Scan statistics for data segmentation of stochastic processes and anomaly detection in large image data

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Philipp Klein, M. Sc.
geb. am 08.12.1988 in Lahnstein

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Claudia Kirch
Dr. Haeran Cho

Eingereicht am: 21.02.2022
Verteidigung am: 16.06.2022

Abstract

Pre-processing is an important step in the analysis of data. For example, it can used to segment time series into stretches with approximately constant means but also for the analysis of image data by e. g. identifying regions containing anomalies. Scan statistics provide a powerful and oftentimes computationally effective tool for both the estimation of structural breaks, so called change points, in time series and the estimation of locations that contain anomalies in image data.

In the first part of this thesis, we present statistical methodology based on scan statistics with the intent of estimating mean changes on a general class of multivariate stochastic processes. More precisely, these processes fulfill strong invariance principles and include e. g. partial sum, renewal and diffusion processes. We introduce a scan statistic based on moving sum (MOSUM) statistics in order to estimate the locations of abrupt mean changes of the processes introduced above. We analyze the behavior of this statistic both in the existence and the absence of change points and provide limit distributions in the case of no change points. We introduce estimators for these change points and show consistency for both the case of linear and sublinear bandwidths. Furthermore, under mild assumptions we are able to show convergence rates and provide the asymptotic distribution for the distance between the change points and their estimators.

In the second part of this thesis, we develop a procedure that aims to identify areas containing potentially dangerous anomalies such as fissures in large scans of concrete blocks as a starting point for further methods, e. g. machine learning algorithms while trying to discard areas with natural anomalies (gravel, air etc.). One can observe that locally, fissures are rectangle-shaped objects with small widths while the natural anomalies resemble bubbles. Therefore, inspired by the use of windows in MOSUM statistics and based on the above described geometric properties of fissures and natural anomalies, we present a MOSUM-type scan statistic using rectangle- and circle-shaped windows in order to detect areas with potentially dangerous anomalies (fissures) and discard areas with natural anomalies. We analyze the performance of our procedure by means of a simulation study and show convergence of our scan statistic to a Gaussian process in the absence of anomalies which is important in order to calculate thresholds for our procedure. Furthermore, we provide limit theorems for scan statistics using shapes for windows that include, but are not not limited to, convex sets.

Zusammenfassung

Die Vorverarbeitung der Daten ist ein wichtiger Schritt im Bereich der Datenanalyse. Sie kann sowohl dazu genutzt werden, Zeitreihen in Abschnitte mit annähernd konstanten Erwartungswerten zu unterteilen, als auch für die Analyse von Bilddaten, um dort z. B. Regionen zu identifizieren, die Anomalien aufweisen. Scan-Statistiken sind wirkungsvolle und oftmals recheneffiziente Methoden sowohl für die Schätzung von Strukturbrüchen, so genannter Change-Points in Zeitreihen, als auch für die Schätzung von Anomalien enthaltenden Regionen in Bilddaten.

Im ersten Teil dieser Arbeit stellen wir statistische Methoden basierend auf Scan-Statistiken vor, mit dem Ziel, Erwartungswertänderungen in einer allgemeinen Klasse multivariater stochastischer Prozesse zu schätzen. Für diese Prozesse fordern wir lediglich, dass sie starke Invarianzprinzipien erfüllen. Dies beinhaltet beispielsweise Partialsummen-, Erneuerungs- und Diffusionsprozesse. Wir stellen eine Scan-Statistik auf Basis so genannter 'moving sum' (MOSUM)-Statistiken vor mit dem Ziel, die genaue Lage von abrupten Mittelwertänderungen in den oben vorgestellten Prozessen zu schätzen. Wir analysieren das Verhalten dieser Statistik sowohl im Falle der Existenz als auch der Abwesenheit von Strukturbrüchen und leiten Grenzverteilungen der Statistik im Falle der Abwesenheit von Strukturbrüchen her. Ferner stellen wir auf der Scan-Statistik basierende Schätzer für die Strukturbrüche vor und zeigen Konsistenzresultate sowohl für lineare als auch für sublineare Bandbreiten. Darüber hinaus zeigen wir unter schwachen Voraussetzungen Konvergenzraten und geben die Grenzverteilung für den Abstand zwischen den Strukturbrüchen und ihren Schätzern an.

Im zweiten Teil dieser Arbeit stellen wir ein Verfahren zur Detektion von Rissen in großen Bildern von Betonblöcken vor, welches als Vorverarbeitung für weitere Methoden wie z. B. solche des Maschinellen Lernens dienen kann. Ein weiteres Ziel unseres Verfahren ist, Regionen mit natürlichen Anomalien im Beton wie Lufteinschlüssen, Schotter usw. zu eliminieren. Basierend auf lokalen geometrischen Eigenschaften von Rissen und natürlichen Anomalien, die optisch Blasen ähneln, stellen wir eine MOSUM- Scan-Statistik vor, die kreisförmige und rechteckige Fenster verwendet um Regionen mit potenziell gefährlichen Anomalien wie Rissen zu identifizieren und Regionen mit natürlichen Anomalien wie Lufteinschlüssen, Schotter usw. zu eliminieren. Wir analysieren die Performance unseres Verfahrens in einer Simulationsstudie und zeigen die Konvergenz unserer Statistik gegen das Funktional eines Gauß-Prozesses im Falle der Abwesenheit von Anomalien. Des weiteren zeigen wir Grenzwertsätze für Scan-Statistiken, die als Scan-Fenster eine geometrische Klasse von Mengen verwendet, die konvexe Mengen beinhaltet, aber nicht auf diese beschränkt ist.

Contents

List of Figures

List of Tables

Notation

Vectors, Vector norms

Let $T \in \mathbb{R}$, $q \geq 1$, $p \in \mathbb{N}$, $\boldsymbol{s} = (s_1, \ldots, s_p)' \in \mathbb{R}^p$.

$\boldsymbol{s}'$	Transpose of vector $\boldsymbol{s}$
$\|\boldsymbol{s}\|_q$	q-norm of $\boldsymbol{s}$
$\|\boldsymbol{s}\|_\infty$	Maximum norm of $\boldsymbol{s}$
$\|\boldsymbol{s}\|$	$= \|\boldsymbol{s}\|_2$ (Euclidean norm of $\boldsymbol{s}$)
$\frac{\boldsymbol{s}}{T}$	$\left(\frac{s_1}{T}, \ldots, \frac{s_p}{T}\right)'$
$\lfloor \boldsymbol{s} \rfloor_T$	$\left(\frac{\lfloor s_1 T \rfloor}{T}, \ldots, \frac{\lfloor s_p T \rfloor}{T}\right)'$

Matrices, Matrix norms

Let $m, p \in \mathbb{N}$, $q \in [1, \infty) \cup \{\infty\}$, $\boldsymbol{A} \in \mathbb{R}^{m \times p}$.

$\boldsymbol{A}'$	Transpose of matrix $\boldsymbol{A}$
$\boldsymbol{A}^{-1}$	Inverse matrix of $\boldsymbol{A}$ for invertible $\boldsymbol{A} \in \mathbb{R}^{p \times p}$
$\boldsymbol{A}^{1/2}$	Hermitian positive definite square root for positive definite $\boldsymbol{A}$
$\boldsymbol{A}^{-1/2}$	Hermitian positive definite square root for positive definite $\boldsymbol{A}^{-1}$
$\mathrm{diag}(a_1, \ldots, a_p)$	$\mathbb{R}^{p \times p}$-valued diagonal matrix with $a_1, \ldots, a_p$ on diagonal
I_m	$m \times m$ identity matrix
$\|\boldsymbol{A}\|_q$	Operator norm of $\boldsymbol{A}$ induced by $q-$vector norm
$\|\boldsymbol{A}\|$	$= \|\boldsymbol{A}\|_2$

Sets

Let $p \in \mathbb{N}$, $A, B \subset \mathbb{R}^p$, $\boldsymbol{s} \in \mathbb{R}^p$.

A', A^c	Complement of A
$\lambda(A)$	Lebesgue measure of A
$\ell(\partial A)$	Perimeter of $A \subset \mathbb{R}^2$
$A \Delta B$	Symmetric difference $(A \setminus B) \cup (B \setminus A)$
$\sum_{i=1}^{m} A_i$	$\bigcup_{i=1}^{m} A_i$ for disjoint $A_1, \ldots, A_m$
$A_{\boldsymbol{s}}$, $A(\boldsymbol{s})$, $A + \boldsymbol{s}$	$\{\boldsymbol{x} \in \mathbb{R}^p : \boldsymbol{x} - \boldsymbol{s} \in A\}$
A^δ	$\left\{\boldsymbol{x} \in \mathbb{R}^p : \inf_{\boldsymbol{s} \in A} \|\boldsymbol{x} - \boldsymbol{s}\| \leq \delta\right\}$
$\mathrm{conv}(A)$	Convex hull of A

Arithmetic operations

Let $p \in \mathbb{N}$, $c \in \mathbb{R}$, $\boldsymbol{s} = (s_1, \ldots, s_p)'$, $\boldsymbol{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$.

$c\boldsymbol{s}$, $\boldsymbol{s}c$	$(cs_1, \ldots, cs_p)'$
$c\,\mathbb{N}^n$	$\{(ci_1, \ldots, ci_p)' \vert (i_1, \ldots, i_p) \in \mathbb{N}^p\}$
$\boldsymbol{s} \odot \boldsymbol{t}$	$(s_1 t_1, \ldots, s_p t_p)'$
$c^{\boldsymbol{s}}$	$(c^{s_1}, \ldots, c^{s_p})'$

Probability theory

Let $p \in \mathbb{N}$. Let X, Y, $X_1, X_2, \ldots$ be random variables. Let $\mathbf{X}$ be a random vector in $\mathbb{R}^p$. Let $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$.

$\mathbb{E}[\mathbf{X}]$	Expected value of random variable/vector $\mathbf{X}$
$\mathrm{Var}[X]$	Variance of X
$\mathrm{Cov}[\mathbf{X}]$	Covariance matrix X
$\mathrm{Cov}[X, Y]$	Covariance of random variables X and Y
$\mathcal{L}(X)$	Distribution of X
F_X, $F_{\mathcal{L}(X)}$	Cumulative distribution function (CDF) of distribution $\mathcal{L}(X)$
F_X^{-1}, $F_{\mathcal{L}(X)}^{-1}$	Quantile function of distribution $\mathcal{L}(X)$
$\sim$	'distributed as'
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Normal distribution with mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$
Φ	CDF of $\mathcal{N}(0, 1)$
$\Gamma(s, \lambda)$	Gamma distribution with shape $s > 0$, rate $\lambda > 0$
$(X_{\boldsymbol{t}})_{\boldsymbol{t} \in I}$	Stochastic process over an index set I
$X \overset{\mathcal{D}}{=} Y$	$\mathcal{L}(X) = \mathcal{L}(Y)$
$X_n \to X$ *a.s.*	Almost sure convergence of X_n to X as $n \to \infty$
$X_n \xrightarrow{P} X$	Stochastic convergence of X_n to X as $n \to \infty$
$X_n \xrightarrow{\mathcal{D}} X$	Convergence in distribution of random variables
$(X_{\boldsymbol{t},n})_{\boldsymbol{t} \in I} \xrightarrow{w} (X_{\boldsymbol{t}})_{\boldsymbol{t} \in I}$	Weak convergence of stochastic processes in some functional space with a specified norm
$\mathcal{C}([0,1]^p)$	Space of bounded continuous functions on $[0,1]^p$
$\mathcal{D}([0,1]^p)$	p-dimensional 'equivalent' to the space of cádlág (right-continuous with left limits) functions $\mathcal{D}([0,1])$. For further details, see Section A.2.

(Stochastic) Landau Notation

Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ be deterministic sequences, $b_n > 0$. Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ be (stochastic) sequences with $Y_n > 0$ $a.s.$

$a_n = O(b_n)$	$\exists C > 0$,: $\lvert a_n\rvert \leq C b_n \ \forall\ n \in \mathbb{N}$
$a_n = o(b_n)$	$a_n / b_n \to 0$ as $n \to \infty$
$X_n = O_P(Y_n)$	$\lim_{C\to\infty} \limsup_{n\to\infty} \mathrm{P}\left(\lvert X_n\rvert / Y_n \geq C\right) = 0$
$X_n = o_P(Y_n)$	$X_n / Y_n \xrightarrow{P} 0$ as $n \to \infty$

Miscellaneous

Let $p \in \mathbb{N}$, $c \geq 0$, $\boldsymbol{s} = (s_1, \ldots, s_p)'$, $\boldsymbol{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$, $f : \mathbb{R} \to \mathbb{R}$, $B \subset \mathbb{R}$.

$\mathbb{1}_A(\cdot)$	Indicator function of set A
$[\boldsymbol{s}, \boldsymbol{t}]$	$[s_1, t_1] \times \ldots \times [s_p, t_p]$ with $[s_i, t_i] := [t_i, s_i]$ for $t_i < s_i$
$[\boldsymbol{s}, \boldsymbol{s} + \gamma]$	$[s_1, s_1 + \gamma] \times \ldots \times [s_p, s_p + \gamma]$
$\underset{x\in B}{\operatorname{argmax}}\ f(x)$	$\min\left\{ y \in B \lvert f(y) = \max_{x\in B} f(x) \right\}$

Preface

The detection and the estimation of the location of anomalies is an important problem in many practical applications. It is of special interest but not limited to medical application when trying to find anomalies in DNA sequences in order to make assertions about cancer progression (Olshen et al. (2004), Niu and Zhang (2012)) or the detection and treatment of cancer (McInerney and Terzopoulos (1996)). Another field of interest is in neurophysiology where it is important to divide neuronal firing patterns, so called spike trains, into stretches with approximately constant firing rates in order to make assertions about the firing patterns on a local level (Grün, Diesmann, and Aertsen (2002), Schneider (2008), Messer et al. (2014)). Other fields of interest are for example in astrophysics when trying to find anomalies in light curves of stars in order to detect exoplanets (see Fisch, Eckley, and Fearnhead (2018)) and more recently the detection of potentially dangerous anomalies such as fissures in building material (see Weise et al. (2015) Baranowski et al. (2019)).
Anomaly detection is of great importance both in time series and image data. Despite being different on first sight, we can model image data as multi-parameter time series and adjust methods intended for time series analysis for the analysis of image data.

The segmentation of time series is an important pre-processing tool: Many stationary models for the analysis assume stationarity of the underlying data. However, in practice it is natural that the model parameters change throughout a time series. To account for this problem, there exist basically two approaches: On the one hand, one can expand the model accounting for different 'states' such as Hidden Markov models do, but estimating the parameters in these types of models can be rather expensive. Therefore, having a computationally efficient pre-processing procedure that splits the time series into parts with approximately constant model parameters allows for a meaningful analysis and interpretation of the time series.
Pre-processing can also be important in the analysis of image data, which can be seen by the example given in this thesis when trying to identify sparse objects in large data: Our goal is to identify potentially dangerous anomalies such as fissures in large 3D CT scans of concrete blocks. Since fissures are 2D objects, they are sparse in the data. Oftentimes, algorithms such as machine learning methods are applied on the whole data in order to trace the fissures. As these algorithms are computationally expensive and the objects they search for are sparse, this can be highly ineffective. Therefore, discarding areas without fissures with a computationally effective procedure allows for a much more efficient search for fissures by computationally more expensive procedures in smaller areas.

Scan statistics have proven to be powerful tools in change point analysis for which they were first introduced by Page (1954) and are frequently used both for testing and estimation of change points. Recently, they have also been used for testing of the

existence of anomalies in image data (see e. g. Haiman and Preda (2006), Kabluchko (2011), Sharpnack and Arias-Castro (2016)). Moving sum (MOSUM) statistics were first introduced by Bauer and Hackl (1980) and have proven to be computationally efficient tools for the estimation of change points and anomalies both in time series and image data.

Contributions

In this thesis, based on the goals to estimate change points in renewal processes and to estimate the locations of potentially dangerous anomalies like fissures in concrete, we present MOSUM-type scan statistics for both the time series and image data in order to estimate and detect anomalies in these data. We analyze and show theoretic results for more general classes and demonstrate the performance of our procedures in small simulation studies for three-dimensional renewal processes and image data with a fissure, respectively.

Part I

To this date, many procedures aiming at detecting or estimating finitely many change points in univariate stochastic processes exist, but there is very little literature on the localization of change points in multivariate processes such as renewal processes. In the first part, we present a procedure based on MOSUM statistics in order to estimate change points in a general class of multivariate stochastic processes fulfilling strong invariance principles – a class that includes, but is not limited to multivariate renewal processes. In particular, we study a procedure that compares the increments of processes in symmetric intervals around time points, allowing the interval length to grow linearly or sublinearly in the length of the process. We present estimators for the change points introduced by Meier, Cho, and Kirch (2021); Cho and Kirch (2021+) based on sufficiently large and isolated maxima that allow for the detection of changes even under mild violations of model assumptions. We then study the theoretic properties of the statistic and estimator and derive localization rates for the distance between change points and the respective estimators.
With the exception of Section 5, Part I has been published as a joint paper (Kirch and Klein (2021)).

Part II

There exist a variety of methods for the detection of anomalies in image data, such as EM algorithms, deformable models and Machine Learning methods like neural networks. However, these methods are computationally expensive and some anomalies like fissures in construction material are sparse and pre-processing steps may be needed in order to apply the above mentioned methods.
In the second part, we present such a pre-processing step for the detection of fissures

in concrete data using MOSUM statistics. In particular, one can observe that on a local scale, a part of a fissure resembles a rectangle with a small width. Therefore, we use circle-shaped windows with a small rectangle inscribed and compare the average gray value in the inner rectangle with the average value in the remaining part of a circle. Furthermore, in our procedure, we account for the fact that concrete is a heterogeneous material and therefore contains natural, 'bubble'-shaped anomalies such as air, gravel etc. We extend our procedure to MOSUM statistics that use windows possessing shapes from a class of sets that includes, but is not limited to convex sets. We show limit results for functionals of statistics that use different windows, where we assume that those functionals are Lipschitz-continuous. This allows for the application of the limit theorem in a general setting that includes statistics based on general local differences. Thus, it can be applied to detect general anomalies.

Part I
Moving sum data segmentation for stochastic processes based on invariance

1 Introduction

Change point analysis aims at detecting and localizing structural breaks in time series data and has a long tradition in statistics, dating back to Page (1954). It has broad applications in a variety of fields such as neurophysiology (see Messer et al. (2014)), genomics (compare Olshen et al. (2004), Niu and Zhang (2012), Li, Munk, and Sieling (2016), Chan and Chen (2017)), finance (see Aggarwal, Inclan, and Leal (1999), Cho and Fryzlewicz (2012)), astrophysics (see Fisch, Eckley, and Fearnhead (2018)) or oceanographics (see Killick et al. (2010)).

Early literature focuses on the detection of a single change point in a univariate time series, primarily in the mean (see Csörgö and Horvàth (1997) for an overview). Recently, a main interest in research is the detection and estimation of multiple change points in high-dimensional data (see Horváth and Rice (2014), Cho and Kirch (2020) for overviews).

Generally, data segmentation methods can roughly be split up in two approaches: The first approach was introduced by Yao (1988) in the context of i.i.d. normally distributed data. It uses the Schwarz' criterion and aims at optimizing suitable objective functions. Kühn (2001) extended this approach to processes in a setting closely related to the one in this thesis while focusing on univariate processes with a finite number of change points. Further approaches include e. g. least-squares (Yao and Au (1989)) or the quasi-likelihood-function (Braun, Braun, and Müller (2000)). Generally, such approaches are computationally expensive, such that there is another body of work proposing fast algorithms e. g. using dynamic programming (Killick, Fearnhead, and Eckley (2012), Maidstone et al. (2017)).

A second approach is based on hypothesis testing, where e. g. binary segmentation introduced by Vostrikova (1981) recursively uses tests constructed for the at-most-one-change situation. Several problems arise including the fact that detection power can be poor if the set of change points is unfavorable, such that several extensions have been proposed in the literature such as circular binary segmentation (Olshen et al. (2004)) or wild binary segmentation (Fryzlewicz (2014)).

Connection to existing work

Another class of methods use moving sum (MOSUM) statistics which were first introduced by Bauer and Hackl (1980). MOSUM-based methods were initially used for testing of multiple change points, see e. g. Hušková and Slabỳ (2001) for an approach using permutation tests. During the last two decades, research interest has shifted to the multiple change problem aiming at segmenting the data into stationary stretches often focusing on changes in the mean e. g. of i.i.d. Gaussian data (Cho and Kirch (2020)), in a Hidden Markov framework (Eichinger and Kirch (2018)) or for changes in autoregressive time series (Yau and Zhao (2016)). Recently, Messer et al. (2014) with a bottom-up-approach, and Meier, Cho, and Kirch (2021) and Cho and Kirch (2021+) with a localized pruning approach have proposed two-step data segmentation procedures based on multiscale MOSUM statistics.

We adopt a MOSUM approach to localize multiple changes in multivariate renewal processes where the analysis of neuronal firing patterns, so called spike trains, is a very prominent example where data segmentation methods for renewal processes are useful. Indeed, many methods, e. g. Grün, Diesmann, and Aertsen (2002) or Schneider (2008) use local approaches applied on segments with approximately constant intensity to model the data. Furthermore, it is of great interest to study the joint behavior of spike trains, compare e. g. Perkel, Gernstein, and Moore (1967), Brown and Mitra (2004) and Grün and Rotter (2010). Chen, Chen, and Ding (2019) use non-parametric methods to detect change points in neuropixel data, which consists of a large amount of neuronal firing patterns, in order to make meaningful assertions about the whole or parts of the data. In particular, they study firing patterns in several different brain areas and make assertions on possible coordination between regions based on their change point patterns. Messer et al. (2014) propose a MOSUM multiscale procedure to detect changes in the firing intensity assuming that the firing patterns follow renewal processes with piecewise constant intensity. Our work extends their results in several ways: We show consistency of the change point estimators and derive the corresponding localization rates both in the case of linear and sublinear bandwidths. Furthermore, although our main focus lies in the estimation of multiple changes in multivariate renewal processes, we extend the results to a general class of multivariate processes fulfilling strong invariance principles. This class includes multivariate partial sum, renewal and diffusion processes. A univariate version of that model with at-most-one change point has been considered by Horváth and Steinebach (2000) and Kühn and Steinebach (2002). A univariate version for finitely many change points has been considered by Kühn (2001) where consistency for the number of change points has been shown. Those results are now extended to include MOSUM methodology for the estimation of a possibly unbounded number of change points in a multivariate setting, where we achieve a minimax optimal separation rate in addition to a minimax optimal localization rate (for the change point estimators) in case of a bounded number of change points as well as for Wiener processes with drift (see Remark 4.2 below). Our results also lay the foundations for the analysis of a two-step procedure as in Cho and Kirch (2021+). With the exception of Section 5, this part has been published as a joint paper Kirch and Klein (2021).

1.1 Outline

In Subsection 2.1, we introduce the multiple change point model we consider followed by some examples of processes fulfilling the model in Subsection 2.2. In Section 3, we describe how to estimate change points based on MOSUM statistics: First, we introduce the MOSUM statistics in Subsection 3.1, before presenting the estimators for the structural breaks in 3.2. In Subsection 3.3 we derive some asymptotic results for the MOSUM statistics that are required for threshold selection and can also be used in a testing context. In Section 4 we show that the corresponding data segmentation procedure is consistent. Finally, we derive the localization rates in

addition to the corresponding asymptotic distribution of the change point estimators for local changes. In Section 5, we extend our MOSUM statistics to the setting of asymmetric bandwidths and show consistency of the corresponding estimators. In Section 6, we present some results from a small simulation study.

2 Multiple change point problem

While our initial motivation for this work was the estimation of changes in renewal processes, we introduce a more general model that also includes partial sum and certain diffusion processes. We derive the theoretic results for this general model.

2.1 Model

Consider $P < \infty$ stochastic processes $\left(\mathbf{R}_{t,T}^{(j)}\right)_{0\leq t\leq T}$ of dimension p with (unknown) drift $(\boldsymbol{\mu}_T^{(j)} \cdot t)$ and (unknown) covariance $(\boldsymbol{\Sigma}_{j,T} \cdot t)$ fulfilling regularity assumptions specified in Assumption 2.1 below. These P processes can be thought of as background processes with only one of them being active at each time in the sense of driving the increments of our observation process. Consequently, at each time point we only observe the active process and do not know the exact structure of any of these processes. To elaborate, for $c_\ell < t \leq c_{\ell+1}$ we observe

$$\begin{aligned}\mathbf{Z}_{t,T} &= \left(\mathbf{R}_{t,T}^{(c_{\ell+1})} - \mathbf{R}_{c_\ell,T}^{(c_{\ell+1})}\right) + \sum_{j=1}^{\ell}\left(\mathbf{R}_{c_j,T}^{(c_j)} - \mathbf{R}_{c_{j-1},T}^{(c_j)}\right), \qquad (2.1)\\ \mathbf{Z}_{0,T} &= 0,\end{aligned}$$

where $0 = c_0 < c_1 < \ldots < c_{q_T} < c_{q_T+1} = T$ are the unknown change points and the number of change points q_T can be bounded or unbounded. The upper index $^{(c_j)}$ at the process $\mathbf{R}_{\cdot,T}$ indicates (with a slight abuse of notation) the active process between the $(j-1)$-th and the j-th change point. We define the change in drift between two neighboring regimes by

$$\mathbf{d}_{i,T} := \boldsymbol{\mu}_T^{(c_{i+1})} - \boldsymbol{\mu}_T^{(c_i)} \neq 0 \quad \text{for all } i = 1, \ldots, q_T, \qquad (2.2)$$

where $\mathbf{d}_{i,T}$ is bounded but we allow for $\mathbf{d}_{i,T} \to 0$ as long as the convergence is slow enough (see Assumption 3.1). For ease of notation we frequently drop the dependency on T for the above quantities in the following. The aim of data segmentation involves the consistent estimation of the number and location of the change points as well as the derivation of the corresponding localization rates.

We assume that the underlying processes $\left(\mathbf{R}_{t,T}^{(j)}\right)_{0\leq t\leq T}$, $j = 1, \ldots, P$, fulfill the following joint invariance principle towards Wiener processes. If the underlying processes are independent, then this simplifies to the validity of an invariance principle for each of these P processes.

Assumption 2.1.
Denote the joint process by $\mathbf{R}_{t,T} = \left(\mathbf{R}_{t,T}^{(1)\prime}, \ldots, \mathbf{R}_{t,T}^{(P)\prime}\right)'$ *as well the joint drift by* $\boldsymbol{\mu}_T = \left(\boldsymbol{\mu}_T^{(1)\prime}, \ldots, \boldsymbol{\mu}_T^{(P)\prime}\right)'$, *where* $'$ *indicates the matrix transpose. For every* $T > 0$ *there exist* $(p \cdot P)$*-dimensional Wiener processes* $\mathbf{W}_{t,T}$ *with covariance matrix* $\boldsymbol{\Sigma}_T$ *and*

$$\boldsymbol{\Sigma}_T^{(i)} = (\boldsymbol{\Sigma}_T(l,k))_{l,k=p\,(i-1)+1,\ldots,p\,i}$$

with

$$\left\|\boldsymbol{\Sigma}_T^{(i)}\right\| = O(1), \qquad \left\|\boldsymbol{\Sigma}_T^{(i)^{-1}}\right\| = O(1),$$

such that, possibly after a change of probability space, it holds that for some sequence $\nu_T \to 0$

$$\sup_{0\le t\le T} \|\widetilde{\mathbf{R}}_{t,T} - \mathbf{W}_{t,T}\| = \sup_{0\le t\le T} \|\left(\mathbf{R}_{t,T} - \boldsymbol{\mu}_T\, t\right) - \mathbf{W}_{t,T}\| = O_P\left(T^{\frac{1}{2}}\,\nu_T\right),$$

where $\widetilde{\mathbf{R}}_{t,T} = \mathbf{R}_{t,T} - \boldsymbol{\mu}_T\, t$ *denotes the centered process.*

The covariance matrix $\boldsymbol{\Sigma}_T^{(i)}$ relates to the i-th underlying process $\{\mathbf{R}_{t,T}^{(i)}\}$ and plays an important role in the below limit results. On the other hand, the cross-dependence between different driving processes does not influence these limit results because at each time only one process actively influences the observed process and the increments of the joint process are asymptotically independent due to the joint invariance principle.

The assumption on the norm of the covariance matrices is equivalent to the smallest eigenvalue of $\boldsymbol{\Sigma}_T^{(i)}$ being bounded in addition to being bounded away from zero – both uniformly in T, compare also Corollary B.3. In many situations, the covariance matrices will not depend on T, in which case this assumption is automatically fulfilled under positive definiteness. The convergence rate ν_T in the invariance principle typically depends on the number of moments that exist. Roughly speaking, the more moments the original process has, the faster ν_T converges (see Section 2.2 for some examples).

The corresponding univariate model with at most one change was first considered by Horváth and Steinebach (2000) and further used in a single-change setting by Steinebach (2000), Kirch and Steinebach (2006) and Gut and Steinebach (2002; 2009). Kühn and Steinebach (2002) make use of the Schwarz information criterion for the estimation of the number of change points in a related univariate framework with a bounded number of change points. Using information criteria is computationally much more expensive with quadratic computational complexity if compared to MOSUM procedures with linear computational complexity as proposed in this thesis.

2.2 Examples

In this section, we give three important examples fulfilling the above model assumptions, namely partial sum-processes, renewal processes and integrals of diffusion processes including Ornstein-Uhlenbeck and Wiener processes with drift. A detailed analysis of the stochastic behavior of estimators obtained by MOSUM procedures for detecting mean changes in (univariate) renewal processes extending the work by Messer et al. (2014) was the original motivation for this work and is covered by this much broader framework.

2.2.1 Partial-Sum-Processes

This first example extends the classical multiple changes in the mean model:

Let $\left(\mathbf{X}_k^{(i)}\right)_{k\in\mathbb{N}}$ be a time series with $\mathbb{E}\left[\mathbf{X}_k^{(i)}\right] = 0$ and $\mathrm{Cov}\left[\mathbf{X}_k^{(i)}\right] = I_p$ for all $i = 1, \ldots, P$, $k \in \mathbb{N}$. Let

$$\mathbf{R}_t^{(i)} = \sum_{j=1}^{\lfloor t \rfloor} \left(\boldsymbol{\mu}^{(i)} + {\boldsymbol{\Sigma}_T^{(i)}}^{1/2} \mathbf{X}_j^{(i)}\right).$$

The corresponding process fulfills Assumption 2.1 in a wide range of situations. For example, Einmahl (1987) shows the validity in the case that $\mathbf{X}_1, \mathbf{X}_2, \ldots$ with $\mathbf{X}_j = \left(\mathbf{X}_j^{(1)}, \ldots, \mathbf{X}_j^{(P)}\right)'$ are i.i.d. with $\mathbb{E}\left[\|\mathbf{X}_1\|^{2+\delta}\right] < \infty$ for some $\delta > 0$ resulting in a rate of $O_P(T^{1/(2+\delta)})$ in Assumption 2.1 (and thus, $\nu_T = T^{-\delta/(4+2\delta)}$). Additionally, Kuelbs and Philipp (1980) state an invariance principle for mixing random vectors in Theorem 4, and there are many corresponding univariate results under many different weak-dependency formulations.

For $\mathbf{X}^{(i)} = \mathbf{X}^{(1)}$ (and $\boldsymbol{\Sigma}^{(i)} = \boldsymbol{\Sigma}^{(1)}$) for all i, then we are back to the classical multiple mean change problem that has been considered in many papers in particular for the univariate situation, see e. g. the recent survey papers by Fearnhead and Rigaill (2020) or Cho and Kirch (2020).

2.2.2 Renewal and some related point processes

The second example aims at finding structural breaks in the rates of renewal and some related point processes:

We consider P independent sequences of p-dimensional point processes that are related to renewal processes in the following way: For each $i = 1, \ldots, P$ we start with $\tilde{p} \geq p$ independent renewal processes $\widetilde{R}_{t,j}^{(i)}$, $j = 1, \ldots, \tilde{p}$, from which we derive a p-dimensional point process $\mathbf{R}_t^{(i)} = \mathbf{B}^{(i)} \left(\widetilde{R}_{t,1}^{(i)}, \ldots, \widetilde{R}_{t,\tilde{p}}^{(i)}\right)'$, where $\mathbf{B}^{(i)}$ is a $(p \times \tilde{p})$ -matrix with non-negative integer-valued entries. By Lemma 4.2 in Steinebach and Eastwood (1996) Assumption 2.1 is fulfilled for a block-diagonal $\boldsymbol{\Sigma}_T$ with

$$\boldsymbol{\Sigma}_T^{(i)} = \mathbf{B}^{(i)}\mathbf{D}\left(\frac{\boldsymbol{\sigma}^2(i)}{\boldsymbol{\mu}^3(i)}\right)\mathbf{B}^{(i)\prime},$$
$$\text{with } \mathbf{D}\left(\frac{\boldsymbol{\sigma}^2(i)}{\boldsymbol{\mu}^3(i)}\right) = \text{diag}\left(\frac{\sigma_1^2(i)}{\mu_1^3(i)}, \ldots, \frac{\sigma_{\tilde{p}}^2(i)}{\mu_{\tilde{p}}^3(i)}\right),$$

where $\mu_j(i)$ and $\sigma_j^2(i)$ are the mean and variance of the corresponding inter-event times. If for the corresponding inter-event times possess $2+\delta$-th moments, then the rate in Assumption 2.1 is given by $O_P(T^{1/(2+\delta)})$ (and thus $\nu_T = T^{-\delta/(4+2\delta)}$). Steinebach and Eastwood (1996) and Csenki (1979) consider $\tilde{p} = p$ but use inter-event times that are dependent for $j = 1, \ldots, p$. In such a situation, the invariance principle in Assumption 2.1 still holds if the intensities are the same across components with $\Sigma_T^{(i)} = \Sigma_{\text{IET}}^{(i)}/\mu_1^3(i)$, where $\Sigma_{\text{IET}}^{(i)}$ is the covariance of the vector of inter-event times – a setting that we adopt in the simulation study. If the intensities differ, then by Steinebach and Eastwood (1996) an invariance principle towards a Gaussian process can still be obtained, but this is no longer a multivariate Wiener process. While each component is a Wiener process, the increments from one component may depend on the past of another. Many of the below results can still be derived in such a situation, however, such a model does not seem to be very realistic for most applications as the stochastic behavior of the increments of one component depends on the lagged behavior of the other components, where the lag increases with time. While a lagged dependence is realistic in many situations, in most situations one would expect this lagged-dependence to be constant across time.

Messer et al. (2014) consider this model for univariate renewal processes with varying variance. They propose a multiscale procedure based on MOSUM statistics related to those we will discuss in the next section using linear bandwidths. In Messer et al. (2017), they extend the procedure to processes with weak dependencies. They show convergence in distribution of the MOSUM statistics to functionals of Wiener processes similar to the results that we obtain and analyze the behavior of the signal term in Messer and Schneider (2017). However, they have not derived any consistency results for the change point estimators. In this work, we extend their results to sublinear bandwidths and prove the consistency of the corresponding estimators as well as their localization rates.

2.2.3 Diffusion processes

Clearly, switching between independent (or components of a multivariate) Brownian motion with drift is included in this framework. Additionally, Heunis (2003) and Mihalache (2011) derive invariance principles in the context of diffusion processes including Ornstein-Uhlenbeck processes among others. Let $(\mathbf{X}_t)_{t\geq 0}$ be a stochastic process in $\mathbb{R}^N$ satisfying a stochastic differential equation (SDE)

$$d\mathbf{X}_t = \boldsymbol{\mu}\left(\mathbf{X}_t\right)dt + \boldsymbol{\Sigma}\left(\mathbf{X}_t\right)d\mathbf{B}_t$$

with respect to an n-dimensional standard Wiener process $(\mathbf{B}_t)_{t\geq 0}$ and let $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ be globally Lipschitz-continuous. Under some conditions on $f : \mathbb{R}^N \to \mathbb{R}^p$, as given by Heunis (2003), relating to $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, which in particular guarantee that the function f applied to the (invariant) diffusion results in a centered process, there exists a p-dimensional Wiener process $(\mathbf{W}_t)_{t\geq 0}$ and some $\eta > 0$ such that

$$\left\| \int_0^T f(\mathbf{X}_s)\ ds - \mathbf{W}_T \right\| = O\left(T^{1/2-\eta}\right),$$

where $(\mathbf{X}_t)_{t\geq 0}$ either is a solution to the SDE with fixed starting value $\mathbf{X}_0 = y_0$ or a strictly stationary solution with respect to an invariant distribution.
Furthermore, in the case of a one-dimensional stochastic diffusion process, Mihalache (2011) showed for some L^2-functions fulfilling constraints depending on $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ that there exists a strong invariance principle for the integrals of diffusion processes with a rate of $O((T \log_2 T)^{1/4} \sqrt{\log T})$ a. s. (and thus $\nu_T = T^{-1/4}(\log_2 T)^{1/4}\sqrt{\log T}$).

3 Data segmentation procedure

Now, we are ready to introduce a MOSUM-based data segmentation procedure for stochastic processes following the above model:

3.1 Moving sum statistics

By assumption the drifts of the two active processes to the left and right of a change point differ, see (2.2); on the other hand, in a stationary stretch away from any change point the drift is the same. Because the difference in drift can be estimated by a difference of increments, we propose the following moving sum (MOSUM) statistic that is based on the moving difference of increments with bandwidth $h = h_T$

$$\begin{aligned} \mathbf{M}_t = \mathbf{M}_{t,T,h_T}(\mathbf{Z}) &= \frac{1}{\sqrt{2h}} \left[(\mathbf{Z}_{t+h} - \mathbf{Z}_t) - (\mathbf{Z}_t - \mathbf{Z}_{t-h})\right] \\ &= \frac{1}{\sqrt{2h}} \left(\mathbf{Z}_{t+h} - 2\mathbf{Z}_t + \mathbf{Z}_{t-h}\right). \end{aligned} \tag{3.1}$$

If there is no change, then this difference will fluctuate around zero. On the other hand close to a change point, this difference will be different from zero. Ideally, the bandwidth should be chosen to be as large as possible (to get a better estimate obtained from a larger 'effective sample size' of the order h). On the other hand, the increments should not be contaminated by a second change as this can lead to situations where the change point can no longer be reliably localized by the signal. This observation is reflected in the following assumptions on the bandwidths:

Assumption 3.1. *For ν_T as in Assumption* (2.1) *the bandwidth $h < T/2$ fulfills*

$$\frac{\nu_T^2\, T\, \log T}{h} \to 0.$$

Furthermore, the first and last change point are detectable and the i-th change point is isolated in the sense that

$$\begin{aligned} &h \leq \min\left\{c_1, T - c_{q_T}\right\}, \\ &h \leq \frac{1}{2}\,\Delta_i, \qquad \textit{where } \Delta_i = \min(c_{i+1} - c_i, c_i - c_{i-1}) \end{aligned} \tag{3.2}$$

for $i = 2, \ldots, q_T - 1$.
Additionally, the signal needs to be large enough to be detectable by this bandwidth, i. e.

$$\frac{\|\mathbf{d}_i\|^2\, h}{\log\left(\frac{T}{h}\right)} \to \infty. \tag{3.3}$$

for all $i = 1, \ldots, q_T$. Combining (3.2) and (3.3) shows that – with an appropriate bandwidth h – changes are detectable as soon as

$$\frac{\|\mathbf{d}_i\|^2\, \Delta_i}{\log\left(\frac{T}{\Delta_i}\right)} \to \infty. \tag{3.4}$$

In case of the classical mean change model as in Subsection 2.2.1 this is known to be the minimax-optimal separation rate that cannot be improved (see Proposition 1 of Arias-Castro, Candes, and Durand (2011)).

