)^{\nu+1/2}} d\omega,
\end{aligned} \tag{A.8}$$

where, identity (c) is implied by *Bochner theorem* (cf. Stein (1999), Chapter 2) and (d) is immediate consequence of Assumption 3.1. Notice that

$$1+\theta_n^2 - 2\theta_n\cos\left(\frac{\omega}{n}\right) \leq (1-\theta_n)^2 + 2\theta_n\left(1-\cos\left(\frac{\omega}{n}\right)\right) \leq \frac{1}{n^2} + 2\left(1-\cos\left(\frac{\omega}{n}\right)\right) = \frac{1}{n^2} + \left[\frac{\omega}{n}\operatorname{sinc}\left(\frac{\omega}{2n}\right)\right]^2.$$

Let $\xi = p - (\nu + 1/2) < 1$. Henceforth, for any $R > 0$,

$$\begin{aligned}
\frac{2\pi n^{2\nu}}{C_K}\operatorname{var}\left(W_{p+1}\right) &\leq n^{2\nu}\int_{\mathbb{R}} \frac{\left\{\frac{1}{n^2} + \left[\frac{\omega}{n}\operatorname{sinc}\left(\frac{\omega}{2n}\right)\right]^2\right\}^p}{\left(1+\omega^2\right)^{\nu+1/2}} d\omega = \int_{\mathbb{R}} \frac{\left\{1/n^2 + \left[\omega\operatorname{sinc}\left(\frac{\omega}{2}\right)\right]^2\right\}^p}{\left(1/n^2+\omega^2\right)^{\nu+1/2}} d\omega \\
&= \int_{-R}^{R} \frac{\left\{1/n^2 + \left[\omega\operatorname{sinc}\left(\frac{\omega}{2}\right)\right]^2\right\}^p}{\left(1/n^2+\omega^2\right)^{\nu+1/2}} d\omega + \int_{|\omega|\geq R} \frac{\left\{1/n^2 + \left[\omega\operatorname{sinc}\left(\frac{\omega}{2}\right)\right]^2\right\}^p}{\left(1/n^2+\omega^2\right)^{\nu+1/2}} d\omega
\end{aligned}$$

$$\overset{(e)}{\le} \int_{-R}^{R} \left(1/n^2+\omega^2\right)^{\xi} d\omega + \int_{|\omega|\ge R} \frac{\left\{1/n^2+\left[\omega \operatorname{sinc}\left(\frac{\omega}{2}\right)\right]^2\right\}^p}{\left(1/n^2+\omega^2\right)^{\nu+1/2}} d\omega$$
$$\overset{(f)}{\le} \int_{-R}^{R} \left(1/n^2+\omega^2\right)^{\xi} d\omega + 5^p \int_{|\omega|\ge R} |\omega|^{-(2\nu+1)} d\omega \overset{(g)}{\le} 3R^3 + 5^p \frac{R^{-2\nu}}{\nu}. \tag{A.9}$$

Inequality (e) follows form the fact that $\sup_{\omega\in\mathbb{R}} |\operatorname{sinc}(\omega/2)| \le 1$. In order to justify (f), observe that $|\omega \operatorname{sinc}(\omega/2)| \le 2$ for any $\omega \in \mathbb{R}$. Thus, for large enough n and $|\omega| \ge R$, we get

$$\frac{\left\{1/n^2+\left[\omega \operatorname{sinc}\left(\frac{\omega}{2}\right)\right]^2\right\}^p}{\left(1/n^2+\omega^2\right)^{\nu+1/2}} \le |\omega|^{-(2\nu+1)}\left(1/n^2+4\right)^p \le 5^p |\omega|^{-(2\nu+1)}.$$

Note that there is some $n_0 := n_0(R,\nu)$ such that $\sup_{\omega\in\mathbb{R}}\left(1/n^2+\omega^2\right)^{\xi} \le 3/2R^2$ for all $n > n_0$. This immediately entails inequality (g).

Finally, minimizing the obtained upper bound in (A.9) over $R > 0$, we get

$$\operatorname{var}\left(W_{p+1}\right) \le CC_K n^{-2\nu}\left(1+\frac{1}{\nu}\right) \tag{A.10}$$

for some universal constant $C > 0$. Thus, there is another strictly positive universal constant, C', for which $\max_{1\le k\le n} \operatorname{var}(W_k) = n^{-2\nu} \vee \operatorname{var}\left(W_{p+1}\right) \le C'C_K n^{-2\nu}\left(1+\frac{1}{\nu}\right)$. So, (A.7) implies that

$$\zeta_t^{\top}(\Sigma_n)^{-1}\zeta_t \gtrsim \frac{n^{2\nu}}{C_K\left(1+\frac{1}{\nu}\right)}. \tag{A.11}$$

The combination of (A.3) and (A.11) completes our proof. □

Proof of Theorem 3.2. The proof proceeds in a similar manner as that of the preceding theorem, in the sense that it is required to show that inequality (A.3) holds. Let $\theta_n = \exp\left(-\frac{\rho^2}{n^2}\right)$. A_n represents the inverse of the Cholesky factorization of Σ_n. For any $k \le j$ and $q \in [0,1]$, $G(k,j;q)$ denotes the following rational function.

$$G(k,j;q) = \prod_{\ell=j-k+1}^{j}\left(1-q^{\ell}\right)\left[\prod_{\ell=1}^{k}\left(1-q^{\ell}\right)\right]^{-1},$$

and $G(k,j;1) = \binom{j}{k}$. $G(k,j;q)$ is usually referred to Gaussian binomial coefficients in the combinatorics literature. Finally, let $U_{n,t} := A_n\zeta_t$. Similar to (A.3), the aim is to obtain a universal lower bound on $\zeta_t^{\top}(\Sigma_n)^{-1}\zeta_t$ for $t \in \mathcal{C}_{n,\alpha}$. Observe that, $\zeta_t^{\top}(\Sigma_n)^{-1}\zeta_t = \left\|U_{n,t}\right\|_{\ell_2}^2$.

In order to achieve a tight lower bound on $\left\|U_{n,t}\right\|_{\ell_2}$, it is pivotal to study the non-asymptotic behavior of the entries of A_n. According to Proposition 1 of Loh and Lam (2000), the entries of A_n are given by

$$(A_n)_{jk} = \left(-\sqrt{\theta_n}\right)^{(j-k)} \frac{G(k-1,j-1;\theta_n)}{\sqrt{\prod_{\ell=1}^{j-1}\left(1-\theta_n^{\ell}\right)}} \mathbb{1}_{\{j\ge k\}}.$$

Since $\frac{\ell\rho^2}{n^2}$ tends to 0 as n gets large for any $\ell \in \{0,\ldots,n\}$ and $\lim_{x\searrow 0}\frac{1-e^{-x}}{x} = 1$, we get

$$\left[\prod_{\ell=1}^{j-1}\left(1-\theta_n^{\ell}\right)\right]^{-1} = \left[\prod_{\ell=1}^{j-1}\left(1-\exp\left(-\frac{\ell\rho^2}{n^2}\right)\right)\right]^{-1} \asymp \frac{1}{(j-1)!}\left(\frac{n}{\rho}\right)^{2(j-1)}. \tag{A.12}$$

Direct calculations show that $G(k-1,j-1;\theta_n) \asymp \binom{j-1}{k-1}$ for any θ_n in a small neighborhood of 1 and $j,k \in \{1,\ldots,n\}$. Thus,

$$\left(-\sqrt{\theta_n}\right)^{(j-k)} G(k-1,j-1;\theta_n) \asymp (-1)^{(j-k)}\binom{j-1}{k-1}. \tag{A.13}$$

The asymptotic identities (A.12) and (A.13) come in handy to analyze $\|U_n\|_{\ell_2}^2$:

$$\begin{aligned}\|U_n\|_{\ell_2}^2 &\geq \sum_{j=t+1}^{n}(U_n)_j^2 = \sum_{j=t+1}^{n}\left[\sum_{k=t+1}^{j}(A_n)_{jk} - \sum_{k=1}^{t}(A_n)_{jk}\right]^2\\ &\asymp \sum_{j=t+1}^{n}\frac{1}{(j-1)\,!}\left(\frac{n}{\rho}\right)^{2(j-1)}\left[\sum_{k=t+1}^{j}(-1)^{(j-k)}\binom{j-1}{k-1} - \sum_{k=1}^{t}(-1)^{(j-k)}\binom{j-1}{k-1}\right]^2\\ &= \sum_{j=t+1}^{n}\frac{1}{(j-1)\,!}\left(\frac{n}{\rho}\right)^{2(j-1)}\left[\sum_{k=1}^{j}(-1)^{(j-k)}\binom{j-1}{k-1} - 2\sum_{k=1}^{t}(-1)^{(j-k)}\binom{j-1}{k-1}\right]^2\\ &= \sum_{j=t+1}^{n}\frac{1}{(j-1)\,!}\left(\frac{n}{\rho}\right)^{2(j-1)}\left[0-2(-1)^j\binom{j-1}{t}\right]^2 \asymp \sum_{j=t+1}^{n}\frac{\binom{j-1}{t}^2}{(j-1)\,!}\left(\frac{n}{\rho}\right)^{2(j-1)}.\end{aligned}$$