3.2 Change point estimators

Lemma 3.1. *The MOSUM statistic $\mathbf{M}_t = \mathbf{m}_t + \mathbf{\Lambda}_t$ as in (3.1) decomposes into a piecewise linear signal term $\mathbf{m}_t = \mathbf{m}_{t,h,T}$ and a centered noise term $\mathbf{\Lambda}_t = \mathbf{\Lambda}_{t,h,T}$ with*

$$\sqrt{2h}\,\mathbf{m}_t = \begin{cases} (h - t + c_i)\,\mathbf{d}_i, & \textit{for } c_i < t \leq c_i + h, \\ 0, & \textit{for } c_i + h < t \leq c_{i+1} - h, \\ (h + t - c_{i+1})\,\mathbf{d}_{i+1}, & \textit{for } c_{i+1} - h < t \leq c_{i+1}, \end{cases} \tag{3.5}$$

$$\begin{aligned} \sqrt{2h}\,\mathbf{\Lambda}_t &= \sqrt{2h}\,\mathbf{\Lambda}_t(\widetilde{\mathbf{R}}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.6) \\ &= \begin{cases} \widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_{t}^{(c_{i+1})} + \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_i)} + \widetilde{\mathbf{R}}_{t-h}^{(c_i)}, & \textit{for } c_i < t \leq c_i + h, \\ \widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_{t}^{(c_{i+1})} + \widetilde{\mathbf{R}}_{t-h}^{(c_{i+1})}, & \textit{for } c_i + h < t \leq c_{i+1} - h, \\ \widetilde{\mathbf{R}}_{t+h}^{(c_{i+2})} - \widetilde{\mathbf{R}}_{c_{i+1}}^{(c_{i+2})} + \widetilde{\mathbf{R}}_{c_{i+1}}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_{t}^{(c_{i+1})} + \widetilde{\mathbf{R}}_{t-h}^{(c_{i+1})}, & \textit{for } c_{i+1} - h < t \leq c_{i+1}, \end{cases} \end{aligned}$$

where $\widetilde{\mathbf{R}}_t := \mathbf{R}_t - t\boldsymbol{\mu}$ for $i = 0, \ldots, q_T$ and the upper index c_j denotes the active regime between the $(j-1)$-th and j-th change point (with a slight abuse of notation).

Proof. For $c_i < t \leq c_i + h$ it holds that

$$\begin{aligned}
\sqrt{2h}\mathbf{M}_t &= \mathbf{Z}_{t+h} - 2\mathbf{Z}_t + \mathbf{Z}_{t-h} \\
&= \mathbf{R}_{t+h}^{(c_{i+1})} - 2\mathbf{R}_t^{(c_{i+1})} + \mathbf{R}_{c_i}^{(c_{i+1})} - \mathbf{R}_{c_i}^{(c_i)} + \mathbf{R}_{t-h}^{(c_i)} \\
&= \left(\widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})} + \boldsymbol{\mu}^{(c_{i+1})}(t+h)\right) - 2\left(\widetilde{\mathbf{R}}_t^{(c_{i+1})} + \boldsymbol{\mu}^{(c_{i+1})}t\right) + \left(\widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} + \boldsymbol{\mu}^{(c_{i+1})}c_i\right) \\
&\quad - \left(\widetilde{\mathbf{R}}_{c_i}^{(c_i)} + \boldsymbol{\mu}^{(c_i)}c_i\right) + \left(\widetilde{\mathbf{R}}_{t-h}^{(c_i)} + \boldsymbol{\mu}^{(c_i)}(t-h)\right) \\
&= \sqrt{2h}\left(\mathbf{m}_t + \boldsymbol{\Lambda}_t\right).
\end{aligned}$$

The decompositions for $c_i + h < t \leq c_{i+1} - h$ and $c_{i+1} - h < t \leq c_{i+1}$ follow analogously. □

The signal term is a piecewise linear function that takes its extrema at the change points and is 0 outside h-intervals around the change points. Additionally, the noise term is asymptotically negligible compared to the signal term (see Theorem 3.2 for the corresponding theoretic statement and Figure 3.1 for an illustrative example).

This motivates the following data segmentation procedure that considers local extrema which are big enough (in absolute value) as change point estimators:

For a suitable threshold $\beta = \beta_{h,T}$ (see Section 3.3 for a detailed discussion) we define *significant* time points, where a point t^* is *significant* if

$$\mathbf{M}_{t^*}' \, \widehat{\mathbf{A}}_{t^*}^{-1} \, \mathbf{M}_{t^*} \geq \beta. \tag{3.7}$$

$\widehat{\mathbf{A}}_{t^*}$ is a symmetric positive definite matrix that may depend on the data and fulfills

Assumption 3.2.

$$\sup_{h \leq t \leq T-h} \left\| \widehat{\mathbf{A}}_{t,T}^{-1} \right\| = O_P(1), \qquad \sup_{i=1,\ldots,q_T} \sup_{|t-c_i| \leq h} \left\| \widehat{\mathbf{A}}_{t,T} \right\| = O_P(1).$$

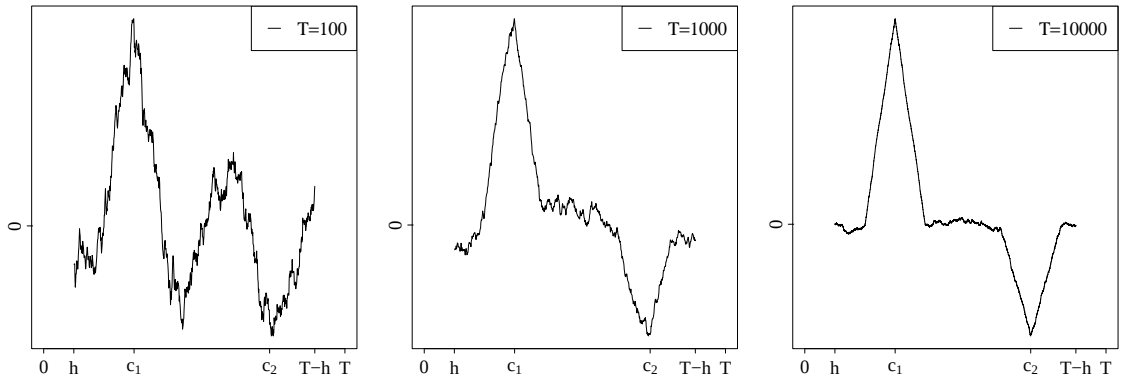

Figure 3.1: Univariate MOSUM statistic with $T = 100,\, 1\,000,\, 10\,000$ (from left to right), where the noise term (fluctuating around the signal) becomes smaller and smaller relative to the signal term.

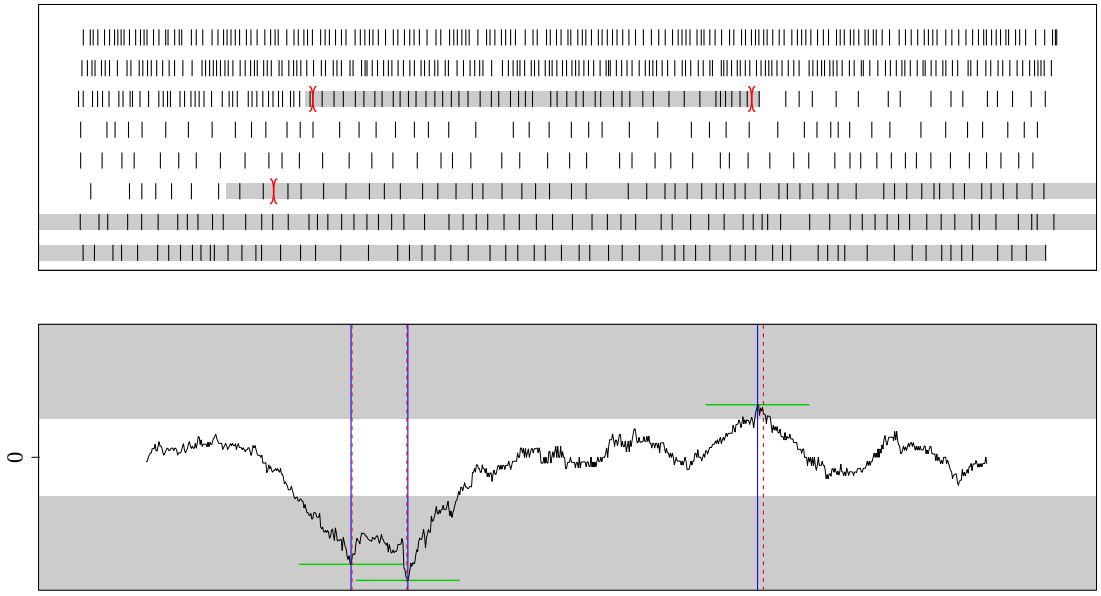

Figure 3.2: In the upper panel, the observed event times of a univariate renewal process with three change points (i. e. four stationary segments) are displayed (where the plot needs to be read like a text: It starts in the upper row on the left, then continues in the first row and jumps to the second row and so on). The gray and white regions mark the estimated segmentation of the data while the red intervals mark the true segmentation.
In the lower panel, the corresponding MOSUM statistic with (relative) bandwidth $h/T = 0.07$ is displayed. The gray areas are the regions where the threshold ($\alpha = 0.05$ as in Remark 3.1) is exceeded (in absolute value). The blue solid lines indicate the change point estimates obtained as local extrema that fall within the gray area (making them *significant*). The true change points are indicated by the red dashed lines. The green horizontal lines denote ηh-environments around the estimators.

A good (non data-driven) choice fulfilling this assumption is given by

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t,T} = \boldsymbol{\Sigma}_T^{(c_i)} \tag{3.8}$$

for $c_{i-1} < t \leq c_i$, which guarantees scale-invariance of the procedure and allows for nicely interpretable thresholds (see Section 3.3). The latter remains true for estimators as long as they fulfill

$$\sup_{i=1,\ldots,q_T} \sup_{|t-c_i|>h} \left\| \widehat{\boldsymbol{\Sigma}}_{t,T}^{-1/2} - \boldsymbol{\Sigma}_t^{-1/2} \right\| = o_P\left(\left(\log \frac{T}{h} \right)^{-1} \right) \tag{3.9}$$

in addition to the above boundedness assumptions. In particular, this permits local estimators that are consistent only away from change points but contaminated by the change in a local environment thereof. The latter is typically the case for covariance

estimators, think e. g. of the sample variance contaminated by a change point. In order to not reduce detection power in small samples, it is beneficial if the estimator is additionally consistent directly at the change point, which is also achievable (see e. g. Eichinger and Kirch (2018)).

Typically, there are intervals of significant points (due to the continuity of the signal) such that only local extrema of such intervals actually indicate a change point. To define what a local extremum is, we require a tuning parameter $0 < \eta < 1$. This parameter defines the locality requirement on the extremum, where a point t^* is a local extremum if it is the leftmost point (for the sake of tie-breaking) to maximize the absolute MOSUM statistic within its ηh-environment, i. e. if

$$t^* = \underset{t^*-\eta h \leq t \leq t^*+\eta h}{\operatorname{argmax}} \|\mathbf{M}_t\|. \tag{3.10}$$

The threshold β distinguishes between *significant* and *spurious* local extrema that are purely associated with the noise term. The set of all significant local extrema is the set of change point estimators with its cardinality an estimator for the number of the change points.

Figure 3.2 shows an example illustrating these ideas: Away from the change points the MOSUM statistic fluctuates around zero (within the white area that is beneath the threshold in absolute value) while it falls within the gray area close to the change points – making corresponding local extrema significant. Furthermore, the statistic does not need to return to the white area in order to have all changes estimated, as can be seen between the first and second change point. This is one of the major advantages of the η-criterion based on *significant* local maxima as described here (in comparison to the ϵ-criterion originally investigated by Eichinger and Kirch (2018) in the context of mean changes, see also the discussion in Meier, Cho, and Kirch (2021)). Nevertheless, results for the ϵ-criterion can be obtained along the lines of our proofs below.

3.3 Threshold selection

As pointed out above we need to choose a threshold $\beta = \beta_{h,T}$ that can distinguish between significant and spurious local extrema. The following theorem gives the magnitudes of signal as well as noise terms:

Theorem 3.2. *Let Assumptions 2.1, 3.1 and 3.2 hold.*

(a) For the signal $\mathbf{m}_t$ *with* $c_i - h < t < c_i + h$*, it holds that*

$$\mathbf{m}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{m}_t \geq \frac{1}{2\|\widehat{\mathbf{A}}_t\|} \frac{(h - |t - c_i|)^2}{h} \|\mathbf{d}_i\|^2.$$

At other time points the signal term is equal to zero.

(b) For the noise term it holds for $q_T = 0$*, i. e. in the no-change situation*

(i) for a linear bandwidth $h = \gamma T$ with $0 < \gamma < 1/2$

$$\sup_{\gamma T \le t \le T - \gamma T} \mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t \xrightarrow{\mathcal{D}} \sup_{\gamma \le s \le 1-\gamma} \frac{1}{2\gamma} \left(\mathbf{B}_{s+\gamma} - 2\mathbf{B}_s + \mathbf{B}_{s-\gamma}\right)' \left(\mathbf{B}_{s+\gamma} - 2\mathbf{B}_s + \mathbf{B}_{s-\gamma}\right),$$

where $\mathbf{B}$ denotes a multivariate standard Wiener process.
In particular, the squared noise term is of order $O_P(1)$ in this case.

(ii) for a sublinear bandwidth $h/T \to 0$ but Assumption 3.1 fulfilled, it holds that

$$a\left(\frac{T}{h}\right) \sup_{h \le t \le T-h} \sqrt{\mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t} - b\left(\frac{T}{h}\right) \xrightarrow{\mathcal{D}} E,$$

where E follows a Gumbel distribution with $\mathrm{P}\left(E \le x\right) = e^{-2e^{-x}}$ and

$$a(x) = \sqrt{2 \log x}$$
$$b(x) = 2 \log x + \frac{p}{2} \log \log x + \log \frac{3}{2} - \log \Gamma\left(\frac{p}{2}\right).$$

In particular, the above squared noise term is of order $O_P\left(\log(T/h)\right)$ in this case.

The assertions remain true if an estimator for the covariance is used fulfilling (3.9) *uniformly over all $h \le t \le T - h$.*

(c) In the situation of multiple change points, it holds that

$$\sup_{h \le t \le T-h} \|\mathbf{\Lambda}_t\| = O_P(\sqrt{\log(T/h)}).$$

Proof. (a) Because $\widehat{\mathbf{A}}_t$ is symmetric and positive definite, the maximum eigenvalue is given by $\|\widehat{\mathbf{A}}_t\|$ by Corollary B.3. By Lemma B.1, it follows that the minimal eigenvalue of $\widehat{\mathbf{A}}_t^{-1}$ is given by $1/\|\widehat{\mathbf{A}}_t\|$. By Lemma B.4 and (3.5) it follows that

$$\mathbf{m}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{m}_t \ge \frac{1}{\|\widehat{\mathbf{A}}_t\|} \|\mathbf{m}_t\|^2 = \frac{1}{\|\widehat{\mathbf{A}}_t\|} \frac{(h - |c_i - t|)^2}{2h} \|\mathbf{d}_i\|^2.$$

(b) Denote by $\mathbf{\Lambda}_t(\mathbf{W}_t)$ the MOSUM statistics defined in (3.1) with $\{\mathbf{Z}_t\}$ there replaced by $\{\mathbf{W}_t\}$. Since $q_T = 0$, it holds that $\mathbf{Z}_t = \mathbf{R}_t = \mathbf{R}_t^{(1)} = \widetilde{\mathbf{R}}_t + t\boldsymbol{\mu}$ and $\boldsymbol{\mu} = \boldsymbol{\mu}^{(1)}$. By the invariance principle from Assumption 2.1 it holds by Assumption 3.1 and the triangle inequality that

$$
\begin{aligned}
&\sup_{h\le t\le T-h} \|\mathbf{\Lambda}_t - \mathbf{\Lambda}_t(\mathbf{W}_t)\| \\
=& \frac{1}{\sqrt{2h}} \sup_{h\le t\le T-h} \|(\mathbf{Z}_{t+h} - 2\mathbf{Z}_t + \mathbf{Z}_{t-h}) - (\mathbf{W}_{t+h} - 2\mathbf{W}_t + \mathbf{W}_{t-h})\| \\
=& \frac{1}{\sqrt{2h}} \sup_{h\le t\le T-h} \big\|(\widetilde{\mathbf{R}}_{t+h} + (t+h)\boldsymbol{\mu} - 2\widetilde{\mathbf{R}}_t - 2t\boldsymbol{\mu} + \widetilde{\mathbf{R}}_{t-h} + (t-h)\boldsymbol{\mu}) \\
&-(\mathbf{W}_{t+h} - 2\mathbf{W}_t + \mathbf{W}_{t-h})\| \\
=& \frac{1}{\sqrt{2h}} \sup_{h\le t\le T-h} \big\|(\widetilde{\mathbf{R}}_{t+h} - 2\widetilde{\mathbf{R}}_t + \widetilde{\mathbf{R}}_{t-h}) - (\mathbf{W}_{t+h} - 2\mathbf{W}_t + \mathbf{W}_{t-h})\big\| \\
\le& \frac{4}{\sqrt{2h}} \sup_{0\le t\le T} \big\|\widetilde{\mathbf{R}}_t - \mathbf{W}_t\big\| \\
=& O_P\left(\frac{T^{1/2}\nu_T}{\sqrt{h}}\right) = o_P\left(\sqrt{\log(T/h)}^{-1}\right). \qquad (3.11)
\end{aligned}
$$

(i) Let $(\mathbf{B}_t)_{t\ge 0} = \left(\mathbf{\Sigma}_T^{-1/2}\mathbf{W}_t\right)_{t\ge 0}$ be a multivariate standard Wiener process. By the self-similarity that $(\sqrt{c}\mathbf{B}_t)_{t\ge 0} \overset{\mathcal{D}}{=} (\mathbf{B}_{ct})_{t\ge 0}$ for $c > 0$ it follows with the transformation $s = t/T$ that

$$
\begin{aligned}
&\sup_{\gamma T\le t\le T-\gamma T} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right) \\
=& \sup_{\gamma T\le t\le T-\gamma T} \frac{1}{2\gamma T} \left(\mathbf{B}_{t+\gamma T} - 2\mathbf{B}_t + \mathbf{B}_{t-\gamma T}\right)' \left(\mathbf{B}_{t+\gamma T} - 2\mathbf{B}_t + \mathbf{B}_{t-\gamma T}\right) \\
=& \sup_{\gamma\le s\le 1-\gamma} \frac{1}{2\gamma T} \left(\mathbf{B}_{(s+\gamma)T} - 2\mathbf{B}_{sT} + \mathbf{B}_{(s-\gamma)T}\right)' \left(\mathbf{B}_{(s+\gamma)T} - 2\mathbf{B}_{sT} + \mathbf{B}_{(s-\gamma)T}\right) \\
\overset{\mathcal{D}}{=}& \sup_{\gamma\le s\le 1-\gamma} \frac{1}{2\gamma} \left(\mathbf{B}_{s+\gamma} - 2\mathbf{B}_s + \mathbf{B}_{s-\gamma}\right)' \left(\mathbf{B}_{s+\gamma} - 2\mathbf{B}_s + \mathbf{B}_{s-\gamma}\right).
\end{aligned}
$$

The assertion then follows from (3.11) and Slutzky's theorem.

(ii) Let $(\mathbf{B}_t)_{t\ge 0} = \left((B_{t,1}, \ldots, B_{t,p})'\right)_{t\ge 0}$ be as above. Similar to (i) it follows by the self-similarity of the multivariate Wiener process with the transformation $s = t/h - 1$ that

$$
\begin{aligned}
&\sup_{h\le t\le T-h} \sqrt{\mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)} = \sup_{h\le t\le T-h} \left\|\mathbf{\Sigma}_T^{-1/2} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)\right\| \\
=& \sup_{h\le t\le T-h} \left\|\frac{1}{\sqrt{2h}} \left(\mathbf{B}_{t+h} - 2\mathbf{B}_t + \mathbf{B}_{t-h}\right)\right\| \\
=& \sup_{0\le s\le \frac{T}{h}-2} \left\|\frac{1}{\sqrt{2h}} \left(\mathbf{B}_{(s+2)h} - 2\mathbf{B}_{(s+1)h} + \mathbf{B}_{sh}\right)\right\| \\
=& \sup_{0\le s\le \frac{T}{h}-2} \left\|\frac{1}{\sqrt{2}} \left(\mathbf{B}_{s+2} - 2\mathbf{B}_{s+1} + \mathbf{B}_s\right)\right\|.
\end{aligned}
$$

W.l.o.g. let $a \geq 0$. The process

$$\left(\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right)_{s\geq 0} = \left(\frac{1}{\sqrt{2}}\left(\mathbf{B}_{s+2} - 2\mathbf{B}_{s+1} + \mathbf{B}_s\right)\right)_{s\geq 0}$$

is a stationary p-dimensional Gaussian process. The componentwise covariance functions are given by

$$\begin{aligned}
& 2\operatorname{Cov}\left[\mathbf{\Lambda}_{a,i}\left(\mathbf{B}_a\right), \mathbf{\Lambda}_{0,i}\left(\mathbf{B}_0\right)\right] \\
= \; & \operatorname{Cov}\left[B_{a+2,i} - B_{a+1,i}, B_{2,i} - B_{1,i}\right] - \operatorname{Cov}\left[B_{a+2,i} - B_{a+1,i}, B_{1,i} - B_{0,i}\right] \\
& - \operatorname{Cov}\left[B_{a+1,i} - B_{a,i}, B_{2,i} - B_{1,i}\right] + \operatorname{Cov}\left[B_{a+1,i} - B_{a,i}, B_{1,i} - B_{0,i}\right] \\
= \; & \begin{cases} 2 - (a+1) - (a+1-1) + (1-a), & \text{for } 0 \leq a < 1 \\ -2 + a, & \text{for } 1 \leq a < 2 \\ 0, & \text{for } a \geq 2 \end{cases} \\
= \; & \begin{cases} 2 - 3a, & \text{for } 0 \leq a < 1 \\ -2 + a, & \text{for } 1 \leq a < 2 \\ 0, & \text{for } a \geq 2 \end{cases}
\end{aligned}$$

with analogous results for $a < 0$. Therefore,

$$\operatorname{Cov}\left[\mathbf{\Lambda}_{s+a,i}\left(\mathbf{B}_{s+a}\right), \mathbf{\Lambda}_{s,i}\left(\mathbf{B}_s\right)\right] = \begin{cases} 1 - \frac{3}{2}|a|, & \text{for } 0 \leq |a| < 1 \\ -1 + \frac{1}{2}|a|, & \text{for } 1 \leq |a| < 2 \\ 0, & \text{for } |a| \geq 2. \end{cases}$$

Therefore it follows by Lemma 3.1 and Remark 3.1 of Steinebach and Eastwood (1996) (see also Lemma A.1) with $\alpha = 1$, $C = 3/2$ that

$$a\left(\frac{T}{h}\right) \sup_{0\leq s\leq \frac{T}{h}} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| - b\left(\frac{T}{h}\right) \xrightarrow{\mathcal{D}} E.$$

Since by the triangle inequality

$$\sup_{\frac{T}{h}-2\leq s\leq \frac{T}{h}} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| \overset{\mathcal{D}}{=} \sup_{0\leq s\leq 2} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| \leq \frac{4}{\sqrt{2}} \sup_{0\leq s\leq 4} \|\mathbf{B}_s\| = O_P(1) = \; o_P\left(\frac{b\left(T/h\right)}{a\left(T/h\right)}\right),$$

it follows by Lemma A.2 that

$$\begin{aligned}
& a\left(\frac{T}{h}\right) \sup_{h\leq t\leq T-h} \sqrt{\mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)} - b\left(\frac{T}{h}\right) \\
\overset{\mathcal{D}}{=} \; & a\left(\frac{T}{h}\right) \sup_{0\leq s\leq \frac{T}{h}-2} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| - b\left(\frac{T}{h}\right) \xrightarrow{\mathcal{D}} E,
\end{aligned}$$

as well. Since by (3.11),

$$\left| a\left(\frac{T}{h}\right) \sup_{h\leq t\leq T-h} \sqrt{\mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t} - a\left(\frac{T}{h}\right) \sup_{h\leq t\leq T-h} \sqrt{\mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)} \right| = o_P(1),$$

the assertion follows by Slutzky's theorem.

If we replace $\mathbf{\Sigma}_T$ by an estimator $\widehat{\mathbf{\Sigma}}_t$ fulfilling (3.9), the above assertions remain true: It holds by (b), (3.9) and Lemma B.2 that

$$\begin{aligned}
&\sup_{h\le t\le T-h}\left(\left\|\mathbf{\Sigma}_T^{-1/2}\mathbf{\Lambda}_t\right\|-\left\|\widehat{\mathbf{\Sigma}}_t^{-1/2}\mathbf{\Lambda}_t\right\|\right)\\
=&\sup_{h\le t\le T-h}\left(\left\|\widehat{\mathbf{\Sigma}}_t^{-1/2}\mathbf{\Lambda}_t+\left(\mathbf{\Sigma}_T^{-1/2}-\widehat{\mathbf{\Sigma}}_t^{-1/2}\right)\mathbf{\Lambda}_t\right\|-\left\|\widehat{\mathbf{\Sigma}}_t^{-1/2}\mathbf{\Lambda}_t\right\|\right)\\
\le&\sup_{h\le t\le T-h}\left\|\left(\mathbf{\Sigma}_T^{-1/2}-\widehat{\mathbf{\Sigma}}_t^{-1/2}\right)\mathbf{\Lambda}_t\right\|\le\sup_{h\le t\le T-h}\left\|\mathbf{\Sigma}_T^{-1/2}-\widehat{\mathbf{\Sigma}}_t^{-1/2}\right\|\sup_{h\le t\le T-h}\|\mathbf{\Lambda}_t\|\\
=&\,o_P\left(\left(\log\frac{T}{h}\right)^{-1}\right)O_P\left(\left(\log\frac{T}{h}\right)^{1/2}\right)=o_P\left(\left(\log\frac{T}{h}\right)^{-1/2}\right).
\end{aligned}$$

Analogously we obtain that

$$\sup_{h\le t\le T-h}\left(\left\|\widehat{\mathbf{\Sigma}}_t^{-1/2}\mathbf{\Lambda}_t\right\|-\left\|\mathbf{\Sigma}_T^{-1/2}\mathbf{\Lambda}_t\right\|\right)=o_P\left(\left(\log\frac{T}{h}\right)^{-1/2}\right)$$

and therefore

$$\begin{aligned}
&\left|\sup_{h\le t\le T-h}\sqrt{\mathbf{\Lambda}_t'\mathbf{\Sigma}_T^{-1}\mathbf{\Lambda}_t}-\sup_{h\le t\le T-h}\sqrt{\mathbf{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t^{-1}\mathbf{\Lambda}_t}\right|\\
=&\left|\sup_{h\le t\le T-h}\left(\left\|\widehat{\mathbf{\Sigma}}_t^{-1/2}\mathbf{\Lambda}_t\right\|-\sup_{h\le t\le T-h}\left\|\mathbf{\Sigma}_T^{-1/2}\mathbf{\Lambda}_t\right\|\right)\right|\\
\le&\sup_{h\le t\le T-h}\left|\left\|\widehat{\mathbf{\Sigma}}_t^{-1/2}\mathbf{\Lambda}_t\right\|-\left\|\mathbf{\Sigma}_T^{-1/2}\mathbf{\Lambda}_t\right\|\right|=o_P\left(\left(\log\frac{T}{h}\right)^{-1/2}\right),
\end{aligned}$$

which shows the assertion.

(c) As the behavior of $\mathbf{\Lambda}_t$ in intervals of length h around change points differs from the behavior of $\mathbf{\Lambda}_t$ away from change points as can be seen in (3.6), we need to analyze $\mathbf{\Lambda}_t$ separately for those two cases. It holds that

$$\sup_{h\le t\le T-h}\|\mathbf{\Lambda}_t\|=\max\left\{\max_{i=1,\dots,q_T}\sup_{c_i-h\le t\le c_i+h}\|\mathbf{\Lambda}_t\|,\max_{i=0,\dots,q_T}\sup_{c_i+h\le t\le c_{i+1}-h}\|\mathbf{\Lambda}_t\|\right\}.$$

By (b) it holds that

$$\begin{aligned}
&\max_{i=0,\dots,q_T}\sup_{c_i+h<t<c_{i+1}-h}\|\mathbf{\Lambda}_t\|=\max_{i=0,\dots,q_T}\sup_{c_i+h<t<c_{i+1}-h}\left\|\widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})}-2\widetilde{\mathbf{R}}_t^{(c_{i+1})}+\widetilde{\mathbf{R}}_{t-h}^{(c_{i+1})}\right\|\\
\le&\max_{i=1,\dots,P}\sup_{h\le t\le T-h}\left\|\widetilde{\mathbf{R}}_{t+h}^{(i)}-2\widetilde{\mathbf{R}}_t^{(i)}+\widetilde{\mathbf{R}}_{t-h}^{(i)}\right\|=O_P\left(\sqrt{\log(T/h)}\right).
\end{aligned}\tag{3.12}$$

Furthermore, we obtain by Proposition A.4 (b) and the triangle inequality that

$$
\begin{aligned}
&\max_{i=1,\dots,q_T} \sup_{c_i \le t \le c_i+h} \|\mathbf{\Lambda}_t\| \\
&= \max_{i=1,\dots,q_T} \frac{1}{\sqrt{2h}} \sup_{c_i \le t \le c_i+h} \left\| \widetilde{\mathbf{R}}_{t+h}^{(c_{i+1})} - 2\widetilde{\mathbf{R}}_{t}^{(c_{i+1})} + \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_i)} + \widetilde{\mathbf{R}}_{t-h}^{(c_i)} \right\| \\
&\le 3 \max_{i=1,\dots,q_T} \sup_{0 \le t \le 2h} \frac{1}{\sqrt{2h}} \left\| \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+t}^{(c_{i+1})} \right\| + \max_{i=1,\dots,q_T} \sup_{0 \le t \le 2h} \frac{1}{\sqrt{2h}} \left\| \widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-t}^{(c_i)} \right\| \\
&= O_P\left(\sqrt{\log\left(2q_T\right)} \right) = O_P\left(\sqrt{\log(T/h)} \right) \qquad (3.13)
\end{aligned}
$$

since by Assumption 3.1 $q_T \le T/(2h)$ and $T^{1/2}\nu_T/\sqrt{2h} = o(1)$.
Similarly we obtain that

$$
\max_{i=1,\dots,q_T} \sup_{c_i - h \le t \le c_i} \|\mathbf{\Lambda}_t\| = O_P\left(\sqrt{\log(T/h)} \right),
$$

which in combination with (3.12) and (3.13) shows the assertion. □

To obtain consistency of the estimators, the threshold needs to be small enough to be asymptotically negligible compared to the squared signal term as in Theorem 3.2 (a) to guarantee that every change is detected with asymptotic probability 1. At the same time, the threshold needs to grow faster than the squared noise term in Theorem 3.2 (c) so that false positives occur with asymptotic probability 0. Hence, both conditions are fulfilled under the following assumption:

Assumption 3.3. *The threshold fulfills:*

$$
\frac{\beta_{h,T}}{h_T \min\limits_{i=1,\dots,q_T} \|\mathbf{d}_i\|^2} \to 0, \qquad \frac{\log \frac{T}{h_T}}{\beta_{h,T}} \to 0 \qquad (T \to \infty).
$$

In particular, larger bandwidths h_T lead to a better detectability of the change point, where due to (3.2) an upper bound related to the distance to the neighboring change points applies. This is also confirmed by the simulation results in Table 6.4.

The following remark introduces a threshold that has a nice interpretation in connection with change point testing:

Remark 3.1. The threshold is often obtained as the asymptotic $1-\alpha_T$-quantile based on the limit result in Theorem 3.2 (b) for some sequence $\alpha_T \to 0$. In this case a choice of

$$
\frac{\left(-\log\log \frac{1}{\sqrt{1-\alpha_T}}\right)^2}{\log \frac{T}{h_T}} = O(1)
$$

similar to Eichinger and Kirch (2018) can replace the slightly stronger lower bound of Assumption 3.3 on the threshold without compromising our theoretical results. In the simulation study in Section 6 we use this threshold with $\alpha_T = 0.05$. This

controls the family-wise error rate at level α_T asymptotically related to testing each time point for a possible change. In fact, Theorem 3.2 shows that such a threshold with a constant sequence α yields an asymptotic test at level α which has asymptotic power one by Theorem 4.1. Tests designed for the at-most-one-change as in Hušková and Steinebach (2000), Hušková and Steinebach (2002) often have a better power, but are not as good at localizing change points (see Figure 1 in Cho and Kirch (2020) for an illustration).

4 Consistency of the segmentation procedure

In this section, we will show consistency of the above segmentation procedure for both the estimators of the number and locations of the change points. Furthermore, we derive localization rates for the estimators of the locations of the change points for some special cases showing that they cannot be improved in general. This is complemented by the observations that these localization rates are indeed minimax-optimal if the number of change points is bounded in addition to observing Wiener processes with drift. Otherwise the generic rates that are obtained based solely on the invariance principle will not be tight in the sense that the proposed procedure can provide better rates than suggested by the invariance principle.

The following theorem shows that the change point estimators defined in (3.10) are consistent for the number and locations of the change points.

Theorem 4.1. *Let Assumptions 2.1, 3.1 - 3.3 hold. Let* $0 < \hat{c}_1 < \ldots < \hat{c}_{\hat{q}_T}$ *be the change point estimators of type* (3.10)*. Then for any* $\tau > 0$ *it holds that*

$$\lim_{T\to\infty} \mathrm{P}\left(\max_{i=1,\ldots,\min(\hat{q}_T,q_T)} |\hat{c}_i - c_i| \leq \tau h, \hat{q}_T = q_T \right) = 1.$$

Proof of Theorem 4.1. Define for $0 < \tau < 1$ the following set

$$S_T = S_T^{(1)} \cap S_T^{(2)} \cap \bigcap_{i=1}^{q_T} \left(S_T^{(3)}(i,\tau) \cap S_T^{(4)}(i,\tau) \right), \tag{4.1}$$

where

$$\begin{aligned}
S_T^{(1)} &= \left\{ \max_{i=0,\ldots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \mathbf{M}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{M}_t < \beta \right\}, \\
S_T^{(2)} &= \left\{ \min_{i=1,\ldots,q_T} \mathbf{M}_{c_i}' \widehat{\mathbf{A}}_{c_i}^{-1} \mathbf{M}_{c_i} \geq \beta \right\}, \\
S_T^{(3)}(i,\tau) &= \bigcap_{k=1}^{\lceil \frac{1}{\tau} \rceil - 1} \left\{ \sup_{c_i - h \leq t \leq c_i - k\tau h} \|\mathbf{M}_t\| < \|\mathbf{M}_{c_i-(k-1)\tau h}\| \right\}, \\
S_T^{(4)}(i,\tau) &= \bigcap_{k=1}^{\lceil \frac{1}{\tau} \rceil - 1} \left\{ \sup_{c_i + k\tau h \leq t \leq c_i + h} \|\mathbf{M}_t\| < \|\mathbf{M}_{c_i+(k-1)\tau h}\| \right\}.
\end{aligned}$$

On $S_T^{(1)}$ there are asymptotically no significant points outside of h-environments of the change points. On $S_T^{(2)}$ there is at least one significant time point for each change point. On $S_T^{(3)}(i,\tau) \cap S_T^{(4)}(i,\tau)$ with $\tau < \eta/2$, there are no local extrema (within the h-environment of c_i) that are outside the interval $(c_i - \tau h, c_i + \tau h)$. Additionally, on $S_T^{(2)} \cap S_T^{(3)}(i,\tau) \cap S_T^{(4)}(i,\tau)$ the global extremum within that interval will be the only significant local extremum within the h-environment of c_i such that

$$\left\{ \max_{i=0,\dots,\min(\hat{q}_T, q_T)} |\hat{c}_i - c_i| \leq \tau h, \hat{q}_T = q_T \right\} \supset S_T.$$

We will show that S_T is an asymptotic one set.

For $S_T^{(1)}$ it holds by Lemma 3.1, Lemma B.2 (ii) and Corollary B.3 (iii) that

$$\begin{aligned}
&\max_{i=0,\dots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \mathbf{M}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{M}_t = \max_{i=0,\dots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \mathbf{\Lambda}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{\Lambda}_t \\
= &\max_{i=0,\dots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \left\| \widehat{\mathbf{A}}_t^{-1/2} \mathbf{\Lambda}_t \right\|^2 \\
\leq &\max_{i=0,\dots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \left\| \widehat{\mathbf{A}}_t^{-1/2} \right\|^2 \max_{i=0,\dots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \|\mathbf{\Lambda}_t\|^2 \\
\leq &\sup_{h\leq t\leq T-h} \left\| \widehat{\mathbf{A}}_t^{-1} \right\| \max_{i=0,\dots,q_T} \sup_{c_i+h<t<c_{i+1}-h} \|\mathbf{\Lambda}_t\|^2 = O_P(1) \cdot O_P\left(\log T/h\right) = O_P\left(\log T/h\right),
\end{aligned}$$

where the second-to-last equality follows from Assumption 3.2 and Theorem 3.2 (c), which by Assumption 3.3 shows that $S_T^{(1)}$ is an asymptotic one set.

For $S_T^{(2)}$ it holds by Corollary B.3 (i) and Lemma B.4 that

$$\begin{aligned}
&\min_{i=1,\dots,q_T} \mathbf{M}_{c_i}' \widehat{\mathbf{A}}_{c_i}^{-1} \mathbf{M}_{c_i} \geq \min_{i=1,\dots,q_T} \frac{1}{\left\| \widehat{\mathbf{A}}_{c_i} \right\|} \|\mathbf{M}_{c_i}\|^2 \geq \min_{i=1,\dots,q_T} \frac{1}{\left\| \widehat{\mathbf{A}}_{c_i} \right\|} \min_{i=1,\dots,q_T} \|\mathbf{M}_{c_i}\|^2 \\
= &\frac{1}{\max_{i=1,\dots,q_T} \left\| \widehat{\mathbf{A}}_{c_i} \right\|} \min_{i=1,\dots,q_T} \|\mathbf{M}_{c_i}\|^2.
\end{aligned}$$

By Assumption 3.2, for each $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that

$$\limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} \left\| \widehat{\mathbf{A}}_{c_i} \right\| > C_\varepsilon \right) < \varepsilon.$$

Furthermore, we obtain by the reverse triangle inequality, Assumption 3.1 and Theorem 3.2 (c) that

$$\begin{aligned}
&\min_{i=1,\dots,q_T} \|\mathbf{M}_{c_i}\| = \min_{i=1,\dots,q_T} \|\mathbf{m}_{c_i} + \mathbf{\Lambda}_{c_i}\| \\
&\geq \min_{i=1,\dots,q_T} \|\mathbf{m}_{c_i}\| - \max_{i=1,\dots,q_T} \|\mathbf{\Lambda}_{c_i}\| = \sqrt{\frac{h}{2}} \min_{i=1,\dots,q_T} \|\mathbf{d}_i\| + O_P\left(\sqrt{\log \frac{T}{h}}\right) \\
&= \sqrt{\frac{h}{2}} \min_{i=1,\dots,q_T} \|\mathbf{d}_i\| + o_P\left(\sqrt{\frac{h}{2}} \min_{i=1,\dots,q_T} \|\mathbf{d}_i\|\right) = \left(\min_{i=1,\dots,q_T} \|\mathbf{d}_i\|\right)(1 + o_P(1))
\end{aligned}$$

and therefore

$$\begin{aligned}
&\min_{i=1,\dots,q_T} \|\mathbf{M}_{c_i}\|^2 \geq \frac{h}{2} \left(\min_{i=1,\dots,q_T} \|\mathbf{d}_i\|^2\right)(1 + o_P(1))^2 \\
&= \frac{h}{2} \left(\min_{i=1,\dots,q_T} \|\mathbf{d}_i\|^2\right)(1 + o_P(1)) = \frac{h}{2} \min_{i=1,\dots,q_T} \|\mathbf{d}_i\|^2 + o_P\left(\frac{h}{2} \min_{i=1,\dots,q_T} \|\mathbf{d}_i\|^2\right),
\end{aligned}$$

which in combination with Assumption 3.3 shows that $S_T^{(2)}$ is an asymptotic one set.

For arbitrary $\tau > 0$ and $S_T^{(3)}(i, \tau)$ it holds by the reverse triangle inequality, Lemma 3.1 and Theorem 3.2 uniformly that

$$\begin{aligned}
&\min_{i=1,\dots,q_T} \left(\left\|\mathbf{M}_{c_i-(k-1)\tau h}\right\| - \sup_{c_i - h \leq t \leq c_i - k\tau h} \|\mathbf{M}_t\|\right) \\
&\geq \min_{i=1,\dots,q_T} \left(\left\|\mathbf{m}_{c_i-(k-1)\tau h}\right\| - \left\|\mathbf{\Lambda}_{c_i-(k-1)\tau h}\right\| - \sup_{c_i - h \leq t \leq c_i - k\tau h} \|\mathbf{m}_t\| - \sup_{c_i - h \leq t \leq c_i - k\tau h} \|\mathbf{\Lambda}_t\|\right) \\
&= \min_{i=1,\dots,q_T} \left(\left\|\mathbf{m}_{c_i-(k-1)\tau h}\right\| - \|\mathbf{m}_{c_i-k\tau h}\|\right) + O_P\left(\sqrt{\log \frac{T}{h}}\right) \\
&= \frac{\tau}{\sqrt{2}} \sqrt{h} \min_{i=1,\dots,q_T} \|\mathbf{d}_i\| + O_P\left(\sqrt{\log \frac{T}{h}}\right),
\end{aligned}$$

thus showing that $\cap_{i=1}^{q_T} S_T^{(3)}(i, \tau)$ is an asymptotic one set. It follows analogously that $\cap_{i=1}^{q_T} S_T^{(4)}(i, \tau)$ is an asymptotic one set. □

The theorem shows in particular that the number of change points is estimated consistently. For the linear bandwidth we additionally get consistency of the change point locations in rescaled time, while for the sublinear bandwidths we already get a convergence rate of h/T towards the rescaled change points.

Under the following stronger assumptions, the localization rates can be improved further:

Assumption 4.1. *(a) It holds for any of the centered processes* $\widetilde{\mathbf{R}}^{(j)}$ *as in* (3.6) *and any value* $\theta_i = \theta_{i,T}$ *(which will be* c_i *or* $c_i \pm h$ *when the assumption is applied) for any sequence* $D_T \geq 1$ *(bounded or unbounded)*

$$\sup_{\frac{D_T}{\|\mathbf{d}_i\|^2} \leq s \leq h} \frac{\sqrt{D_T} \left\| \widetilde{\mathbf{R}}^{(j)}_{\theta_i} - \widetilde{\mathbf{R}}^{(j)}_{\theta_i \pm s} \right\|}{s \|\mathbf{d}_i\|} = O_P(\omega_T).$$

(b) Let now the upper index θ_i *denote the active stretch in the stationary segment* $(\theta_i, \theta_i + s)$ *respectively* $(\theta_i - s, \theta_i)$*. Then, it holds for any sequence* $D_T > 0$

$$\max_{i=1,\ldots,q_T} \sup_{\frac{D_T}{\|\mathbf{d}_i\|^2} \leq s \leq h} \frac{\sqrt{D_T} \left\| \widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i} - \widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i \pm s} \right\|}{s \|\mathbf{d}_i\|} = O_P(\tilde{\omega}_T).$$

The localization rates of the MOSUM procedure are determined by the rates $\omega_n, \tilde{\omega}_n$ which need to be derived for each example separately (at least for the tight ones). In the context of partial sum processes these results are well known. For example, the suprema in (a) are stochastically bounded by the Hájék–Rényi inequality which has been shown for partial sum processes even with weakly dependent errors. In that context, the assertion in (b) is fulfilled with a polynomial rate in q_T (see Cho and Kirch (2021+), Proposition 2.1 (c)(ii)).

Remark 4.1. (a) For Wiener processes with drift, it holds that $\omega_T = 1$ and $\tilde{\omega}_T = \sqrt{\log(q_T)}$ (see Proposition A.4 below).

(b) By the invariance principle in Assumption 2.1, all rates are clearly dominated by $T^{1/2}\nu_T$. However, this is often far too liberal a bound (see Proposition 2.1 in Cho and Kirch (2021+) for some tight bounds in case of partial sum processes).

(c) Often, there exist forward and backwards invariance principles from some arbitrary starting value θ_i for each regime. This is the case for partial sum processes and for (backward and forward) Markov processes due to the Markov property. For renewal processes, this can be shown along the lines of the original proof for the invariance principle (Csörgö, Horváth, and Steinebach (1987)) because the time to the next (previous) event is asymptotically negligible; see also Example 1.2 in Kühn and Steinebach (2002)). In this case, the Hájék–Rényi results for Wiener processes carry over (see Proposition A.4) to the different processes underlying each regime, resulting in $\omega_T = 1$. For the situation with a bounded number of change points this carries over to $\tilde{\omega}_T$.

Theorem 4.2.
Let Assumptions 2.1, 3.1 - 3.3 and 4.1 hold. For $\hat{q}_T < q_T$ *define* $\hat{c}_i = T$ *for* $i = \hat{q}_T + 1, \ldots, q_T$.

(a) For a single change point estimator the following localization rate holds

$$\|\mathbf{d}_i\|^2 |\hat{c}_i - c_i| = O_P\left(\omega_T^2\right).$$

(b) The following uniform rate holds true:

$$\max_{i=1,\ldots,q_T} \|\mathbf{d}_i\|^2 |\hat{c}_i - c_i| = O_P\left(\tilde{\omega}_T^2\right).$$

Proof. We will only show (b) as (a) follows analogously by replacing $\tilde{\omega}_T$ by ω_T and without the maxima and minima. Let S_T be as in (4.1). For $C > 0$ it holds that

$$\begin{aligned}
&\left\{\max_{i=1,\ldots,q_T} \|\mathbf{d}_i\|^2 |\hat{c}_i - c_i| \geq C\tilde{\omega}_T^2\right\} \\
&= \left\{\min_{i=1,\ldots,q_T} \left((\hat{c}_i - c_i) + \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\right) \leq 0, S_T\right\} \cup \left\{\max_{i=1,\ldots,q_T} \left((\hat{c}_i - c_i) - \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\right) \geq 0, S_T\right\} \\
&\quad \cup \left\{\max_{i=1,\ldots,q_T} \|\mathbf{d}_i\|^2 |\hat{c}_i - c_i| \geq C\tilde{\omega}_T^2, S_T^c\right\}.
\end{aligned}$$

By Theorem 4.1, the latter set has asymptotic probability 0. With the substitution $s = t - c_i$ it holds that

$$\begin{aligned}
&\left\{\max_{i=1,\ldots,q_T} \left((\hat{c}_i - c_i) - \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\right) \geq 0, S_T\right\} = \bigcup_{i=1}^{q_T} \left\{(\hat{c}_i - c_i) \geq \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}, S_T\right\} \\
&\subset \bigcup_{i=1}^{q_T} \left\{\sup_{c_i + \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \leq t \leq c_i + h} \|\mathbf{M}_t\|^2 > \sup_{c_i - h \leq t < c_i + \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}} \|\mathbf{M}_t\|^2\right\} \\
&\subset \bigcup_{i=1}^{q_T} \left\{\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \leq s \leq h} 2h\left(\|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2\right) \geq 0\right\}.
\end{aligned}$$

By straightforward algebraic calculus and Lemma 3.1 we obtain for $C\tilde{\omega}_T^2 / \|\mathbf{d}_i\|^2 \leq s \leq h$ that

$$\begin{aligned}
\mathbf{V}_s &= \|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2 \\
&= -\left(\mathbf{m}_{c_i} - \mathbf{m}_{c_i+s} + \mathbf{\Lambda}_{c_i} - \mathbf{\Lambda}_{c_i+s}\right)' \left(\mathbf{m}_{c_i} + \mathbf{m}_{c_i+s} + \mathbf{\Lambda}_{c_i} + \mathbf{\Lambda}_{c_i+s}\right) \\
&= -\frac{1}{2h}\left(D_{1,s}\,\mathbf{d}_i + \mathbf{N}_{1,s}\right)' \left(D_{2,s}\,\mathbf{d}_i + \mathbf{N}_{2,s}\right), \qquad (4.2)
\end{aligned}$$

where $D_{1,s} = s, \quad D_{2,s} = 2h - s,$

$$\begin{aligned}
\mathbf{N}_{1,s} &= 2\left(\widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})}\right) - \left(\widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h}^{(c_i)}\right) - \left(\widetilde{\mathbf{R}}_{c_i+h+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})}\right) \\
\mathbf{N}_{2,s} &= \left(\widetilde{\mathbf{R}}_{c_i+h+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})}\right) + 2\left(\widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})}\right) \\
&\quad - 2\left(\widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)}\right) - \left(\widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h}^{(c_i)}\right).
\end{aligned}$$

Therefore, it holds due to $D_{1,s}D_{2,s} = s(2h - s) > 0$ that

$$
\begin{aligned}
&\mathrm{P}\left(\max_{i=1,\ldots,q_T}(\hat{c}_i - c_i) \geq \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}, S_T\right) \\
\leq\ &\mathrm{P}\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h} 2h\left(\|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2\right) \geq 0\right) \\
=\ &P\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h} -\left(D_{1,s}D_{2,s}\|\mathbf{d}_i\|^2 + D_{1,s}\mathbf{d}_i'\mathbf{N}_{2,s} + D_{2,s}\mathbf{N}_{1,s}'\mathbf{d}_i + \mathbf{N}_{1,s}'\mathbf{N}_{2,s}\right) \geq 0\right) \\
=\ &P\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h} -D_{1,s}D_{2,s}\|\mathbf{d}_i\|^2\left(1 + \frac{\mathbf{d}_i'\mathbf{N}_{2,s}}{D_{2,s}\|\mathbf{d}_i\|^2} + \frac{\mathbf{N}_{1,s}'\mathbf{d}_i}{D_{1,s}\|\mathbf{d}_i\|^2} + \frac{\mathbf{N}_{1,s}'\mathbf{N}_{2,s}}{D_{1,s}D_{2,s}\|\mathbf{d}_i\|^2}\right) \geq 0\right) \\
=\ &P\left(\min_{i=1,\ldots,q_T}\inf_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\left(\frac{\mathbf{d}_i'\mathbf{N}_{2,s}}{D_{2,s}\|\mathbf{d}_i\|^2} + \frac{\mathbf{N}_{1,s}'\mathbf{d}_i}{D_{1,s}\|\mathbf{d}_i\|^2} + \frac{\mathbf{N}_{1,s}'\mathbf{N}_{2,s}}{D_{1,s}D_{2,s}\|\mathbf{d}_i\|^2}\right) \leq -1\right) \\
\leq\ &P\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\left|\frac{\mathbf{d}_i'\mathbf{N}_{2,s}}{D_{2,s}\|\mathbf{d}_i\|^2} + \frac{\mathbf{N}_{1,s}'\mathbf{d}_i}{D_{1,s}\|\mathbf{d}_i\|^2} + \frac{\mathbf{N}_{1,s}'\mathbf{N}_{2,s}}{D_{1,s}D_{2,s}\|\mathbf{d}_i\|^2}\right| \geq 1\right) \\
\leq\ &\mathrm{P}\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\left|\frac{\mathbf{d}_i'\mathbf{N}_{2,s}}{D_{2,s}\|\mathbf{d}_i\|^2}\right| \geq \tfrac{1}{3}\right) \qquad (4.3) \\
&+\mathrm{P}\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\left|\frac{\mathbf{N}_{1,s}'\mathbf{d}_i}{D_{1,s}\|\mathbf{d}_i\|^2}\right| \geq \tfrac{1}{3}\right) \qquad (4.4) \\
&+\mathrm{P}\left(\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\left|\frac{\mathbf{N}_{1,s}'\mathbf{N}_{2,s}}{D_{1,s}D_{2,s}\|\mathbf{d}_i\|^2}\right| \geq \tfrac{1}{3}\right). \qquad (4.5)
\end{aligned}
$$

For (4.3) it holds by Proposition A.4 (b) and Assumption 3.1 combined with $q_T \leq T/(2h)$ that

$$
\begin{aligned}
&\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\frac{\left\|\widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)}\right\|}{D_{2,s}\|\mathbf{d}_i\|} \leq \max_{i=1,\ldots,q_T}\sup_{0\leq s\leq h}\frac{\left\|\widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-s}^{(c_i)}\right\|}{(2h-s)\|\mathbf{d}_i\|} \\
\leq\ &\max_{i=1,\ldots,q_T}\sup_{0\leq s\leq h}\frac{\left\|\widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-s}^{(c_i)}\right\|}{h\|\mathbf{d}_i\|} = O_P\left(\frac{\sqrt{\log(2q_T)}}{\sqrt{h}\|\mathbf{d}_i\|}\right) = O_P\left(\frac{\sqrt{\log(T/h)}}{\sqrt{h}\|\mathbf{d}_i\|}\right) = o_P(1).
\end{aligned}
$$

We obtain analogously that

$$
\max_{i=1,\ldots,q_T}\sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}\leq s\leq h}\frac{\left\|\widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h}^{(c_i)}\right\|}{D_{2,s}\|\mathbf{d}_i\|} = o_P(1),
$$

$$\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\left\| \widetilde{\mathbf{R}}_{c_i+h+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+h}^{(c_i)} \right\|}{D_{2,s} \left\| \mathbf{d}_i \right\|} = o_P(1),$$

$$\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\left\| \widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} \right\|}{D_{2,s} \left\| \mathbf{d}_i \right\|} = o_P(1).$$

Therefore, by the Cauchy-Schwarz inequality, it follows that

$$\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \left| \frac{\mathbf{N}_{1,s}' \mathbf{d}_i}{D_{2,s} \left\| \mathbf{d}_i \right\|^2} \right| \le \max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\|\mathbf{N}_{1,s}\|}{D_{2,s} \left\| \mathbf{d}_i \right\|} = o_P(1)$$

and therefore

$$\mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \left| \frac{\mathbf{d}_i' \mathbf{N}_{2,s}}{D_{2,s} \|\mathbf{d}_i\|^2} \right| \ge \tfrac{1}{3} \right) = o(1). \tag{4.6}$$

For (4.4) it holds by Assumption 4.1 for arbitrary $y > 0$ that

$$\begin{aligned} &\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\left\| \widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} \right\|}{s\|\mathbf{d}_i\|} \ge y \right) \\ = &\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\sqrt{C\tilde{\omega}_T^2} \left\| \widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} \right\|}{s\|\mathbf{d}_i\|} \ge \sqrt{C}\tilde{\omega}_T y \right) = 0. \end{aligned}$$

Analogous results hold for

$$\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\left\| \widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h}^{(c_i)} \right\|}{s \left\| \mathbf{d}_i \right\|}, \max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\left\| \widetilde{\mathbf{R}}_{c_i+h+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})} \right\|}{s \left\| \mathbf{d}_i \right\|}, .$$

Therefore it holds by the Cauchy-Schwarz inequality for (4.4) that

$$\begin{aligned} &\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \left| \frac{\mathbf{N}_{1,s}' \mathbf{d}_i}{D_{1,s} \|\mathbf{d}_i\|^2} \right| \ge \tfrac{1}{3} \right) \\ \le &\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\|\mathbf{N}_{1,s}\|}{D_{1,s} \|\mathbf{d}_i\|} \ge \tfrac{1}{3} \right) = 0 \end{aligned} \tag{4.7}$$

For (4.5) we obtain by (4.6) and (4.7) together with the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \left| \frac{\mathbf{N}_{1,s}' \mathbf{N}_{2,s}}{D_{1,s} D_{2,s} \|\mathbf{d}_i\|^2} \right| \\
\le &\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\|\mathbf{N}_{1,s}\|}{D_{1,s} \|\mathbf{d}_i\|} \max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \frac{\|\mathbf{N}_{2,s}\|}{D_{2,s} \|\mathbf{d}_i\|} = O_P(1) o_P(1) = o_P(1)
\end{aligned}$$

and therefore

$$\mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} \left| \frac{\mathbf{N}_{1,s}' \mathbf{N}_{2,s}}{D_{1,s} D_{2,s} \|\mathbf{d}_i\|^2} \right| \ge \tfrac{1}{3} \right) = o(1). \tag{4.8}$$

By combining (4.6)-(4.8), we obtain that

$$\begin{aligned}
&\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} (\hat{c}_i - c_i) \ge \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}, S_T \right) \\
\le &\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\max_{i=1,\dots,q_T} \sup_{\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \le s \le h} 2h \left(\|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2 \right) \ge 0 \right) = 0.
\end{aligned}$$

We obtain analogously that

$$\lim_{C\to\infty} \limsup_{T\to\infty} \mathrm{P}\left(\min_{i=1,\dots,q_T} (\hat{c}_i - c_i) \le -\frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2}, S_T \right) = 0.$$

Since

$$\begin{aligned}
&\mathrm{P}\left(\max_{i=1,\dots,q_T} \left(|\hat{c}_i - c_i| - \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \right) \ge 0 \right) \\
\le\ &\mathrm{P}\left(\max_{i=1,\dots,q_T} \left((\hat{c}_i - c_i) - \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \right) \ge 0, S_T \right) \\
&+ \mathrm{P}\left(\min_{i=1,\dots,q_T} \left((\hat{c}_i - c_i) + \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \right) \le 0, S_T \right) \\
&+ \mathrm{P}\left(\max_{i=1,\dots,q_T} \left(|\hat{c}_i - c_i| - \frac{C\tilde{\omega}_T^2}{\|\mathbf{d}_i\|^2} \right) \ge 0 S_T^c \right)
\end{aligned}$$

this shows the assertions as S_T^c has asymptotic probability 0 by Theorem 4.1. □

Remark 4.2 (Minimax optimality)**.** We have already mentioned beneath (3.4) that the separation rate given there is minimax optimal (see Proposition 1 of Arias-Castro, Candes, and Durand (2011)).

Minimax optimal localization rates (derived in the context of changes in the mean of univariate time series, which is covered by the partial sum processes in our framework) are known for a few special cases: First, the minimax optimal localization rate for a single change point and in extension also for a bounded number of change points is given by $\omega_T = 1$ in the above notation (see e. g. Lemma 2 in Wang, Yu, and Rinaldo (2020)). In particular this shows that our procedures achieves the minimax optimality in case of a bounded number of change points under weak assumptions (as pointed out in Remark 4.1 (c)). Secondly, the optimal localization rate for unbounded change points under sub-Gaussianity (attained for partial sum process of i.i.d. errors) is given by $\tilde{\omega_T} = \sqrt{\log T}$ (see Proposition 6 in Verzelen et al. (2020) and Proposition 2.3 in Cho and Kirch (2021+)). Indeed, we match this rate for Wiener processes with drift.

The following theorem derives the limit distribution of the change point estimators for local changes which shows in particular that the rates are tight. In principle, this result can be used to obtain asymptotically valid confidence intervals for the change point locations. In case of fixed changes, the limit distribution depends on the underlying distribution of the original process (see Antoch and Hušková (1999) for the case of partial sum processes), where the proof can be done along the same lines. We need the following assumption:

Assumption 4.2. *Let* $\mathbf{d}_i = \mathbf{d}_{i,T} = \|\mathbf{d}_i\|\mathbf{u}_i + o(\|\mathbf{d}_i\|)$ *with* $\|\mathbf{u}_i\| = 1$ *and* $\|\mathbf{d}_{i,T}\| \to 0$. *Assume that* $\mathbf{Y}_s^{(j)} = \mathbf{Y}_s^{(j)}(c_i, D)$ *with*

$$\begin{aligned}
\mathbf{Y}_s^{(1)} &= \widetilde{\mathbf{R}}^{(c_i)}_{c_i - h + \frac{s-D}{\|d_i\|^2}} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i - h - \frac{D}{\|d_i\|^2}}, \\
\mathbf{Y}_s^{(21)} &= \widetilde{\mathbf{R}}^{(c_i)}_{c_i + \frac{s-D}{\|d_i\|^2}} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i - \frac{D}{\|d_i\|^2}}, \qquad \mathbf{Y}_s^{(22)} = \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + \frac{s-D}{\|d_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i - \frac{D}{\|d_i\|^2}}, \\
\mathbf{Y}_s^{(3)} &= \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + h + \frac{s-D}{\|d_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + h - \frac{D}{\|d_i\|^2}}
\end{aligned}$$

fulfill the following multivariate functional central limit theorem for any constant $D > 0$ *in an appropriate space equipped with the supremum norm*

$$\left(\|\mathbf{d}_i\| \, (\mathbf{Y}_s^{(1)}, \mathbf{Y}_s^{(21)}, \mathbf{Y}_s^{(22)}, \mathbf{Y}_s^{(3)})'\right)_{0 \le s \le 2D} \xrightarrow{w} \left(\widetilde{\mathbf{W}}_s\right)_{0 \le s \le 2D},$$

where $\widetilde{\mathbf{W}}$ *is a Wiener process with covariance matrix* $\mathbf{\Xi}$ *(not depending on* D*). For* $-D \le t \le D$ *denote* $\mathbf{W}_t = (\mathbf{W}_t^{(1)}, \mathbf{W}_t^{(21)}, \mathbf{W}_t^{(22)}, \mathbf{W}_t^{(3)})' = \widetilde{\mathbf{W}}_{D+t} - \widetilde{\mathbf{W}}_D$.

By Assumption 3.1 it holds $h\|\mathbf{d}_i\|^2 \to \infty$, such that the distance $h - \frac{2D}{\|d_i\|^2}$ between $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2j)}$ (resp. between $\mathbf{Y}^{(2j)}$ and $\mathbf{Y}^{(3)}$) diverges to infinity. As such for processes with independent increments the processes $\mathbf{Y}^{(1)}$, $(\mathbf{Y}^{(21)}, \mathbf{Y}^{(22)})'$, $\mathbf{Y}^{(3)}$ are independent for T large enough. Additionally, under weak assumptions such as mixing conditions this independence still holds asymptotically in the sense that $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$, $\mathbf{W}^{(3)}$ are independent.

Functional central limit theorems for these processes follow from invariance principles as in Assumption 2.1 with $\mathbf{\Sigma}_T \to \mathbf{\Sigma}$ as long as such invariance principles

still hold with an arbitrary (moving) starting value, which is typically the case (see also Remark 4.1 (c)). As such, it typically holds that $\boldsymbol{\Xi}^{(1)} = \boldsymbol{\Xi}^{(21)} = \boldsymbol{\Sigma}^{(c_i)}$ and $\boldsymbol{\Xi}^{(3)} = \boldsymbol{\Xi}^{(22)} = \boldsymbol{\Sigma}^{(c_{i+1})}$ where $\boldsymbol{\Xi}^j = \mathrm{Cov}(\mathbf{W}_1^{(j)})$ and $\boldsymbol{\Sigma}^{(c_i)}$ is the covariance matrix associated with the regime between the $(i-1)$-th and i-th change point.

The following theorem gives the asymptotic distribution for the change point estimators in case of local change points.

Theorem 4.3.
Let Assumptions 2.1, 3.1 - 3.3, 4.1 (a) with $\omega_T = 1$ and 4.2 hold. For $\hat{q}_T < q_T$ define $\hat{c}_i = T$ for $i = \hat{q}_T + 1, \ldots, q_T$. Let

$$\Psi_s^{(i)} := -|s| + \begin{cases} \mathbf{u}_i' \mathbf{W}_s^{(1)} - 2\,\mathbf{u}_i' \mathbf{W}_s^{(21)} + \mathbf{u}_i' \mathbf{W}_s^{(3)}, & s < 0 \\ \mathbf{u}_i' \mathbf{W}_s^{(1)} - 2\,\mathbf{u}_i' \mathbf{W}_s^{(22)} + \mathbf{u}_i' \mathbf{W}_s^{(3)}, & s \geq 0. \end{cases}$$

Then, for all $i = 1, \ldots, q_T$, it holds that for $T \to \infty$

$$\|\mathbf{d}_i\|^2 (\hat{c}_i - c_i) \xrightarrow{\mathcal{D}} \operatorname{argmax} \left\{ \Psi_t^{(i)} \middle| s \in \mathbb{R} \right\}$$

If there is a fixed number of changes $q_T = q$ with q fixed and a functional central limit theorem as in Assumption 4.2 holds jointly for all q change points, then the result also holds jointly.

Proof. Analogously to the proof of Theorem 4.2 it holds for $C > 0$, $-C \leq x \leq C$ and $i = 1, \ldots, q_T$ that

$$\begin{aligned} & \mathrm{P}\left(\|\mathbf{d}_i\| (\hat{c}_i - c_i) \leq x,\ \|\mathbf{d}_i\| |\hat{c}_i - c_i| \leq C \right) \\ = \ & \mathrm{P}\left(\sup_{-\frac{C}{\|\mathbf{d}_i\|^2} \leq s \leq x} \left(\|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2 \right) \geq \sup_{x < s \leq \frac{C}{\|\mathbf{d}_i\|^2}} \left(\|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2 \right) \right). \end{aligned}$$

By defining $D_{1,s}$, $D_{2,S}$, $D_{1,s}$ and $D_{2,s}$ from (4.2) also for $s < 0$, it holds that

$$\begin{aligned} \mathbf{V}_s &= \|\mathbf{M}_{c_i+s}\|^2 - \|\mathbf{M}_{c_i}\|^2 \\ &= -\left(\mathbf{m}_{c_i} - \mathbf{m}_{c_i+s} + \boldsymbol{\Lambda}_{c_i} - \boldsymbol{\Lambda}_{c_i+s}\right)' \left(\mathbf{m}_{c_i} + \mathbf{m}_{c_i+s} + \boldsymbol{\Lambda}_{c_i} + \boldsymbol{\Lambda}_{c_i+s}\right) \\ &= -\frac{1}{2h} \left(D_{1,s}\,\mathbf{d}_i + \mathbf{N}_{1,s}\right)' \left(D_{2,s}\,\mathbf{d}_i + \mathbf{N}_{2,s}\right), \end{aligned}$$

where

$$\begin{aligned} & D_{1,s} = |s|, \quad D_{2,s} = 2h - |s|, \\ & \mathbf{N}_{1,s} = \begin{cases} 2\left(\widetilde{\mathbf{R}}_{c_i+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i}^{(c_i)}\right) - \left(\widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h}^{(c_i)}\right) & \\ \quad - \left(\widetilde{\mathbf{R}}_{c_i+h+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})}\right), & \text{for } s < 0 \\ 2\left(\widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})}\right) - \left(\widetilde{\mathbf{R}}_{c_i-h+s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i-h}^{(c_i)}\right) & \\ \quad - \left(\widetilde{\mathbf{R}}_{c_i+h+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i+h}^{(c_{i+1})}\right), & \text{for } s \geq 0 \end{cases} \end{aligned}$$

$$\mathbf{N}_{2,s} = \begin{cases} \left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h+s}\right) + 2\left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h+s} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i}\right) \\ - 2\left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i+s} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i-h}\right) - \left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i-h} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i-h+s}\right), & \text{for } s<0 \\ \left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h+s} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h}\right) + 2\left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s}\right) \\ - 2\left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i-h+s}\right) - \left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i-h+s} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i-h}\right), & \text{for } s\geq 0 \end{cases}$$

By Proposition A.4 (b), we obtain that

$$\sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s}\right\| \leq \sup_{0\leq s\leq h} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+h-s}\right\| = O_P\left(\sqrt{h}\right)$$

and similarly

$$\sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i} - \widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i\pm h\pm s}\right\| = O_P\left(\sqrt{h}\right),$$

where θ_i takes values c_i-h, c_i, c_i+h and (with a slight abuse of notation) the upper index $^{(\theta_i)}$ denotes the index of the active regime of the process.
Since furthermore

$$\begin{aligned} &\sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i}\right\| \\ =& \sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-\frac{C}{\|\mathbf{d}_i\|^2}} + \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-\frac{C}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i}\right\| \\ \leq& \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-\frac{C}{\|\mathbf{d}_i\|^2}}\right\| + \sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-\frac{C}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s}\right\| \end{aligned}$$

and since by Assumption 4.2

$$\begin{aligned} &\|\mathbf{d}_i\| \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-\frac{C}{\|\mathbf{d}_i\|^2}}\right\| \xrightarrow{\mathcal{D}} \left\|\widetilde{\mathbf{W}}^{(22)}_C\right\|, \\ &\|\mathbf{d}_i\| \sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i-\frac{C}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s}\right\| \xrightarrow{\mathcal{D}} \sup_{C\leq t\leq 2C} \left\|\widetilde{\mathbf{W}}^{(22)}_t\right\|, \end{aligned}$$

it follows that

$$\sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i+s} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i}\right\| = O_P\left(\frac{1}{\|\mathbf{d}_i\|}\right).$$

Analogously, it follows that

$$\sup_{0\leq s\leq \frac{C}{\|\mathbf{d}_i\|^2}} \left\|\widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i\pm s} - \widetilde{\mathbf{R}}^{(\theta_i)}_{\theta_i}\right\| = O_P\left(\frac{1}{\|\mathbf{d}_i\|}\right),$$

where θ_i takes values $c_i - h, c_i, c_i + h$ and (with a slight abuse of notation) the upper index $^{(\theta_i)}$ denotes the index of the active regime of the process. Therefore,

$$\|\mathbf{N}_{1,s}\| = O_P\left(\frac{1}{\|\mathbf{d}_i\|}\right), \quad \|\mathbf{N}_{2,s}\| = O_P\left(\sqrt{h}\right).$$

Thus, it holds by Assumption 3.1 that

$$\begin{aligned}
\frac{1}{2h} D_{1,s} D_{2,s} \|\mathbf{d}_i\|^2 &= |s| \|\mathbf{d}_i\|^2 - \frac{s^2}{2h} \|\mathbf{d}_i\|^2 = |s| \|\mathbf{d}_i\|^2 + O\left(\frac{1}{h \|\mathbf{d}_i\|^2}\right) \\
&= |s| \|\mathbf{d}_i\|^2 + o(1), \\
\frac{1}{2h} D_{1,s} \mathbf{d}_i' \mathbf{N}_{2,s} &= |s| O_P\left(\frac{\|\mathbf{d}_i\|}{\sqrt{h}}\right) = O_P\left(\frac{1}{\sqrt{h} \|\mathbf{d}_i\|}\right) = o_P(1), \\
\frac{1}{2h} D_{2,s} \mathbf{N}_{1,s}' \mathbf{d}_i &= \left(1 - \frac{|s|}{2h}\right) \mathbf{N}_{1,s}' \mathbf{d}_i = \mathbf{N}_{1,s}' \mathbf{d}_i + o_P(1), \\
\frac{1}{2h} \mathbf{N}_{1,s}' \mathbf{N}_{2,s} &= O_P\left(\frac{1}{h \|\mathbf{d}_i\|}\right) = o_P(1)
\end{aligned}$$

and therefore,

$$\mathbf{V}_s = -|s| \|\mathbf{d}_i\| - \mathbf{N}_{1,s}' \mathbf{d}_i + o_P(1) = -|s| \|\mathbf{d}_i\| - \|\mathbf{d}_i\| \mathbf{N}_{1,s}' \mathbf{u}_i + o_P(1),$$

where $\mathbf{u}_i$ is as in Assumption 4.2 and the $o_P(1)$-term holds uniformly in s. With the substitution $r = s \|\mathbf{d}_i\|^2$ we obtain that

$$\begin{aligned}
\mathbf{V}_r = o_p(1) - |r| & \\
- \|\mathbf{d}_i\| & \begin{cases}
2\left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i + \frac{r}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i}\right)' \mathbf{u}_i - \left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i - h + \frac{r}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i - h}\right)' \mathbf{u}_i & \\
- \left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + h + \frac{r}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + h}\right)' \mathbf{u}_i, & \text{for } r < 0 \\
2\left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + \frac{r}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i}\right)' \mathbf{u}_i - \left(\widetilde{\mathbf{R}}^{(c_i)}_{c_i - h + \frac{r}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_i)}_{c_i - h}\right)' \mathbf{u}_i & \\
- \left(\widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + h + \frac{r}{\|\mathbf{d}_i\|^2}} - \widetilde{\mathbf{R}}^{(c_{i+1})}_{c_i + h}\right)' \mathbf{u}_i, & \text{for } r \geq 0
\end{cases}
\end{aligned}$$

By Assumption 4.2, it holds that $(\mathbf{V}_r)_{-C \leq r \leq C} \xrightarrow{w} \left(\Psi_s^{(i)}\right)_{-C \leq r \leq C}$ on the 'appropriate space equipped with the supremum norm' from Assumption 4.2. Since the supremum is a continuous function on this space it holds that

$$\mathrm{P}\left(\|\mathbf{d}_i\|\,(\hat{c}_i - c_i) \le x, \|\mathbf{d}_i\|\,|\hat{c}_i - c_i| \le C\right) = \mathrm{P}\left(\sup_{-C\le r\le x} \mathbf{V}_r \ge \sup_{x\le r\le C} \mathbf{V}_r\right)$$

$$\longrightarrow \mathrm{P}\left(\sup_{-C\le r\le x} \Psi_r^{(i)} \ge \sup_{x\le r\le C} \Psi_s^{(i)}\right) = \mathrm{P}\left(\operatorname{argmax} \Psi_r^{(i)} \le x | -C \le x \le C\right).$$

By Theorem 4.2, letting $C \to \infty$ proves the assertion. □

Due to the Markov property of Wiener processes, $\{\Psi_t^{(i)} : t > 0\}$ is independent of $\{\Psi_t^{(i)} : t < 0\}$.

Remark 4.3. (a) If $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$, $\mathbf{W}^{(3)}$ are independent which is typically the case (see discussion beneath Assumption 4.2), then $\Psi_t^{(i)}$ simplifies to

$$\Psi_t^{(i)} := -|t| + \begin{cases} \sqrt{\sigma_{(1)}^2 + 4\,\sigma_{(21)}^2 + \sigma_{(3)}^2}\, B_t, & t < 0 \\ \sqrt{\sigma_{(1)}^2 + 4\,\sigma_{(22)}^2 + \sigma_{(3)}^2}\, B_t, & t \ge 0, \end{cases}$$

where B is a (univariate) standard Wiener process and $\sigma_{(j)}^2 = \mathbf{u}_i' \boldsymbol{\Xi}^{(j)} \mathbf{u}_i$. Usually (see discussion beneath Assumption 4.2) $\sigma_{(21)} = \sigma_{(1)}$ and $\sigma_{(22)} = \sigma_{(3)}$ further simplifying the expression. For some examples such as partial sum processes it holds $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ for all t, such that all $\sigma_{(j)}$ coincide. In this case this further simplifies to

$$\Psi_t^{(i)} := -|t| + \sqrt{6}\,\sigma_{(1)}\, B_t.$$

For univariate partial sum processes this result has already been obtained in Theorem 3.3 of Eichinger and Kirch (2018). However, the assumption of $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ is typically not fulfilled for renewal processes because the covariance depends on the changing intensity of the process.

(b) If $\mathbf{W}^{(1)}$, $(\mathbf{W}^{(21)}, \mathbf{W}^{(22)})'$, $\mathbf{W}^{(3)}$ are independent and $\mathbf{M}_t$ in (3.10) is replaced by $\boldsymbol{\Sigma}_t^{-1/2}\mathbf{M}_t$, then the Wiener processes $\mathbf{W}^{(j)}$ are standard Wiener processes, such that $\Psi_t^{(i)}$ simplifies to

$$\Psi_t^{(i)} := -|t| + \sqrt{6}\, B_t.$$

This shows that in this case the limit distribution of $\hat{c}_i - c_i$ only depends on the magnitude of the change $\mathbf{d}_i$ but not on its direction $\mathbf{u}_i$.

Statistically, however, this is difficult to achieve as it requires a uniformly (in t) consistent estimator for the usually unknown covariance matrices $\boldsymbol{\Sigma}_t$.

5 Asymmetric bandwidths

In analogy to Meier, Cho, and Kirch (2021), we define the MOSUM-statistic as

$$\mathbf{M}_t = \mathbf{M}_{t,T,h_{l,T},h_{r,T}}(\mathbf{Z}) = \sqrt{\frac{h_l h_r}{h_l + h_r}} \left[\frac{1}{h_r} \left(\mathbf{Z}_{t+h_r} - \mathbf{Z}_t \right) - \frac{1}{h_l} \left(\mathbf{Z}_t - \mathbf{Z}_{t-h_l} \right) \right]. \tag{5.1}$$

Similar to Section 3, the statistic will fluctuate around 0 in the case of no mean change. On the other hand, the statistic will differ from 0 close to a change point. Ideally, we want to chose the two bandwidths as large as possible (similar to Section 3). On the other hand, the increments shall not be contaminated by a second change since this can lead to situations where the change point can no longer be reliably localized by the signal. Furthermore, we need to impose restrictions on the unbalancedness in order to avoid problems in the asymptotic theory and in practice, as noted by Cho and Kirch (2021+) in Section 4.2. Indeed, we need the following assumptions – that are analogous to Assumption 3.1 – on the bandwidths for a change to be detectable:

Assumption 5.1. *We assume the bandwidths h_l, h_r with $h_l + h_r < T$ to be sufficiently balanced in the sense that there exists $C \geq 1$ such that*

$$\frac{\max\{h_l, h_r\}}{\min\{h_l, h_r\}} \leq C \tag{5.2}$$

for all $T > 0$. We assume that for ν_T as in Assumption (2.1) it holds that

$$\frac{\nu_T^2 \, T \, \log T}{h_l} \to 0. \tag{5.3}$$

Furthermore, the first and last change point are detectable and the i-th change point is isolated in the sense that

$$\begin{aligned} &h_l \leq c_1, \qquad c_{q_T} \leq T - h_r, \\ &h_l + h_r \leq \Delta_i, \qquad \text{where } \Delta_i = \min(c_{i+1} - c_i, c_i - c_{i-1}) \end{aligned}$$

for $i = 2, \ldots, q_T - 1$.
Additionally, the signal needs to be large enough to be detectable by these bandwidths, i. e.

$$\frac{\|\mathbf{d}_i\|^2 \, h_l}{\log\left(\frac{T}{h_l}\right)} \to \infty \tag{5.4}$$

for $i = 1, \ldots, q_T$.