Thus, there are universal constants $C, C' > 0$ and C_0 depending on α and ρ such that

$$\|U_n\|_{\ell_2}^2 \geq C\sum_{j=t+1}^{n}\frac{\binom{j-1}{t}^2}{(j-1)\,!}\left(\frac{n}{\rho}\right)^{2(j-1)} \geq C\frac{\binom{n-1}{t}^2}{(n-1)\,!}\left(\frac{n^2}{\rho^2}\right)^{(n-1)} \overset{(a)}{\geq} C'\left(\frac{n}{t}\right)^t\frac{1}{\sqrt{n}}\left(\frac{en^2}{n\rho^2}\right)^{(n-1)} \overset{(b)}{\geq} (C_0 n)^n.$$

Note that inequality (a) can be shown using *Stirling's formula* and (b) is obvious implication of the fact that $t \leq (1-\alpha)\,n$ (Recall $\mathcal{C}_{n,\alpha}$ from Section 2.1). In summary, we have that

$$|b|\sqrt{\zeta_n^\top(\Sigma_n)^{-1}\zeta_n} \gtrsim |b|\,(C_0 n)^{n/2}.$$

We conclude the proof by appealing to Lemma B.1. □

A.3. Proofs for Section 4

Proof of Theorem 4.1. For simplicity set $c := \sigma^2\rho^{-2\nu}$ and use $\tilde{c}_m$ to represent its estimated quantity $\tilde{\sigma}_m^2\tilde{\rho}_m^{-2\nu}$. Recall that $\tilde{\rho}_m$ is a fixed quantity which has been chosen as the largest possible range parameter in the space Ω. Since c is consistently estimable by the maximum likelihood algorithm, there are vanishing non-negative sequences $\{\tau_m\}_{m=1}^\infty$ and $\{\varepsilon_m\}_{m=1}^\infty$ and $n_0\mathbb{N}$ such that

$$\mathbb{P}(\mathcal{A}_m) := \mathbb{P}\left(\left|\frac{\tilde{c}_m}{c}-1\right| < \varepsilon_m\right) > 1-\tau_m, \quad \forall\, n \geq n_0.$$

As the range of φ_n is $[0,2]$ (See Definition 3.1), we have

$$\varphi_n\left(\tilde{T}_{GLRT}\right) \leq 2\tau_m + \mathbb{E}\left(\varphi_n\left(\tilde{T}_{GLRT}\right) \mid \mathcal{A}_m\right). \tag{A.14}$$

Furthermore, for any $\eta' = \left(\sigma', \tilde{\rho}_m\right) \in \Omega$

$$Z_t\left(\eta'\right) := \frac{\zeta_t^\top\Sigma_n^{-1}\left(\eta'\right)\boldsymbol{X}}{\sqrt{\zeta_t^\top\Sigma_n^{-1}(\eta')\,\zeta_t}}. \tag{A.15}$$

Notice that the Matern covariance matrix associated to η' can be re-parametrized as

$$\Sigma_n\left(\eta'\right) = \left[K\left(\frac{r-s}{n},\eta'\right)\right]_{r,s=1}^n = \sigma'^2\rho'^{-2\nu}\left[\int_{\mathbb{R}}\left[\omega^2+\rho'^{-2}\right]^{-(\nu+1/2)}\exp\left(-j\omega\left(\frac{r-s}{n}\right)\right)d\omega\right]_{r,s=1}^n. \tag{A.16}$$

Notice that the matrix appearing in the second line of (A.16), which will be denoted by $\Gamma_n\left(\rho'\right)$, only depends on ρ'. The following property of $\Gamma_n(\cdot)$ is essential in our proof.

$$\Gamma_n(\rho_1) \preceq \Gamma_n(\rho_2)\,, \quad \forall\ (\rho_1,\rho_2) \in \Omega \text{ with } \rho_1 \leq \rho_2$$

We aim to obtain a sufficient condition on b to control the second term on the right hand side of (A.14) below δ. Similar to the proof of Theorem 3.1, it is necessary to study the two following quantities: 1. variance of $Z_t\left(\eta'\right)$ and 2. expected value of $Z_t\left(\eta'\right)$ under the alternative hypothesis, to control the false alarm and miss detection probabilities. Observe that

$$\begin{aligned}\mathrm{var}Z_t\left(\eta'\right) &= \frac{\zeta_t^\top\Sigma_n^{-1}\left(\eta'\right)\Sigma_n(\eta)\,\Sigma_n^{-1}\left(\eta'\right)\zeta_t}{\zeta_t^\top\Sigma_n^{-1}(\eta')\,\zeta_t} = \frac{\sigma^2\rho^{-2\nu}}{\sigma'^2\tilde{\rho}_m^{-2\nu}}\;\frac{\zeta_t^\top\Gamma_n^{-1}(\tilde{\rho}_m)\,\Gamma_n(\rho)\,\Gamma_n^{-1}(\tilde{\rho}_m)\,\zeta_t}{\zeta_t^\top\Gamma_n^{-1}(\tilde{\rho}_m)\,\zeta_t}\\ &\overset{(a)}{\leq} \frac{1}{1-\varepsilon_m}\,\frac{\zeta_t^\top\Gamma_n^{-1}(\tilde{\rho}_m)\,\Gamma_n(\rho)\,\Gamma_n^{-1}(\tilde{\rho}_m)\,\zeta_t}{\zeta_t^\top\Gamma_n^{-1}(\tilde{\rho}_m)\,\zeta_t} \overset{(b)}{\leq} \frac{1}{1-\varepsilon_m}.\end{aligned} \tag{A.17}$$

Where (a) is an easy implication of the fact that $\eta' \in \mathcal{A}_m$. Furthermore, (b) follows from the fact that $\Gamma(\rho) \preceq \Gamma(\tilde{\rho}_m)$. In other words, $\mathrm{var} Z_t(\eta') < 1/(1-\varepsilon_m)$. Lemma B.1 guarantees the existence of a vanishing sequence $\{\tau'_m\}_{m\in\mathbb{N}}$, which depends on ε_m, such that

$$\mathbb{P}\left(\max_{1\le t\le n} Z_t^2(\tilde{\eta}_m) \ge \left[1+2\left(\log\left(\frac{2n}{\delta}\right)+\sqrt{\log\left(\frac{2n}{\delta}\right)}\right)\right] \mid \mathcal{A}_m\right) \le \frac{\delta}{2}+\tau'_m. \tag{A.18}$$

So, we have controlled type one error from above in (A.18). Now we turn to control the type two error from above. Assume that there is a sudden change in the mean at $\bar{t} \in \mathcal{C}_{n,\alpha}$. According to Lemma B.1, type *II* error is less than $\delta/2$ whenever for any $\eta' \in \mathcal{A}_m$

$$\frac{|b|}{2}\sqrt{\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta')\zeta_{\bar{t}}} \ge \frac{4}{1-\varepsilon_m}\sqrt{\log\left(\frac{2n}{\delta}\right)}. \tag{A.19}$$

Having a lower bound on $\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta')\zeta_{\bar{t}}$ is necessary to make sure that b satisfying (A.19). Notice that

$$\Sigma_n(\eta') = \sigma'^2\tilde{\rho}_m^{-2\nu}\Gamma_n(\tilde{\rho}_m) \preceq (1+\varepsilon_m)\sigma^2\rho^{-2\nu}\Gamma_n(\tilde{\rho}_m) = (1+\varepsilon_m)c\Gamma_n(\tilde{\rho}_m).$$

Thus, (A.19) holds true whenever

$$\frac{|b|}{2}\sqrt{c\zeta_{\bar{t}}^{\top}\Gamma_n^{-1}(\tilde{\rho}_m)\zeta_{\bar{t}}} \ge \frac{4\sqrt{1+\varepsilon_m}}{1-\varepsilon_m}\sqrt{\log\left(\frac{2n}{\delta}\right)}. \tag{A.20}$$

Theorem 4 of Skorokhod and Yadrenko (1973) conveys the equivalence of the associated Gaussian measures to matrices $\Gamma_n(\rho)$ and $\Gamma_n(\tilde{\rho}_m)$. Thus Lemma B.3 ensures the existence of a bounded scalar $B > 1$ for which

$$\zeta_{\bar{t}}^{\top}\Gamma_n^{-1}(\tilde{\rho}_m)\zeta_{\bar{t}} \ge \frac{1}{B}\zeta_{\bar{t}}^{\top}\Gamma_n^{-1}(\rho)\zeta_{\bar{t}}. \tag{A.21}$$

Combining (A.20) and (A.21) yields a sufficient condition on b to control the type two error

$$\frac{|b|}{2}\sqrt{c\zeta_{\bar{t}}^{\top}\Gamma_n^{-1}(\rho)\zeta_{\bar{t}}} = \frac{|b|}{2}\sqrt{\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta)\zeta_{\bar{t}}} \ge \frac{4\sqrt{B(1+\varepsilon_m)}}{1-\varepsilon_m}\sqrt{\log\left(\frac{2n}{\delta}\right)}.$$

In conclusion we employ the lower bound on $\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta)\zeta_{\bar{t}}$ in (A.11) gives the proper rate of detectable b. □