Note that (5.3) and (5.4) automatically also hold for h_r due to (5.2).

5.1 Change point estimators

Analogously to (3.5) and (3.6), the MOSUM statistic $\mathbf{M}_t = \mathbf{m}_t + \mathbf{\Lambda}_t$ decomposes into a piecewise linear signal term $\mathbf{m}_t = \mathbf{m}_{t,h_l,h_r,T}$ and a centered noise term $\mathbf{\Lambda}_t = \mathbf{\Lambda}_{t,h_l,h_r,T}$ with

$$\sqrt{\frac{h_l+h_r}{h_l h_r}}\,\mathbf{m}_t = \sqrt{\frac{h_l+h_r}{h_l h_r}}\,\mathbf{m}_{t,h_l,h_r,T} \tag{5.5}$$
$$= \begin{cases} \frac{1}{h_l}(h_l - t + c_i)\,\mathbf{d}_i, & \text{for } c_i < t \le c_i + h_l, \\ 0, & \text{for } c_i + h_l < t \le c_{i+1} - h_r, \\ \frac{1}{h_r}(h_r + t - c_{i+1})\,\mathbf{d}_{i+1}, & \text{for } c_{i+1} - h_r < t \le c_{i+1}, \end{cases}$$

$$\sqrt{\frac{h_l+h_r}{h_l h_r}}\,\mathbf{\Lambda}_t = \sqrt{\frac{h_l+h_r}{h_l h_r}}\,\mathbf{\Lambda}_t(\widetilde{\mathbf{R}}) \tag{5.6}$$
$$= \begin{cases} \frac{1}{h_r}\left(\widetilde{\mathbf{R}}_{t+h_r}^{(c_{i+1})} - \widetilde{\mathbf{R}}_t^{(c_{i+1})}\right) & \\ -\frac{1}{h_l}\left(\widetilde{\mathbf{R}}_t^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} + \widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{t-h_l}^{(c_i)}\right), & \text{for } c_i < t \le c_i + h_l \\ \frac{1}{h_r}\left(\widetilde{\mathbf{R}}_{t+h_r}^{(c_{i+1})} - \widetilde{\mathbf{R}}_t^{(c_{i+1})}\right) - \frac{1}{h_l}\left(\widetilde{\mathbf{R}}_t^{(c_{i+1})} - \widetilde{\mathbf{R}}_{t-h_l}^{(c_{i+1})}\right), & \text{for } c_i + h_l < t \le c_{i+1} - h_r, \\ \frac{1}{h_r}\left(\widetilde{\mathbf{R}}_{t+h_r}^{(c_{i+2})} - \widetilde{\mathbf{R}}_{c_{i+1}}^{(c_{i+2})} + \widetilde{\mathbf{R}}_{c_{i+1}}^{(c_{i+1})} - \widetilde{\mathbf{R}}_t^{(c_{i+1})}\right) & \\ -\frac{1}{h_l}\left(\widetilde{\mathbf{R}}_t^{(c_{i+1})} - \widetilde{\mathbf{R}}_{t-h_l}^{(c_{i+1})}\right), & \text{for } c_{i+1} - h_r < t \le c_{i+1}, \end{cases}$$

Similar to (3.5), the signal term is a piecewise linear function with extrema at the change points and is 0 outside of intervals $[c_i - h_r, c_i + h_l]$. Since this is similar to our findings in Section 3, we will use an analogous procedure to the one proposed in Section 3: A point t^* is *significant* if

$$\mathbf{M}_{t^*}' \widehat{\mathbf{A}}_{t^*}^{-1} \mathbf{M}_{t^*} \ge \beta, \tag{5.7}$$

where $\beta = \beta_{h_l,h_r,T}$ is a suitable threshold and $\widehat{\mathbf{A}}_{t^*}$ is a symmetric positive definite matrix possibly depending on the data that fulfills

Assumption 5.2.

$$\sup_{h_l \le t \le T - h_r} \left\| \widehat{\mathbf{A}}_{t,T}^{-1} \right\| = O_P(1), \qquad \sup_{i=1,\dots,q_T} \sup_{c_i - h_l \le t \le c_i + h_r} \left\| \widehat{\mathbf{A}}_{t,T} \right\| = O_P(1).$$

We define local extrema analogously to (3.10) by having a tuning parameter $0 < \eta < 1$, where a point t^* is a local extremum if

$$t^* = \operatorname*{argmax}_{t^* - \eta h_l \le t \le t^* + \eta h_r} \|\mathbf{M}_t\|. \tag{5.8}$$

As in Section 3, the threshold β distinguishes between significant and spurious local extrema, and the set of all significant local extrema is the set of change point estimators. Its cardinality is an estimator for the number of change points.

Theorem 5.1. *Let the Assumptions 2.1, 5.1 and 5.2 hold. Denote by* $h_{\min} = \min\{h_l, h_r\}$, $h_{\max} = \max\{h_l, h_r\}$.

(a) For $c_i - h_r < t \leq c_i$ *let* $h = h_r$, *for* $c_i < t < c_i + h_l$ *let* $h = h_l$. *For the signal term* $\mathbf{m}_t$ *it holds that*

$$\mathbf{m}_l' \widehat{\mathbf{A}}_t^{-1} \mathbf{m}_t \geq \frac{1}{(C+1)\|\widehat{\mathbf{A}}_t\|} \frac{(h - |t - c_i|)^2}{h} \|\mathbf{d}_i\|^2 .$$

At other time points the signal term is equal to zero.

(b) For the noise term it holds for $q_T = 0$, *i. e. in the no-change situation*

(i) for linear bandwidths $h_l = \gamma_l T$, $h_r = \gamma_r T$ *with* $0 < \gamma_l + \gamma_r < 1$

$$\begin{aligned} &\sup_{\gamma_l T \leq t \leq T - \gamma_r T} \mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t \\ \xrightarrow{\mathcal{D}} &\frac{\gamma_l \gamma_r}{\gamma_l + \gamma_r} \sup_{\gamma_l \leq s \leq 1 - \gamma_r} \left(\left(\frac{1}{\gamma_r} (\mathbf{B}_{s+\gamma_r} - \mathbf{B}_s) - \frac{1}{\gamma_l} (\mathbf{B}_s - \mathbf{B}_{s-\gamma_l}) \right)' \right. \\ &\left. \left(\frac{1}{\gamma_r} (\mathbf{B}_{s+\gamma_t} - \mathbf{B}_s) - \frac{1}{\gamma_l} (\mathbf{B}_s - \mathbf{B}_{s-\gamma_l}) \right) \right) \end{aligned}$$

where $\mathbf{B}$ *denotes a multivariate standard Wiener process.*
In particular, the squared noise term is of order $O_P(1)$ *in this case.*

(ii) for sublinear bandwidths $h_l/T \to 0$, $h_r/T \to 0$ *but Assumption 5.1 fulfilled and* $h_{\min}/h_{\max} = D > 0$, *it holds that*

$$a\left(\frac{T}{h_{\min}}\right) \sup_{h_l \leq t \leq T - h_r} \sqrt{\mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t} - b\left(\frac{T}{h_{\min}}\right) \xrightarrow{\mathcal{D}} E,$$

where E *follows a Gumbel distribution with* $\mathrm{P}(E \leq x) = e^{-2e^{-x}}$ *and*

$$\begin{aligned} a(x) &= \sqrt{2 \log x} \\ b(x) &= 2 \log x + \frac{p}{2} \log \log x + \log \frac{D^2 + D + 1}{D + 1} - \log \Gamma\left(\frac{p}{2}\right). \end{aligned}$$

In particular, the above squared noise term is of order $O_P(\log(T/h_l))$ *in this case.*

The assertions remain true if an estimator for the covariance is used fulfilling (3.9) *uniformly over all* $h \leq t \leq T - h$.

(c) In the situation of multiple change points, it holds that

$$\sup_{h_l \le t \le T-h_r} \|\mathbf{\Lambda}_t\| = O_P(\sqrt{\log(T/h_l)}).$$

Proof. (a) Because $\widehat{\mathbf{A}}_t$ is symmetric and positive definite, the maximum eigenvalue is given by $\|\widehat{\mathbf{A}}_t\|$ by Corollary B.3. By Lemma B.1, it follows that the minimal eigenvalue of $\widehat{\mathbf{A}}_t^{-1}$ is given by $1/\|\widehat{\mathbf{A}}_t\|$. By straightforward calculations and (5.2) it holds that

$$h_l h_r \ge h \min\{h_l, h_r\}$$

and

$$h_l + h_r = \min\{h_l, h_r\} + \max\{h_l, h_r\} \le (1+C)\min\{h_l, h_r\}.$$

With Lemma B.4 and (5.5) it follows that

$$\begin{aligned}
&\mathbf{m}_t \widehat{\mathbf{A}}_t^{-1} \mathbf{m}_t \ge \frac{1}{\|\widehat{\mathbf{A}}_t\|} \|\mathbf{m}_t\|^2 = \frac{1}{\left\|\widehat{\mathbf{A}}_t\right\|} \frac{h_l h_r}{h_l + h_r} \frac{1}{h^2} (h - |t - c_i|)^2 \|\mathbf{d}_i\|^2 \\
&= \frac{1}{\left\|\widehat{\mathbf{A}}_t\right\|} \frac{1}{(h_l + h_r)} \frac{h_l h_r}{h^2} (h - |t - c_i|)^2 \|\mathbf{d}_i\|^2 \ge \frac{1}{\left\|\widehat{\mathbf{A}}_t\right\|} \frac{\min\{h_l, h_r\}}{(h_l + h_r)h} (h - |t - c_i|)^2 \|\mathbf{d}_i\|^2 \\
&\ge \frac{1}{\|\widehat{\mathbf{A}}_t\|} \frac{\min\{h_l, h_r\}}{(1+C)\min\{h_l, h_r\}} \frac{1}{h} (h - |t - c_i|)^2 \|\mathbf{d}_i\|^2.
\end{aligned}$$

(b) Denote by $\mathbf{\Lambda}_t(\mathbf{W}_t)$ the MOSUM statistics defined in (5.1) with $\{\mathbf{Z}_t\}$ there replaced by $\{\mathbf{W}_t\}$. By Assumption 5.1 it holds that $h_l \le C h_r$ and $h_r \le C h_l$. With the invariance principle from Assumption 2.1 it holds therefore that

$$\begin{aligned}
&\sup_{h_l \le t \le T-h_r} \|\mathbf{\Lambda}_t - \mathbf{\Lambda}_t(\mathbf{W}_t)\| \\
&= \sqrt{\frac{h_l h_r}{h_l + h_r}} \sup_{h_l \le t \le T-h_r} \left\| \left(\frac{1}{h_r} (\mathbf{Z}_{t+h_r} - \mathbf{Z}_t) - \frac{1}{h_l} (\mathbf{Z}_t - \mathbf{Z}_{t-h_l}) \right) \right. \\
&\quad \left. - \left(\frac{1}{h_r} (\mathbf{W}_{t+h_r} - \mathbf{W}_t) - \frac{1}{h_l} (\mathbf{W}_t - \mathbf{W}_{t-h_l}) \right) \right\| \\
&= \sqrt{\frac{h_l h_r}{h_l + h_r}} \sup_{h_l \le t \le T-h_r} \left\| \left(\frac{1}{h_r} \left(\widetilde{\mathbf{R}}_{t+h_r} - \widetilde{\mathbf{R}}_t \right) - \frac{1}{h_l} \left(\widetilde{\mathbf{R}}_t - \widetilde{\mathbf{R}}_{t-h_l} \right) \right) \right. \\
&\quad \left. - \left(\frac{1}{h_r} (\mathbf{W}_{t+h_r} - \mathbf{W}_t) - \frac{1}{h_l} (\mathbf{W}_t - \mathbf{W}_{t-h_l}) \right) \right\| \\
&\le \sqrt{\frac{h_l h_r}{h_l + h_r}} \left(2 \sup_{0 \le t \le T} \frac{1}{h_r} \left\| \widetilde{\mathbf{R}}_t - \mathbf{W}_t \right\| + 2 \sup_{0 \le t \le T} \frac{1}{h_l} \left\| \widetilde{\mathbf{R}}_t - \mathbf{W}_t \right\| \right) \\
&= 2 \sqrt{\frac{h_l + h_r}{h_l h_r}} \sup_{0 \le t \le T} \left\| \widetilde{\mathbf{R}}_t - \mathbf{W}_t \right\| \\
&\le 2 \sqrt{\frac{1+C}{h_l}} \sup_{0 \le t \le T} \left\| \widetilde{\mathbf{R}}_t - \mathbf{W}_t \right\| = O_P\left(\frac{T^{1/2} \nu_T}{\sqrt{h_l}} \right) = o_P\left(\sqrt{\log(T/h_l)}^{-1} \right). \qquad (5.9)
\end{aligned}$$

(i) Let $(\mathbf{B}_t)_{t\geq 0} = \left(\mathbf{\Sigma}_T^{-1/2}\mathbf{W}_t\right)_{t\geq 0}$ be a multivariate standard Wiener process. By the self-similarity of the Wiener process it follows with the transformation $s = t/T$ that

$$\begin{aligned}
&\sup_{\gamma_l T\leq t\leq T-\gamma_r T} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1}\mathbf{\Lambda}_t\left(\mathbf{W}_t\right) \\
&= \frac{\gamma_l\gamma_r T}{\gamma_l+\gamma_r} \sup_{\gamma_l T\leq t\leq T-\gamma_r T} \left(\left(\frac{1}{\gamma_r T}\left(\mathbf{B}_{t+\gamma_r T}-\mathbf{B}_t\right) - \frac{1}{\gamma_l T}\left(\mathbf{B}_t-\mathbf{B}_{t-\gamma_l T}\right)\right)'\right. \\
&\quad \left.\left(\frac{1}{\gamma_r T}\left(\mathbf{B}_{t+\gamma_r T}-\mathbf{B}_t\right) - \frac{1}{\gamma_l T}\left(\mathbf{B}_t-\mathbf{B}_{t-\gamma_l T}\right)\right)\right) \\
&\overset{\mathcal{D}}{=} \frac{\gamma_l\gamma_r}{\gamma_l+\gamma_r} \sup_{\gamma_l\leq s\leq 1-\gamma_r} \left(\left(\frac{1}{\gamma_r}\left(\mathbf{B}_{s+\gamma_r}-\mathbf{B}_s\right) - \frac{1}{\gamma_l}\left(\mathbf{B}_s-\mathbf{B}_{s-\gamma_l}\right)\right)'\right. \\
&\quad \left.\left(\frac{1}{\gamma_r}\left(\mathbf{B}_{s+\gamma_r}-\mathbf{B}_s\right) - \frac{1}{\gamma_l}\left(\mathbf{B}_s-\mathbf{B}_{s-\gamma_l}\right)\right)\right).
\end{aligned}$$

The assertion then follows from (5.9) and Slutzky's theorem.

(ii) Let $(\mathbf{B}_t)_{t\geq 0} = \left((B_{t,1},\ldots,B_{t,p})'\right)_{t\geq 0}$ be as above. If $h_l = h_{\min}$, it follows similar to (i) by the self-similarity of the multivariate Wiener process with the transformation $s = t/h_l - 1$ that

$$\begin{aligned}
&\sup_{h_l\leq t\leq T-h_r} \sqrt{\mathbf{\Lambda}_t(\mathbf{W}_t)'\mathbf{\Sigma}_T^{-1}\mathbf{\Lambda}_t(\mathbf{W}_t)} = \sup_{h_l\leq t\leq T-h_r} \left\|\mathbf{\Sigma}_T^{-1/2}\mathbf{\Lambda}_t(\mathbf{W}_t)\right\| \\
&= \sup_{h_l\leq t\leq T-h_r} \left\|\sqrt{\frac{h_l h_r}{h_l+h_r}}\left(\frac{1}{h_r}\left(\mathbf{B}_{t+h_r}-\mathbf{B}_t\right) - \frac{1}{h_l}\left(\mathbf{B}_t-\mathbf{B}_{t-h_l}\right)\right)\right\| \\
&= \sup_{h_l\leq t\leq T-\frac{h_l}{D}} \left\|\sqrt{\frac{h_l}{D+1}}\left(\frac{D}{h_l}\left(\mathbf{B}_{t+\frac{h_l}{D}}-\mathbf{B}_t\right) - \frac{1}{h_l}\left(\mathbf{B}_t-\mathbf{B}_{t-h_l}\right)\right)\right\| \\
&= \sup_{0\leq s\leq \frac{T}{h_l}-\frac{1}{D}-1} \left\|\frac{1}{\sqrt{D+1}}\left(\frac{D}{\sqrt{h_l}}\left(\mathbf{B}_{h_l\left(s+1+\frac{1}{D}\right)}-\mathbf{B}_{h_l(s+1)}\right) - \frac{1}{\sqrt{h_l}}\left(\mathbf{B}_{h_l(s+1)}-\mathbf{B}_{h_l s}\right)\right)\right\| \\
&\overset{\mathcal{D}}{=} \sup_{0\leq s\leq \frac{T}{h_l}-\frac{1}{D}-1} \left\|\frac{1}{\sqrt{D+1}}\left(D\left(\mathbf{B}_{s+1+\frac{1}{D}}-\mathbf{B}_{s+1}\right) - \left(\mathbf{B}_{s+1}-\mathbf{B}_s\right)\right)\right\| \\
&= \sup_{0\leq s\leq \frac{T}{h_{\min}}-\frac{1}{D}-1} \left\|\frac{1}{\sqrt{D+1}}\left(D\left(\mathbf{B}_{s+1+\frac{1}{D}}-\mathbf{B}_{s+1}\right) - \left(\mathbf{B}_{s+1}-\mathbf{B}_s\right)\right)\right\|.
\end{aligned}$$

If $h_r = h_{\min}$, it holds analogously with the time-reversal property of the multivariate Wiener process that

$$\begin{aligned}
&\sup_{h_l\leq t\leq T-h_r} \sqrt{\mathbf{\Lambda}_t(\mathbf{W}_t)'\mathbf{\Sigma}_T^{-1}\mathbf{\Lambda}_t(\mathbf{W}_t)} \\
&\overset{\mathcal{D}}{=} \sup_{0\leq s\leq \frac{T}{h_{\min}}-\frac{1}{D}-1} \left\|\frac{1}{\sqrt{D+1}}\left(\left(\mathbf{B}_{s+1+\frac{1}{D}}-\mathbf{B}_{s+\frac{1}{D}}\right) - D\left(\mathbf{B}_{s+\frac{1}{D}}-\mathbf{B}_s\right)\right)\right\|
\end{aligned}$$

$$\stackrel{\mathcal{D}}{=} \sup_{0\le s\le \frac{T}{h_{\min}}-\frac{1}{D}-1} \left\| \frac{1}{\sqrt{D+1}} \left(D\left(\mathbf{B}_{s+1+\frac{1}{D}} - \mathbf{B}_{s+1}\right) - \left(\mathbf{B}_{s+1} - \mathbf{B}_s\right)\right)\right\|.$$

W.l.o.g. let $a \ge 0$. The process

$$\left(\mathbf{\Lambda}_s(\mathbf{B}_s)\right)_{s\ge 0} = \left(\frac{1}{\sqrt{D+1}} \left(D\left(\mathbf{B}_{s+1+\frac{1}{D}} - \mathbf{B}_{s+1}\right) - \left(\mathbf{B}_{s+1} - \mathbf{B}_s\right)\right)\right)_{s\ge 0}$$

is a stationary p-dimensional Gaussian process. The componentwise covariance functions are given by

$$\begin{aligned}
&(D+1)\operatorname{Cov}\left[\mathbf{\Lambda}_{a,i}\left(\mathbf{B}_a\right), \mathbf{\Lambda}_{0,i}\left(\mathbf{B}_0\right)\right] \\
&= D^2 \operatorname{Cov}\left[B_{a+1+\frac{1}{D},i} - B_{a+1,i}, B_{1+\frac{1}{D},i} - B_{1,i}\right] - D\operatorname{Cov}\left[B_{a+1+\frac{1}{D},i} - B_{a+1,i}, B_{1,i} - B_{0,i}\right] \\
&\quad - D\operatorname{Cov}\left[B_{a+1,i} - B_{a,i}, B_{1+\frac{1}{D},i} - B_{1,i}\right] \\
&\quad + \operatorname{Cov}\left[B_{a+1,i} - B_{a,i}, B_{1,i} - B_{0,i}\right] \\
&= \begin{cases} D - D^2 a - Da + 1 - a, & \text{for } 0 \le a < 1 \\ -Da, & \text{for } 1 \le a < \frac{1}{D} \\ -D\left(1 + \frac{1}{D} - a\right), & \text{for } \frac{1}{D} \le a < 1 + \frac{1}{D} \\ 0, & \text{for } a \ge 1 + \frac{1}{D} \end{cases}
\end{aligned}$$

with analogous results for $a < 0$. Therefore,

$$\begin{aligned}
&\operatorname{Cov}\left[\mathbf{\Lambda}_{s+a,i}\left(\mathbf{B}_{s+a}\right), \mathbf{\Lambda}_{s,i}\left(\mathbf{B}_s\right)\right] \\
&= \begin{cases} 1 - \frac{D^2+D+1}{D+1}|a|, & \text{for } 0 \le |a| < 1 \\ \frac{-D|a|}{D+1}, & \text{for } 1 \le |a| < \frac{1}{D} \\ -\frac{D}{D+1}\left(1 + \frac{1}{D} - |a|\right), & \text{for } \frac{1}{D} \le |a| < 1 + \frac{1}{D} \\ 0, & \text{for } |a| \ge 1 + \frac{1}{D}. \end{cases}
\end{aligned}$$

Therefore it follows by Lemma 3.1 and Remark 3.1 of Steinebach and Eastwood (1996) (see also Lemma A.1) with $\alpha = 1$, $C = \frac{D^2+D+1}{D+1}$ that

$$a\left(\frac{T}{h_{\min}}\right) \sup_{0\le s\le \frac{T}{h_{\min}}} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| - b\left(\frac{T}{h_{\min}}\right) \xrightarrow{\mathcal{D}} E.$$

Since by the triangle inequality

$$\begin{aligned}
&\sup_{\frac{T}{h_{\min}}-1-\frac{1}{D}\le s\le \frac{T}{h_{\min}}} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| \stackrel{\mathcal{D}}{=} \sup_{0\le s\le 1+\frac{1}{D}} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| \\
&\le \frac{2D+2}{\sqrt{D+1}} \sup_{0\le s\le 2+\frac{2}{D}} \|\mathbf{B}_s\| = O_P(1) = \; o_P\left(\frac{b\left(T/h_{\min}\right)}{a\left(T/h_{\min}\right)}\right),
\end{aligned}$$

it follows by Lemma A.2 that

$$\begin{aligned} & a\left(\frac{T}{h_{\min}}\right) \sup_{h_l \leq t \leq T-h_r} \sqrt{\mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)} - b\left(\frac{T}{h_{\min}}\right) \\ \stackrel{\mathcal{D}}{=} & a\left(\frac{T}{h_{\min}}\right) \sup_{0 \leq s \leq \frac{T}{h_{\min}}-1-\frac{1}{D}} \left\|\mathbf{\Lambda}_s\left(\mathbf{B}_s\right)\right\| - b\left(\frac{T}{h_{\min}}\right) \xrightarrow{\mathcal{D}} E, \end{aligned}$$

as well. Since by (5.9),

$$\left| a\left(\frac{T}{h_{\min}}\right) \sup_{h_l \leq t \leq T-h_r} \sqrt{\mathbf{\Lambda}_t' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t} - a\left(\frac{T}{h_{\min}}\right) \sup_{h_l \leq t \leq T-h_r} \sqrt{\mathbf{\Lambda}_t\left(\mathbf{W}_t\right)' \mathbf{\Sigma}_T^{-1} \mathbf{\Lambda}_t\left(\mathbf{W}_t\right)} \right| = o_P(1),$$

the assertion follows by Slutzky's theorem.
The proof that the assertion remains true if we replace $\mathbf{\Sigma}_T$ by an estimator $\widehat{\mathbf{\Sigma}}_T$ fulfilling Assumption 5.2 is completely analogous to the corresponding statement in Theorem 3.2 (b).

(c) It holds that

$$\sup_{h_l \leq t \leq T-h_l} \left\|\mathbf{\Lambda}_t\right\| = \max\left\{ \max_{i=1,\ldots,q_T} \sup_{c_i - h_r \leq t \leq c_i + h_l} \left\|\mathbf{\Lambda}_t\right\|, \max_{i=0,\ldots,q_T} \sup_{c_i + h_l < t < c_{i+1} - h_r} \left\|\mathbf{\Lambda}_t\right\| \right\}.$$

By (b) it holds that

$$\begin{aligned} & \max_{i=0,\ldots,q_T} \sup_{c_i + h_l < t < c_{i+1} - h_r} \left\|\mathbf{\Lambda}_t\right\| \\ = & \max_{i=0,\ldots,q_T} \sup_{c_i + h_l < t < c_{i+1} - h_r} \sqrt{\frac{h_l h_r}{h_l + h_r}} \left\| \frac{1}{h_r}\left(\widetilde{\mathbf{R}}_{t+h_r}^{(c_{i+1})} - \widetilde{\mathbf{R}}_t^{(c_{i+1})}\right) - \frac{1}{h_l}\left(\widetilde{\mathbf{R}}_t^{(c_{i+1})} - \widetilde{\mathbf{R}}_{t-h_l}^{(c_{i+1})}\right) \right\| \\ \leq & \max_{i=1,\ldots,P} \sup_{h_l \leq t \leq T-h_r} \sqrt{\frac{h_l h_r}{h_l + h_r}} \left\| \frac{1}{h_r}\left(\widetilde{\mathbf{R}}_{t+h_r}^{(i)} - \widetilde{\mathbf{R}}_t^{(i)}\right) - \frac{1}{h_l}\left(\widetilde{\mathbf{R}}_t^{(i)} - \widetilde{\mathbf{R}}_{t-h_l}^{(i)}\right) \right\| \\ = & O_P\left(\sqrt{\log(T/h_l)}\right). \end{aligned} \tag{5.10}$$

Furthermore, on noting that by Assumption 5.1, h_l, h_r and $h_l + h_r$ are of the same order, we obtain by Proposition A.4 (b) and the triangle inequality analogously to the corresponding proof in Theorem 3.2 that

$$\begin{aligned} & \max_{i=1,\ldots,q_T} \sup_{c_i \leq t \leq c_i + h_l} \left\|\mathbf{\Lambda}_t\right\| \\ = & \max_{i=1,\ldots,q_T} \sqrt{\frac{h_l h_r}{h_l + h_r}} \\ & \sup_{c_i \leq t \leq c_i + h_l} \left\| \frac{1}{h_r}\left(\widetilde{\mathbf{R}}_{t+h_r}^{(c_{i+1})} - \widetilde{\mathbf{R}}_t^{(c_{i+1})}\right) - \frac{1}{h_l}\left(\widetilde{\mathbf{R}}_t^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} + \widetilde{\mathbf{R}}_{c_i}^{(c_i)} - \widetilde{\mathbf{R}}_{t-h_l}^{(c_i)}\right) \right\| \end{aligned}$$

$$
\begin{aligned}
&\leq 2 \max_{i=1,\ldots,q_T} \sup_{0\leq s\leq h_l+h_r} \sqrt{\frac{h_l h_r}{h_l+h_r}} \frac{1}{h_r} \left\| \widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} \right\| \\
&\quad + \max_{i=1,\ldots,q_T} \sup_{0\leq s\leq h_l} \sqrt{\frac{h_l h_r}{h_l+h_r}} \frac{1}{h_l} \left\| \widetilde{\mathbf{R}}_{c_i+s}^{(c_{i+1})} - \widetilde{\mathbf{R}}_{c_i}^{(c_{i+1})} \right\| \\
&\quad + \max_{i=1,\ldots,q_T} \sup_{0\leq s\leq h_l} \sqrt{\frac{h_l h_r}{h_l+h_r}} \frac{1}{h_l} \left\| \widetilde{\mathbf{R}}_{c_i-s}^{(c_i)} - \widetilde{\mathbf{R}}_{c_i}^{(c_i)} \right\| \\
&= O_P\left(\sqrt{\log(2q_T)}\right) = O_P\left(\sqrt{\log(T/h_l)}\right) \qquad (5.11)
\end{aligned}
$$

since by Assumption 3.1 $q_T \leq T/(h_l+h_r)$ and $T^{1/2}\nu_T/\sqrt{h_l} = o(1)$.
Similarly we obtain that

$$
\max_{i=1,\ldots,q_T} \sup_{c_i-h_r\leq t\leq c_i} \|\mathbf{\Lambda}_t\| = O_P\left(\sqrt{\log(T/h_l)}\right),
$$

which in combination with (5.10) and (5.11) shows the assertion. □

Assumption 5.3. *The threshold fulfills:*

$$
\frac{\beta_{h_l,h_r,T}}{h_{l,T} \min\limits_{i=1,\ldots,q_T} \|\mathbf{d}_i\|^2} \to 0, \qquad \frac{\log \frac{T}{h_{l,T}}}{\beta_{h_l,h_r,T}} \to 0 \qquad (T\to\infty).
$$

Theorem 5.2. *Let Assumptions 2.1, 5.1 – 5.3 hold and let $h_l/h_r = D$ for some $0 < D < \infty$. Let $0 < \hat{c}_1 < \ldots < \hat{c}_{\hat{q}_T}$ be the change point estimators of type* (5.8)*. Then for any $\tau > 0$ it holds*

$$
\lim_{T\to\infty} \mathrm{P}\left(\max_{i=1,\ldots,\min(\hat{q}_T,q_T)} |\hat{c}_i - c_i| \leq \tau \min\left\{h_l, h_r\right\}, \hat{q}_T = q_T \right) = 1.
$$

Proof. Assume that $h_l \leq h_r$ as the proof in the other case is analogous. The proof is analogous to the proof of Theorem 4.1. First, define for $0 < \tau < 1$ the following set

$$
S_T = S_T^{(1)} \cap S_T^{(2)} \cap \bigcap_{i=1}^{q_T} \left(S_T^{(3)}(i,\tau) \cap S_T^{(4)}(i,\tau) \right),
$$

where

$$
\begin{aligned}
S_T^{(1)} &= \left\{ \max_{i=0,\ldots,q_T} \sup_{c_i+h_l<t<c_{i+1}-h_r} \mathbf{M}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{M}_t < \beta \right\}, \\
S_T^{(2)} &= \left\{ \min_{i=1,\ldots,q_T} \mathbf{M}_{c_i}' \widehat{\mathbf{A}}_{c_i}^{-1} \mathbf{M}_{c_i} \geq \beta \right\}, \\
S_T^{(3)}(i,\tau) &= \bigcap_{k=1}^{\lceil \frac{1}{D\tau} \rceil - 1} \left\{ \sup_{c_i-h_r\leq t\leq c_i-k\tau h_l} \|\mathbf{M}_t\| < \|\mathbf{M}_{c_i-(k-1)\tau h_l}\| \right\}, \\
S_T^{(4)}(i,\tau) &= \bigcap_{k=1}^{\lceil \frac{1}{\tau} \rceil - 1} \left\{ \sup_{c_i+k\tau h_l\leq t\leq c_i+h_l} \|\mathbf{M}_t\| < \|\mathbf{M}_{c_i+(k-1)\tau h_l\|}\| \right\}.
\end{aligned}
$$

On $S_T^{(1)}$ there are asymptotically no significant points outside of $[c_i - h_r, c_i + h_l]$. On $S_T^{(2)}$ there is at least one significant time point for each change point. On $S_T^{(3)}(i,\tau) \cap S_T^{(4)}(i,\tau)$ with $\tau < \eta/2$, there are no local extrema within $[c_i - h_r, c_i + h_l]$ that are outside the interval $(c_i - \tau h_l, c_i + \tau h_l)$. Additionally, on $S_T^{(2)} \cap S_T^{(3)}(i,\tau) \cap S_T^{(4)}(i,\tau)$ the global extremum within that interval will be the only significant local extremum within $[c_i - h_r, c_i + h_l]$ such that

$$\left\{ \max_{i=1,\ldots,\min(\hat{q}_T, q_T)} |\hat{c}_i - c_i| \leq \tau \min\left\{h_l, h_r\right\}, \hat{q}_T = q_T \right\} \supset S_T.$$

We will show that S_T is an asymptotic one set.

Analogously to the proof of Theorem 4.1 it holds by Assumption 5.2 and Theorem 5.1 (c) that

$$\max_{i=0,\ldots,q_T} \sup_{c_i + h_l < t < c_{i+1} - h_r} \mathbf{M}_t' \widehat{\mathbf{A}}_t^{-1} \mathbf{M}_t \leq \sup_{h_l \leq t \leq T - h_r} \left\| \widehat{\mathbf{A}}_t^{-1} \right\| \max_{i=0,\ldots,q_T} \sup_{c_i + h_l < t < c_{i+1} - h_r} \left\| \mathbf{\Lambda}_t \right\|^2$$
$$= O_P\left(\log T / h_l\right),$$

Therefore, by Assumption 5.3, $S_T^{(1)}$ is an asymptotic one set.

For $S_T^{(2)}$ we obtain analogously to Theorem 5.1 that

$$\min_{i=1,\ldots,q_T} \mathbf{M}_{c_i}' \widehat{\mathbf{A}}_{c_i}^{-1} \mathbf{M}_{c_i} \geq \min_{i=1,\ldots,q_T} \frac{1}{\left\| \widehat{\mathbf{A}}_{c_i} \right\|} \min_{i=1,\ldots,q_T} \left\| \mathbf{M}_{c_i} \right\|^2.$$

By Assumption 3.2, for each $\varepsilon > 0$, there exists $C_\varepsilon >$ such that

$$\limsup_{T \to \infty} \mathrm{P}\left(\min_{i=1,\ldots,q_T} \left\| \widehat{\mathbf{A}}_{c_i} \right\|^{-1} < C_\varepsilon \right) < \varepsilon.$$

Furthermore, we obtain analogously to Theorem 5.1 by Assumption 5.1 and $h_r = h_l / D$ that

$$\begin{aligned}
&\min_{i=1,\ldots,q_T} \left\| \mathbf{M}_{c_i} \right\|^2 = \min_{i=1,\ldots,q_T} \left\| \mathbf{m}_{c_i} + \mathbf{\Lambda}_{c_i} \right\|^2 \\
\geq\ & \min_{i=1,\ldots,q_T} \left\| \mathbf{m}_{c_i} \right\| \left(\min_{i=1,\ldots,q_T} \left\| \mathbf{m}_{c_i} \right\| - 2 \max_{i=1,\ldots,q_T} \left\| \mathbf{\Lambda}_{c_i} \right\| \right) \\
=\ & \sqrt{\frac{h_l h_r}{h_l + h_r}} \min_{i=1,\ldots,q_T} \left\| \mathbf{d}_i \right\| \left(\sqrt{\frac{h_l h_r}{h_l + h_r}} \min_{i=1,\ldots,q_T} \left\| \mathbf{d}_i \right\| - O_P\left(\sqrt{\log \frac{T}{h_l}} \right) \right) \\
=\ & \frac{h_l}{(D+1)} \min_{i=1,\ldots,q_T} \left\| \mathbf{d}_i \right\|^2 + o_P\left(h_l \min_{i=1,\ldots,q_T} \left\| \mathbf{d}_i \right\|^2 \right),
\end{aligned}$$

which in combination with Assumption 5.3 shows that $S_T^{(2)}$ is an asymptotic one set.

For arbitrary $\tau > 0$ and $S_T^{(3)}(i,\tau)$ it holds by Theorem 5.3 uniformly that

$$
\begin{aligned}
&\min_{i=1,\ldots,q_T} \left(\left\| \mathbf{M}_{c_i-(k-1)\tau h_l} \right\| - \sup_{c_i-h_r \leq t \leq c_i-k\tau h_l} \|\mathbf{M}_t\| \right) \\
&= \min_{i=1,\ldots,q_T} \left(\left\| \mathbf{m}_{c_i-(k-1)\tau h_l} \right\| - \|\mathbf{m}_{c_i-k\tau h_l}\| \right) + O_P\left(\sqrt{\log \frac{T}{h_l}} \right) \\
&= \tau \sqrt{\frac{h_l h_r}{h_l + h_r}} \min_{i=1,\ldots,q_T} \|\mathbf{d}_i\| + O_P\left(\sqrt{\log \frac{T}{h_l}} \right) \\
&= \tau \sqrt{\frac{h_l}{D+1}} \min_{i=1,\ldots,q_T} \|\mathbf{d}_i\| + O_P\left(\sqrt{\log \frac{T}{h_l}} \right),
\end{aligned}
$$

thus showing that $\cap_{i=1}^{q_T} S_T^{(3)}(i,\tau)$ is an asymptotic one set. It follows analogously that $\cap_{i=1}^{q_T} S_T^{(4)}(i,\tau)$ is an asymptotic one set. □

6 Simulation study

We want to analyze the performance of our procedure by means of a simulation study with a focus on detection rates for change points. In particular, we want to compare the detection rates for different choices for the matrix $\widehat{\mathbf{A}}_t$ as in (3.7) and also analyze the performance of our procedure under violations of the model assumptions in Assumption 3.1.

We have mentioned in the Preface that our initial goal was the estimation of change points in (multivariate) renewal processes. As noted in Section 2.2.2, multivariate renewal processes with the same intensities across all components fulfill the invariance principle from Assumption 2.1. Therefore, we analyze the performance of our procedure on this particular type of process. More precisely, we analyze three-dimensional renewal processes with $T = 1600$, where the increments of the inter-event times for each component are $\Gamma-$distributed with intensity changes at 250, 500, 900 and 1150, where the expected time μ between events is given by 1.3, 0.9, 0.6, 0.8 and 1.3. We use a bandwidth of $h = 120$ and the parameter $\eta = 0.75$. Smaller values of η as suggested by Meier, Cho, and Kirch (2021) for partial sum processes tend to produce duplicate change point estimators by having two or more significant local maxima for each change point if the variance is too large (as can be seen in Table 6.1 below), while larger values of η lead to slightly worse detection rates. For a single-bandwidth MOSUM procedure as suggested here, this should be avoided but can be relaxed if a post-processing procedure is applied as e. g. by Cho and Kirch (2021+) for partial sum processes.