Proof of Theorem 4.2. As the proof has much in common with the proof of Theorem 4.1, we skip the algebraic details to avoid repetition. $\tilde{\eta}_m = (\tilde{\sigma}_m, \tilde{\rho}_m)$ stands for an estimate of the unknown parameters $\eta = (\sigma, \rho)$. Note that in this case η is consistently estimable, i.e. there are two vanishing sequences ε_m and τ_m, $m \in \mathbb{N}$ and a large enough $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}(\mathcal{A}_m) := \mathbb{P}(|\rho-\tilde{\rho}_m| \vee |\sigma-\tilde{\sigma}_m| < \varepsilon_m) > 1-\tau_m, \quad \forall\, n \ge n_0.$$

Recall from (A.14) that $\varphi_n\left(\tilde{T}_{GLRT}\right) \le 2\tau_m + \mathbb{E}\left(\varphi_n\left(\tilde{T}_{GLRT}\right) \mid \mathcal{A}_m\right)$. We use $\hat{K}_{\eta'}(\omega)$ to represent the spectral density of the powered exponential covariance function associated to $\eta' = (\sigma', \rho')$. That is

$$\hat{K}_{\eta'}(\omega) = \int_{\mathbb{R}} \sigma'^2 \exp\left(-\left|\frac{x}{\rho'}\right|^{\beta} - jx\omega\right)dx.$$

It is trivial that $\hat{K}_{\eta'}(\cdot)$ is a strictly positive, continuous function of both ω and η'. Furthermore, due to the compactness of the parameter space Ω, $\hat{K}_{\eta'}$ is uniformly continuous with respect to η'. Thus, there is another vanishing sequence ε'_m (depending on ε_m) such that

$$\left\|\frac{\hat{K}_{\eta'}}{\hat{K}_{\eta}} - 1\right\|_{\infty} \le \varepsilon'_m, \quad \forall\, \eta' = (\sigma', \rho') \text{ with } |\rho-\rho'| \vee |\sigma-\sigma'| < \varepsilon_m.$$

So, the following inequality holds for any $\omega \in \mathbb{R}$, and any η' satisfying $|\rho-\rho'| \vee |\sigma-\sigma'| < \varepsilon_m$.

$$\hat{K}_{\eta}(\omega)(1-\varepsilon'_m) \le \hat{K}_{\eta'}(\omega) \le \hat{K}_{\eta}(\omega)(1+\varepsilon'_m). \tag{A.22}$$

Note that (A.22) can be easily translated in terms of the covariance matrices $\Sigma_n(\eta)$ and $\Sigma_n(\eta')$.

$$(1-\varepsilon'_m)\Sigma_n(\eta) \preceq \Sigma_n(\eta') \preceq (1+\varepsilon'_m)\Sigma_n(\eta). \tag{A.23}$$

Now using similar techniques as (A.17) and (A.18), we can bound the variance of $Z_t(\eta')$.

$$\mathrm{var} Z_t(\eta') = \frac{\zeta_t^{\top}\Sigma_n^{-1}(\eta')\Sigma_n(\eta)\Sigma_n^{-1}(\eta')\zeta_t}{\zeta_t^{\top}\Sigma_n^{-1}(\eta')\zeta_t} \le \frac{1}{1-\varepsilon'_m}\frac{\zeta_t^{\top}\Sigma_n^{-1}(\eta')\Sigma_n(\eta')\Sigma_n^{-1}(\eta')\zeta_t}{\zeta_t^{\top}\Sigma_n^{-1}(\eta')\zeta_t} = \frac{1}{1-\varepsilon'_m},$$

and control the type one error below $\delta/2+\tau'_m$ for a vanishing sequence τ'_m (depending on ε'_m). Now we introduce a sufficient condition on b to keep the type two error below $\delta/2$. Using Lemma B.1 b should satisfy

$$\frac{|b|}{2}\sqrt{\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta')\,\zeta_{\bar{t}}} \geq \frac{4}{1-\varepsilon'_m}\sqrt{\log\left(\frac{2n}{\delta}\right)}$$

for any η' with $\left|\rho-\rho'\right| \vee \left|\sigma-\sigma'\right| < \varepsilon_m$. It follows from (A.23) that

$$\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}\left(\eta'\right)\zeta_{\bar{t}} \geq \left(1-\varepsilon'_m\right)\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta)\,\zeta_{\bar{t}}.$$

So, we can have a slightly stronger restriction on b by combining the last two inequalities.

$$\frac{|b|}{2}\sqrt{\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta)\,\zeta_{\bar{t}}} \geq \frac{4}{\left(1-\varepsilon'_m\right)^2}\sqrt{\log\left(\frac{2n}{\delta}\right)}. \tag{A.24}$$

Notice that as m increases we have $4/\left(1-\varepsilon'_m\right)^2 < 5$. Thus, (A.24) holds when $n \geq n_0$ (for some large enough n_0) and

$$\frac{|b|}{2}\sqrt{\zeta_{\bar{t}}^{\top}\Sigma_n^{-1}(\eta)\,\zeta_{\bar{t}}} \geq 5\sqrt{\log\left(\frac{2n}{\delta}\right)}. \tag{A.25}$$

Inequality (A.25) is same as the sufficient condition on b for the case of known covariance parameters, which leads to the same detection rate. $\square$

Proof of Theorem 4.3. Define $c_1 := \sigma_1^2\rho_1^{-2\nu_1}$ and use $\tilde{c}_m$ to represent its estimated quantity $\tilde{\sigma}_m^2\tilde{\rho}_m^{-2\nu}$. Notice that $\tilde{\rho}_m$ is a fixed quantity which has been chosen as the largest possible range parameter in the space Ω and $\tilde{\sigma}_m$ is the estimate of σ_1 using the misspecified Matern model. The consistency of $\tilde{c}_m$ means that there are vanishing non-negative sequences $\{\tau_m\}_{m=1}^{\infty}$ and $\{\varepsilon_m\}_{m=1}^{\infty}$ and $n_0\mathbb{N}$ such that

$$\mathbb{P}\left(\mathcal{A}_m\right) := \mathbb{P}\left(\left|\frac{\tilde{c}_m}{c_1}-1\right| < \varepsilon_m\right) > 1-\tau_m, \quad \forall\, n \geq n_0.$$

Thus using a similar technique as Eq. (A.14), we get

$$\varphi_n\left(\tilde{T}_{GLRT}\right) \leq 2\tau_m + \mathbb{E}\left(\varphi_n\left(\tilde{T}_{GLRT}\right) \mid \mathcal{A}_m\right).$$

Furthermore, define $Z_t\left(\eta'\right)$ for any $\eta' = (\sigma, \tilde{\rho}_m) \in \Omega$ exactly same as Eq. (A.15) (recall $\Sigma_n\left(\eta'\right)$ from Eq. (A.16)). Notice that the true covariance matrix of $\boldsymbol{X}$ is given by

$$S_n\left(\sigma_1,\ldots,\sigma_p,\rho_1,\ldots,\rho_p\right) := \sum_{r=1}^{p}\sigma_r^2\rho_r^{-2\nu_r}\left[\int_{\mathbb{R}}\left[\omega^2+\rho_r^{-2}\right]^{-(\nu_r+1/2)}\exp\left(-j\omega\left(\frac{t-s}{n}\right)\right)d\omega\right]_{s,t=1}^{n}.$$

The variance of $Z_t\left(\eta'\right)$ (for a fixed, arbitrary η') is given by

$$\mathrm{var}Z_t\left(\eta'\right) = \frac{\zeta_t^{\top}\Sigma_n^{-1}\left(\eta'\right)S_n\left(\sigma_1,\ldots,\sigma_p,\rho_1,\ldots,\rho_p\right)\Sigma_n^{-1}\left(\eta'\right)\zeta_t}{\zeta_t^{\top}\Sigma_n^{-1}\left(\eta'\right)\zeta_t}.$$

Finally define $\eta := (\sigma_1, \rho_1)$. Corollary 4.1 guarantees the existence of a bounded constant C_1 such that

$$\mathrm{var}Z_t\left(\eta'\right) \leq C\frac{\zeta_t^{\top}\Sigma_n^{-1}\left(\eta'\right)\Sigma_n(\eta)\,\Sigma_n^{-1}\left(\eta'\right)\zeta_t}{\zeta_t^{\top}\Sigma_n^{-1}\left(\eta'\right)\zeta_t}, \quad \forall\,\eta' = (\sigma, \tilde{\rho}_m) \in \Omega.$$

Thus, applying the same argument as Eq. (A.17) implies that $\mathrm{var}Z_t\left(\eta'\right) \leq C/(1-\varepsilon_m)$ for any $\eta' = (\sigma, \tilde{\rho}_m) \in \Omega$. We skip the rest of the proof as it is exactly same as the proof of Theorem 4.1. $\square$

A.4. Proofs for Section 5

Proof of Theorem 5.1. Choose $t \in C_{n,\alpha}$ and define

$$U_t^{\star} := \sqrt{\frac{t\,(n-t)}{n^2}}\left(\frac{1}{n-t}\sum_{k=t+1}^{n}X_k - \frac{1}{t}\sum_{k=1}^{n}X_k\right).$$

Moreover set

$$R_{n,\delta} = \sqrt{n\left(1+2\log\left(\frac{2n\,(1-2\alpha)}{\delta}\right)+2\sqrt{\log\left(\frac{2n\,(1-2\alpha)}{\delta}\right)}\right)}.$$

Note that under the null hypothesis, $U_t^\star$ is a zero mean random variable and

$$\begin{aligned}\lim_{n\to\infty} \operatorname{var}\left(U_t^\star\right) &\overset{(a)}{=} \lim_{n\to\infty} \frac{t\,(n-t)}{n^2} \int_{-\infty}^{\infty} \frac{\hat{K}\,(\omega)}{2\pi} \left| \frac{1}{n-t} \sum_{k=t+1}^{n} \exp\left(-jk\omega/n\right) - \frac{1}{t} \sum_{k=1}^{n} \exp\left(-jk\omega/n\right) \right|^2 d\omega \\ &= \lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{\hat{K}\,(\omega)}{2\pi} \left| \sqrt{\frac{\beta}{1-\beta}} \sum_{k=t+1}^{n} \frac{\exp\left(-jk\omega/n\right)}{n} - \sqrt{\frac{1-\beta}{\beta}} \sum_{k=1}^{n} \frac{\exp\left(-jk\omega/n\right)}{n} \right|^2 d\omega \\ &\overset{(a)}{=} \int_{-\infty}^{\infty} \frac{\hat{K}\,(\omega)}{2\pi} \left| \sqrt{\frac{\beta}{1-\beta}} \int_{\beta}^{1} e^{-j\omega u} du - \sqrt{\frac{1-\beta}{\beta}} \int_{0}^{\beta} e^{-j\omega u} du \right|^2 d\omega = \int_{-\infty}^{\infty} \frac{\hat{K}\,(\omega)\, G_\beta\,(\omega)}{2\pi} d\omega,\end{aligned}$$

where

$$\begin{aligned}G_\beta\,(\omega) &:= \left[(1-\beta)\operatorname{sinc}\left(\frac{\beta\omega}{2}\right)\right]^2 + \left[\beta \operatorname{sinc}\left(\frac{(1-\beta)\,\omega}{2}\right)\right]^2 \\ &\quad + 4\beta\,(1-\beta)\operatorname{sinc}\left(\frac{\beta\omega}{2}\right)\operatorname{sinc}\left(\frac{(1-\beta)\,\omega}{2}\right)\sin^2\left(\frac{\omega}{2}\right).\end{aligned} \tag{A.26}$$

The identity (a) is implied by Bochner Theorem and (b) follows from the dominated convergence theorem. It is easy to see that $\left\|G_\beta\right\|_\infty \le 1$ and so $\lim_{n\to\infty}\operatorname{var}\left(U_n^\star\right) \le 1$ by the triangle inequality. Moreover, Lemma B.5 shows that the achieved upper bound on $\sigma_n^2 = \operatorname{var}\left(U_n^\star\right)$ is tight up to some constant whenever $\hat{K}$ has a uniformly bounded derivative. Namely, there is a universal constant $c \in (0, 1)$ such that $c \le \lim_{n\to\infty}\operatorname{var}\left(U_n^\star\right) \le 1$ for any $\beta \in (0, 1)$. Let $R_{n,\delta}^\star = R_{n,\delta}^2/n$. Thus

$$\mathbb{P}\left(T = 1 \mid \mathbb{H}_0\right) = \mathbb{P}\left(\max_{t\in\mathcal{C}_{n,\alpha}} |U_t| \ge R_{n,\delta} \mid \mathbb{H}_0\right) = \mathbb{P}\left(\max_{t\in\mathcal{C}_{n,\alpha}} \left|U_t^\star\right|^2 \ge R_{n,\delta}^\star \mid \mathbb{H}_0\right). \tag{A.27}$$

For any $t \in \mathcal{C}_{n,\alpha}$, $\left|U_t^\star\right|^2$ is a (non-normalized) χ_1^2 random variable, as $\sigma_n^2 \le 1$. Moreover $\left|\mathcal{C}_{n,\alpha}\right| = n\,(1-2\alpha)$. So the part (a) of Lemma B.1 says that

$$\mathbb{P}\left(\max_{t\in\mathcal{C}_{n,\alpha}} \left|U_t^\star\right|^2 \ge R_{n,\delta}^\star \mid \mathbb{H}_0\right) \le \frac{\delta}{2}.$$

Now we turn to control the miss detection probability. Without loss of generality assume that $b > 0$. Choose an arbitrary $t \in \mathcal{C}_{n,\alpha}$. A line of algebra shows that

$$\mathbb{E}\left(U_t^\star \mid \mathbb{H}_{1,t}\right) \ge b\sqrt{\alpha\,(1-\alpha)}. \tag{A.28}$$

Eq. (5.1) on b implies that $\mathbb{E}\left(U_t^\star \mid \mathbb{H}_{1,t}\right) \ge 4\sqrt{\log\left(2n\,(1-2\alpha)\,/\delta\right)}$. In other words, given a sudden jump at t, $\left|U_s^\star\right|^2,\ s \in \mathcal{C}_{n,\alpha}$ are non-central χ_1^2 random variables satisfying the conditions of the part (b) of Lemma B.1. Hence

$$\mathbb{P}\left(T = 0 \mid \mathbb{H}_{1,t}\right) = \mathbb{P}\left(\max_{s\in\mathcal{C}_{n,\alpha}} \left|U_s^\star\right|^2 \le R_{n,\delta}^\star \mid \mathbb{H}_{1,t}\right) \le \frac{\delta}{2}. \quad \square \tag{A.29}$$

A.5. Proofs for Section 6

Proof of Theorem 6.1. We follow the standard method for bounding the Bayes risk from below. Observe that

$$\begin{aligned}\inf_T \varphi_n\,(T) &= 1 - \sup_T \inf_{t\in\mathcal{C}_{n,\alpha}} \left[\mathbb{P}\left(T = 0 \mid \mathbb{H}_0\right) - \mathbb{P}\left(T = 0 \mid \mathbb{H}_{1,t}\right)\right] \\ &\ge 1 - \inf_{t\in\mathcal{C}_{n,\alpha}} \sup_T \left|\mathbb{P}\left(T = 0 \mid \mathbb{H}_0\right) - \mathbb{P}\left(T = 0 \mid \mathbb{H}_{1,t}\right)\right| \overset{(a)}{\ge} 1 - \inf_{t\in\mathcal{C}_{n,\alpha}} H\left(\mathbb{P}_0, \mathbb{P}_{1,t}\right),\end{aligned}$$

where (a) follows from inequality 2.27 in Tsybakov (2009). So, it suffices to show that $\inf_{t\in\mathcal{C}_{n,\alpha}} H^2\left(\mathbb{P}_0, \mathbb{P}_{1,t}\right) \le (1-\delta)^2$. A few lines of straightforward algebra on the explicit form of Hellinger distance of Gaussian measures indicates that $\inf_T \varphi_n\,(T) \ge \delta$, whenever

$$b^2 \inf_{t\in\mathcal{C}_{n,\alpha}} \zeta_t^\top \left(\Sigma_n\right)^{-1} \zeta_t \le 32 \log\left(\frac{1}{\delta\,(2-\delta)}\right). \tag{A.30}$$

Henceforth, it is enough to obtain a tight upper bound on $\inf_{t\in\mathcal{C}_{n,\alpha}} \zeta_t^\top \left(\Sigma_n\right)^{-1} \zeta_t$.

Let $\sigma = 1$ and choose $\rho > 0$ by $\rho^{-2p+1} = \frac{C_K' \Gamma(p-1/2)}{\sqrt{4\pi}\,\Gamma(p)}$. Furthermore, let $\hat{F}_{\rho,p,\sigma} : \mathbb{R} \mapsto \mathbb{R}$ denote the Matern spectral density parametrized by p, ρ and σ as (3.2). Note that ρ is well defined due to the first condition in Assumption 6.1. Define $\xi_t \in \mathbb{R}^n$

by $\xi_t(k) = \mathbb{1}_{\{k>t\}}$ and let $\xi'_t = \xi_t - \zeta_t$ for any $t \in \mathcal{C}_{n,\alpha}$. Moreover, let $\theta_n = \exp(-1/n)$ and $S_t = \{t+1, \ldots, n\}$. Finally, define the covariance matrix $\Psi_n \in \mathbb{R}^{n\times n}$ by $\Psi_n = \left[F_{\rho,p,\sigma}\left((r-s)/n\right)\right]_{r,s=1}^n$. Observe that

$$\begin{aligned}\zeta_t^\top(\Sigma_n)^{-1}\zeta_t &= 2\left(\xi_t^\top(\Sigma_n)^{-1}\xi_t + \xi_t^{'\top}(\Sigma_n)^{-1}\xi'_t\right) - \mathbb{1}_n^\top(\Sigma_n)^{-1}\mathbb{1}_n \\ &\le 4\left(\xi_t^\top(\Sigma_n)^{-1}\xi_t \vee \xi_t^{'\top}(\Sigma_n)^{-1}\xi'_t\right).\end{aligned} \tag{A.31}$$

We aim to prove that there is a constant $C := C(p) > 0$ for which $\xi_t^\top(\Sigma_n)^{-1}\xi_t \le Cn^{2p-1}$. The same upper bound can be obtained for $\xi_t^{'\top}(\Sigma_n)^{-1}\xi'_t$ in an analogous manner.