In contrast to partial sum processes, it is natural for renewal processes that the variances change with the intensity. Therefore we consider the following three scenarios: (i) standard deviations of constant value 0.7 (referred to as `constvar`), (ii) standard deviations being $5/6\mu$ (referred to as `smallvar`) and (iii) multivariate Poisson processes (referred to as `Poisson`).

We consider both the case of independence and dependence between the three components. In the latter case, we generate for each regime i an independent (in

time) sequence of Γ-distributed inter-event times $Y_j = Y_j^{(i)}$, $j = 1, 2, 3$, with a correlation of 0.2 (for all pairs) as $Y_j = X_j + X_4$, where (with a slight abuse of notation) $X_j \sim \Gamma(s, \lambda)$ for $j = 1, 2, 3$ and $X_4 \sim \Gamma(s^*, \lambda)$ for appropriate values of s, s^* and λ resulting in the above intensities and standard deviations for each regime (see Remark A.3 (a) for details on the parameters).

In the simulations, we use a threshold as in Remark 3.1 with $\alpha_T = 0.05$. By Section 2.2.2 and (3.8) it holds that $\mathbf{\Sigma}_t = \mathrm{Cov}\left[(Y_1, Y_2, Y_3)'\right] / \mathbb{E}\left[Y_1\right]^3$ while we use the following choices for the matrix $\widehat{\mathbf{A}}_t$ as in (3.7): (A) Diagonal matrix with locally estimated variances $\widehat{\mathbf{\Sigma}}_t(j, j)$ on the diagonal, $j = 1, 2, 3$, (B) with the true variances $\mathbf{\Sigma}_t(j, j)$ on the diagonal and (C) in case of dependent components (non-diagonal) true covariance matrix $\mathbf{\Sigma}_t$. While only (A) is of relevance in applications, this allows us to understand the influence of estimating the variance on the procedure. For dependent data, the distinction between (B) and (C) is important for applications, because a good enough estimator (resulting in a reasonable estimator for the inverse) is often not available for the full covariance matrix as in (C) for moderately high or high dimensions, while it is much less problematic to estimate (B). In (A) the variances at location t are estimated as

$$\widehat{\mathbf{\Sigma}}_t(j, j) = \min\left\{\frac{\widehat{\sigma}^2_{j,-}(t)}{\widehat{\mu}^3_{j,-}(t)}, \frac{\widehat{\sigma}^2_{j,+}(t)}{\widehat{\mu}^3_{j,+}(t)}\right\}, \tag{6.1}$$

where $\widehat{\sigma}^2_{j,\pm}(t)$ and $\widehat{\mu}_{j,\pm}(t)$ are the sample variance and sample mean, respectively, based on the inter-event times of the j-th component within the windows $(t - h, t]$ for $'-'$ respectively $(t, t + h]$ for $'+'$. The first and last inter-event times that have been censored by the window are not included. Using the minimum of the left and right local estimators takes into account that the variance can (and typically will) change with the intensity which has already been discussed by Meier, Cho, and Kirch (2021) in the context of partial sum processes. The results of the simulation study can be found in Table 6.2, where we consider a change point to be detected if there is an estimator in the interval $[c_i - h, c_i + h]$. Additional significant local maxima in such an interval are called *duplicate* change point estimators, while additional significant local maxima outside any of these intervals are called *spurious*.

	dupl., `c.var` $\eta = 0.4$	dupl., `c.var` $\eta = 0.75$	dupl., `s.var` $\eta = 0.4$	dupl., `s.var` $\eta = 0.75$	dupl., `pois` $\eta = 0.4$	dupl., `pois` $\eta = 0.75$
indep., type (A)	0.0798	0.0024	0.1046	0.0033	0.1559	0.0077
indep., type (B)	0.0484	0.0007	0.0496	0.0004	0.0526	0.0014
dep., type (A)	0.1057	0.0027	0.1505	0.0055	0.1880	0.0091
dep., type (B)	0.0779	0.0008	0.0814	0.0018	0.0755	0.0020
dep., type (C)	0.0512	0.0019	0.0494	0.0017	0.0349	0.0013

Table 6.1: Comparison of the average number of duplicate estimators for $\eta = 0.4$ and $\eta = 0.75$.

(a) `constvar`: Constant standard deviation of 0.7, $\eta = 0.75$.

Change point at	250	500	900	1150	spurious	duplicate
independent, type (A)	1	0.9998	0.9434	1	0.0251	0.0024
independent, type (B)	0.9974	0.9789	0.6271	1	0.0035	0.0007
dependent, type (A)	0.9998	0.9991	0.9219	1	0.0344	0.0027
dependent, type (B)	0.9916	0.9610	0.6351	0.9997	0.0074	0.0008
dependent, type (C)	0.9522	0.8485	0.3670	0.9984	0.0055	0.0019

(b) `smallvar`: Standard deviation of 5/6 the expected time between events, $\eta = 0.75$.

Change point at	250	500	900	1150	spurious	duplicate
independent, type (A)	0.9831	1	0.9735	1	0.0313	0.0033
independent, type (B)	0.9368	1	0.9309	0.9999	0.0038	0.0004
dependent, type (A)	0.9711	0.9999	0.9556	0.9998	0.0386	0.0055
dependent, type (B)	0.9207	0.9986	0.9094	0.9987	0.0073	0.0018
dependent, type (C)	0.7494	0.9890	0.7210	0.9908	0.0052	0.0017

(c) `Poisson`: Standard deviation equal to the expected time between events, $\eta = 0.75$.

Change point at	250	500	900	1150	spurious	duplicate
independent, type (A)	0.9054	0.9971	0.8710	0.9983	0.0445	0.0077
independent, type (B)	0.7366	0.9852	0.7188	0.9885	0.0028	0.0014
dependent, type (A)	0.8818	0.9924	0.8418	0.9939	0.0528	0.0091
dependent, type (B)	0.7166	0.9764	0.6978	0.9761	0.0054	0.0020
dependent, type (C)	0.4602	0.8934	0.4289	0.9007	0.0048	0.0013

Table 6.2: Detection rates for each change point as well as the average number of spurious and duplicate estimators for different distributions of the inter-event times.

The procedure performs well throughout all simulations with high detection rate, few spurious and very few duplicate estimators. The results improve further for smaller variance, in which case the signal-to-noise ratio is better.

When the diagonal matrix with the estimated variance is being used, the detection power is larger in all cases than when the true variance is being used. In case of the changes at location 900 this is a substantial improvement, such that the use of this local variance estimator can help boost the signal significantly. This comes at the cost of having an increased but still reasonable amount of spurious and duplicate change point estimators.

This effect stems from using the minimum in (6.1), which was introduced to gain detection power if the variance changes with the intensity. Additionally, the use of the true (asymptotic) covariance matrix leads to worse results than only using the corresponding diagonal matrix, which is due to the fact that the theoretical signal term is smaller when using the true (asymptotic) covariance matrix in this example (compare Remark A.2). From a statistical perspective this is advantageous because

(a) `constvar`: Constant standard deviation of 0.7, $\eta = 0.75$, $h = 120$.

Change point at	250	500	900	1150	spurious	duplicate
dependent, type (A)	1	1	0.9764	1	0.0666	0.0068
dependent, type (B)	0.9997	0.9936	0.6515	1	0.0037	0.0007
dependent, type (C)	1	0.9999	0.9539	1	0.0049	0.0004

(b) `smallvar`: Standard deviation of 5/6 the expected time between events, $\eta = 0.75$.

Change point at	250	500	900	1150	spurious	duplicate
dependent, type (A)	0.9955	1	0.9905	1	0.0772	0.0077
dependent, type (B)	0.9741	1	0.9628	1	0.0042	0.0006
dependent, type (C)	0.9995	1	0.9989	1	0.0045	0.0003

(c) `Poisson`: Standard deviation equal to the expected time between events, $\eta = 0.75$.

Change point at	250	500	900	1150	spurious	duplicate
dependent, type (A)	0.9535	0.9994	0.9192	0.9998	0.1187	0.0203
dependent, type (B)	0.7781	0.9964	0.7433	0.9973	0.0054	0.0012
dependent, type (C)	0.9824	1	0.9776	1	0.0059	0.0008

Table 6.3: Detection rates for each change point as well as the average number of spurious and duplicate estimators for different distributions of the inter-event times.

the local estimation of the inverse of a covariance matrix in moderately large or large dimensions is a very hard problem leading to a loss in precision, while the diagonal elements are far less difficult to estimate consistently.

However, in other examples, using the full covariance matrix can also lead to better behavior, namely if the theoretical signal term is bigger in that case (compare Remark A.2). The results for one such example can be found in Table 6.3. Here, the inter-event times are (with a slight abuse of notation) $Y_j = X_j + \sum_{1 \leq k < j} X_{k,j} - \sum_{j < k \leq 3} X_{j,k}$ $j = 1, 2, 3$ where the $X_j = X_j^{(i)}$ are sequences of independent in time $\Gamma(s, \lambda)$-distributed random variables. The $X_{j,k} = X_{j,k}^{(i)}$ are sequences of independent in time $\mathcal{N}(0, s_1^2)$-distributed random variables with s, λ, s_1 appropriately chosen such that the distributions of the inter-event times have the above average intensities, standard deviations and that the correlation between each dimension is -0.2 (see Remark A.3 (b) for details on the parameters).

Furthermore, we illustrate the performance of our procedure in the case that Assumption 3.1 is violated, and in particular (3.2) that the bandwidth is less than half the distance to the next change point: We analyze three-dimensional renewal processes with $T = 1600$, where the increments of the inter-event times for each component are Γ-distributed with intensity changes at 250, 500 and 600, where the expected time μ between events is given by 1.3, 0.9, 0.6 and 0.8. We use bandwidths of $h = 60, 90, 120$ and the parameter $\eta = 0.75$. While for the change point at 250 all bandwidths fulfill the assumption, this is true for neither of the other two change

(a) `constvar`: Constant standard deviation of 0.7, $\eta = 0.75$.

CP at	250	500	600	spurious	duplicate	Dist. 500	Dist. 600
h=60	0.9847	0.9690	0.6528	0.2436	0.0073	5.52	8.47
h=90	0.9996	0.9970	0.7957	0.1044	0.0025	4.92	8.16
h=120	1	0.9984	0.6368	0.0561	0.0002	9.94	21.43

(b) `smallvar`: Standard deviation of 5/6 the expected time between events, $\eta = 0.75$.

CP at	250	500	600	spurious	duplicate	Dist. 500	Dist. 600
h=60	0.7534	0.9592	0.6476	0.2689	0.0100	5.28	7.55
h=90	0.9273	0.9978	0.8461	0.1025	0.0054	4.89	7.28
h=120	0.9846	0.9987	0.7237	0.0546	0.0020	10.01	19.66

(c) `Poisson`-distributed inter-event times, $\eta = 0.75$.

CP at	250	500	600	spurious	duplicate	Dist. 500	Dist. 600
h=60	0.5904	0.8494	0.4698	0.3807	0.0129	7.05	9.36
h=90	0.7838	0.9702	0.6696	0.1457	0.0077	6.65	9.29
h=120	0.9070	0.9807	0.5798	0.0724	0.0046	11.43	20.76

Table 6.4: Detection rates for each change point, average number of spurious and duplicate estimators for different distributions of the inter-event times as well as the average distances of the change point estimators closest to the true change points in the intervals $[c_i - \min\{h, (c_i - c_{i-1})/2\}, c_i + \min\{h, (c_{i+1} - c_i)/2\}]$ for $c_i = 500, 600$, respectively.

points with the bandwidth $h = 120$ being larger than the distance between these two points. We use the same three scenarios for the standard deviations of the inter-event times as above. We assume independence between the components and for the matrix $\widehat{\mathbf{A}}_t$, we consider only choice (A) – a matrix with locally estimated variances $\widehat{\mathbf{\Sigma}}_t(j,j)$ on the diagonal, $j = 1, 2, 3$. The results of the simulation study can be found in Table 6.4, where we consider a change point to be detected if there was an estimator in the interval $[c_i - \min\{h, (c_i - c_{i-1})/2\}, c_i + \min\{h, (c_{i+1} - c_i)/2\}]$.

Clearly, the procedure is performing well even when the model assumptions are mildly violated, as for $h = 60$ and $h = 90$ and the last two change points. For $h = 120$, the detection rates for the change point at 500 slightly increases but the average distance of the estimator to the true change point becomes much larger. For the change at 600, additionally the detection rate clearly decreases. On the other hand, as long as Assumption 3.1 holds (as for the first change point) or is only mildly violated (as for the last two change points and the two smaller bandwidths), the detection rate increases with larger bandwidth while at the time the average distances between the estimator and the corresponding true change point decreases. This is due to an increased signal-to-noise ratio due to the larger bandwidths (corresponding to a larger sample size in classical two-sample testing).

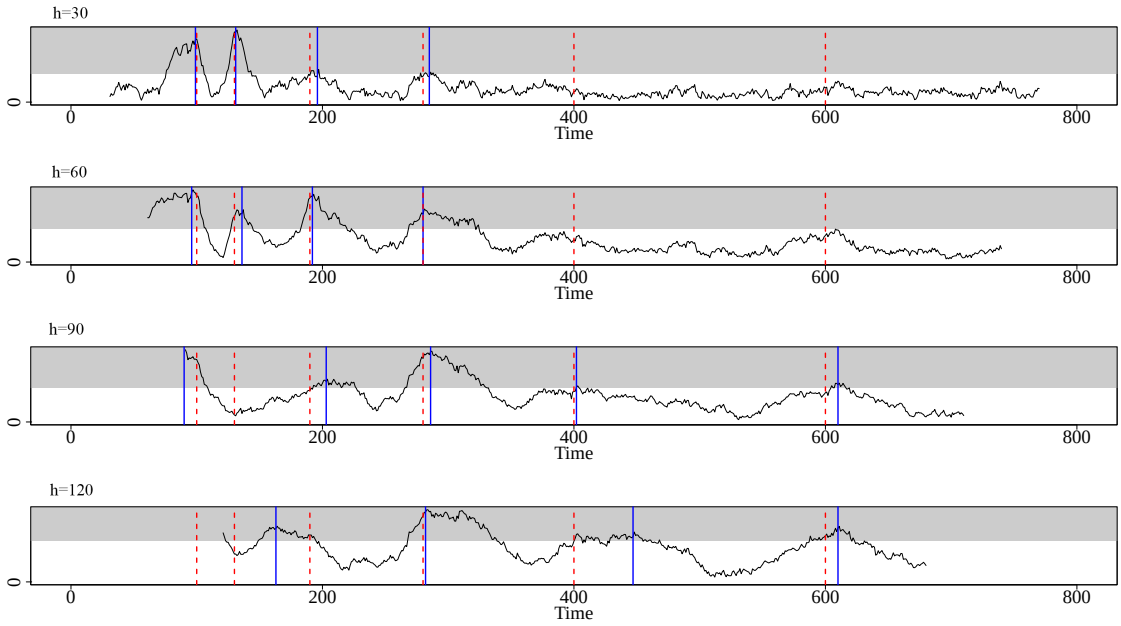

Figure 6.1: MOSUM statistics with bandwidths of $h = 30, 60, 90, 120$ (top to bottom) for a three-dimensional renewal process with *multiscale* changes with increasing distance between change points in combination with decreasing magnitude of the changes in intensity. The dashed vertical lines indicate the location of the true changes, while the solid lines indicate the change point estimators.
In this multiscale situation no single bandwidth can detect all changes: The changes to the left are well estimated by smaller bandwidth, the ones in the middle by medium-sized bandwidths and the one to the right by the largest bandwidth.

In the above situation the changes are *homogeneous* in the sense that the smallest change in intensity is still large enough compared to the smallest distance to neighboring change points (for a detailed definition we refer to Cho and Kirch (2021+), Definition 2.1, or Cho and Kirch (2020), Definition 2.1). In particular, this guarantees that all changes can be detected with a single bandwidth only. In some applications with *multiscale* signals, where frequent large changes as well as small isolated changes are present, this is no longer the case as Figure 6.1 shows. In such cases, several bandwidths need to be used and the obtained candidates are pruned down in a second step (see Cho and Kirch (2021+) for an information criterion based approach for partial sum processes and Messer et al. (2014) for a bottom-up-approach for renewal processes). Similarly, if the distance to the neighboring change points is unbalanced MOSUM procedures with asymmetric bandwidths as suggested by Meier, Cho, and Kirch (2021) may be necessary.

7 Conclusions

In this part, we have considered a class of multivariate processes that fulfill a uniform strong invariance principle, possibly after a change of probability space. We have assumed that the processes switch possibly infinitely many times between finitely many regimes, with each switch inducing a change in the drift. This setup includes several important examples, including multivariate partial sum processes, diffusion processes and renewal processes. In order to localize these changes, we have extended the work of Eichinger and Kirch (2018) and Messer et al. (2014) and proposed a single-bandwidth procedure using MOSUM statistics in order to estimate changes, allowing for local changes. We have been able to show consistency for the estimators. Further, we have been able to derive (uniform) localization rates in the form of exact convergence rates, which are indeed minimax-optimal. In the simulation study, our procedure has performed well even under mild violations of our model assumptions. Furthermore, it has shown that minimum-type variance estimators tend to have higher detection rates than when using the true variances and covariance matrices, which in practice is very useful as the covariance matrix oftentimes is hard to estimate. One drawback of the procedure is the use of a single bandwidth. In practice, the identification of the optimal bandwidth turns out to be rather difficult as pointed out e. g. by Cho and Kirch (2021+) and Messer et al. (2014): On the one hand, one wants to choose a large bandwidth in order to have maximal power, while on the other hand choosing too large a bandwidth may lead to misspecification or nonidentification of changes. Furthermore, as can be seen in the simulation study, in a multiscale change point situation (see Definition 2.1 of Cho and Kirch (2021+)) no single bandwidth can detect all change points. Therefore, one future topic of interest is the extension of the proposed procedure to a true multiscale setup as in Cho and Kirch (2021+).

Part II
Anomaly detection based on scan statistics in large image data

8 Introduction

The detection of anomalies in the construction material of buildings like houses, bridges etc. is of great importance to guarantee the stability and the safety of people using them. In particular, for buildings made of concrete, it is important to detect fissures in the substance. Over the past couple of years, CT scans have been used in order to analyze the structure of materials, compare e. g. Weise et al. (2015) or Baranowski et al. (2019), thus allowing for the scanning and analysis of large data.
Early methods for the detection of anomalies, especially in concrete, focus on the analysis of the gray value distribution (see e. g. Acosta, Figueroa, and Mullen (1992)). Another approach for anomaly detection in 2D data is image processing with algorithms (see Ito, Aoki, and Hashimoto (2002), Tang and Gu (2013)). Those methods include a variety of methods such as Template Matching (Roseman (2003)), algorithms to detect minimal paths (Amhaz et al. (2015)) and percolation based on the Hessian matrix (Yamaguchi and Hashimoto (2010)). For three-dimensional data, filtering methods like Frangi filters (see Frangi et al. (2000), Wirjadi et al. (2014)), sheet filters (Sato et al. (2000)) and algorithms to find minimal paths (Müsebeck et al. (2020)) are used. See Ehrig et al. (2011), Paetsch et al. (2012), Barisin et al. (2021) for a comparison of methods. Furthermore, see Barisin et al. (2021) for a comprehensive overview over methods for both 2D and 3D data.
Furthermore, Machine and Deep Learning methods like convolutional neural networks (see Ronneberger, Fischer, and Brox (2015), Çiçek et al. (2016) and Badrinarayanan, Kendall, and Cipolla (2016)) and random forests (Furat et al. (2019), Shi et al. (2016)) are frequently used for the segmentation of both 2D and 3D image data. However, since concrete is a heterogeneous material with various types of anomalies, and since fissures are sparse in the data while varying in size and changing direction (see Figure 8.1), a large number of training data is required to correctly fit

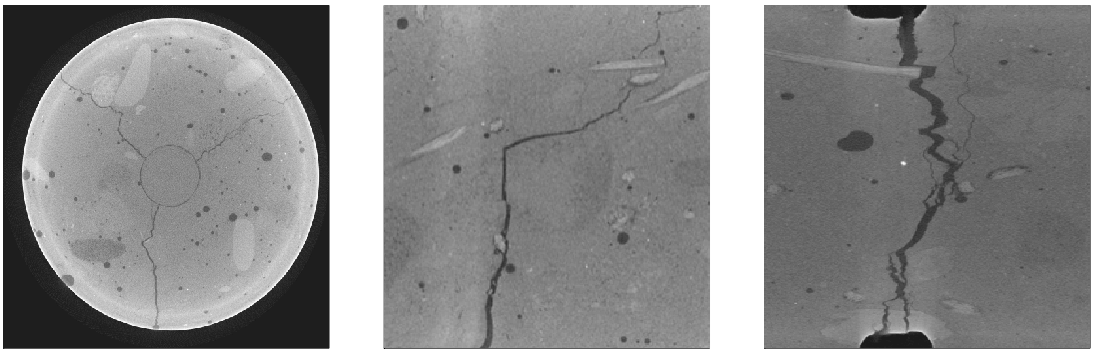

Figure 8.1: 2D-slices of 3D CT scans of cracked concrete blocks. The concrete samples are provided by Martin Kiesche and Christian Caspari, Department of Civil Engineering, University of Kaiserslautern. The CT imaging was performed by Franz Schreiber, Fraunhofer ITWM, Kaiserslautern.

these methods to the data. Furthermore, for large image data, these methods are computationally too expensive to be applied on the whole data set, but they can be applied to smaller regions of the data.
Another computationally cheaper approach for the testing of the existence of anomalies is the use of scan statistics. They were first introduced by Kulldorff (1997) using a Likelihood-ratio test in order to detect clusters in spatial data. Under the assumption of independence of the noise, there exist methods using 2D windows (see Haiman and Preda (2006)), hypercubes with arbitrary dimension (see Kabluchko (2011) for a single-window approach and Jiang (2002); Sharpnack and Arias-Castro (2016) for multiple window approaches) and spatial scan statistics for point processes (Glaz, Naus, and Wallenstein (2001)). Under the assumption of a weak invariance principle, Bucchia (2014) introduced a test for an epidemic change. Proksch, Werner, and Munk (2018) and Munk et al. (2020) introduced tests on inverse regression models with independent, but not necessarily identically distributed noise. Other approaches include the use of EM algorithms (see Moon et al. (2006)), spatial scan statistics (see Kulldorff (2016)) and deformable models (see McInerney and Terzopoulos (1996)). See also Glaz, Naus, and Wallenstein (2001) for an overview over scan statistics in general.

8.1 Outline

In Section 9, we motivate the scanning procedure introduced in Section 11 by the example of a concrete scan containing an artificial fissure. In Section 10, we derive limit theorems for scan statistics based on convex sets with linear size. In Section 11, we propose a scanning procedure for the detection of fissures in concrete image data on 2D slices of 3D images. Based on the geometric properties of fissures and aggregate, we use a combination of scan statistics on rectangles and circular segments in order to find areas that potentially contain fissures. Furthermore, we illustrate the procedure based on 2D images with artificial cracks. In Section 12 we analyze the performance of our procedure in a simulation study.

9 Motivation

The detection of anomalies in concrete is of great importance to ensure the stability of buildings. However, as can be seen in Figure 8.1, concrete is a heterogeneous material that consists of cement and various aggregates such as gravel, sand etc. and often contains small pores of air. These natural anomalies are by no means dangerous, but rather help increase the stability of the material. In Sections 10 and 11, we will develop a procedure based on MOSUM scan statistics with the goal of identifying regions that contain potentially dangerous anomalies such as fissures (see e. g. the middle panel of Figure 9.1) while simultaneously discarding regions with no or only natural anomalies (see e. g. the right panel of Figure 9.1). In this section, we will motivate the specific windows used in the scan statistic that will be introduced in

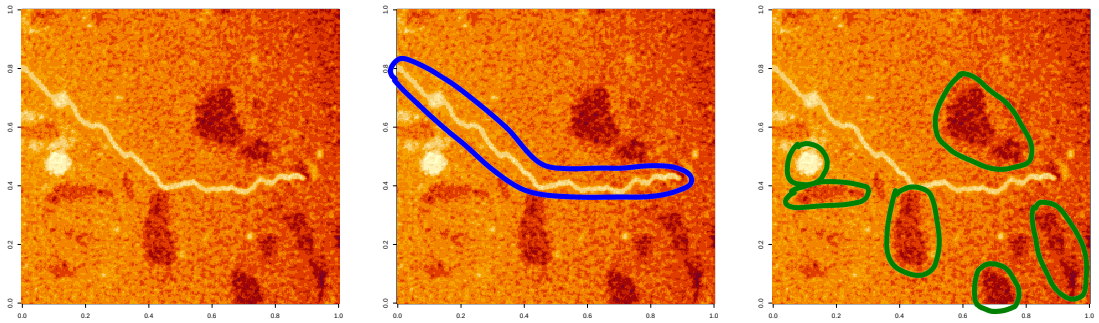

Figure 9.1: The left panel shows a 2D slice of a concrete block containing an artificial fissure. The middle and right panels highlight the fissure and the natural anomalies, respectively. For the sake of better visibility, we have converted the images from a black-to-white (as in Figure 8.1) to a yellow-to-red color scale. All raw images with artificial fissures were provided by Franziska Müsebeck (TU Kaiserslautern).

(11.1) - (11.3) using the example of a 2D slice of a 3D CT scan of a block of concrete with an artificial fissure. The mathematical details are postponed to Section 11.

In the middle panel of Figure 9.1, we can observe that globally in 2D, fissures are lines with small width and changing direction. Locally, as displayed in the upper left panel of Figure 9.2, they resemble rectangles with small width. They have a high contrast to the neighboring environment (mathematically: large difference in gray values). Therefore, as can be seen in the upper left panel of Figure 9.2, we use circle-shaped windows that have an inscribed rectangle with small width in order to detect fissures. More specifically, we compare the average gray value in the rectangle with the average gray values in the circle segments and maximize over multiple angles of the rectangle in order to account for the changes in direction (for a mathematical description see (11.1)). It can be seen in the lower left panel of Figure 9.2 that indeed, fissures are enhanced by this type of statistic. Unfortunately, the edges of the natural anomalies are enhanced, as well. On the other hand, we can observe in the right panel of Figure 9.1 that in 2D, natural anomalies geometrically resemble *bubbles* (from here on, we will refer to the natural anomalies as bubbles). On the edges, they have a high contrast to their neighboring environment while there is little contrast inside the bubbles. As can be seen in the upper middle panel of Figure 9.2, we use circle-shaped windows that are split in half and compare the average gray values in the two semicircles in order to enhance the edges of bubbles. We maximize over multiple angles to account for the change in direction of the edges (for a mathematical description see (11.2)). As can be seen in the lower middle panel of Figure 9.2, this type of statistic indeed enhances the edges of bubbles. By subtracting the latter values intended to enhance the edges of the bubbles from the values from the statistic aiming to enhance fissures (for a mathematical description see (11.3)), our method achieves both an enhancement of the regions close to the fissure and the diminishing of the bubbles. This can be seen for this example in the lower right panel of Figure 9.2.

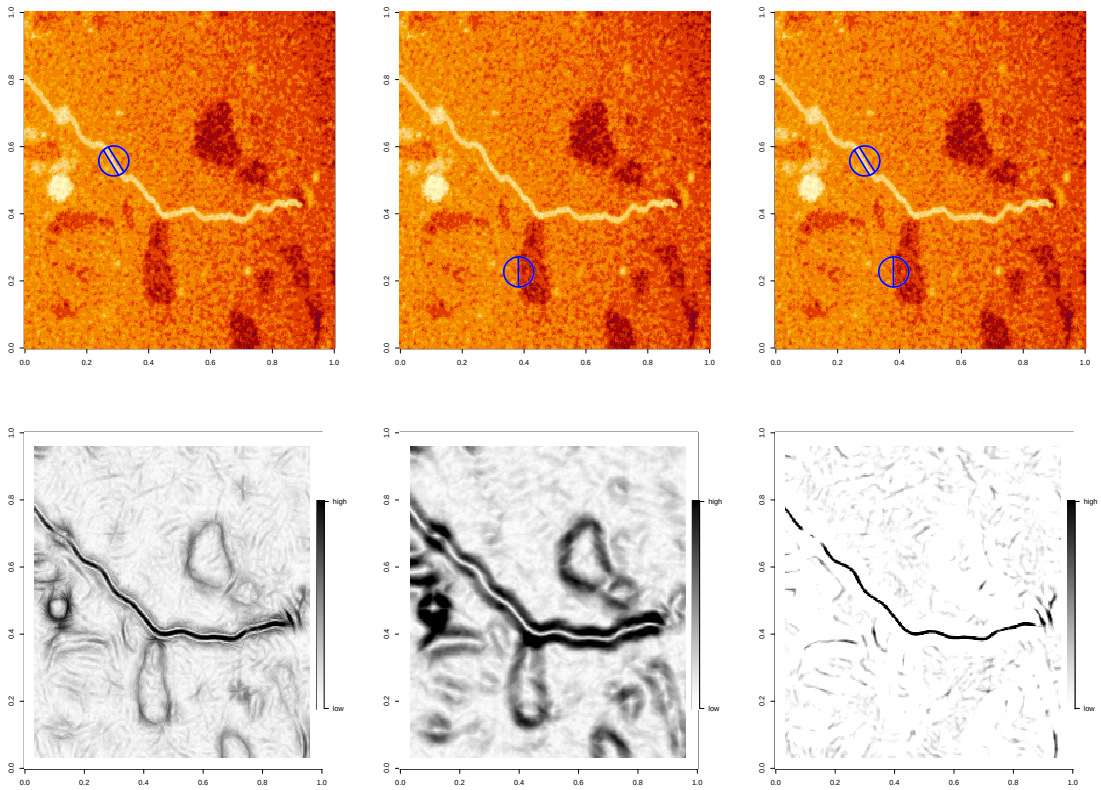

Figure 9.2: The upper left panel illustrates the local geometric properties of fissures. In the lower left panel, a heatmap of the scan statistic in (11.1) is displayed, showing that in this example, (11.1) enhances the fissure and to a lesser extent the edges of bubbles and fissures. The upper middle panel illustrates the local geometric properties of bubbles. In the lower middle panel, a heatmap of the scan statistic in (11.2) is displayed, showing that in this example, (11.2) enhances the edges of bubbles and fissures. The upper right panel illustrates the local geometric properties of fissures and bubbles. In the lower right panel, a heatmap of the combined scan statistic (11.3) is displayed, showing that in this example, the statistic enhances the fissure and eliminates most of the noise.

10 Limit theorems of some MOSUM scan statistics

As motivated in Section 9, we use a combination of scan statistics on rectangles and circles in order to detect fissures in the image data. In particular, we use properly rescaled averages of the gray values of an image over specific areas (e. g. small rectangles and circle segments) as an input for our scan statistic. In the following, we will show limit theorems for these types of rescaled sums over a class of sets in $\mathbb{R}^p$ that in 2D includes, but is not limited to convex sets and in particular circles and rectangles.

Definition 10.1. Let $p \in \mathbb{N}$, $T \in \mathbb{N}$. Let $(Y_{\boldsymbol{k},T})_{\boldsymbol{k}\in\mathbb{Z}^p}$ be a sequence of random variables with $\mathbb{E}[Y_{\boldsymbol{k},T}] = \mu_{\boldsymbol{k},T} \in \mathbb{R}$ and let $A \subset \mathbb{R}^p$. For $\boldsymbol{s} = (s_1, \ldots, s_p)' \in \mathbb{R}^p$, $\boldsymbol{k} = (k_1, \ldots, k_p)' \in \mathbb{Z}^p$, let

$$\begin{aligned} A(\boldsymbol{s}) = A + \boldsymbol{s} = & \{\boldsymbol{x} \in \mathbb{R}^p \,|\boldsymbol{x} - \boldsymbol{s} \in A\} \\ \lfloor \boldsymbol{s} \rfloor_T = & \left(\frac{\lfloor s_1 T \rfloor}{T}, \ldots, \frac{\lfloor s_p T \rfloor}{T}\right)' \\ \frac{\boldsymbol{k}}{T} = & \left(\frac{k_1}{T}, \ldots, \frac{k_p}{T}\right)' \\ S_A\left(\lfloor \boldsymbol{s} \rfloor_T\right) = S_A\left(Y; \lfloor \boldsymbol{s} \rfloor_T\right) = & \sum_{\frac{\boldsymbol{k}}{T} \in A\left(\lfloor \boldsymbol{s} \rfloor_T\right)} Y_{\boldsymbol{k}}. \end{aligned}$$

Furthermore, let $\gamma > 0$. For $\boldsymbol{s} = (s_1, \ldots, s_p)'$, $\boldsymbol{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$ denote by

$$\begin{aligned} [\boldsymbol{s}, \boldsymbol{s} + \boldsymbol{t}) = & \; [s_1, s_1 + t_1) \times \ldots \times [s_p, s_p + t_p) \\ [\boldsymbol{s}, \boldsymbol{s} + \gamma) = & \; [s_1, s_1 + \gamma) \times \ldots \times [s_p, s_p + \gamma), \end{aligned}$$

$p-$dimensional hyperrectangles and -cubes, where $[s_i, s_i + t_i) := [s_i + t_i, s_i)$ if $t_i < 0$. Also denote by

$$\boldsymbol{s} \odot \boldsymbol{t} = (s_1 t_1, \ldots, s_p t_p)'$$

the pointwise product of $\boldsymbol{s}$ and $\boldsymbol{t}$.

In the context of image data with anomalies as described above, it is reasonable to assume that, potentially after some centering, for some anomaly $A \subset [0,1]^p$, the signal is constant across the anomaly, i. e. $\mu_{\boldsymbol{k},T} = \mu_{A,T} \neq 0$ for $\boldsymbol{k}/T \in A$, while it is zero outside of anomalies. Also, with regard to the multiple mean change problem for partial sum processes, it is reasonable to assume that $Y_{\boldsymbol{k},T} = \mu_{\boldsymbol{k},T} + \epsilon_{\boldsymbol{k}}$ for some noise sequence $(\epsilon_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{Z}^p}$.
Analogously to Part I, we drop the dependency on T for ease of notation except in situations where it helps to clarify the argument.

For scan statistics based on p-dimensional hyperrectangles, there exists some theory on the distribution of scan statistics: Kabluchko (2011), Arias-Castro, Donoho, and Huo (2005) and Sharpnack and Arias-Castro (2016) analyze various multiscale procedures with sublinear bandwidths for i.i.d. Gaussian noise. Haiman and Preda (2006) analyze the behavior for positive integer-valued noise while Jaruskova and Piterbarg (2011), Zemlys (2008) assume independent noise with mean 0 and existing variance. Bucchia (2014) studies a CUSUM-type scan statistic under the assumption of an existing functional central limit theorem on the noise.
Under the assumption of an existing functional central limit theorem on the $(\epsilon_{\boldsymbol{k}})$, we are able to derive the limit process of $(T^{-p/2} S_A\left(\lfloor \boldsymbol{s} \rfloor_T\right))_{\boldsymbol{s}\in[0,1]^p}$ under the null situation of no anomalies:

Theorem 10.1. *Let* $\boldsymbol{a} = (a_1, \ldots, a_p)' \in (0, 1/2)^p$ *exist such that* $A = (-\boldsymbol{a}, \boldsymbol{a}] = (-a_1, a_1] \times \ldots \times (-a_p, a_p]$. *Let* $Y_{\boldsymbol{k},T} = \mu_{\boldsymbol{k},T} + \epsilon_{\boldsymbol{k}}$ *with* $\mu_{\boldsymbol{k},T} \in \mathbb{R}$, $\mathbb{E}[\epsilon_{\boldsymbol{k}}] = 0$ *and* $0 < \mathrm{Var}[\epsilon_{\boldsymbol{k}}] < \infty$. *Furthermore let* $S_A(\lfloor \boldsymbol{s} \rfloor_T) = S_A(\epsilon; \lfloor \boldsymbol{s} \rfloor_T)$ *be as in Definition 10.1. Let there exist a p-parameter Wiener process* $(W_{\boldsymbol{s}})_{\boldsymbol{s} \in [0,1]^p}$ *(compare Definition A.3) and some* $\sigma > 0$ *such that*

$$\left(\frac{1}{T^{p/2}} \sum_{k_1=1}^{\lfloor s_1 T \rfloor} \cdots \sum_{k_p=1}^{\lfloor s_p T \rfloor} \epsilon_{k_1,\ldots,k_p} \right)_{\boldsymbol{s} \in [0,1]^p} \xrightarrow{w} \sigma (W_{\boldsymbol{s}})_{\boldsymbol{s} \in [0,1]^p} \tag{10.1}$$

on $\mathcal{D}([0,1]^p)$ *and that for any* $\boldsymbol{c} = (c_1, \ldots, c_p)' \in \mathbb{Z}^p$, *and any bounded hyperrectangle* $I \subset \mathbb{R}^p$, *it holds that*

$$\frac{1}{T^{p/2}} \sup_{\boldsymbol{s} \in I} \left| \sum_{k_1=1}^{\lfloor s_1 T \rfloor} \cdots \sum_{k_p=1}^{\lfloor s_p T \rfloor} \epsilon_{k_1,\ldots,k_p} - \sum_{k_1=1}^{\lfloor s_1 T \rfloor + c_1} \cdots \sum_{k_p=1}^{\lfloor s_p T \rfloor + c_p} \epsilon_{k_1,\ldots,k_p} \right| = o_P(1). \tag{10.2}$$

Then it holds with

$$\Lambda_{\boldsymbol{s}} = \sigma \left(\sum_{\mathbf{d}=(d_1,\ldots,d_p)' \in \{0,1\}^p} (-1)^{\sum_i d_i} W_{\boldsymbol{s} + (-1)^{\mathbf{d}} \odot \boldsymbol{a}} \right)$$

that

$$\left(\frac{1}{T^{p/2}} S_A(\epsilon; \lfloor \boldsymbol{s} \rfloor_T) \right)_{\boldsymbol{s} \in [0,1]^p} \xrightarrow{w} (\Lambda_{\boldsymbol{s}})_{\boldsymbol{s} \in [0,1]^p}$$

on $\mathcal{D}([0,1]^p)$ *as* $T \to \infty$.

Proof. Let $\boldsymbol{s} = (s_1, \ldots, s_p)' \in [0,1]^p$. Since by definition

$$A(\lfloor \boldsymbol{s} \rfloor_T) = (\lfloor \boldsymbol{s} \rfloor_T - \boldsymbol{a}, \lfloor \boldsymbol{s} \rfloor_T + \boldsymbol{a}],$$

it holds by the definition of $S_A(\epsilon; \lfloor \boldsymbol{s} \rfloor_T)$ in Definition 10.1 that

$$\begin{aligned} S_A(\epsilon; \lfloor \boldsymbol{s} \rfloor_T) &= \sum_{k_1 = \lfloor (\lfloor s_1 \rfloor_T - a_1) T \rfloor + 1}^{\lfloor (\lfloor s_1 \rfloor_T + a_1) T \rfloor} \cdots \sum_{k_p = \lfloor (\lfloor s_p \rfloor_T - a_p) T \rfloor + 1}^{\lfloor (\lfloor s_p \rfloor_T + a_p) T \rfloor} \epsilon_{k_1,\ldots,k_p} \\ &= \sum_{(d_1,\ldots,d_p) \in \{0,1\}^p} (-1)^{\sum_i d_i} \sum_{k_1=1}^{\lfloor (\lfloor s_1 \rfloor_T + (-1)^{d_1} a_1) T \rfloor} \cdots \sum_{k_p=1}^{\lfloor (\lfloor s_p \rfloor_T + (-1)^{d_p} a_p) T \rfloor} \epsilon_{k_1,\ldots,k_p}, \end{aligned} \tag{10.3}$$

where we can see the last equality as follows: We need to distinguish between two cases: If $(l_1, \ldots, l_p)' \in \mathbb{N}^p$ with $\lfloor (\lfloor s_i \rfloor_T - a_i) T \rfloor < l_i \leq \lfloor (\lfloor s_i \rfloor_T + a_i) T \rfloor$ for all i, then $\epsilon_{l_1,\ldots,l_p}$ is only counted for $(d_1, \ldots, d_p)' = (0, \ldots, 0)'$ in (10.3).
If there exists some $j = 1, \ldots, p$ such that $l_j \leq \left\lfloor \left(\lfloor s_j \rfloor_T - a_j \right) T \right\rfloor$, then $\epsilon_{l_1,\ldots,l_p}$ is counted in both $(d_1, \ldots, d_{j-1}, 0, d_{j+1}, \ldots, d_p)'$ and $(d_1, \ldots, d_{j-1}, 1, d_{j+1}, \ldots, d_p)'$ with

$d_i \in \{0,1\}$ appropriate for $i \neq j$. Since $(-1)^{\sum_{i\neq j} d_i} = -(-1)^{1+\sum_{i\neq j} d_i}$, $\epsilon_{l_1,\ldots,l_p}$ gets canceled out in all of these summands. This telescoping sum is also in the spirits of Nelsen (2006), Def. 2.10.1. for the volume of p-dimensional hyperrectangles.
Furthermore, since $s_i - \lfloor s_i \rfloor_T = s_i - \lfloor s_i T \rfloor / T \leq 1/T$, it holds for $d \in \{0,1\}$ that $\left|\left\lfloor (s_i + (-1)^d a_i)T \right\rfloor - \left\lfloor (\lfloor s_i \rfloor_T + (-1)^d a_i)T \right\rfloor\right| \leq 2$. Thus, it holds by (10.2) for all $(d_1,\ldots,d_p)' \in \{0,1\}^p$ uniformly in $\boldsymbol{s}$ that

$$\left| \frac{1}{T^{p/2}} \sum_{k_1=1}^{\lfloor (\lfloor s_1 \rfloor_T + (-1)^{d_1} a_1)T \rfloor} \cdots \sum_{k_p=1}^{\lfloor (\lfloor s_p \rfloor_T + (-1)^{d_p} a_p)T \rfloor} \epsilon_{k_1,\ldots,k_p} \right.$$
$$\left. - \frac{1}{T^{p/2}} \sum_{k_1=1}^{\lfloor (s_1 + (-1)^{d_1} a_1)T \rfloor} \cdots \sum_{k_p=1}^{\lfloor (s_p + (-1)^{d_p} a_p)T \rfloor} \epsilon_{k_1,\ldots,k_p} \right| = o_P(1).$$

The assertion follows by (10.1) and the continuous mapping theorem. □

As noted earlier, the scan statistic that we use for the detection of fissures and bubbles is a combination of scan statistics on rectangles, circles and circle segments by some Lipschitz-continuous function (see Section 11 for more details and Figure 11.1 for an illustration of the sets used). Therefore, in the following, we will establish a limit theorem for $\left(F\left(1/T^{p/2} S_{A_1}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, 1/T^{p/2} S_{A_P}(\lfloor \boldsymbol{s} \rfloor_T)\right)\right)_{\boldsymbol{s}\in[0,1]^p}$ on $\mathcal{D}([0,1]^p)$, where $A_1,\ldots,A_P$ come from the following class of sets in $\mathbb{R}^p$ that in 2D includes all convex sets (compare Lemma 10.4):

Assumption 10.1. *Let $A \subset \mathbb{R}^p$. For $\gamma > 0$, $\boldsymbol{s} \in \mathbb{R}^p$ let*

$$A^{(\gamma)}(\boldsymbol{s}) = \bigcup_{\boldsymbol{y}\in[0,\gamma]^p} A(\boldsymbol{s}+\boldsymbol{y}).$$

(a) For any $\boldsymbol{s} \in \mathbb{R}^p$ it holds that

$$\left| \left\{ \frac{\boldsymbol{k}}{T} \in A(\boldsymbol{s}) \right\} \right| = T^p \lambda(A) + O(T^{p-1}).$$

(b) There exists $\gamma_0 > 0$ and $C > 0$ such that for all $0 < \gamma \leq \gamma_0$,

$$\lambda\left(A^{(\gamma)}(\boldsymbol{s})\right) - \lambda(A) \leq C\gamma.$$

Remark 10.1. If A fulfills Assumption 10.1 and $\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \gamma \leq \gamma_0$, then $\lambda\left(A(\boldsymbol{s}) \Delta A(\boldsymbol{t})\right) \leq 2C\gamma$: Since $\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \gamma$, there exists $\boldsymbol{x} \in \mathbb{R}^p$ such that $\boldsymbol{s}, \boldsymbol{t} \in [\boldsymbol{x}, \boldsymbol{x}+\gamma]$. Therefore, $A(\boldsymbol{s}), A(\boldsymbol{t}) \subset A^{(\gamma)}(\boldsymbol{x})$ and it follows by Assumption 10.1 (b) that

$$\lambda\left(A(\boldsymbol{s}) \Delta A(\boldsymbol{t})\right) = \lambda\left(A(\boldsymbol{s}) \setminus A(\boldsymbol{t})\right) + \lambda\left(A(\boldsymbol{t}) \setminus A(\boldsymbol{s})\right)$$
$$\leq \lambda\left(A^{(\gamma)}(\boldsymbol{x}) \setminus A(\boldsymbol{t})\right) + \lambda\left(A^{(\gamma)}(\boldsymbol{x}) \setminus A(\boldsymbol{s})\right) \leq 2C\gamma.$$

By Theorem A.6 (also compare Wichura (1969), Theorem 2), it is sufficient to show the convergence of the finite-dimensional distributions (fidis) of

$$\left(F\left(1/T^{p/2}S_{A_1}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, 1/T^{p/2}S_{A_P}(\lfloor \boldsymbol{s} \rfloor_T)\right)\right)_{\boldsymbol{s}\in[0,1]^p}$$

and the fulfilling of some modulus of continuity. In order to prove these properties, we will first show the convergence of the fidis and the modulus of continuity for

$$\left(\left(1/T^{p/2}S_{A_1}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, 1/T^{p/2}S_{A_P}(\lfloor \boldsymbol{s} \rfloor_T)\right)'\right)_{\boldsymbol{s}\in[0,1]^p}.$$

Theorem 10.2.
Let $A_1, \ldots, A_P \subset \mathbb{R}^p$ fulfill Assumption 10.1. Let $(\epsilon_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{Z}^p}$ be a sequence of i.i.d. random variables with $\mathbb{E}[\epsilon_{\boldsymbol{k}}] = 0$, $\mathbb{E}[\epsilon_{\boldsymbol{k}}^2] = \sigma^2 \in (0, \infty)$ and let $\mathbb{E}[|\epsilon_{\boldsymbol{k}}|^r] < \infty$ for some $r > 2p$. Let $Y_{\boldsymbol{k},T} = \mu_{\boldsymbol{k},T} + \epsilon_{\boldsymbol{k}}$ with $\mu_{\boldsymbol{k},T} \in \mathbb{R}$, let $S_A(\lfloor \boldsymbol{s} \rfloor_T) = S_A(\epsilon; \lfloor \boldsymbol{s} \rfloor_T)$ be as in Definition 10.1 and let

$$\mathbb{S}_T(\boldsymbol{s}) = \left(\frac{1}{T^{p/2}}S_{A_1}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, \frac{1}{T^{p/2}}S_{A_P}(\lfloor \boldsymbol{s} \rfloor_T)\right)'.$$

Then the following properties hold:

(i) There exists a P-dimensional centered Gaussian process $(\mathbb{W}(\boldsymbol{s}))_{\boldsymbol{s}\in[0,1]^p} = ((W_1(\boldsymbol{s}), \ldots, W_P(\boldsymbol{s}))')_{\boldsymbol{s}\in[0,1]^p}$ with

$$\mathrm{Cov}\left[W_i(\boldsymbol{s}), W_j(\boldsymbol{t})\right] = \sigma^2 \lambda\left(A_i(\boldsymbol{s}) \cap A_j(\boldsymbol{t})\right)$$

such that for all $n \in \mathbb{N}$, $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^p$ it holds that

$$\left(\mathbb{S}_T(\boldsymbol{t}_1), \ldots \mathbb{S}_T(\boldsymbol{t}_n)\right) \xrightarrow{\mathcal{D}} \left(\mathbb{W}(\boldsymbol{t}_1), \ldots, \mathbb{W}(\boldsymbol{t}_n)\right).$$

(ii) For any $x > 0$ it holds that

$$\lim_{\delta\to 0} \limsup_{T\to\infty} \mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta} \|\mathbb{S}_T(\boldsymbol{s}) - \mathbb{S}_T(\boldsymbol{t})\|_\infty \geq x\right) = 0.$$

Proof. (i) By the theorem of Cramér-Wold (compare Billingsley (1995), p. 383), it is sufficient to show that for arbitrary $a_{1,1}, \ldots, a_{n,1}, \ldots, a_{1,P}, \ldots, a_{n,P} \in \mathbb{R}$,

$$\sum_{i=1}^{n}\sum_{j=1}^{P} a_{i,j}\frac{1}{T^{p/2}}S_{A_j}(\lfloor \boldsymbol{t}_i \rfloor_T)$$

converges in distribution to a normal limit with the variance of

$$\mathrm{Var}\left[\sum_{i=1}^{n}\sum_{j=1}^{P} a_{i,j}W_j(\boldsymbol{t}_i)\right].$$

For the ease of notation, denote by $A^1 = A$, $A^0 = A'$. For ease of notation we are also renumbering the $A_j(\boldsymbol{t}_i)$ successively with a slight abuse of notation in the following way: For $i = 1, \ldots, n$, $j = 1, \ldots, P$ let $A_{(j-1)n+i-1} = A_j(\boldsymbol{t}_i)$ and let $m = nP$.
For $l = 0, \ldots, 2^m - 1$ let $l = \sum_{i=0}^{m-1} l_i 2^i$ be the unique binary representation of l with $l_i \in \{0, 1\}$ and let

$$M_l = M_{\sum_{i=0}^{m-1} l_i 2^i} = \bigcap_{i=0}^{m-1} A_i^{l_i}.$$

Since we are using the unique binary representation for l, the sets M_l are pairwise disjoint: For $k \neq l$, there exists $i = 0, \ldots, m-1$ such that $k_i \neq l_i$. It follows that $A_i^{k_i} \cap A_i^{l_i} = \emptyset$ and therefore $M_k \cap M_l = \emptyset$. Furthermore, it holds that

$$\begin{aligned} &(a) \quad \sum_{l=0}^{2^m-1} M_l = \mathbb{R}^2 \\ &(b) \quad A_i = \sum_{l:l_i=1} M_l \end{aligned} \tag{10.4}$$

for $i = 0, \ldots, m-1$.

Proof of (a): Let $\boldsymbol{x} \in \mathbb{R}^p$. For $i = 0, \ldots, m-1$, let $l_i = \mathbb{1}_{\{\boldsymbol{x} \in A_i\}}$. Then $\boldsymbol{x} \in A_i^{l_i}$ for all i, hence

$$\boldsymbol{x} \in \bigcap_{i=0}^{m-1} A_i^{l_i} = M_{\sum_{i=0}^{m-1} l_i 2^i}.$$

Proof of (b): We first show that $A_i \subset \sum_{l:l_i=1} M_l$. Let $\boldsymbol{x} \in A_i$. For $k = 0, \ldots, m-1$, $k \neq i$ let $l_k = \mathbb{1}_{\{\boldsymbol{x} \in A_k\}}$. Then,

$$\boldsymbol{x} \in \left(\bigcap_{k \neq i} A_k^{l_k} \right) \cap A_i = M_{\sum_{k=0}^{m-1} l_k 2^k}$$

with $l_i = 1$.
We then show that $A_i \supset \sum_{l:l_i=1} M_l$. By definition, it holds for l with $l_i = 1$ that

$$M_l = A_i \cap \left(\bigcap_{k \neq i} A_k^{l_k} \right) \subset A_i$$

and the assertion follows immediately.

Since the M_l are pairwise disjoint, it holds by the central limit theorem and Assumption 10.1 (a) that

$$\left(\frac{1}{T^{p/2}} \sum_{\frac{\boldsymbol{k}}{T} \in M_0} \epsilon_{\boldsymbol{k}}, \ldots, \frac{1}{T^{p/2}} \sum_{\frac{\boldsymbol{k}}{T} \in M_{2^m-1}} \epsilon_{\boldsymbol{k}} \right)' \xrightarrow{\mathcal{D}} \left(\widetilde{W}_0, \ldots, \widetilde{W}_{2^m-1} \right)',$$

where the $\widetilde{W}_l$, are pairwise independent $\mathcal{N}(0, \sigma^2\lambda(M_l))$-distributed random variables for $l = 0, \dots, 2^m - 1$. For $i = 0, \dots, m-1$ denote by

$$W_i = \sum_{l:l_i=1} \widetilde{W}_l.$$

By (10.4), it holds that W_i is $\mathcal{N}(0, \sigma^2\lambda(A_i))$-distributed.
Let $a_{1,1}, \dots, a_{n,1}, \dots, a_{1,P}, \dots, a_{n,P} \in \mathbb{R}$. Analogously to above, we are renumbering the $a_{i,j}$ for the ease of notation with a slight abuse of notation in the following way: For $i = 1, \dots, n$, $j = 1, \dots, P$ with $a_{(j-1)n+i-1} = a_{i,j}$ that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{P} a_{i,j}\frac{1}{T^{p/2}}S_{A_j}(\boldsymbol{t}_i) = \sum_{i=1}^{n}\sum_{j=1}^{P} a_{i,j}\frac{1}{T^{p/2}}\sum_{\frac{\boldsymbol{k}}{T}\in A_j(\boldsymbol{t}_i)}\epsilon_{\boldsymbol{k}} = \sum_{i=0}^{m-1} a_i\frac{1}{T^{p/2}}\sum_{\frac{\boldsymbol{k}}{T}\in A_i}\epsilon_{\boldsymbol{k}} \\
&= \sum_{i=0}^{m-1} a_i \sum_{l:l_i=1}\frac{1}{T^{p/2}}\sum_{\frac{\boldsymbol{k}}{T}\in M_l}\epsilon_{\boldsymbol{k}} \xrightarrow{\mathcal{D}} \sum_{i=0}^{m-1} a_i \sum_{l:l_i=1}\widetilde{W}_l = \sum_{i=0}^{m-1} a_i W_i.
\end{aligned}$$

We have shown the convergence with $S_{A_j}(\boldsymbol{t}_i)$ instead of $S_{A_j}(\lfloor \boldsymbol{t}_i \rfloor_T)$. However, this does not present a problem: For $j = 1, \dots, P$ and $\boldsymbol{s} \in [0,1]^p$, we obtain that

$$\begin{aligned}
&\left|\frac{1}{T^{p/2}}S_{A_j}(\boldsymbol{s}) - \frac{1}{T^{p/2}}S_{A_j}(\lfloor \boldsymbol{s} \rfloor_T)\right| = \frac{1}{T^{p/2}}\left|\sum_{\frac{\boldsymbol{k}}{T}\in A_j(\boldsymbol{s})}\epsilon_{\boldsymbol{k}} - \sum_{\frac{\boldsymbol{k}}{T}\in A_j(\lfloor \boldsymbol{s} \rfloor_T)}\epsilon_{\boldsymbol{k}}\right| \\
&\leq \frac{1}{T^{p/2}}\sum_{\frac{\boldsymbol{k}}{T}\in A_j(\boldsymbol{s})\Delta A_j(\lfloor \boldsymbol{s} \rfloor_T)}\left|\epsilon_{\boldsymbol{k}}\right|. \qquad (10.5)
\end{aligned}$$

Since $\|\boldsymbol{s} - \lfloor \boldsymbol{s} \rfloor_T\|_\infty < 1/T$, it holds by Remark 10.1 that $\lambda(A_j(\boldsymbol{s})\Delta A_j(\lfloor \boldsymbol{s} \rfloor_T)) = O(1/T)$ and therefore, the number of independent summands with identical variance in (10.5) by Assumption 10.1 (a) is of order $O(T^{p-1})$. Thus, it holds by Tschebyscheff's inequality that

$$\frac{1}{T^{p/2}}S_{A_j}(\boldsymbol{s}) - \frac{1}{T^{p/2}}S_{A_j}(\lfloor \boldsymbol{s} \rfloor_T) = o_P(1).$$

and the assertion follows.

Proof of (ii):
First note that for arbitrary $\boldsymbol{s}, \boldsymbol{t} \in [0,1]^p$, it holds that

$$\|\mathbb{S}_T(\boldsymbol{s}) - \mathbb{S}_T(\boldsymbol{t})\|_\infty = \max_{i=1,\dots,P}\left|\frac{1}{T^{p/2}}S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_T\right) - \frac{1}{T^{p/2}}S_{A_i}\left(\lfloor \boldsymbol{t} \rfloor_T\right)\right|.$$

Therefore, showing

$$\lim_{\delta\to 0}\limsup_{T\to\infty}\mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta}\|\mathbb{S}_T(\boldsymbol{s}) - \mathbb{S}_T(\boldsymbol{t})\|_\infty \geq x\right) = 0$$

for all $x > 0$ is equivalent to showing

$$\lim_{\delta\to 0}\limsup_{T\to\infty} \mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta}\left|\frac{1}{T^{p/2}}S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_T\right)-\frac{1}{T^{p/2}}S_{A_i}\left(\lfloor \boldsymbol{t}\rfloor_T\right)\right|\geq x\right)=0.$$

for all $x > 0$, $i = 1, \ldots, P$.

For any $\delta > 0$, define $k_0 = k_0(\delta)$ such that $2^{-(k_0+1)} < \delta \leq 2^{-k_0}$. Then letting $\delta \to 0$ is equivalent to letting $k_0 \to \infty$. Let $m = m(T) = \lceil (2p/r + \varepsilon) \log_2 T \rceil$ for some sufficiently small $\varepsilon > 0$ such that $2^m = o(T)$ as $2p < r$. Since in the following, we are only going to use sets of the form $A^{(2^{-k})}$, we will write $A^k = A^{(2^{-k})}$ for the ease of notation. For $\boldsymbol{s}, \boldsymbol{t} \in [0,1]^p$ with $\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \delta$, we can write

$$\begin{aligned}
&S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_T\right) - S_{A_i}\left(\lfloor \boldsymbol{t}\rfloor_T\right)\\
&= S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_T\right) - S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right) && (10.6)\\
&\quad + S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right) - S_{A_i^{m-2}}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right) && (10.7)\\
&\quad + \sum_{k=k_0}^{m-1}\left(S_{A_i^{k-1}}\left(\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}\right) - S_{A_i^{k-2}}\left(\lfloor \boldsymbol{s}\rfloor_{2^k}\right)\right) && (10.8)\\
&\quad + S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{s}\rfloor_{2^{k_0}}\right) - S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{t}\rfloor_{2^{k_0}}\right) && (10.9)\\
&\quad + \sum_{k=k_0}^{m-1}\left(S_{A_i^{k-2}}\left(\lfloor \boldsymbol{t}\rfloor_{2^k}\right) - S_{A_i^{k-1}}\left(\lfloor \boldsymbol{t}\rfloor_{2^{k+1}}\right)\right)\\
&\quad + S_{A_i^{m-2}}\left(\lfloor \boldsymbol{t}\rfloor_{2^m}\right) - S_{A_i}\left(\lfloor \boldsymbol{t}\rfloor_{2^m}\right)\\
&\quad + S_{A_i}\left(\lfloor \boldsymbol{t}\rfloor_{2^m}\right) - S_{A_i}\left(\lfloor \boldsymbol{t}\rfloor_T\right).
\end{aligned}$$

Note that this telescoping series consists of four separate elements: In the first step (10.6), the step function $S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_T\right)$ on a grid of spacing $1/T$ is replaced by the corresponding step function $S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)$ on a grid of spacing $2^{-m} > 1/T$. In the second step (10.7), A_i is replaced by a larger set of Lebesgue measure $\lambda(A_i) + O(2^{-m})$. In the third step (10.8) we continue to coarsen the grid and expand the sets around $\boldsymbol{s}$ until the distance between adjacent grid points is more than $\|\boldsymbol{s}-\boldsymbol{t}\|_\infty$.
In the fourth step (10.9), we move from $\boldsymbol{s}$ to $\boldsymbol{t}$ on sets of the form $A_i^{k_0-2}$. Then steps 3 to 1 are reversed for $\boldsymbol{t}$. In order to show the assertion, we will analyze (10.6)-(10.9) separately.

For (10.6): For T large enough (by assumption, $2^m = o(T)$), it holds that

$$\left\|\lfloor \boldsymbol{s}\rfloor_T - \lfloor \boldsymbol{s}\rfloor_{2^m}\right\|_\infty \leq \left\|\lfloor \boldsymbol{s}\rfloor_T - \boldsymbol{s}\right\|_\infty + \left\|\boldsymbol{s} - \lfloor \boldsymbol{s}\rfloor_{2^m}\right\|_\infty \leq 1/T + 2^{-m} \leq 2\cdot 2^{-m}. \quad (10.10)$$

Therefore, it holds by Remark 10.1 with a suitable constant $C > 0$ that

$$\begin{aligned}
&\lambda\left[A_i\left(\lfloor \boldsymbol{s}\rfloor_T\right) \setminus A_i\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\right] + \lambda\left[A_i\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right) \setminus A_i\left(\lfloor \boldsymbol{s}\rfloor_T\right)\right]\\
&= \lambda\left[A_i\left(\lfloor \boldsymbol{s}\rfloor_T\right) \Delta A_i\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\right] \leq C2^{-m}.
\end{aligned}$$

Therefore, we obtain by Markov's inequality, Assumption 10.1 (a) and Lemma A.3 with suitable constants $C_r, C_2 > 0$ that

$$
\begin{aligned}
&\mathrm{P}\left(\left|\frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_T\right) - \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)\right| \geq x\right) \\
&= \mathrm{P}\left(\frac{1}{T^{p/2}} \left| \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right)} \epsilon_{\boldsymbol{k}} - \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)} \epsilon_{\boldsymbol{k}} \right| \geq x\right) \\
&= \mathrm{P}\left(\frac{1}{T^{p/2}} \left| \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)} \epsilon_{\boldsymbol{k}} - \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right)} \epsilon_{\boldsymbol{k}} \right| \geq x\right) \\
&\leq \mathrm{P}\left(\frac{1}{T^{p/2}} \left| \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)} \epsilon_{\boldsymbol{k}} \right| \geq \frac{x}{2}\right) + \mathrm{P}\left(\frac{1}{T^{p/2}} \left| \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right)} \epsilon_{\boldsymbol{k}} \right| \geq \frac{x}{2}\right) \\
&\leq \frac{\mathbb{E}\left[\left| \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)} \epsilon_{\boldsymbol{k}} \right|^r\right]}{T^{rp/2}\left(\frac{x}{2}\right)^r} + \frac{\mathbb{E}\left[\left| \sum_{\frac{\boldsymbol{k}}{T} \in A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right)} \epsilon_{\boldsymbol{k}} \right|^r\right]}{T^{rp/2}\left(\frac{x}{2}\right)^r} \\
&\leq 2^r C_r \frac{\left(T^p \lambda\left[A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)\right] + O(T^{p-1})\right)^{r/2}}{T^{rp/2} x^r} \\
&\quad + 2^r C_r \frac{\left(T^p \lambda\left[A_i\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) \setminus A_i\left(\lfloor \boldsymbol{s} \rfloor_T\right)\right] + O(T^{p-1})\right)^{r/2}}{T^{rp/2} x^r} \\
&\leq C_2 \frac{T^{rp/2} 2^{-mr/2}}{T^{rp/2}} = C_2 2^{-mr/2} \leq C_2 T^{-\frac{r}{2}\left(\frac{2p}{r} + \varepsilon\right)} = C_2 T^{-(p + r\varepsilon/2)},
\end{aligned}
$$

since by definition, $m = \lceil (2p/r + \varepsilon) \log_2 T \rceil$ and thus, $2^{-m} \leq T^{-(2p/r+\varepsilon)}$. We will now consider the supremum over $\boldsymbol{s} \in [0,1]^p$ of the above expression that only depends on $\boldsymbol{s}$ via $\lfloor \boldsymbol{s} \rfloor_T$ and $\lfloor \boldsymbol{s} \rfloor_{2^m}$. Thus, it is by (10.10) sufficient to take the supremum over $\lfloor \boldsymbol{s} \rfloor_T$ with $\boldsymbol{s} \in [0,1]^p$ and over all $\lfloor \boldsymbol{s} \rfloor_{2^m}$ with $\|\lfloor \boldsymbol{s} \rfloor_{2^m} - \lfloor \boldsymbol{s} \rfloor_T\| \leq 2 \cdot 2^{-m}$ for fixed $\lfloor \boldsymbol{s} \rfloor_T$. The first supremum is taken over the set $\{\lfloor \boldsymbol{s} \rfloor_T | \boldsymbol{s} \in [0,1]^p\}$, which has cardinality $(T+1)^p \leq 2^p T^p$. The second supremum is taken over a set that by Lemma C.1 (a) has finite cardinality as the number of points of the form $\boldsymbol{k} 2^{-m}$, $\boldsymbol{k} \in \mathbb{Z}^p$ in the hypercube $[\lfloor \boldsymbol{s} \rfloor_T - 2 \cdot 2^{-m}, \lfloor \boldsymbol{s} \rfloor_T + 2 \cdot 2^{-m}]$ is finite. Therefore, we obtain for any $x > 0$ with a suitable constant $C_3 > 0$ potentially depending on x by subadditivity that

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{\boldsymbol{s} \in [0,1]^p} \left|\frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_T\right) - \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)\right| \geq x\right) \\
&\leq \sum_{\lfloor \boldsymbol{s} \rfloor_T \colon\ \boldsymbol{s} \in [0,1]^p} \sum_{\substack{\lfloor \boldsymbol{s} \rfloor_{2^m} \colon\ \boldsymbol{s} \in [0,1]^p, \\ \left\|\lfloor \boldsymbol{s} \rfloor_T - \lfloor \boldsymbol{s} \rfloor_{2^m}\right\|_\infty \leq 2 \cdot 2^{-m}}} \mathrm{P}\left(\left|\frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_T\right) - \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right)\right| \geq x\right) \\
&\leq C_3 T^p T^{-(p + r\varepsilon/2)} = C_3 T^{-r\varepsilon/2} \to 0.
\end{aligned} \tag{10.11}
$$

as $T \to \infty$.

For (10.7), it holds by Assumption 10.1 (b) that

$$\lambda\left(A^{m-2}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\setminus A\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\right)\leq 4C2^{-m}.$$

Therefore, we obtain analogously to (10.11) with suitable constants $C_r, C_2 > 0$ that

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\boldsymbol{s}\in[0,1]^p}\left|\frac{1}{T^{p/2}}S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)-\frac{1}{T^{p/2}}S_{A_i^{m-2}}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\right|\geq x\right)\\
\leq&\sum_{\lfloor \boldsymbol{s}\rfloor_{2^m}:\ \boldsymbol{s}\in[0,1]^p}\mathrm{P}\left(\left|\frac{1}{T^{p/2}}S_{A_i}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)-\frac{1}{T^{p/2}}S_{A_i^{m-2}}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\right|\geq x\right)\\
\leq& 2^p\cdot 2^{pm}C_r\frac{\left(T^p\lambda\left(A^{m-2}\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\setminus A\left(\lfloor \boldsymbol{s}\rfloor_{2^m}\right)\right)+O(T^{p-1})\right)^{r/2}}{T^{pr/2}x^r}\\
\leq& C_2 2^{pm}2^{-mr/2}=C_2 2^{m(p-r/2)}\to 0
\end{aligned}\tag{10.12}$$

as $T\to\infty$ since $r>2p$ and $m\to\infty$ as $T\to\infty$.

For (10.8), for any $k=k_0,\ldots,m-1$ and any $\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}$, $\lfloor \boldsymbol{s}\rfloor_{2^k}$ is uniquely determined: If $\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}=\boldsymbol{l}2^{-(k+1)}=(\boldsymbol{l}/2)2^{-k}$ for $\boldsymbol{l}\in\mathbb{Z}^p$, then $\lfloor \boldsymbol{s}\rfloor_{2^k}=\lfloor \boldsymbol{l}/2\rfloor 2^{-k}$, where $\lfloor \boldsymbol{l}/2\rfloor$ denotes the componentwise integer part of $\boldsymbol{l}/2$. By construction it holds that

$$\left\|\lfloor \boldsymbol{s}\rfloor_{2^k}-\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}\right\|_\infty\leq 2^{-(k+1)}<2^{-(k-1)}=2^{-(k-2)}-2^{-(k-1)}.$$

hence by construction,

$$\lfloor \boldsymbol{s}\rfloor_{2^k}\leq\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}\leq\lfloor \boldsymbol{s}\rfloor_{2^k}+2^{-(k+1)}\leq\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}+2^{-(k+1)}\leq\lfloor \boldsymbol{s}\rfloor_{2^k}+2^{-(k-2)}.$$

Therefore

$$\left[\lfloor \boldsymbol{s}\rfloor_{2^{k+1}},\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}+2^{-(k-1)}\right]\subset\left[\lfloor \boldsymbol{s}\rfloor_{2^k},\lfloor \boldsymbol{s}\rfloor_{2^k}+2^{-(k-2)}\right]$$

and by definition of the sets $A^{(\gamma)}$ in Assumption 10.1,

$$A^{k-1}\left(\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}\right)\subset A^{k-2}\left(\lfloor \boldsymbol{s}\rfloor_{2^k}\right).$$

It follows by Assumption 10.1 (b) that

$$\lambda\left(A^{k-2}\left(\lfloor \boldsymbol{s}\rfloor_{2^k}\right)\setminus A^{k-1}\left(\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}\right)\right)\leq\lambda\left(A^{k-2}\left(\lfloor \boldsymbol{s}\rfloor_{2^k}\right)\setminus A\left(\lfloor \boldsymbol{s}\rfloor_{2^{k+1}}\right)\right)\leq 4C2^{-k}.$$

Furthermore, for $x>0$, if δ is small enough (and therefore, k_0 large enough), $x\geq\sum_{k\geq k_0}2^{k(p/(2r)-1/4)}$ since $p/(2r)-1/4<0$ by $r>2p$. Analogously to (10.11), we obtain by σ-subadditivity, Markov's inequality, Assumption 10.1 (a) and Lemma A.3 with suitable constants $C_r, C_2>0$ that

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{\boldsymbol{s}\in[0,1]^p}\left|\sum_{k=k_0}^{m-1}\left(\frac{1}{T^{p/2}}S_{A_i^{k-1}}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}\right)-\frac{1}{T^{p/2}}S_{A_i^{k-2}}\left(\lfloor\boldsymbol{s}\rfloor_{2^k}\right)\right)\right|\geq x\right)\\
&\leq \mathrm{P}\left(\sum_{k=k_0}^{m-1}\sup_{\boldsymbol{s}\in[0,1]^p}\left|\left(\frac{1}{T^{p/2}}S_{A_i^{k-1}}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}\right)-\frac{1}{T^{p/2}}S_{A_i^{k-2}}\left(\lfloor\boldsymbol{s}\rfloor_{2^k}\right)\right)\right|\geq\sum_{k=k_0}^{m-1}2^{k\left(\frac{p}{2r}-\frac{1}{4}\right)}\right)\\
&\leq \mathrm{P}\left(\bigcup_{k=k_0}^{m-1}\left\{\sup_{\boldsymbol{s}\in[0,1]^p}\left|\frac{1}{T^{p/2}}S_{A_i^{k-1}}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}\right)-\frac{1}{T^{p/2}}S_{A_i^{k-2}}\left(\lfloor\boldsymbol{s}\rfloor_{2^k}\right)\right|\geq 2^{k\left(\frac{p}{2r}-\frac{1}{4}\right)}\right\}\right)\\
&\leq \sum_{k=k_0}^{m-1}\mathrm{P}\left(\sup_{\boldsymbol{s}\in[0,1]^p}\left|\frac{1}{T^{p/2}}S_{A_i^{k-1}}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}\right)-\frac{1}{T^{p/2}}S_{A_i^{k-2}}\left(\lfloor\boldsymbol{s}\rfloor_{2^k}\right)\right|\geq 2^{k\left(\frac{p}{2r}-\frac{1}{4}\right)}\right)\\
&\leq \sum_{k=k_0}^{m-1}\sum_{\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}:\ \boldsymbol{s}\in[0,1]^p}\mathrm{P}\left(\left|\frac{1}{T^{p/2}}S_{A_i^{k-1}}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}\right)-\frac{1}{T^{p/2}}S_{A_i^{k-2}}\left(\lfloor\boldsymbol{s}\rfloor_{2^k}\right)\right|\geq 2^{k\left(\frac{p}{2r}-\frac{1}{4}\right)}\right)\\
&\leq \sum_{k=k_0}^{m-1}\left(2^{k+1}+1\right)^p C_r\frac{\left(T^p\lambda\left(A^{k-2}\left(\lfloor\boldsymbol{s}\rfloor_{2^k}\right)\setminus A^{k-1}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k+1}}\right)\right)+O(T^{p-1})\right)^{r/2}}{T^{pr/2}2^{kr\left(\frac{p}{2r}-\frac{1}{4}\right)}}\\
&\leq C_2\sum_{k=k_0}^{m-1}2^{kp}\cdot 2^{-kr/2}\cdot 2^{-kr\left(\frac{p}{2r}-\frac{1}{4}\right)}=C_2\sum_{k=k_0}^{m-1}2^{\frac{2p-r}{4}k}\leq C_2\sum_{k=k_0}^{\infty}2^{\frac{2p-r}{4}k}\to 0 \qquad (10.13)
\end{aligned}
$$

as $\delta\to 0$ (and therefore $k_0\to\infty$) since $r>2p$.

For (10.9), it holds for $\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta$ and due to $\delta\leq 2^{-k_0}$ by definition of k_0 that

$$
\begin{aligned}
&\|\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}-\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\|_\infty\leq\|\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}-\boldsymbol{s}\|_\infty+\|\boldsymbol{s}-\boldsymbol{t}\|_\infty+\|\boldsymbol{t}-\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\|_\infty\\
&\leq 3\cdot 2^{-k_0}\leq 2^{-(k_0-2)}. \qquad (10.14)
\end{aligned}
$$

since by definition of k_0, $2^{-(k_0+1)}<\delta\leq 2^{-k_0}$. Therefore, by Lemma C.1 (c), there exists

$$
\boldsymbol{x}\in\left[\lfloor\boldsymbol{s}\rfloor_{2^{k_0}},\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}+2^{-(k_0-2)}\right]\cap\left[\lfloor\boldsymbol{t}\rfloor_{2^{k_0}},\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}+2^{-(k_0-2)}\right]
$$

and by definition of A^k, it holds that $A(\boldsymbol{x})\subset A^{k_0-2}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}\right)\cap A^{k_0-2}\left(\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\right)$. Therefore, it follows by Assumption 10.1 (b) that

$$
\begin{aligned}
&\lambda\left(A^{k_0-2}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}\right)\Delta A^{k_0-2}\left(\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\right)\right)\\
&=\lambda\left(A^{k_0-2}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}\right)\right)+\lambda\left(A^{k_0-2}\left(\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\right)\right)-2\lambda\left(A^{k_0-2}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}\right)\cap A^{k_0-2}\left(\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\right)\right)\\
&\leq\lambda\left(A^{k_0-2}\left(\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}\right)\right)+\lambda\left(A^{k_0-2}\left(\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\right)\right)-2\lambda\left(A(\boldsymbol{x})\right)\leq C^*2^{-k_0}
\end{aligned}
$$

with a suitable constant $C^*>0$. Analogously to (10.11), we now consider the supremum over $\boldsymbol{s},\boldsymbol{t}\in[0,1]^p$ with $\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta$ of the below expression that only depends on $\boldsymbol{s},\boldsymbol{t}$ via $\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}$ and $\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}$. Thus, it is by (10.14) sufficient to take the supremum over $\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}$ with $\boldsymbol{s}\in[0,1]^p$ and over all $\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}$ with $\|\lfloor\boldsymbol{s}\rfloor_{2^{k_0}}-\lfloor\boldsymbol{t}\rfloor_{2^{k_0}}\|\leq$

$3 \cdot 2^{-k_0}$ for fixed $\lfloor \boldsymbol{s} \rfloor_{2^{k_0}}$.
The first supremum is taken over the set $\{\lfloor \boldsymbol{s} \rfloor_{2^{k_0}} | \boldsymbol{s} \in [0,1]^p\}$, which has cardinality $(2^{k_0}+1)^p \leq 2^p 2^{pk_0}$. The second supremum is taken over a set that by Lemma C.1 (a) has finite cardinality as the number of points of the form $\boldsymbol{k}2^{-k_0}$, $\boldsymbol{k} \in \mathbb{Z}^p$ in the hypercube $\left[\lfloor \boldsymbol{s} \rfloor_{2^{k_0}} - 3 \cdot 2^{-k_0}, \lfloor \boldsymbol{s} \rfloor_{2^{k_0}} + 3 \cdot 2^{-k_0}\right]$ is finite. Thus we obtain for $x > 0$ with σ-subadditivity and suitable constants $C_r, C_2 > 0$ that

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \delta} \left| \frac{1}{T^{p/2}} S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{s} \rfloor_{2^{k_0}}\right) - \frac{1}{T^{p/2}} S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{t} \rfloor_{2^{k_0}}\right) \right| \geq x \right) \\
\leq & \sum_{\lfloor \boldsymbol{s} \rfloor_{2^{k_0}} :\ \boldsymbol{s} \in [0,1]^p} \sum_{\substack{\lfloor \boldsymbol{t} \rfloor_{2^{k_0}} :\ \boldsymbol{t} \in [0,1]^p, \\ \left\| \lfloor \boldsymbol{s} \rfloor_{2^{k_0}} - \lfloor \boldsymbol{t} \rfloor_{2^{k_0}} \right\|_\infty \leq 3 \cdot 2^{-k_0}}} \mathrm{P}\left(\frac{1}{T^{p/2}} \left| S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{s} \rfloor_{2^{k_0}}\right) - S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{t} \rfloor_{2^{k_0}}\right) \right| \geq x \right) \\
\leq & 2^{pk_0} C_r \frac{\left(T^p \lambda\left(A^{k_0-2}\left(\lfloor \boldsymbol{s} \rfloor_{2^{k_0}}\right) \Delta A^{k_0-2}\left(\lfloor \boldsymbol{t} \rfloor_{2^{k_0}}\right)\right) + O(T^{p-1}) \right)^{r/2}}{T^{pr/2} x^r} \\
\leq & C_2 2^{pk_0} 2^{-k_0 r/2} = C_2 2^{k_0(p-r/2)} \to 0
\end{aligned} \tag{10.15}
$$

as $\delta \to 0$ (and therefore $k_0 \to \infty$) since $r > 2p$.