We first show that

$$\left(\frac{\hat{K}}{\hat{F}_{\rho,p,\sigma}} - 1\right) \in \mathbb{L}^2(\mathbb{R}). \tag{A.32}$$

Let M represent the finite lim sup in the second condition of Assumption 6.1. Without loss of generality, we can assume that $\beta < 2$ in Assumption 6.1. Using a few lines of algebra one can find a bounded positive scalar M for which the following inequality holds.

$$\begin{aligned}\limsup_{\omega\to\infty}\left|\omega^\beta\left(\frac{\hat{K}(\omega)}{\hat{F}_{\rho,p,\sigma}(\omega)} - 1\right)\right| &= \limsup_{\omega\to\infty}\left|\omega^\beta\left(\frac{\hat{K}(\omega)\,\omega^{2p}}{C'_K}\left(1+\frac{1}{\rho^2\omega^2}\right)^p - 1\right)\right| \le \limsup_{\omega\to\infty}\left|\omega^\beta\left(\frac{\hat{K}(\omega)\,\omega^{2p}}{C'_K} - 1\right)\right| \\ &\quad + \limsup_{\omega\to\infty}\left|\omega^\beta\left[\frac{\hat{K}(\omega)\,\omega^{2p}}{C'_K}\left(\left(1+\frac{1}{\rho^2\omega^2}\right)^p - 1\right)\right]\right| \\ &\overset{(a)}{=} M + \frac{2p\rho^{-2}}{C'_K}\limsup_{\omega\to\infty}\hat{K}(\omega)\,|\omega|^{2p-2+\beta} \overset{(b)}{=} M.\end{aligned}$$

Notice that, identity (a) follows from Assumption 6.1 and first order Taylor expansion of $(1+x)^p$ for infinitesimal $x > 0$. Moreover, (b) follows from the combination of $\beta < 2$ and the first condition in Assumption 6.1. Namely, there is $R > 0$ such that

$$\left|\frac{\hat{K}(\omega)}{\hat{F}_{\rho,p,\sigma}(\omega)} - 1\right| \le \frac{2M}{|\omega|^\beta}, \quad \forall\ |\omega| \ge R,$$

which substantiates (A.32) as $\beta > 1/2$.

It is known (4.31, Chapter *III*, Ibragimov and Rozanov, 1978) that there is a function $\phi \in \mathbb{L}^2(\mathbb{R})$ with bounded support such that $\hat{F}_{\rho,p,\sigma}(\omega) \asymp \left|\hat{\phi}(\omega)\right|^2$ as $|\omega| \to \infty$. Theorem 4 of Skorokhod and Yadrenko, (1973) implies that the associated zero mean Gaussian measures to spectral densities $\hat{K}$ and $\hat{F}_{\rho,p,\sigma}$ are equivalent. Based upon Lemma B.3, there exists a constant $h \in (0, \infty)$ such that

$$\frac{1}{h} \le \left|\lim_{n\to\infty}\frac{\xi_t^\top(\Sigma_n)^{-1}\xi_t}{\xi_t^\top(\Psi_n)^{-1}\xi_t}\right| \le h.$$

So, it suffices to show that $\xi_t^\top(\Psi_n)^{-1}\xi_t \le C'n^{2p-1}$ for some appropriately chosen $C' > 0$ depending on h and C. Letting $\nu = p - 1/2$ and recalling A_n, W and D_n form the proof of Theorem 3.1, we have

$$\xi_t^\top(\Psi_n)^{-1}\xi_t = (A_n\xi_t)^\top D_n^{-1}(A_n\xi_t) \overset{(b)}{\le} \frac{\|A_n\xi_t\|_{\ell_2}^2}{\lambda_{\min}(D_n(S_t, S_t))}. \tag{A.33}$$

Note that inequality (b) is inferred from $\operatorname{supp}(A_n\xi_t) = S_t$. Applying a similar technique as (A.5), we get

$$\begin{aligned}\|A_n\xi_t\|_{\ell_2}^2 &= (n-t-p)(1-\theta_n)^p + \sum_{k=1}^p\left(\sum_{j=0}^{k-1}\binom{p}{j}(-\theta_n)^j\right)^2 \le n(1-\theta_n)^p + 2\sum_{k=1}^p\left(\sum_{j=0}^{k-1}\binom{p}{j}(-1)^j\right)^2 \\ &\le 2\binom{2p-2}{p-1} + n(1-\theta_n)^p \le n^{-(p-1)} + 2(2e)^{p-1} \le (2e)^p.\end{aligned} \tag{A.34}$$

So, $\xi_t^\top(\Psi_n)^{-1}\xi_t \le (2e)^{2p}[\lambda_{\min}(D_n(S_t, S_t))]^{-1}$.

Next, we control the smallest eigenvalue of $D_n(S_t, S_t)$ from the below. We first control the diagonal entries from below. Note that all the diagonal entries of $D_n(S_t, S_t)$ are the same and given by (cf. (A.8))

$$\begin{aligned} Q &= \int_{\mathbb{R}} \frac{\hat{F}_{\rho,p,\sigma}(\omega)}{2\pi}\left[1+\theta_n^2-2\theta_n\cos(\omega/n)\right]^p d\omega \overset{(c)}{\propto} \int_{\mathbb{R}} \rho^{-2\nu}\left(\frac{1}{\rho^2}+\omega^2\right)^{-p}\left[1+\theta_n^2-2\theta_n\cos(\omega/n)\right]^p d\omega \\ &= n^{-2\nu}\int_{\mathbb{R}}\left[\frac{(1-\theta_n)^2+4\theta_n\sin^2(\omega/2\rho)}{1/n^2+\omega^2}\right]^p d\omega \overset{(d)}{\geq} \frac{n^{-2\nu}\rho^{-2p}}{2}\int_{\mathbb{R}}[\mathrm{sinc}(\omega/2\rho)]^{2p}d\omega = C'_\rho n^{-2\nu}, \end{aligned} \tag{A.35}$$

where (c) is obtained from (3.2) and the inequality (d) follows from the fact that for any $\gamma \in (0, 1)$ (here we put $\gamma = 2^{-\frac{1}{p}}$), there is $n_0(\gamma)$ such that for any $n \geq n_0(\gamma)$,

$$\frac{(1-\theta_n)^2+4\theta_n\sin^2(\omega/2\rho)}{1/n^2+\omega^2} \geq \frac{\gamma}{\rho^2}[\mathrm{sinc}(\omega/2\rho)]^2.$$

We skip the proof of this inequality due to the simplicity.

Now, let $\Xi := D_n(S_t, S_t)/Q$. The combination of (A.33), (A.34) and (A.35) shows that

$$\xi_t^\top(\Psi_n)^{-1}\xi_t \leq \frac{C_0 n^{-2\nu}}{\lambda_{\min}(\Xi)} \quad\Rightarrow\quad \xi_t^\top(\Sigma_n)^{-1}\xi_t \leq \frac{C'_0 n^{2\nu}}{\lambda_{\min}(\Xi)} = \frac{C'_0 n^{2p-1}}{\lambda_{\min}(\Xi)},$$

for some constants, $C_0(p)$ and C'_0 depending on C_0, h and K. It can be shown using identity 1.2 of Bolin and Lindgren (2011) that there is some integrable function $g : [-\pi, \pi] \mapsto \mathbb{R}$ with $m_g := \operatorname{essinf}(g) > 0$ such that Ξ is a $p-$banded correlation matrix, i.e. $\Xi(r, s) = 0$ for $|r-s| \geq p$, and $\Xi = \mathcal{T}_n(f)$. It remains to note that Lemma 6 of Gray (2006) implies that $\lambda_{\min}(\Xi) > m_g$ for any n, which concludes the proof. □

Appendix B. Auxiliary results

This section contains several technical results needed in Appendix A.

Lemma B.1. *Let $\sigma_0 \geq 1$ and $n \geq 2$. Let $\mathfrak{z} \in \mathbb{R}^n$ be a Gaussian random vector with $\mathbb{E}\mathfrak{z} = \mu$ and $\mathrm{var}\mathfrak{z}_k \leq \sigma_0^2$ for any $1 \leq k \leq n$. Moreover, let $R_n = 1 + 2\left(\log\left(\frac{2n}{\delta}\right) + \sqrt{\log\left(\frac{2n}{\delta}\right)}\right)$. For any $\delta \in (0, 1)$ and any $n \in \mathbb{N}$, the following results hold.*

1. *If $\mu = 0$, then $\mathbb{P}\left[\max_{1\leq j\leq n}\mathfrak{z}_j^2 \geq \sigma_0^2 R_n\right] \leq \frac{\delta}{2}$.*
2. *If $\max_{1\leq j\leq n}|\mu_j| \geq 4\sigma_0\sqrt{\log\left(\frac{2n}{\delta}\right)}$, then $\mathbb{P}\left[\max_{1\leq j\leq n}\mathfrak{z}_j^2 \leq \sigma_0^2 R_n\right] \leq \frac{\delta}{2}$.*

Proof. For brevity, let $\sigma_j = \mathrm{var}\mathfrak{z}_j$, $j = 1, \ldots, n$. Notice that $\left(\frac{\mathfrak{z}_j}{\sigma_j}\right)^2$ are standard χ_1^2 random variables, for any $j = 1, \ldots, n$. Lemma 8.1 in Birgé (2001) implies that $\mathbb{P}\left(\mathfrak{z}_j^2 \geq \sigma_j^2 R_n\right) \leq \frac{\delta}{2n}$. Thus, $\mathbb{P}\left(\mathfrak{z}_j^2 \geq \sigma_0^2 R_n\right) \leq \frac{\delta}{2n}$ due to $\sigma_j \leq \sigma_0$. We conclude the proof of the first part by a union bound argument. Now, we turn to prove the second part. Define $k := \arg\max_{1\leq j\leq n}|\mu_j|$. It is easy to verify that $R_n \leq 4\log\left(\frac{2n}{\delta}\right)$. Observe that

$$\mathbb{P}\left[\max_{1\leq j\leq n}\mathfrak{z}_j^2 \leq \sigma_0^2 R_n\right] \leq \mathbb{P}\left[\frac{\mathfrak{z}_k^2}{\sigma_k^2} \leq \left(\frac{\sigma_0}{\sigma_k}\right)^2 R_n\right] \leq \mathbb{P}\left[\frac{\mathfrak{z}_k^2}{\sigma_k^2} \leq 4\left(\frac{\sigma_0}{\sigma_k}\right)^2\log\left(\frac{2n}{\delta}\right)\right].$$

Moreover, $\frac{\mathfrak{z}_k^2}{\sigma_k^2}$ is a non-central χ_1^2 random variables with non-centrality parameter $B_k := \left|\frac{\mu_k}{\sigma_k}\right|$. The lower bound condition on $|\mu_k|$ implies that $B_k \geq 4\frac{\sigma_0}{\sigma_k}\sqrt{\log\left(\frac{2n}{\delta}\right)}$. We finish the proof by the following inequality,

$$\mathbb{P}\left[\frac{\mathfrak{z}_k^2}{\sigma_k^2} \leq 4\left(\frac{\sigma_0}{\sigma_k}\right)^2\log\left(\frac{2n}{\delta}\right)\right] \overset{(a)}{\leq} \mathbb{P}\left[\frac{\mathfrak{z}_k^2}{\sigma_k^2} \leq 1 + B_k^2 - 2\sqrt{\left(1+2B_k^2\right)\log\left(\frac{2}{\delta}\right)}\right] \overset{(b)}{\leq} \frac{\delta}{2}.$$

In order to demonstrate inequality (a), we need to show that $1 + B_k^2 - 2\sqrt{\left(1+2B_k^2\right)\log\left(\frac{2}{\delta}\right)} \geq 4\left(\frac{\sigma_0}{\sigma_k}\right)^2\log\left(\frac{2n}{\delta}\right)$ which can be shown by obvious inequality $\sigma_0/\sigma_k \geq 1$ and a few lines of algebra. Inequality (b) can be inferred from Lemma 8.1 of Birgé (2001). □

Proposition B.1 (*Kantorovich Inequality (p. 452, Horn and Johnson, 2012)*). *Let $\Sigma \in \mathbb{R}^{n\times n}$ be a non-singular covariance matrix and let $V \in \mathbb{R}^n$ be a non-zero vector. Then, $V^\top\Sigma^{-1}V \geq \frac{\|V\|_{\ell_2}^4}{V^\top\Sigma V}$.*

Lemma B.2. *Let $\delta \in (0,2)$, $d \in (0,\infty)$ and define $K : \mathbb{R} \mapsto \mathbb{R}$ by $K(r) = \sigma^2 \exp\left(-\left|\frac{r}{\rho}\right|^{\delta}\right)$. Then,*

$$\lim_{\omega\to\infty} \hat{K}(\omega)\,|\omega|^{1+\delta} = C_\delta(\rho,\sigma) := \frac{\sigma^2 \delta \Gamma(\delta) \sin\left(\frac{\pi\delta}{2}\right)}{\pi \rho^{\delta}}.$$

Proof. Obviously $C_\delta(\rho,\sigma) = \sigma^2 C_\delta(\rho,1)$, so without loss of generality assume that $\sigma = 1$. Moreover $K(r)$ is of index δ as $|r| \to 0$, i.e. $\lim_{|r|\to 0} \frac{1-K(r\lambda)}{1-K(r)} = \lambda^\delta \ \forall\ \lambda > 0$. The Tauberian Theorem (p. 35, Stein, 1999) says that

$$\lim_{\omega\to\infty} [1-K(1/\omega)]^{-1} \int_\omega^\infty \hat{K}(u)\,du = \frac{\Gamma(\delta)\sin\left(\frac{\pi\delta}{2}\right)}{\pi} = \frac{C_\delta(\rho,1)\,\rho^\delta}{\delta}. \tag{B.1}$$

Moreover, the first order Taylor expansion of e^{-x} is at 0, implies that $[1-K(1/\omega)]^{-1}(\rho\omega)^{-\delta} \to 1$ as $\omega \to 0$. Thus, (B.1) can be rewritten by last limiting identity and applying L'Hospital's rule.

$$C_\delta(\rho,1) = \lim_{\omega\to\infty} \delta\rho^{-\delta}(\rho\omega)^\delta \delta\omega^\delta \int_\omega^\infty \hat{K}(u)\,du = \lim_{\omega\to\infty} \delta\omega^\delta \int_\omega^\infty \hat{K}(u)\,du = \lim_{\omega\to\infty} \hat{K}(\omega)\,|\omega|^{1+\delta}. \quad \square$$

The following lemma is probably well-known in the literature of Gaussian processes (e.g. the identity 2 of Stein (1999) (p.112) is analogous but not exactly same as the part (*a*) of Lemma B.3). Because of the absence of direct references, we include and prove the following result in this section.

Lemma B.3. *Let G_i, $i = 1,2$ be two zero mean stationary Gaussian process in $[0,1]$ associated to covariance functions K_i, $i = 1,2$, respectively. For any $n \in \mathbb{N}$, define two positive definite covariance matrices by $\Sigma_n := \left[K_1\left(\frac{r-s}{n}\right)\right]$ and $\Psi_n := \left[K_2\left(\frac{r-s}{n}\right)\right]$. If G_1 and G_2 induce equivalent measures on the Hilbert space of $\mathbb{L}^2([0,1])$, then there exists a scalar $B \in [1,\infty)$ for which*

1. $\frac{1}{B} \le \lim_{n\to\infty}\inf_{v\neq \mathbf{0}_n} \frac{v^\top \Sigma_n v}{v^\top \Psi_n v} \le \lim_{n\to\infty}\sup_{v\neq \mathbf{0}_n} \frac{v^\top \Sigma_n v}{v^\top \Psi_n v} \le B.$
2. $\frac{1}{B} \le \lim_{n\to\infty}\inf_{v\neq \mathbf{0}_n} \frac{v^\top \Sigma_n^{-1} v}{v^\top \Psi_n^{-1} v} \le \lim_{n\to\infty}\sup_{v\neq \mathbf{0}_n} \frac{v^\top \Sigma_n^{-1} v}{v^\top \Psi_n^{-1} v} \le B.$

Proof. We use $\mathbb{P}_i$, $i = 1,2$ to denote the probability measures with respect to G_i, $i = 1,2$, respectively. Abusing the notation, $\boldsymbol{X} \in \mathbb{R}^n$ represents the random vector generated by sampling Gaussian process at $\{k/n\}_{k=1}^n$ for any $n \in \mathbb{N}$. We prove the existence of a finite scalar B_1 for which $\lim_{n\to\infty}\sup_{v\neq\mathbf{0}_n} \frac{v^\top \Sigma_n v}{v^\top \Psi_n v} \le B_1$. Assume toward contradiction that $\lim_{n\to\infty}\sup_{v\neq\mathbf{0}_n} \frac{v^\top \Sigma_n v}{v^\top \Psi_n v}$ tends to infinity. So, there is a sequence of non-zero vectors $\{v_n \in \mathbb{R}^n\}_{n=1}^\infty$ such that

$$\limsup_{n\to\infty} \frac{v_n^\top \Sigma_n v_n}{v_n^\top \Psi_n v_n} = \infty. \tag{B.2}$$

Consider the measurable event $\mathrm{E}_n = \left[|\langle v_n, \boldsymbol{X}\rangle| \ge \sqrt{v_n^\top \Sigma_n v_n}\right]$. Simple calculations show that

$$\mathbb{P}_1(\mathrm{E}_n) = Q(1), \quad \mathbb{P}_2(\mathrm{E}_n) = Q\left(\sqrt{\frac{v_n^\top \Sigma_n v_n}{v_n^\top \Psi_n v_n}}\right), \tag{B.3}$$

in which $Q(\cdot)$ stands for the Q-function, i.e. $Q(r) = \int \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) \mathbb{1}(|x| \ge r)\,dx$. Combining (B.2) and (B.3) leads to $\limsup_{n\to\infty} \frac{\mathbb{P}_1(\mathrm{E}_n)}{\mathbb{P}_2(\mathrm{E}_n)} = \infty$ which contradicts the absolute continuity of $\mathbb{P}_1$ with respect to $\mathbb{P}_2$. One can show using the same technique that there is $B_2 \in (1,\infty)$ such that

$$\frac{1}{B_2} \le \lim_{n\to\infty}\inf_{v\neq\mathbf{0}_n} \frac{v^\top \Sigma_n v}{v^\top \Psi_n v}.$$

We conclude the proof by choosing $B = B_1 \vee B_2$. Now, we turn to substantiate the second claim. Pick a non-zero vector $v \in \mathbb{R}^n$. According to Lemma B.4, there is a suitably chosen n-dimensional vector u (the inner product of u and v is necessarily 1) such that

$$\frac{v^\top \Sigma_n^{-1} v}{v^\top \Psi_n^{-1} v} = v^\top \Sigma_n^{-1} v u^\top \Psi_n u = \frac{u^\top \Psi_n u}{\max_{\langle\omega,v\rangle=1} \omega^\top \Sigma_n \omega} \le \frac{u^\top \Psi_n u}{u^\top \Sigma_n u} \overset{(a)}{\le} B.$$