By subadditivity and (10.11)-(10.13) and (10.15), we obtain for any $x > 0$ that

$$
\begin{aligned}
&\lim_{\delta \to 0} \limsup_{T \to \infty} \mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \delta} \left| \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_T\right) - \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{t} \rfloor_T\right) \right| \geq 7x \right) \\
\leq & \lim_{\delta \to 0} \limsup_{T \to \infty} 2\,\mathrm{P}\left(\sup_{\boldsymbol{s} \in [0,1]^p} \left| \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_T\right) - \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) \right| \geq x \right) \\
&+ \lim_{\delta \to 0} \limsup_{T \to \infty} 2\,\mathrm{P}\left(\sup_{\boldsymbol{s} \in [0,1]^p} \left| \frac{1}{T^{p/2}} S_{A_i}\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) - \frac{1}{T^{p/2}} S_{A_i^{m-2}}\left(\lfloor \boldsymbol{s} \rfloor_{2^m}\right) \right| \geq x \right) \\
&+ \lim_{\delta \to 0} \limsup_{T \to \infty} 2\,\mathrm{P}\left(\sup_{\boldsymbol{s} \in [0,1]^p} \left| \sum_{k=k_0}^{m-1} \frac{1}{T^{p/2}} S_{A_i^{k-1}}\left(\lfloor \boldsymbol{s} \rfloor_{2^{k+1}}\right) - \frac{1}{T^{p/2}} S_{A_i^{k-2}}\left(\lfloor \boldsymbol{s} \rfloor_{2^k}\right) \right| \geq x \right) \\
&+ \lim_{\delta \to 0} \limsup_{T \to \infty} \mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \delta} \left| \frac{1}{T^{p/2}} S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{s} \rfloor_{2^{k_0}}\right) - \frac{1}{T^{p/2}} S_{A_i^{k_0-2}}\left(\lfloor \boldsymbol{t} \rfloor_{2^{k_0}}\right) \right| \geq x \right) = 0
\end{aligned}
$$

thus showing the assertion. □

With this Theorem, we can prove a functional central limit theorem for $\left(F\left(\mathbb{S}_T(\boldsymbol{s})\right)\right)_{\boldsymbol{s} \in [0,1]^p}$ on $\mathcal{D}([0,1]^p)$, where F is a Lipschitz-continuous function.

Theorem 10.3. *Let $(\epsilon_{\boldsymbol{k}})_{\boldsymbol{k} \in \mathbb{Z}^p}$ be a sequence of i.i.d. random variables with $\mathbb{E}[\epsilon_{\boldsymbol{k}}] = 0$, $\mathbb{E}[\epsilon_{\boldsymbol{k}}^2] = \sigma^2 \in (0, \infty)$ and $\mathbb{E}[|\epsilon_{\boldsymbol{k}}|^r] < \infty$ for some $r > 2p$. Let $Y_{\boldsymbol{k},T} = \mu_{\boldsymbol{k},T} + \epsilon_{\boldsymbol{k}}$ with $\mu_{\boldsymbol{k},T} \in \mathbb{R}$, define $S_A(\lfloor \boldsymbol{s} \rfloor_T) = S_A(\epsilon; \lfloor \boldsymbol{s} \rfloor_T)$ as in Definition 10.1 and let*

$$
\mathbb{S}_T(\boldsymbol{s}) = \left(\frac{1}{T^{p/2}} S_{A_1}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, \frac{1}{T^{p/2}} S_{A_P}(\lfloor \boldsymbol{s} \rfloor_T) \right)'.
$$

Furthermore, let $F : \mathbb{R}^P \to \mathbb{R}$ *be a Lipschitz-continuous function with Lipschitz-constant* $C_F > 0$*, i. e.*

$$|F(\boldsymbol{x}) - F(\boldsymbol{y})| \leq C_F \left\|\boldsymbol{x} - \boldsymbol{y}\right\|_\infty$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^p$*. Then, there exists a* P*-dimensional centered Gaussian process* $(Z(\boldsymbol{s}))_{\boldsymbol{s}\in[0,1]^p} = ((Z_1(\boldsymbol{s}), \ldots, Z_P(\boldsymbol{s}))')_{\boldsymbol{s}\in[0,1]^p}$ *with*

$$\operatorname{Cov}\left[Z_i(\boldsymbol{s}), Z_j(\boldsymbol{t})\right] = \sigma^2 \lambda\left(A_i(\boldsymbol{s}) \cap A_j(\boldsymbol{t})\right)$$

such that

$$\left(F\left(\mathbb{S}_T(\boldsymbol{s})\right)\right)_{\boldsymbol{s}\in[0,1]^p} \xrightarrow{w} \left(F\left(Z(\boldsymbol{s})\right)\right)_{\boldsymbol{s}\in[0,1]^p}$$

on $\mathcal{D}([0,1]^p)$ *equipped with the topology induced by the maximum norm on* $\mathcal{D}([0,1]^p)$ *(called* U-topology *by Wichura (1969), compare Definition A.2).*

Proof. By Theorem A.6, we need to show (i) that for all $n \in \mathbb{N}$, $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^p$

$$\left(F\left(\mathbb{S}_T(\boldsymbol{t}_1)\right), \ldots, F\left(\mathbb{S}_T(\boldsymbol{t}_n)\right)\right)' \xrightarrow{\mathcal{D}} \left(F(Z(\boldsymbol{t}_1)), \ldots, F(Z(\boldsymbol{t}_n))\right)'$$

and (ii) for all $x > 0$, it holds that

$$\lim_{\delta\to 0} \limsup_{T\to\infty} \mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta} \left|F\left(\mathbb{S}_T(\boldsymbol{s})\right) - F\left(\mathbb{S}_T(\boldsymbol{t})\right)\right| \geq x\right) = 0.$$

Proof of (i): By Theorem 10.2 it holds that

$$\left(\mathbb{S}_T(\boldsymbol{t}_1), \ldots \mathbb{S}_T(\boldsymbol{t}_n)\right) \xrightarrow{\mathcal{D}} \left(Z(\boldsymbol{t}_1), \ldots, Z(\boldsymbol{t}_n)\right).$$

Denote for $n \in \mathbb{N}$, $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^p$ by $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n} : \left(\mathcal{D}\left([0,1]^p\right)\right)^P \to \left(\mathbb{R}^P\right)^n$, $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(f) = (f(\boldsymbol{t}_1), \ldots, f(\boldsymbol{t}_n))'$ the projection mapping from $\left(\mathcal{D}\left([0,1]^p\right)\right)^P$ to $\left(\mathbb{R}^P\right)^n$. Since the U-topology by Wichura (1969) is induced by the maximum norm on $\mathcal{D}([0,1]^p)$, it follows by Lemma A.7 that $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(\cdot)$ is continuous if $\left(\mathcal{D}\left([0,1]^p\right)\right)^P$ is equipped with a proper metric. Furthermore, as F is continuous, $F^{(n)} : \left(\mathbb{R}^P\right)^n \to \mathbb{R}^n$, $F^{(n)}(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n) = (F(\boldsymbol{x}_1), \ldots, F(\boldsymbol{x}_n))$ is continuous as well. Since $F^{(n)} \circ \pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}$ as a composition of continuous functions is continuous, it follows by the continuous mapping theorem that

$$\begin{aligned}
&\left(F\left(\mathbb{S}_T(\boldsymbol{t}_1)\right), \ldots, F\left(\mathbb{S}_T(\boldsymbol{t}_n)\right)\right)' = \left(F^{(n)} \circ \pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}\right)\left(\mathbb{S}_T(\boldsymbol{t})\right) \\
&\xrightarrow{\mathcal{D}} \left(F^{(n)} \circ \pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}\right)\left(Z(\boldsymbol{t})\right) = \left(Z(\boldsymbol{t}_1), \ldots, Z(\boldsymbol{t}_n)\right)'
\end{aligned}$$

as $T \to \infty$.

Proof of (ii) : By the Lipschitz-continuity of F, we have for $x > 0$ that

$$\mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta} \left|F\left(\mathbb{S}_T(\boldsymbol{s})\right) - F\left(\mathbb{S}_T(\boldsymbol{t})\right)\right| \geq x\right) \leq \mathrm{P}\left(\sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty<\delta} \left\|\mathbb{S}_T(\boldsymbol{s}) - \mathbb{S}_T(\boldsymbol{t})\right\|_\infty \geq \frac{x}{C_F}\right)$$

and the assertion follows immediately from Theorem 10.2 (ii). □

Lemma 10.4. *Bounded convex sets $A \subset \mathbb{R}^2$ with $0 < \lambda(A) < \infty$ fulfill Assumption 10.1.*

Proof. (a) is fulfilled by Lemma C.3.
For (b) let $\|\boldsymbol{s} - \boldsymbol{t}\|_\infty < \gamma$ and denote for $\delta > 0$

$$A^\delta = \left\{ \boldsymbol{y} \in \mathbb{R}^2 \,\middle|\, \inf_{\boldsymbol{x} \in A} \|\boldsymbol{x} - \boldsymbol{y}\| < \delta \right\}.$$

Since $\|\boldsymbol{s} - \boldsymbol{t}\| < \sqrt{2}\gamma$ and $A_{\boldsymbol{s}}, A_{\boldsymbol{t}}$ can be obtained from each other by shift, it follows that $A_{\boldsymbol{s}} \subset A_{\boldsymbol{t}}^{\sqrt{2}\gamma}$ and $A_{\boldsymbol{t}} \subset A_{\boldsymbol{s}}^{\sqrt{2}\gamma}$. Therefore, we obtain by Lemma C.2 that

$$\lambda\left(\bigcup_{\boldsymbol{y} \in [0,\gamma)^2} A_{\boldsymbol{y}} \right) - \lambda(A) \leq \lambda\left(A^{\sqrt{2}\gamma}\right) - \lambda\left(A\right) = \ell(\partial A) \cdot \sqrt{2}\gamma + \pi \cdot 2\gamma^2 \leq C\gamma$$

with suitable $C > 0$, since obviously, $\ell(\partial A) < \infty$ as A is bounded (compare also Billingsley (1999), Appendix M17 equation (29)). □

11 Moving window procedure for the detection of fissures in concrete

Our goal is to identify areas that potentially contain fissures and eliminate bubbles in concrete data (as displayed in Figure 8.1). As already motivated in Section 9, a key feature of a fissure is that locally, the contrast between a rectangle covering a part of the fissure and its' neighboring environment is large in the sense that there is a large difference in the gray values. On the other hand, a key feature of the edge of a bubble is that the contrast between two equally large neighboring environments (e. g. hemispheres) is large. This idea is illustrated in the upper right panel of Figure 9.1 and in the upper left panel of Figure 11.1.
Based on these observations, we choose suitable convex sets in $\mathbb{R}^2$. We define statistics similar to Definition 10.1 on these sets and combine them to a single statistic in order to estimate the location of the fissures (for an illustration of these sets, see Figure 11.1 below): Let A be a circle of diameter $d \in (0,1)$ centered around 0. For some angle $\alpha \in [0, \pi)$, we split this circle by having a 'strip' of width $h \in (0, d)$ turned by α (see Figure 11.1 (a)). We denote this inner 'strip' by $A^{(1,\alpha)}$ and the two remaining segments by $A^{(2,\alpha)}$ and $A^{(3,\alpha)}$, respectively (see Figure 11.1 (b)). In a second step, we split A into semicircles $A^{(4,\alpha)}$ and $A^{(5,\alpha)}$ turned by α (see Figure 11.1 (c)). Denote by $(Y_{\boldsymbol{k},T})_{\boldsymbol{k} \in \{1,\ldots,T\}^2}$ the process of gray values and let $S_A\left(\lfloor \boldsymbol{s} \rfloor_T\right) = S_A\left(Y; \lfloor \boldsymbol{s} \rfloor_T\right)$ be as in Definition 10.1. For $i = 1, \ldots, 5$, $s \in [d/2, 1 - d/2]^2$ define

$$\begin{aligned}
\overline{S}_{A^{(i,\alpha)}}\left(\lfloor \boldsymbol{s} \rfloor_T\right) = \overline{S}_{A^{(i,\alpha)}}\left(Y; \lfloor \boldsymbol{s} \rfloor_T\right) &= \frac{T}{\left|\left\{ \frac{\boldsymbol{k}}{T} \in A^{(i,\alpha)}\left(\lfloor \boldsymbol{s} \rfloor_T\right) \right\}\right|} S_{A^{(i,\alpha)}}\left(\lfloor \boldsymbol{s} \rfloor_T\right) \\
&= \frac{T}{\left|\left\{ \frac{\boldsymbol{k}}{T} \in A^{(i,\alpha)}\left(\lfloor \boldsymbol{s} \rfloor_T\right) \right\}\right|} \sum_{\frac{\boldsymbol{k}}{T} \in A^{(i,\alpha)}\left(\lfloor \boldsymbol{s} \rfloor_T\right)} Y_{\boldsymbol{k},T}.
\end{aligned}$$

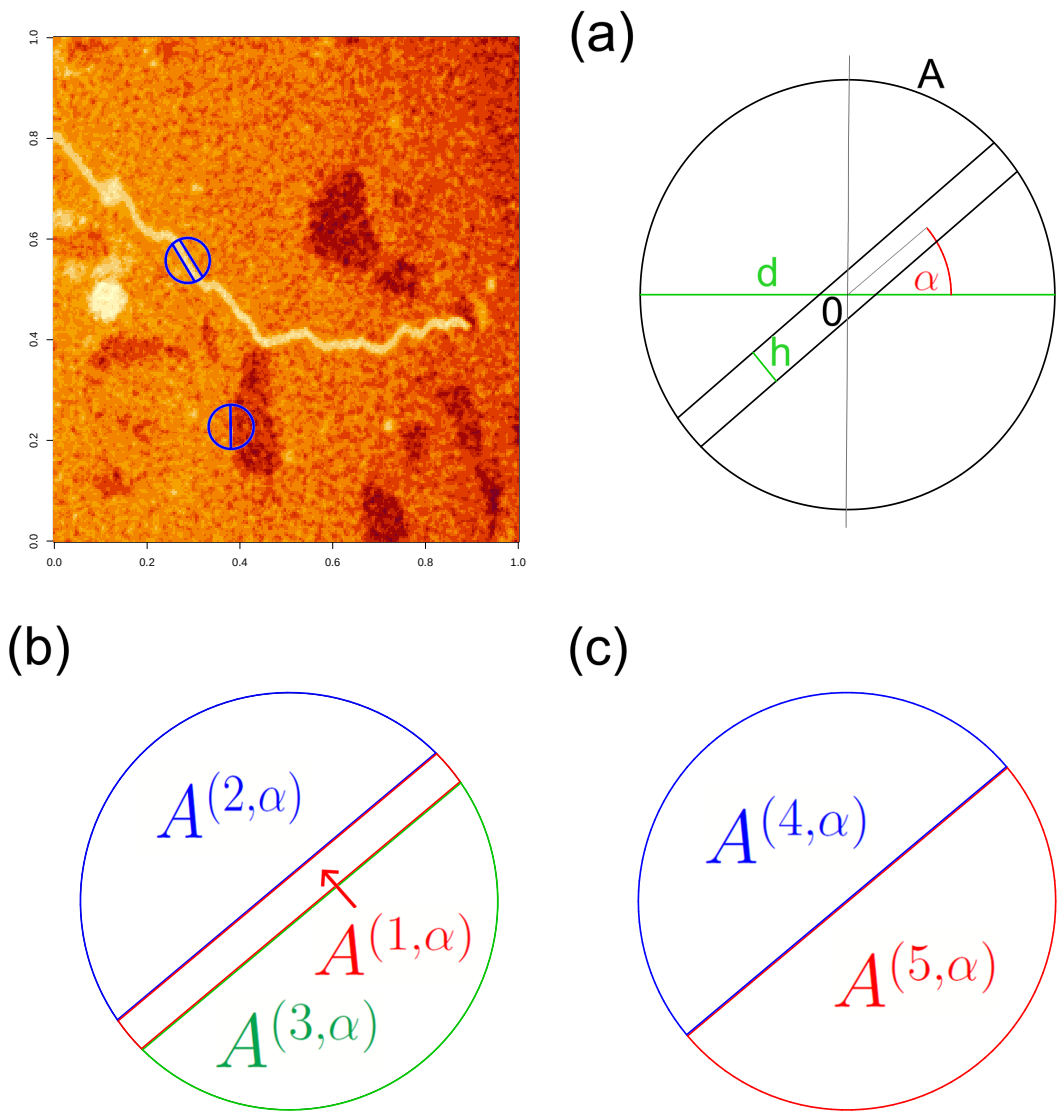

Figure 11.1: In the upper left panel, a 2D slice of a 3D CT scan of a concrete block with an artificially added fissure is displayed. The blue split circles display characteristics of fissures and the edges of bubbles, respectively. Panels (a)-(c) illustrate the sets that we use in our scan statistic introduced in (11.1) - (11.3).

Corollary 11.1. *Let* $Y_{\boldsymbol{k},T} = \mu_{\boldsymbol{k},T} + \epsilon_{\boldsymbol{k}}$, *where* $\mu_{\boldsymbol{k},T} \in \mathbb{R}$ *and* $(\epsilon_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{N}^2}$ *is a sequence of i.i.d. random variables with* $\mathbb{E}[\epsilon_{\boldsymbol{k}}] = 0$, $\mathbb{E}[\epsilon_{\boldsymbol{k}}] = \sigma^2 \in (0,\infty)$ *and* $\mathbb{E}[|\epsilon_{\boldsymbol{k}}|^r] < \infty$ *for some* $r > 4$. *Then it holds uniformly in* $\boldsymbol{s} \in [d/2, 1-d/2]^2$ *that for all* $i = 1,\dots,5$,

$$\overline{S}_{A^{(i,\alpha)}}(\epsilon; \lfloor \boldsymbol{s} \rfloor_T) = \frac{1}{T\lambda\left(A^{(i,\alpha)}\right)} S_{A^{(i,\alpha)}}(\epsilon; \lfloor \boldsymbol{s} \rfloor_T) + o_P(1).$$

Proof. By Theorem 10.2, it holds that

$$\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2}\left|\frac{1}{T}S_A\left(\epsilon;\lfloor\boldsymbol{s}\rfloor_T\right)\right| = O_P(1).$$

By Lemma C.3, it holds that

$$\left|T^2\lambda(A^{(i,\alpha)}) - \left|\left\{\frac{\boldsymbol{k}}{T}\in A^{(i,\alpha)}\left(\lfloor\boldsymbol{s}\rfloor_T\right)\right\}\right|\right| = O_P(T).$$

Therefore, it holds that

$$\begin{aligned}
&\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2}\left|\overline{S}_A\left(\epsilon;\lfloor\boldsymbol{s}\rfloor_T\right) - \frac{1}{T\lambda\left(A^{(i,\alpha)}\right)}S_A\left(\epsilon;\lfloor\boldsymbol{s}\rfloor_T\right)\right| \\
\leq &\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2}\left|\frac{T^2}{\left|\left\{\frac{\boldsymbol{k}}{T}\in A^{(i,\alpha)}\left(\lfloor\boldsymbol{s}\rfloor_T\right)\right\}\right|} - \frac{1}{\lambda\left(A^{(i,\alpha)}\right)}\right|\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2}\left|\frac{1}{T}S_A\left(\epsilon;\lfloor\boldsymbol{s}\rfloor_T\right)\right| \\
= &\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2}\frac{\left|T^2\lambda\left(A^{(i,\alpha)}\right) - \left|\left\{\frac{\boldsymbol{k}}{T}\in A^{(i,\alpha)}\left(\lfloor\boldsymbol{s}\rfloor_T\right)\right\}\right|\right|}{\lambda\left(A^{(i,\alpha)}\right)\left|\left\{\frac{\boldsymbol{k}}{T}\in A^{(i,\alpha)}\left(\lfloor\boldsymbol{s}\rfloor_T\right)\right\}\right|}\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2}\left|\frac{1}{T}S_A\left(\epsilon;\lfloor\boldsymbol{s}\rfloor_T\right)\right| \\
= &\, O_P\left(\frac{T}{T^2}\right)\cdot O_P(1) = O_P\left(\frac{1}{T}\right).
\end{aligned}$$

□

Since a fissure can change directions, we need to consider multiple angles when trying to detect the fissure. Since $A^{(1,\alpha)} = A^{(1,\alpha+k\pi)}$, $A^{(2,\alpha)} = A^{(3,\alpha+(2k+1)\pi)}$ and $A^{(4,\alpha)} = A^{(5,\alpha+(2k+1)\pi)}$ for all $k\in\mathbb{Z}$, we only need to consider angles in $[0,\pi)$ in the following statistic. Therefore, for $T>0$, $0\leq\alpha_1<\ldots<\alpha_P<\pi$, $\boldsymbol{s}\in[d/2,1-d/2]^2$, define

$$\begin{aligned}
\overline{S}^{(12,\alpha)}(\boldsymbol{s}) &= \overline{S}_{A^{(1,\alpha)}}\left(\lfloor\boldsymbol{s}\rfloor_T\right) - \overline{S}_{A^{(2,\alpha)}}\left(\lfloor\boldsymbol{s}\rfloor_T\right) \\
\overline{S}^{(13,\alpha)}(\boldsymbol{s}) &= \overline{S}_{A^{(1,\alpha)}}\left(\lfloor\boldsymbol{s}\rfloor_T\right) - \overline{S}_{A^{(3,\alpha)}}\left(\lfloor\boldsymbol{s}\rfloor_T\right) \\
\overline{S}^{(45,\alpha)}(\boldsymbol{s}) &= \overline{S}_{A^{(4,\alpha)}}\left(\lfloor\boldsymbol{s}\rfloor_T\right) - \overline{S}_{A^{(5,\alpha)}}\left(\lfloor\boldsymbol{s}\rfloor_T\right) \\
M_1(\boldsymbol{s}) = M_1(\boldsymbol{s},T) &= \max_{i=1\ldots,P}\frac{1}{\hat{\sigma}_{1,i,T}(\boldsymbol{s})}\min\left\{\left|\overline{S}^{(12,\alpha_i)}(\boldsymbol{s})\right|,\left|\overline{S}^{(13,\alpha_i)}(\boldsymbol{s})\right|\right\} \qquad (11.1)\\
M_2(\boldsymbol{s}) = M_2(\boldsymbol{s},T) &= \max_{i=1\ldots,P}\frac{1}{\hat{\sigma}_{2,i,T}(\boldsymbol{s})}\left|\overline{S}^{(45,\alpha_i)}(\boldsymbol{s})\right| \qquad (11.2)\\
M(\boldsymbol{s}) = M(\boldsymbol{s},T) &= \max\left\{M_1(\boldsymbol{s}) - M_2(\boldsymbol{s}),0\right\} \qquad (11.3)
\end{aligned}$$

where $\hat{\sigma}_{1,i,T},\hat{\sigma}_{2,i,T}$ are suitable estimators for σ. In the simulation study in Section 12, we consider three types of variance estimators: One 'minimum'-type estimator similar to (6.1), but properly adapted to our situation, one local robust estimator and one global robust estimator. We refer to Section 12 for more details.
M consists of two separate terms: For a given angle α in (11.1), we compare the (rescaled) average in the inner 'strip' $A^{(1,\alpha)}$ with the (rescaled) averages in the two

sections $A^{(2,\alpha)}$ and $A^{(3,\alpha)}$ and then take the minimum of the two differences. Subsequently, we maximize over all considered angles $\alpha_1, \ldots, \alpha_P$. As illustrated in the second to left column of Figure 11.2, this type of statistic enhances fissures and to a lesser extent the edges of bubbles and fissures. On the other hand, in (11.2), we compare the (rescaled) averages of the two semicircles $A^{(4,\alpha)}$ and $A^{(5,\alpha)}$ and maximize over $\alpha_1, \ldots, \alpha_P$. As illustrated in the second to right column of Figure 11.2, this type of statistic enhances the edges of the bubbles and fissures.
Therefore, as illustrated in the right column of Figure 11.2, a combination of $M_1(\boldsymbol{s})$ from (11.1) and $M_1(\boldsymbol{s})$ from (11.2) to $M(\boldsymbol{s})$ from (11.3) leads to an enhancement of fissures and elimination of most of the other parts of the image. Since by Lemma D.1 all functions used are Lipschitz-continuous, we can apply Theorems and 10.2 and 10.3 to obtain a limit process for $(M(\boldsymbol{s}, T))_{\boldsymbol{s}\in[d/2,1-d/2]^2}$ in the case of no anomalies:

Theorem 11.2. *Let $(\epsilon_{\boldsymbol{k}})_{\boldsymbol{k}\in\mathbb{Z}^2}$ be a sequence of i.i.d. random variables with $\mathbb{E}[\epsilon_{\boldsymbol{k}}] = 0$, $\mathbb{E}[\epsilon_{\boldsymbol{k}}^2] = \sigma^2 \in (0,\infty)$ and let $\mathbb{E}[|\epsilon_{\boldsymbol{k}}|^r] < \infty$ for some $r > 4$. Let $Y_{\boldsymbol{k},T} = \mu_{\boldsymbol{k},T} + \epsilon_{\boldsymbol{k}}$ with $\mu_{\boldsymbol{k},T} \in \mathbb{R}$. Let $\hat{\sigma}_{1,i,T}(\boldsymbol{s})$, $\hat{\sigma}_{2,i,T}(\boldsymbol{s})$ be uniformly consistent estimators for σ in the sense that*

$$\sup_{\boldsymbol{s}\in[d/2,1-d/2]^2} |\hat{\sigma}_{1,i,T}(\boldsymbol{s}) - \sigma| = o_P(1), \qquad \sup_{\boldsymbol{s}\in[d/2,1-d/2]^2} |\hat{\sigma}_{2,i,T}(\boldsymbol{s}) - \sigma| = o_P(1)$$

for all $i = 1, \ldots, P$. There exists a $5P$-dimensional centered Gaussian process

$$\left((W_{1,1}(\boldsymbol{s}), \ldots, W_{1,P}(\boldsymbol{s}), \ldots, W_{5,1}(\boldsymbol{s}), \ldots, W_{5,P}(\boldsymbol{s}))'\right)_{\boldsymbol{s}\in[d/2,1-d/2]^2}$$

with

$$\operatorname{Cov}[W_{i,j}(\boldsymbol{s}), W_{k,l}(\boldsymbol{t})] = \frac{\lambda\left(A^{(i,\alpha_j)}(\boldsymbol{s}) \cap A^{(k,\alpha_l)}(\boldsymbol{t})\right)}{\lambda\left(A^{(i,\alpha_j)}\right)\lambda\left(A^{(k,\alpha_l)}\right)}$$

such that in the case of $\mu_{\boldsymbol{k},T} = 0$ for all $\boldsymbol{k}, T$ and

$$\begin{aligned}\Gamma(\boldsymbol{s}) = \max\Big\{0, \max_{i=1,\ldots,P} \min\left\{|W_{1,i}(\boldsymbol{s}) - W_{2,i}(\boldsymbol{s})|, |W_{1,i}(\boldsymbol{s}) - W_{3,i}(\boldsymbol{s})|\right\} \\ - \max_{i=1,\ldots,P} |W_{4,i}(\boldsymbol{s}) - W_{5,i}(\boldsymbol{s})|\Big\},\end{aligned}$$

it holds that

$$(M(\boldsymbol{s}, T))_{\boldsymbol{s}\in[d/2,1-d/2]^2} \xrightarrow{w} (\Gamma(\boldsymbol{s}))_{\boldsymbol{s}\in[d/2,1-d/2]^2}$$

in $\mathcal{D}([d/2, 1-d/2]^2)$ as $T \to \infty$.

Proof. Since the $A^{(i,\alpha)} \subset \mathbb{R}^2$ are convex, they fulfill Assumption 10.1 by Lemma 10.4. Therefore, it holds with $\overline{S}_{A^{(i,\alpha_j)}}(\lfloor \boldsymbol{s} \rfloor_T) = \overline{S}_{A^{(i,\alpha_j)}}(\epsilon; \lfloor \boldsymbol{s} \rfloor_T)$ by Theorem 10.2 that

$$\begin{aligned}&\left(\overline{S}_{A^{(1,\alpha_1)}}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, \overline{S}_{A^{(1,\alpha_P)}}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, \overline{S}_{A^{(5,\alpha_1)}}(\lfloor \boldsymbol{s} \rfloor_T), \ldots, \overline{S}_{A^{(5,\alpha_P)}}(\lfloor \boldsymbol{s} \rfloor_T)\right)_{\boldsymbol{s}} \\ \xrightarrow{w}\ &\left(\sigma\left(W_{1,1}(\boldsymbol{s}), \ldots, W_{1,P}(\boldsymbol{s}), \ldots, W_{5,1}(\boldsymbol{s}), \ldots, W_{5,P}(\boldsymbol{s})\right)\right)_{\boldsymbol{s}}\end{aligned}$$

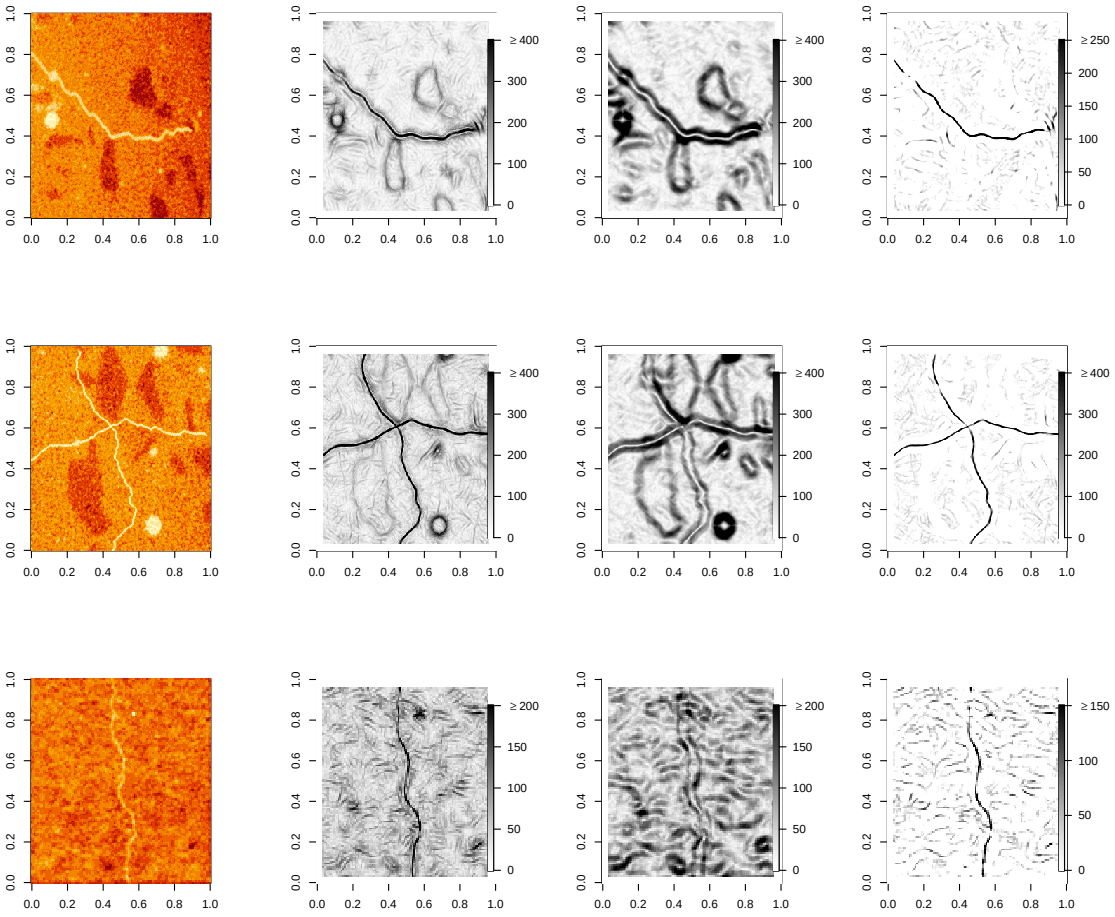

Figure 11.2: In the left column, various 2D slices of 3D CT scans from concrete blocks with fissures are shown. In the second to left column, the corresponding heat maps of (11.1) are displayed, showing that in these examples, (11.1) enhances fissures and to a lesser extent the edges of bubbles and fissures. In the second to right column, the corresponding heat maps of (11.2) are displayed, showing that in these examples, (11.2) enhances the edges of bubbles and fissures. In the right column, the corresponding heat maps of the combined statistic in (11.3) are displayed, showing that in these examples, the statistic enhances fissures and eliminates most of the noise.

in $(\mathcal{D}([d/2, 1-d/2]^2))^{5P}$. Since by Lemma D.1, maximum, minimum, addition, subtraction and the absolute value are Lipschitz-continuous and $\hat{\sigma}_{1,i,T}(\boldsymbol{s})$, $\hat{\sigma}_{2,i,T}(\boldsymbol{s})$ are uniformly consistent estimators for σ, the assertion follows by Theorem 10.3 and Slutzky's theorem. □

We need a threshold $\beta = \beta_{d,h,T}$ in order to determine areas that can potentially contain fissures. We distinguish between *significant* points $\boldsymbol{s} \in [d/2, 1-d/2]^2$, for

which

$$M(\boldsymbol{s}, T) \geq \beta_{d,h,T}$$

and *non-significant* points. The significant points can then be used by post-processing procedures like machine learning algorithms in order to estimate the exact paths of the fissures. Note that usually, if one point is significant, then a whole environment around this point is significant in analogy to Part I as a single pixel of a fissure influences multiple points of the statistic.
A common choice for the threshold is obtained as the asymptotic $(1 - \alpha_T)$-quantile of the maximum of the statistic defined in (11.3) for some sequence $\alpha_T \to 0$, which can be obtained by Theorem 11.2. In the simulation studies, we use $\alpha_T = 0.05$. This guarantees that all points above the threshold are significant at a global level α_T (in the usual testing sense), while also ensuring that the global false positive rate is at most α_T.
Note that to the best of our knowledge, the limit distribution in Theorem 11.2 cannot be obtained explicitly. Therefore, we estimated $\beta_{d,h,T}$ in the simulation study below by scanning fields with no anomalies with the above statistics.

12 Simulation study and Data analysis

If no prior knowledge on the direction of a fissure exists, it is natural to choose the angles $\alpha_1, \ldots, \alpha_P$ equidistantly on $[0°, 180°)$ as illustrated in Figure 12.1 in order to have a 'best worst-case scenario' for the misspecification of the true direction of a fissure. We want to know how many angles P are needed to detect a fissure at a 'reasonably high' rate. On the one hand, it is important to control P for computational efficiency, especially in large image data. On the other hand, parts of a fissure are enhanced better if the misspecification of the angle for that part is lower. In practice, with fissures changing direction, this typically means that the fissure is enhanced better with higher P as can be seen in Figure 12.2.
If $\alpha_1, \ldots, \alpha_P$ are chosen equidistantly on $[0°, 180°)$, the direction of a fissure can be misspecified by an angle of $(90/P)°$ in the worst case (see Figure 12.1). Now denote by Δ the difference between the angle of the inner strip and a (part of a) fissure. The fact that for P angles, the worst-case misspecification is $(90/P)°$ in turn implies that if the detection rates are 'reasonably high' for Δ, then at most $\lceil 90/\Delta \rceil$ angles are needed to a (part of a) fissure at a 'reasonably high' rate. Therefore, it is plausible to make assertions on the choice of P based on the detection rates w.r.t. the misspecification Δ.
Furthermore, it is important to have computationally efficient, consistent estimators $\hat{\sigma}_{1,i,T}(\boldsymbol{s})$ and $\hat{\sigma}_{2,i,T}(\boldsymbol{s})$ for σ in (11.1) and (11.2), respectively, as an overestimation of σ leads to worse detection rates, while an underestimation leads to too many spurious significant pixels. Furthermore, we want to have estimators that are robust against outliers as the real data contains bubbles. Thus, it is interesting to study the detection rates w.r.t. various choices of $\hat{\sigma}_{1,i,T}(\boldsymbol{s})$ and $\hat{\sigma}_{2,i,T}(\boldsymbol{s})$.

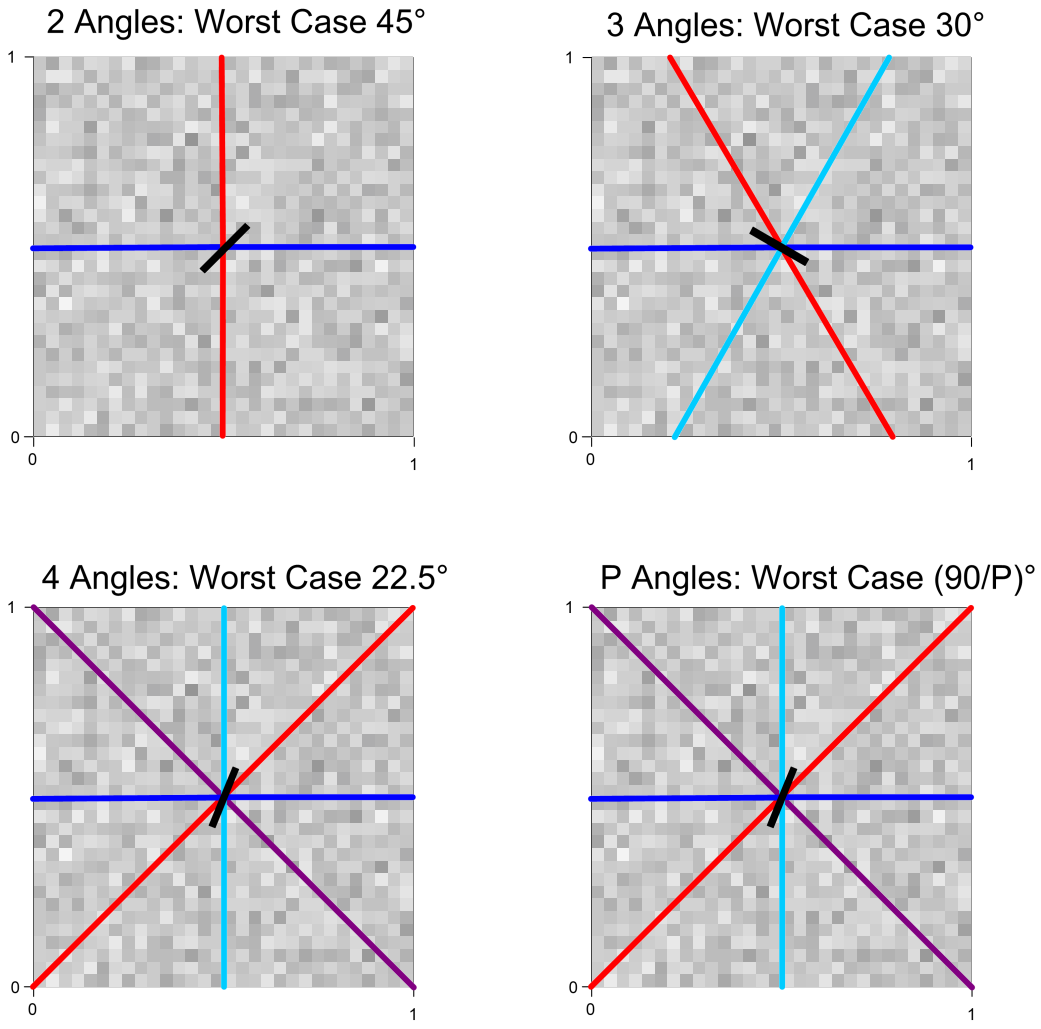

Figure 12.1: This graphic schematically illustrates a worst-case scenario for the direction of a fissure for various values of P if $\alpha_1, \ldots, \alpha_P$ are chosen equidistantly on $[0°, 180°)$. The fissure is schematically illustrated as a thin black rectangle, while the colored lines represent the angles of the scan sets from figure 11.1.