Note that the inequality (*a*) is obtained from the first part of this lemma. Taking supremum over all non-zero $v \in \mathbb{R}^n$ and $n \in \mathbb{N}$ terminates the proof. $\square$

Lemma B.4. *Let $\Sigma \in \mathbb{R}^{n\times n}$ be a non-singular covariance matrix and let $\omega \in \mathbb{R}^n$ be a non-zero vector. Then,*

$$\left(\omega^\top \Sigma^{-1} \omega\right)^{-1} = \min_{\langle v,\omega\rangle=1} v^\top \Sigma v. \tag{B.4}$$

Proof. Since the optimization problem in (B.4) is a convex program with continuously differentiable objective function and constraint, so its minimal value can be obtained solving the KKT equations. That is, there are $\hat{\lambda} \geq 0$ and $\hat{\upsilon}$ such that

$$2\Sigma\hat{\upsilon} - \hat{\lambda}\omega = 0, \quad \hat{\lambda}\left(\langle \hat{\upsilon}, \omega\rangle - 1\right) = 0.$$

Solving the above set of equations yields, $\hat{\upsilon} = \frac{\Sigma^{-1}\omega}{\omega^\top \Sigma^{-1}\omega}$. The desired result will be established by replacing $\hat{\upsilon}$ into the right hand side of (B.4). □

Lemma B.5. *Let K be a covariance function such that $\left\|\hat{K}'\right\|_\infty < \infty$ and define $G_\beta : \mathbb{R} \mapsto [0, 1]$ by (A.26). Then, there is a universal constant $c > 0$ such that*

$$\inf_{\beta\in(0,1)} \int_{-\infty}^{\infty} \hat{K}(\omega)\, G_\beta(\omega)\, d\omega \geq c.$$

Proof. Observe that for any $\omega \in \mathbb{R}, G_\beta(\omega)$ is a quadratic function of β in the compact interval $[0, 1]$ and $\lim_{n\to\infty} \left\|G_{\beta_n} - G_\beta\right\|_\infty = 0$ for any convergent sequence $\beta_n \to \beta$. This property implies that

$$\inf_{\beta\in(0,1)} \int_{-\infty}^{\infty} \hat{K}(\omega)\, G_\beta(\omega)\, d\omega \geq \frac{1}{2}\left[\inf_{\beta\in(0,1),\ |\beta-1/2|\geq r} \int_{-\infty}^{\infty} \hat{K}(\omega)\, G_\beta(\omega)\, d\omega \wedge \int_{-\infty}^{\infty} \hat{K}(\omega)\, G_{0.5}(\omega)\, d\omega\right] \tag{B.5}$$

for some sufficiently small $r > 0$. Observe that, $G_\beta(0) = (1-2\beta)^2 > 0$ for $\beta \neq 1/2$. The differentiability of G_β and $\hat{K}(\omega)$ implies the existence of a non-degenerate open interval $\mathcal{I}_\beta$ centered at 0 such that,

$$\inf_{\omega\in\mathcal{I}_\beta} \hat{K}(\omega)\, G_\beta(\omega) \geq \frac{(1-2\beta)^2\hat{K}(0)}{2} \quad\Rightarrow\quad \int_{-\infty}^{\infty} \frac{\hat{K}(\omega)\, G_\beta(\omega)}{2\pi} d\omega \geq \frac{(1-2\beta)^2\hat{K}(0)}{4\pi}\left|\mathcal{I}_\beta\right|.$$

Notice that $\inf_{|\beta-1/2|\geq r}(1-2\beta)^2\left|\mathcal{I}_\beta\right| > 0$. So, we just need to show that the corresponding term to $\beta = 1/2$ on the right hand side of (B.5) is strictly positive. For $\beta = 1/2, G_\beta(\omega) = [\mathrm{sinc}(\omega/4)\sin(\omega/2)]^2$ and so

$$\int_{-\infty}^{\infty} \frac{\hat{K}(\omega)\, G_\beta(\omega)}{2\pi} d\omega \geq \int_{-2\pi}^{2\pi} \frac{\hat{K}(\omega)\, [\mathrm{sinc}(\omega/4)\sin(\omega/2)]^2}{2\pi} d\omega \overset{(b)}{\geq} \frac{2}{\pi^3}\int_{-2\pi}^{2\pi} \hat{K}(\omega)\sin^2(\omega/2)\, d\omega \overset{(c)}{>} 0.$$

Note that (b) is a consequence of monotonicity of $\mathrm{sinc}(\cdot)$ in the interval $(0, \pi/2)$ and inequality (c) follows from the combination of $\left|\hat{K}'(0)\right| < \infty$ and $\hat{K}(0) > 0$. □

Appendix C. Change-point detection in the increasing domain regime

In this section we briefly investigate the detection rate of both CUSUM and GLRT algorithms in the increasing domain setting. Due to the space constraint, the proofs of all the results appearing in this section are omitted. We refer the reader to Keshavarz et al. (2017) for the detailed proofs. Before proceeding further, we present the covariance structure of our Gaussian process model in the increasing domain framework.

In the increasing domain setting, $G - \mathbb{E}G$ is a mean-zero Gaussian process in $\mathcal{D} = [0, \infty)$ and $\mathcal{D}_n = \{1, 2, \ldots\}$. Define $\mathrm{cov}(X_1, X_k) = f_k$ for any k, in which $\{f_m\}_{m=0}^\infty$ is an absolutely summable sequence with $f_0 = 1$. Due to the stationarity assumption, $\Sigma_\mathbb{N} := \mathrm{cov}\left(\{X_k\}_{k=1}^\infty\right)$ is an infinite symmetric Toeplitz matrix. We view $\{X_k\}_{k=1}^n$ as the observed part of an infinite stationary time series, $\{X_k\}_{k=1}^\infty$. Accordingly, the covariance matrix of $\{X_k\}_{k=1}^n$, denoted by Σ_n, is a symmetric (truncated) Toeplitz matrix.

It is a known fact (Chapter 4, Gray, 2006) that there is a symmetric and almost surely (with respect to *Lebesgue* measure) positive function, $f : [-\pi, \pi] \mapsto \mathbb{R}$ such that $\Sigma_\mathbb{N} = \mathcal{T}_\mathbb{N}(f)$. Thus $\Sigma_n = \mathcal{T}_n(f)$. For studying the asymptotic properties of the change detection algorithm, certain regularity conditions are required on f.

Assumption C.1. $f : [-\pi, \pi] \mapsto \mathbb{R}$ is a real symmetric function such that

(a) There are two positive universal scalars, $0 < m_f \leq M_f < \infty$ such that

$$m_f := \inf_{\omega\in[-\pi,\pi]} f(\omega) \leq M_f := \sup_{\omega\in[-\pi,\pi]} f(\omega).$$

(b) There exist positive constants c and λ such that

$$|f_k| \leq c(1+k)^{-(1+\lambda)}. \tag{C.1}$$

The first condition regarding the infimum of f is necessary to have a positive definite infinite covariance matrix, i.e., $\nu^\top \Sigma_\mathbb{N}\nu > 0$ for any non-zero $\nu \in \mathbb{R}^\mathbb{N}$. Moreover, the polynomial decay of f_k's as stated in (C.1) is a sufficient condition to ensure that f can be equivalently expressed by its Fourier series. Such condition is common in the non-asymptotic analysis of Toeplitz matrices (see, e.g., Gray, 2006). We now present a result describing the detection rate of the GLRT algorithm with known covariance matrix (see Eq. (2.4)).

Theorem C.1. *Let $\delta \in (0, 1)$ and suppose that $\Sigma_n = \mathcal{T}_n(f)$ in which f admits Assumption C.1 for some positive scalars c and λ. There exist $n_0 \in \mathbb{N}$, $C > 0$ (depending only on c and λ) and $R_{n,\delta} > 0$ such that for any $n \geq n_0$, if*

$$|b| \geq C\sqrt{f(0)\, n^{-1} \log\left(\frac{n(1-2\alpha)}{\delta}\right)}, \tag{C.2}$$

then

$$\varphi_n(T_{GLRT}) \leq \delta.$$

Some comments are in order. First, the threshold $R_{n,\delta}$ in Theorem C.1 is chosen in exactly the same way as in the fixed domain setting, as given by Eq. (3.6). Second, in contrast to the fixed domain setting, the dependence structure for G no longer plays the central role in the characterization of detection performance. In particular, $f(0)$ is the only factor in (C.2) that captures the correlation in the samples, but this scalar quantity evidently has an insignificant effect: the asymptotic behavior of GLRT remains the same (up to some constant factor) for different Gaussian processes satisfying Assumption C.1. A related observation that arises by comparing between (3.5) and (C.2) is that the correlation structure of observations, which is encapsulated into ν or $f(0)$, and the quantities encoding the marginal density information such as n have been completely decoupled in the rate of the GLRT in the increasing domain. An examination of the proof reveals that the decoupling effect in the increasing domain setting arises due to the short-range correlation assumption ($\mathrm{cov}(X_r, X_s) \to 0$ polynomially in $|r-s|$). It follows that as n increases the correlation for most pairs of observed sample become negligible.

Now we aim to study the CUSUM test whose formulation is given in Eq. (1.2).

Theorem C.2. *Let $\delta \in (0, 1)$, and $\mathcal{C}_{n,\alpha} = [\alpha n, (1-\alpha)n] \cap \mathbb{N}$. Assume that f satisfies Assumption C.1 for some c and λ. There are $n_0 = n_0(f)$ and scalar $C(\lambda, c) > 0$, such that if $n \geq n_0$ and*

$$|b| \geq C\sqrt{\frac{f(0)}{n\alpha(1-\alpha)} \log\left(\frac{n(1-2\alpha)}{\delta}\right)}, \tag{C.3}$$

then

$$\varphi_n(T_{\mathrm{CUSUM}}) \leq \delta.$$

The presented detection rates in Theorems C.1 and C.2 reveal that the both CUSUM and GLRT exhibit similar detection performance in the increasing domain setting. However, according to the numerical studies in Section 7, the GLRT slightly outperforms the CUSUM test, especially in the presence of strong long range dependence.

In the sequel we give a condition on jump size $|b|$ according to which no algorithm in the increasing domain can properly detect the existence of a shift in the mean.

Theorem C.3. *Let $\delta \in (0, 2)$, and $\mathcal{C}_{n,\alpha} = [\alpha n, (1-\alpha)n]$. Suppose that $\Sigma_n = \mathcal{T}_n(f)$ in which f satisfies Assumption C.1. There exist $n_0 := n_0(f)$ and $C > 0$ such that if $n \geq n_0$ and*

$$|b| \leq C\sqrt{\frac{(1+\vartheta)\, f(0) \log\left(\frac{1}{\delta(2-\delta)}\right)}{\alpha n}},$$

then for any test T,

$$\varphi_n(T) \geq \delta.$$

The direct comparison between the detection rate of both CUSUM (in Theorem C.2) and GLRT (see Theorem C.1) test with the above result indicates the minimax optimality (up to some order $\log n$ term) of both of these procedures in the increasing domain setting.

References

Adak, S., 1998. Time-dependent spectral analysis of non-stationary time series. J. Amer. Statist. Assoc. 93 (444), 1488–1501.

Addario-Berry, L., Broutin, N., Devroye, L., Lugosi, G., 2010. On combinatorial testing problems. Ann. Statist. 38 (5), 3063–3092.

Anderes, E., 2010. On the consistent separation of scale and variance for Gaussian random fields. Ann. Statist. 870–893.

Antoch, J., Hušková, M., Prášková, Z., 1997. Effect of dependence on statistics for determination of change. J. Statist. Plann. Inference 60 (2), 291–310.

Arias-Castro, E., Candès, E.J., Durand, A., 2011. Detection of an anomalous cluster in a network. Ann. Statist. (Feb), 278–304.

Aue, A., Horváth, L., 2013. Structural breaks in time series. J. Time Series Anal. 34 (1), 1–6.

Bertrand, P., Raphael, P., Fhima, M., Guillin, A., 2011. Off-line detection of multiple change points by the filtered derivative with p-value method. Sequential Anal. 30 (2), 172–207.

Birgé, L., 2001. An Alternative Point of View on Lepski's Method. In: Lecture Notes-Monograph Series, pp. 113–133.

Bolin, D., Lindgren, F., 2011. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. Ann. Appl. Stat. 5 (1), 523–550.

Butucea, C., Ingster, Y.I., 2013. Detection of a sparse submatrix of a high-dimensional noisy matrix. Bernoulli 19 (5B), 2652–2688.
Chandola, V., Vatsavai, R.R., (2011) A gaussian process based online change detection algorithm for monitoring periodic time series, in: SDM, pp. 95–106.
Chernoff, H., Zacks, S., 1964. Estimating the current mean of a normal distribution which is subjected to changes in time. Ann. Math. Statist. 35 (3), 999–1018.
Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. Ann. Statist. 25 (1), 1–37.
Deshayes, J., Picard, D., 1985. Off-line statistical analysis of change-point models using non parametric and likelihood methods. In: Detection of Abrupt Changes in Signals and Dynamical Systems. Springer, Berlin Heidelberg, pp. 103–168.
Gabriel, E., Allard, D., Bacro, J.N., 2011. Estimating and testing zones of abrupt change for spatial data. Stat. Comput. 21 (1), 107–120.
Gillet, O., Essid, S., Richard, G., 2007. On the correlation of automatic audio and visual segmentations of music videos. IEEE Trans. Circuits Syst. Video Technol. 17 (3), 347–355.
Gray, R.M., 2006. Toeplitz and Circulant Matrices: A Review, now publishers inc.
Horn, R.A., Johnson, C.R., 2012. Matrix Analysis. Cambridge University Press.
Horváth, L., 1997. Detection of changes in linear sequences. Ann. Inst. Statist. Math. 49 (2), 271–283.
Horváth, L., Hušková, M., 2012. Change point detection in panel data. J. Time Series Anal. 33 (4), 631–648.
Horváth, L., Rice, G., 2014. Extensions of some classical methods in change point analysis. TEST 23 (2), 219–255.
Hu, X., Qiu, H., Iyer, N., 2007. Multivariate change detection for time series data in aircraft engine fault diagnostics. In: ISIC. IEEE International Conference on Systems, Man and Cybernetics. IEEE, pp. 2484–2489.
Ibragimov, I.A., Rozanov, Y.A., 1978. Gaussian Random Processes. Springer.
Ivanoff, B.G., Merzbach, E., 2010. Optimal detection of a change-set in a spatial Poisson process. Ann. Appl. Probab. 640–659.
Kaufman, C.G., Shaby, B.A., 2013. The role of the range parameter for estimation and prediction in geostatistics. Biometrika 100 (2), 473–484.
Kawahara, Y., Sugiyama, M., Change-point detection in time-series data by direct density-ratio estimation, in: Proceedings of SIAM International Conference on Data Mining, SDM, 2009, pp. 389–400.
Keshavarz, H., Scott, C., Nguyen, X.L., 2017. Optimal change point detection in Gaussian processes. arXiv preprint arXiv:1506.01338v2.
Kokoszka, P., Leipus, R., 1998. Change-point in the mean of dependent observations. Statist. Probab. Lett. 40 (4), 385–393.
Lai, T.L., 1998. Information bounds and quick detection of parameter changes in stochastic systems. IEEE Trans. Inform. Theory 44 (7), 2917–2929.
Last, M., Shumway, R., 2008. Detecting abrupt changes in a piecewise locally stationary time series. J. Multivariate Anal. 99 (2), 191–214.
Lavielle, M., 2005. Using penalized contrasts for the change-point problem. Signal Process. 85 (8), 1501–1510.
Levy-leduc, C., 2007. Catching change-points with lasso. In: Advances in Neural Information Processing Systems, NIPS, pp. 617–624.
Liu, J., Yuan, L., Ye, J., 2010. An efficient algorithm for a class of fused lasso problems. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 323–332.
Loh, W.L., Lam, T.K., 2000. Estimating structured correlation matrices in smooth Gaussian random field models. Ann. Statist. 28 (3), 880–904.
MacNeill, I.B., 1974. Test for change of parameter at unknown time and distributions of some related func- tionals on Brownian motion. Ann. Statist. 2, 950-962.
Noh, H.Y., Rajagopal, R., Kiremidjian, A.S., 2012. Damage diagnosis algorithm using a sequential change point detection method with an unknown distribution for damage. In: SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring. International Society for Optics and Photonics, vol. 8345.
Perron, P., Vogelsang, T.J., 1992. Testing for a unit root in a time series with a changing mean: corrections and extensions. J. Bus. Econom. Statist. 10 (4), 467–470.
Rao, T.S., Rao, S.S., Rao, C.R., 2012. HandBook of Statistics: Time Series Analysis, Methods and Applications. Vol. 30. Elsevier.
Rencova, M., 2009. Change Point Detection in Temperature Series (Doctoral dissertation), Czech Technical University.
Rigaill, G., 2010. Pruned dynamic programming for optimal multiple change-point detection, arXiv preprint arXiv:1004.0887.
Shao, X., Zhang, X., 2010. Testing for change points in time series. J. Amer. Statist. Assoc. 105 (491).
Shen, X., Huang, H.C., Cressie, N., 2002. Nonparametric hypothesis testing for a spatial signal. J. Amer. Statist. Assoc. 97 (460), 1122–1140.
Skorokhod, A.V., Yadrenko, M.I., 1973. On absolute continuity of measures corresponding to homogeneous Gaussian fields. Theory Probab. Appl. 18 (1), 27–40.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer.
Tartakovsky, A.G., Rozovskii, B.L., Blazek, R.B., Kim, H., 2006. A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. IEEE Trans. Signal Process. 54 (9), 3372–3382.
Tsybakov, A.B., 2009. Introduction To Non-Parametric Estimation. Vol. 11. Springer, New York.
Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114 (1), 106–115.
Wu, W.Y., Lim, C.Y., Xiao, Y., 2013. Tail estimation of the spectral density for a stationary Gaussian random field. J. Multivariate Anal. 116, 74–91.
Yao, Y.C., Davis, R.A., 1986. The asymptotic behaviour of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. Indian J. Statist. Ser. A 339–353.
Ying, Z., 1991. Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. J. Multivariate Anal. 36 (2), 280–296.
Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Amer. Statist. Assoc. 99 (465), 250–261.