12.1 Data and method

We analyze images with 100×100 pixels containing a single fissure F_w of relative length 0.3 turned by 50° against the x-axis having relative widths of $w = 0.01, 0.02$, respectively (see the left two panels of Figure 12.3 and the whole Figure 12.4 for examples). We assume that the values of the pixels are given by $Y_{i,j} = \mu_{i,j} + \epsilon_{i,j}$, where $(\epsilon_{i,j})_{i,j=1,\ldots,100}$ is a sequence of i.i.d. standard-normally distributed random variables while $\mu_{i,j}$ is constant across the fissure and 0 outside of the fissure. We use the statistic $M(\boldsymbol{s})$ from (11.3) with a circle of diameter of 0.1 and a 'strip' of width h for a single angle α. Therefore, in the notation of (11.1) - (11.3), we have for given

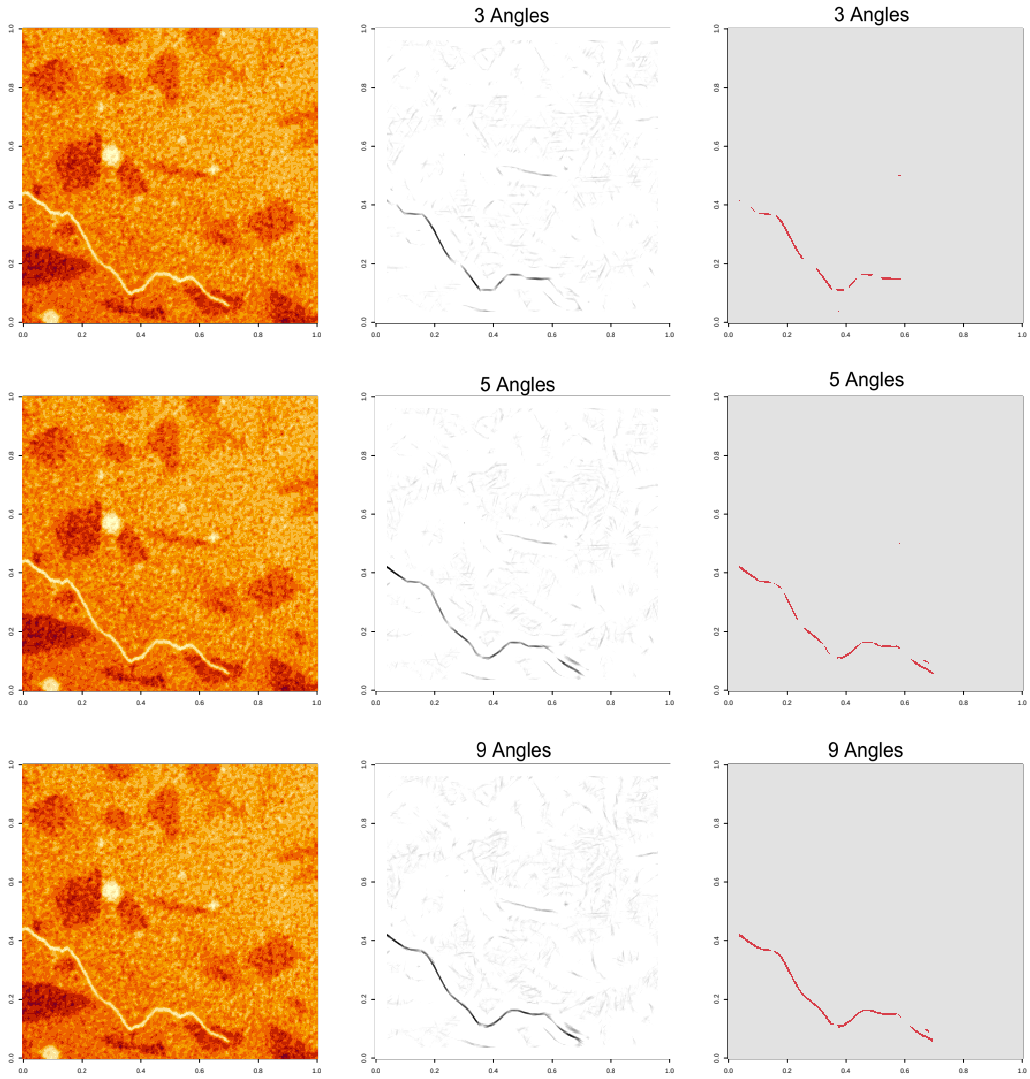

Figure 12.2: In the left column, a 2D slice of a 3D CT scan of a concrete block with fissures are displayed. In the middle column, the corresponding heat maps for 3, 5 and 9 angles from top to bottom of (11.3) are displayed. In the right column, the corresponding significant pixels for 3, 5 and 9 angles from top to bottom are displayed. As expected, the enhancement of the fissure and the number of significant pixels close to the fissure increase in the number of angles.

h and α that

$$\begin{aligned} & M(\boldsymbol{s}) = M(\boldsymbol{s}, h, \alpha) \\ = \; & \max\left\{0, \frac{1}{\hat{\sigma}_{1,\alpha}(\boldsymbol{s})} \min\left\{\left|\overline{S}^{(12,\alpha)}(\boldsymbol{s})\right|, \left|\overline{S}^{(13,\alpha)}(\boldsymbol{s})\right|\right\} - \frac{1}{\hat{\sigma}_{2,\alpha,T}(\boldsymbol{s})}\left\{\left|\overline{S}^{(45,\alpha)}(\boldsymbol{s})\right|\right\}\right\}. \end{aligned}$$

We consider the following parameters, where the angles are given in degrees instead of radians:

Name	variable	values
width of the fissure	w	$0.01, 0.02$
signal of the fissure	μ	$1, 1.5, 2, 2.5, 3$
width of inner strip	h	$0.01, 0.02, 0.03, 0.04$
misspecification of angle	Δ	$0, 5, \ldots, 25$

One of our main objectives is to determine, how well various estimators for the standard deviation perform at separating signal from noise. In order analyze this, we fix the rate of spurious significant pixels at 5% for each estimator $\hat{\sigma}$ by determining thresholds w.r.t. $\hat{\sigma}$: For given h, we determine the threshold $\beta_h(\hat{\sigma})$ via simulation: We generate 10 000 images with 100×100 i.i.d. standard-normally distributed pixels. Then we scan each of the fields with the above statistic $M(\boldsymbol{s}, h, 0)$ using $\hat{\sigma}$ as estimator for σ and choose $\beta_h(\hat{\sigma})$ as the empirical 95%-quantile of the maximal values of each image.
In order to analyze the performance of our procedure, we generate for each signal-width-combination (μ, w) 1 000 images containing a single fissure of width w and signal μ. Subsequently, we scan these images with $M(\boldsymbol{s}, h, \alpha)$ using $\hat{\sigma}$ as estimator for σ.

Definition 12.1. We call a pixel $\boldsymbol{k}/T$ *adjacent* to the fissure if the intersection between the circle $A(\boldsymbol{k}/T)$ centered around $\boldsymbol{k}/T$ and the fissure is non-empty. See the right side of Figure 12.3 for an illustration.
We determine the fissure to be *detected* if for at least one pixel $\boldsymbol{k}/T$ adjacent to the fissure, it holds that $M(\boldsymbol{k}/T, h, \alpha) \geq \beta_h(\hat{\sigma})$ when using $\hat{\sigma}$ as estimator for σ.

Note that this definition of 'detection' fits our objective of identifying regions that potentially contain fissures. Identifying a single pixel that is potentially influenced by a fissure is sufficient and can be used e. g. as the starting point for machine learning methods in order to trace the fissure.

12.2 Variance estimators

In this section, we study, how good various estimators for the standard deviations are at separating signal from noise. Denote for some set A by $\hat{\sigma}^2(A)$ the classic sample variance of the values $(Y_{\boldsymbol{k}})_{\boldsymbol{k}/T \in A}$ in A. Furthermore, denote by $\hat{q}_A(\cdot)$ the function of empirical quantiles of $(Y_{\boldsymbol{k}})_{\boldsymbol{k}/T \in A}$ and let $\Phi(\cdot)$ be the CDF of the standard normal distribution. Denote by $A^{(i,\alpha)}$ the circle segments introduced at the start of Section 11 (for a visual display see Figure 11.1).
Local minimum-type estimators have proven to yield good results in change point estimation (see e. g. Meier, Cho, and Kirch (2021)) and have also yielded good results in the simulation study in the first part of this thesis in Section 6. Therefore, we use a version of minimum-type estimators adapted to the two-parameter scenario:

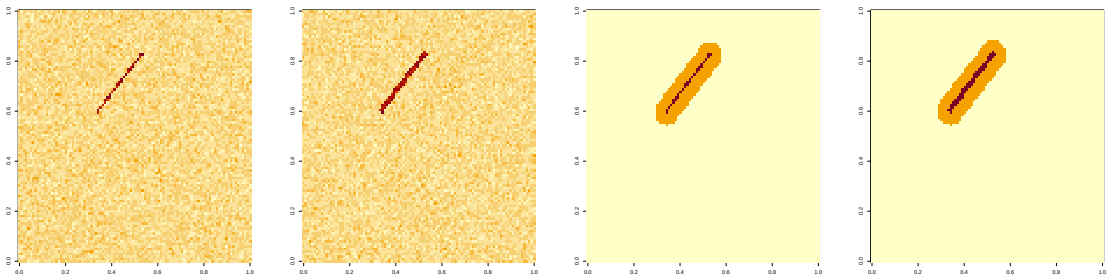

Figure 12.3: The two graphics on the left side show images of 100×100 pixel containing fissures of relative width 0.01 (left image) and 0.02 (right image) with large signal-to-noise-ratios. The two graphics on the right side schematically display the fissures and their 'adjacent' pixel – i. e. those for which intersection between the circle $A(\boldsymbol{k}/T)$ centered around $\boldsymbol{k}/T$ and the fissure is non-empty. The fissures are the dark red areas while the adjacent pixel are the ones colored in orange. We consider the fissure to be detected if the statistic is above the threshold for any of the red and orange pixel.

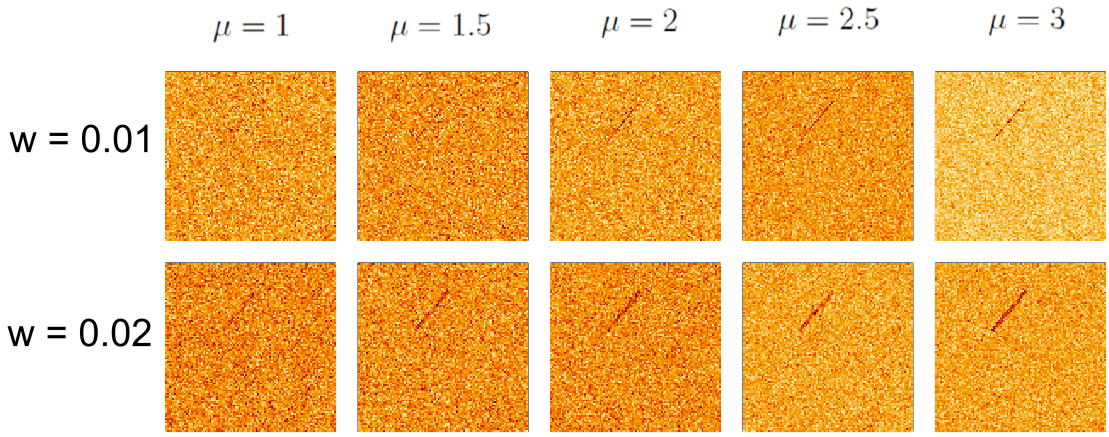

Figure 12.4: The graphic shows images with all signal-width-combinations (μ, w) that we consider in this simulation study.

(A) Minimum-type estimator:

$$\hat{\sigma}^2_{1,\alpha}(\boldsymbol{s}) = \min\left\{\hat{\sigma}^2(A^{(2,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T)), \hat{\sigma}^2(A^{(3,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T))\right\},$$
$$\hat{\sigma}^2_{2,\alpha}(\boldsymbol{s}) = \min\left\{\hat{\sigma}^2(A^{(4,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T)), \hat{\sigma}^2(A^{(5,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T))\right\}.$$

As already noticed in Section 6, one drawback of minimum-type estimators, is that they tend to find more spurious maxima than other types of estimators.

Since one of our main goals is the elimination of areas without fissures, while having to account for bubbles, we also analyze the performance of our procedure using a robust variance estimator, both on a local and on a global level. We use an estimator based on the empirical interquartile range that was initially introduced by Silverman (1986) as an estimator for the optimal bandwidth in kernel density estimation:

(B) Local robust estimator:

$$\hat{\sigma}_{1,\alpha}(\boldsymbol{s}) = \frac{\hat{q}_{A^{(2,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T) \cup A^{(3,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T)}\left(\frac{3}{4}\right) - \hat{q}_{A^{(2,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T) \cup A^{(3,\alpha)}(\lfloor \boldsymbol{s} \rfloor_T)}\left(\frac{1}{4}\right)}{\Phi^{-1}\left(\frac{3}{4}\right) - \Phi^{-1}\left(\frac{1}{4}\right)},$$

$$\hat{\sigma}_{2,\alpha}(\boldsymbol{s}) = \frac{\hat{q}_{A(\lfloor \boldsymbol{s} \rfloor_T)}\left(\frac{3}{4}\right) - \hat{q}_{A(\lfloor \boldsymbol{s} \rfloor_T)}\left(\frac{1}{4}\right)}{\Phi^{-1}\left(\frac{3}{4}\right) - \Phi^{-1}\left(\frac{1}{4}\right)}.$$

(C) Global robust estimator:

$$\hat{\sigma}_{1,\alpha}(\boldsymbol{s}) = \hat{\sigma}_{2,\alpha}(\boldsymbol{s}) = \frac{\hat{q}_{[0,1]^2}\left(\frac{3}{4}\right) - \hat{q}_{[0,1]^2}\left(\frac{1}{4}\right)}{\Phi^{-1}\left(\frac{3}{4}\right) - \Phi^{-1}\left(\frac{1}{4}\right)}.$$

In the context of independent, identically $\mathcal{N}(\mu, \sigma^2)$ random variables, Silverman's estimator $\hat{\sigma}_{IQR,n} = (\hat{F}_n^{-1}(3/4) - \hat{F}_n^{-1}(1/4))/(\Phi^{-1}(3/4) - \Phi^{-1}(1/4))$ is consistent: Due to the continuity of the normal distribution, and Glivenko-Cantelli's theorem, $\hat{F}_n^{-1}(p) \to F^{-1}_{\mathcal{N}(\mu,\sigma^2)}(p) = \mu + \sigma\Phi^{-1}(p)$ for all $p \in (0,1)$.
We compare the results for (A)-(C) with the results for (D), where we use the true variance $\sigma^2 = 1$.

As it can be seen in Figures 12.5 and 12.6, it shows that across all parameter combinations, the global robust estimator (C) performs best among the three estimators (A)-(C) and that the detection rates are virtually identical with the detection rates of the true variance (D). We can also see that in the non-trivial cases (i. e. detection rates away from 0 or 1), (C) usually performs a lot better than (A) and (B). It is straightforward to see, why the detection rates for (C) and (D) are virtually identical: Under the assumption of normality, (C) is consistent. In each image, 10 000 pixels are involved in the calculation of (C) and as (C) is robust against outliers, (C) is usually very close to the true variance. A possible explanation for the dominance of (C) over (A) and (B) is the massive discretization of the sets $A^{(i,\alpha)}$, as shown for $A^{(1,\alpha)} - A^{(3,\alpha)}$ in Figure E.1. We conjecture that due to the much higher volatility of (A) and (B), some pixels that are significant by estimator (C) are not significant anymore by (A) and (B).
For our purpose, it is advantageous that the detection rates are highest for (C) as (C) is a consistent, robust estimators of σ and due to being a global estimator, is computationally much more efficient than (A) and (B).

12.3 Analysis of the number of angles

We have elaborated in the introduction of this section, that it is plausible to choose the angles $\alpha_1, \ldots, \alpha_P$ equidistantly on the interval $[0°, 180°)$ and that in then, if for a misspecification of Δ, we obtain 'reasonably high' (up to the reader's preferences) detection rates, we need at most $P = \lceil 90/\Delta \rceil$ in order to detect the fissure at a 'reasonably high' rate. As can be seen in Figures 12.5 and 12.6 as well as in

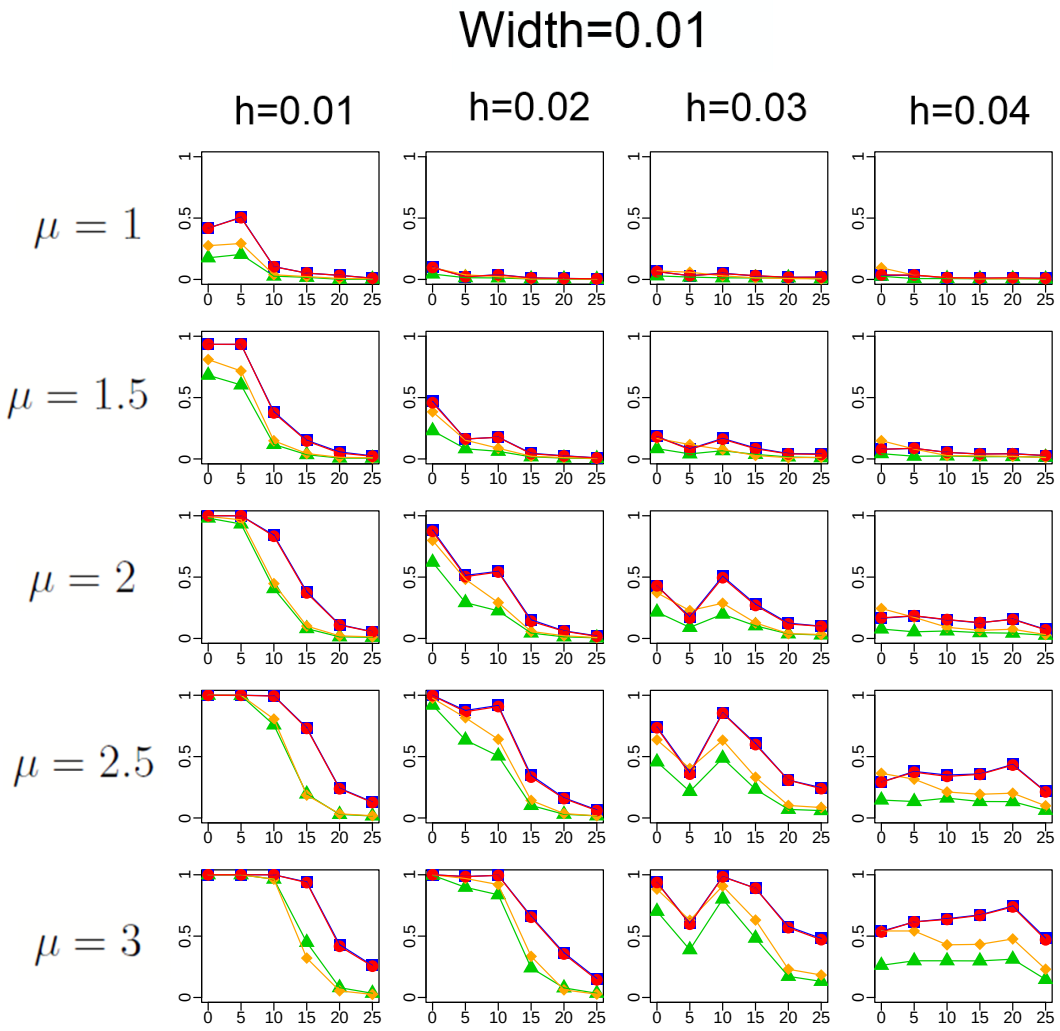

Figure 12.5: Comparison of the detection rates for all combinations of signal, bandwidth, width of the fissure and variance estimators. The yellow diamond-shaped points indicate the detection rates for the minimum-type estimator (A). The green triangle-shaped points indicate the detection rates for the local robust estimator (B). The red bullet-shaped points indicate the detection rates for the global robust estimator (C). The blue square-shaped points indicate the detection rates for the use of the true variance (D). On the x-axis, the misspecification of the angle of the fissure is displayed, on the y-axis the detection rates are displayed.

Table 12.1, when using the global robust estimator (C), our procedure performs reasonably well. Even for moderate signal-to-noise ratios (the fissure can barely be observed by visual inspection, see Figure 12.4), we are able to detect fissures at rates of at least 70% with at most 9 angles. In most cases, even 6 angles are sufficient for detection rates of at least 70%. Our procedure also proves to be robust against slight misspecifications of the width of the fissure. In particular, 9 angles are sufficient for detection rates of $\geq 70\%$ when $w = h = 0.01$ and $\mu \geq 2$ as well as for

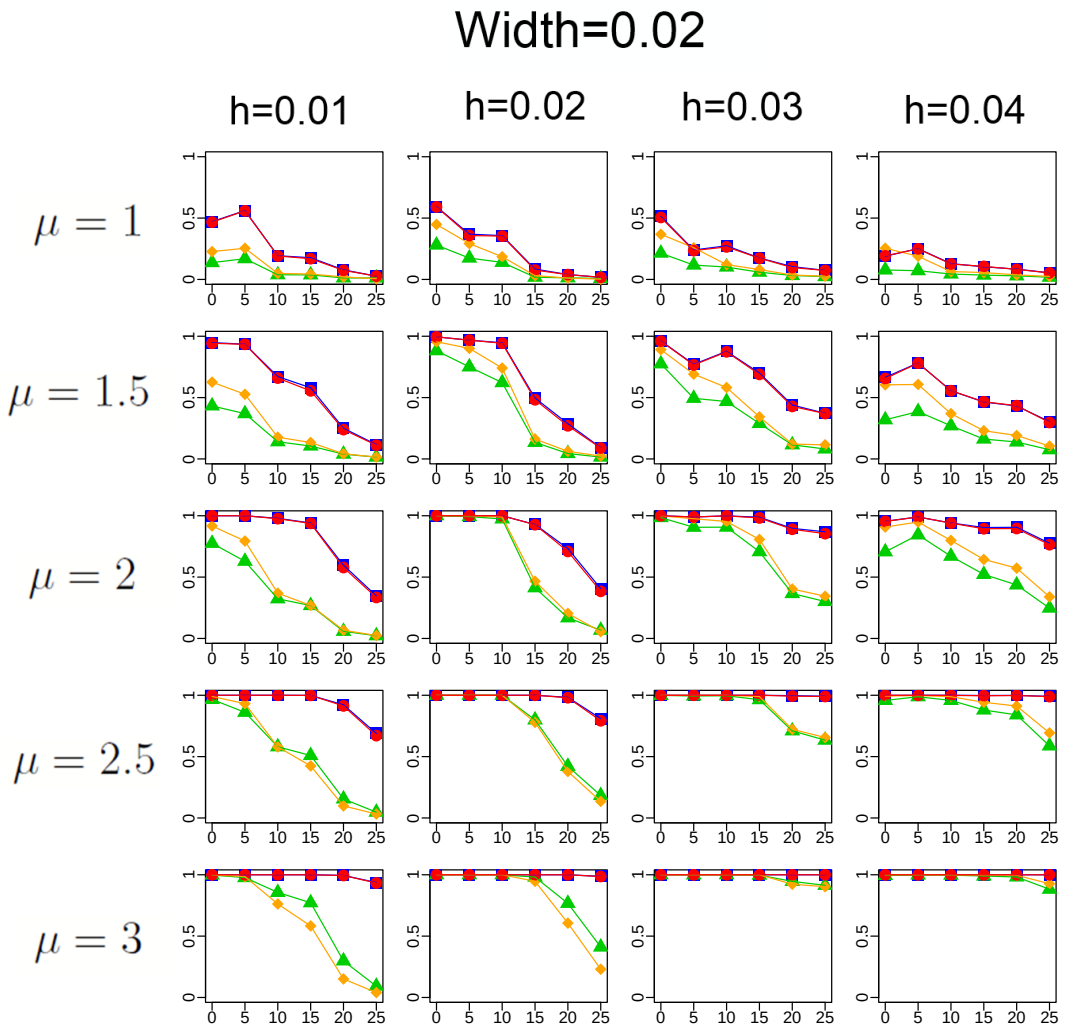

Figure 12.6: Comparison of the detection rates for all combinations of signal, bandwidth, width of the fissure and variance estimators. The yellow diamond-shaped points indicate the detection rates for the minimum-type estimator (A). The green triangle-shaped points indicate the detection rates for the local robust estimator (B). The red bullet-shaped points indicate the detection rates for the global robust estimator (C). The blue square-shaped points indicate the detection rates for the use of the true variance (D). On the x-axis, the misspecification of the angle of the fissure is displayed, on the y-axis the detection rates are displayed.

$w = h = 0.02$ and $\mu \geq 1.5$. If $w = 0.02$ and $\mu \geq 2$, then 9 angles are sufficient for any h, showing the robustness against misspecification. This is consistent with our observations in real data, where even for thin fissures that can barely be observed by visual inspection, 9 angles are sufficient to have some pixels adjacent to the fissure significant. In most cases, even 5 angles are sufficient to have some pixels adjacent to the fissure significant, see Figure 12.7 for examples.

(a) $w = 0.01$

μ	1.5	2	≥ 2.5	≥ 2.5
h	0.01	0.01	0.01	0.02
Δ	≤ 5	≤ 10	≤ 15	≤ 10
P	18	9	6	9

(b) $w = 0.02$

μ	1.5	1.5	2	2	2	2.5	2.5	3
h	0.01, 0.04	0.02, 0.03	0.01	0.02	0.03, 0.04	0.01	0.02, 0.03, 0.04	all
Δ	≤ 5	≤ 10	≤ 15	≤ 20	≤ 25	≤ 20	≤ 25	≤ 25
P	18	9	6	5	4	5	4	4

Table 12.1: The table shows all combinations (w, μ, Δ, h) for which the detection rates are higher than 70%, excluding the combinations for which the detection rate did not decline in Δ. The cases of misspecification of h are colored in blue.

12.4 Miscellaneous results

An argument can be made that one should consider the fissure to be 'detected' only if for at least one pixel $\boldsymbol{k}/T$ belonging to the fissure, it holds that $M(\boldsymbol{k}/T, h, \alpha) \geq \beta_h$ as in this case a pixel belonging, and not just close to the fissure would potentially serve as the starting point for further procedures like machine-learning algorithms. The detection rates in that case however only marginally differ from the detection rates when considering pixels that are adjacent to the fissure in the sense of Definition 12.1. The results can be found in Figures E.2 and E.3.
Another observation is that somewhat counterintuitively, in some cases the detection rates do not decline in Δ, most notably in the case of $w = 0.01$ and $h = 0.03$, where there is a sharp increase from $\Delta = 5$ to $\Delta = 10$ and also for $h = 0.04$ and $\mu \geq 2.5$, where the detection rates increase from $\Delta = 0$ to $\Delta = 20$. We conjecture that this effect comes from the discretization and due to the fact that the width of the 'inner strip' is large compared to the diameter of the circle, see Figure E.1 for a visualization.

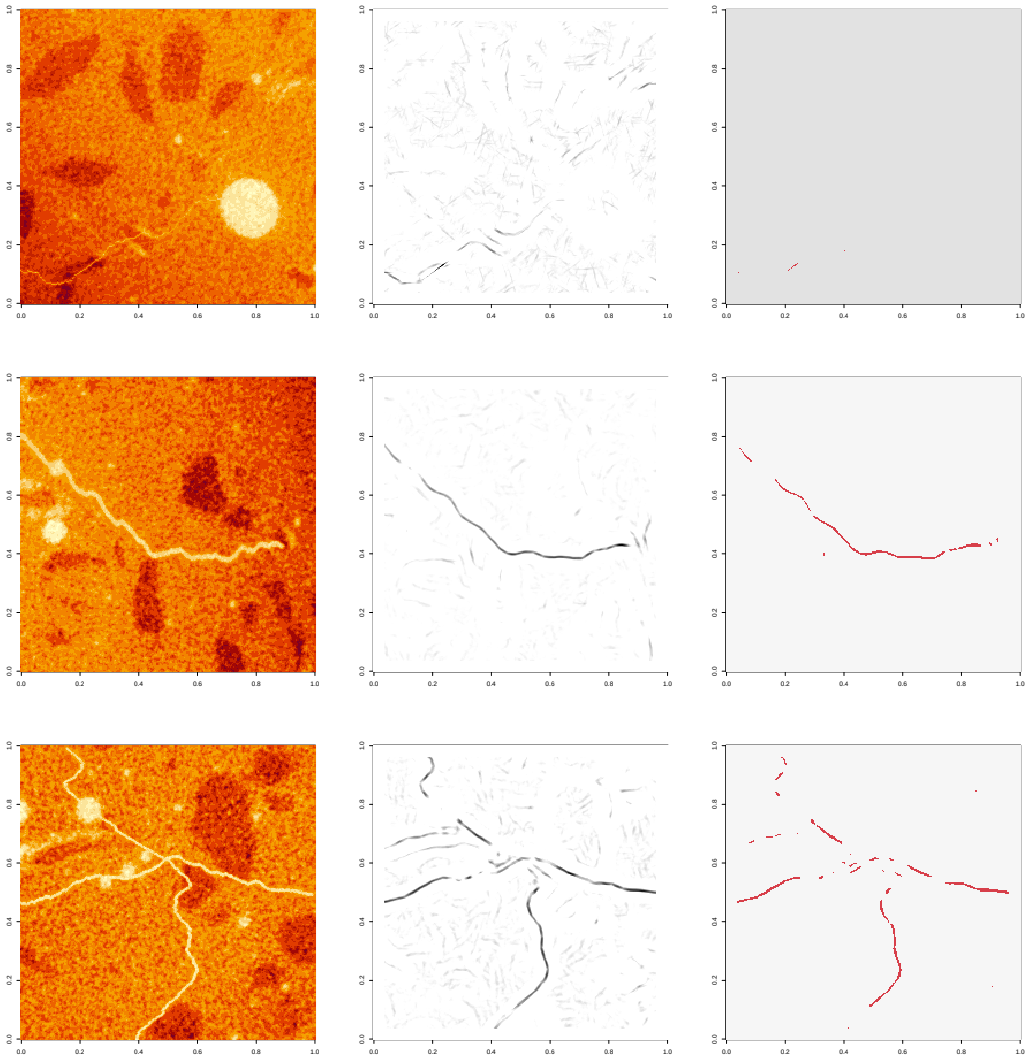

Figure 12.7: In the left column, 2D slices of a 3D CT scan of a concrete block with fissures are displayed. In the middle column, the corresponding heat maps for 5 angles of (11.3) are displayed. In the right column, the corresponding significant pixels for 5 are displayed. It shows that even for moderate signal-to-noise ratios, fissures are detected with 5 angles.

13 Conclusions

In this part of the thesis, our goal has been the development of a pre-processing procedure for the detection of areas that potentially contain fissures in concrete. Based on the observation of geometric properties of fissures and bubbles, and by the idea of one-parameter MOSUM statistics, we have introduced a scan statistic using a combination of rectangle- and circle segment-shaped windows in order to do so and identify areas that can then be used e. g. as starting points for machine learning

algorithms in order to trace the fissure. The simulation study demonstrates that our procedure performs well at detecting pixels close to a fissure even in the case of a small signal-to-noise ratio and also in the case of slight misspecification of the width of the fissure. It also shows that a global robust variance estimator based on Silverman's rule of thumb performs best among a number of variance estimators.
Furthermore, we have been able to show a weak convergence result for scan statistics using a general class of windows including two-dimensional convex sets towards a functional of a Gaussian process.
One drawback of the procedure introduced in Section 11 is the use of a single set of bandwidths for the width of the underlying rectangle and circle as typically in practice, the width of a fissure is unknown and even different parts of the fissure may vary in their widths – even though the theory in Section 10 already allows the maximization over finitely many bandwidths. Furthermore, we have only considered windows with linear bandwidths. Therefore, one future point of research is the extension to a multiple window approach using infinitely many windows with potentially sublinear bandwidths as e. g. by Sharpnack and Arias-Castro (2016) for hyperrectangles. Also note that the concrete image data is three-dimensional. While our mathematical theory includes the 3D case, a numerical analysis of the data set is yet to be conducted, which comes with additional computational challenges.

Appendix

A Stochastic properties

Lemma A.1. *(Compare Lemma 3.1 of Steinebach and Eastwood (1996))*
Let $(\mathbf{X}_t)_{t\geq 0}$ be a separable $\mathbb{R}^p$-valued stationary Gaussian process with independent components and covariance functions $r_1, \ldots, r_p$ fulfilling

$$\begin{aligned} r_i(a) &= 1 - C_i|a|^\alpha + o\left(|a|^\alpha\right) \quad \text{as } a \to 0 \\ r_i(a) &= o\left(1/\log a\right) \quad \text{as } a \to \infty \quad \text{for all } i = 1, \ldots, p. \end{aligned}$$

for some $0 < \alpha \leq 2$, $C_1, \ldots, C_p > 0$. Then there exists a constant $H = H_{p,\alpha}(C_1, \ldots, C_P)$ such that with

$$\begin{aligned} a(x) &= \sqrt{2\log x} \\ b(x) &= 2\log x + (p/2 - 1 + 1/\alpha)\log\log x - \log\left(2^{1-1/\alpha} H^{-1}\Gamma(p/2)\right) \end{aligned}$$

it holds that

$$a(x) \sup_{0\leq t\leq x} \|\mathbf{X}_t\| - b(x) \xrightarrow{\mathcal{D}} E, \quad x \to \infty$$

where $\mathrm{P}\left(E \leq y\right) = e^{-2e^{-y}}$.

Furthermore, if $C_1 = \ldots = C_p = C$, then $H = H_{p,\alpha}(C_1, \ldots, C_P) = C^{1/\alpha} H_\alpha$, where it is known that $H_1 = 1$, $H_2 = \pi^{-1/2}$.

Lemma A.2. *Let $(a_n)_{n\geq 1}$, $(b_n)_{n\geq 1}$ be sequences of positive real numbers and $b_n \to \infty$ as $n \to \infty$. Let $(X_n)_n$, $(Y_n)_n$ be sequences of random variables such that $Y_n = o_P(b_n/a_n)$. Then it holds for any $x \in \mathbb{R}$ that*

$$\lim_{n\to\infty} \left|\mathrm{P}\left(a_n X_n - b_n \leq x\right) - \mathrm{P}\left(a_n \max\left\{X_n, Y_n\right\} - b_n \leq x\right)\right| = 0$$

as $n \to \infty$.

Proof. It is trivial that $\mathrm{P}\left(a_n X_n - b_n \leq x\right) \geq \mathrm{P}\left(a_n \max\left\{X_n, Y_n\right\} - b_n \leq x\right)$.
It holds that

$$\begin{aligned} \mathrm{P}\left(a_n \max\left\{X_n, Y_n\right\} - b_n \leq x\right) = {} & \mathrm{P}\left(a_n X_n - b_n \leq x\right) + \mathrm{P}\left(a_n Y_n - b_n \leq x\right) \\ & - \mathrm{P}\left(\left\{a_n X_n - b_n \leq x\right\} \cap \left\{a_n Y_n - b_n \leq x\right\}\right) \\ \geq {} & \mathrm{P}\left(a_n X_n - b_n \leq x\right) + \mathrm{P}\left(a_n Y_n - b_n \leq x\right) - 1 \end{aligned}$$

Since $b_n \to \infty$, there exists $n_0 \in \mathbb{N}$ such that $x/b_n \geq -1/2$ for all $n \geq n_0$. Therefore

$$\mathrm{P}\left(a_n Y_n - b_n \leq x\right) = \mathrm{P}\left(\frac{a_n}{b_n} Y_n \leq \frac{x}{b_n} + 1\right) \geq \mathrm{P}\left(\frac{a_n}{b_n} Y_n \leq \frac{1}{2}\right) \to 1$$

as $n \to \infty$ and the assertion follows. □

Theorem A.3. *(Compare e. g. Lin and Bai (2010), 9.4, equation (45))*
Let $\epsilon_1, \epsilon_2 \ldots$ be i.i.d. such that $\mathbb{E}[\epsilon_1] = 0$ and let there exist $r > 2$ such that $\mathbb{E}[|\epsilon_1|^r] < \infty$. Then there exists $C_r > 0$ such that

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} \epsilon_i\right|^r\right] \leq C_r n^{r/2}.$$

Proposition A.4. *Let Assumption 2.1 hold with a rate of convergence as in Assumption 3.1 with the notation of Assumption 4.1. Let $0 < \xi_T \leq h_T$ and $D_T \geq 1$ be arbitrary sequences (bounded or unbounded).*

(a) The following bounds hold for the Wiener processes as in Assumption 2.1:

$$(i) \quad \max_{i=1,\ldots,q_T} \sup_{0\leq t\leq \xi_T} \frac{1}{\sqrt{\xi_T}} \|\mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i \pm t}^{(\theta_i)}\| = O_P\left(\sqrt{\log(2q_T)}\right),$$

$$(ii) \quad \sup_{\frac{D_T}{\|\mathbf{d}_i\|^2}\leq s\leq h_T} \frac{\sqrt{D_T}\left\|\mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i \pm s}^{(\theta_i)}\right\|}{s\,\|\mathbf{d}_i\|} = O_P(1),$$

$$(iii) \quad \max_{i=1,\ldots,q_T} \sup_{\frac{D_T}{\|\mathbf{d}_i\|^2}\leq s\leq h_T} \frac{\sqrt{D_T}\left\|\mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i \pm s}^{(\theta_i)}\right\|}{s\,\|\mathbf{d}_i\|} = O_P\left(\sqrt{\log(2q_T)}\right),$$

where the upper index $^{(\theta_i)}$ denotes the active stretch in the stationary segment $(\theta_i, \theta_i + s)$, respectively $(\theta_i - s, \theta_i)$.

(b) The bound in (i) carries over to the centered increments of the original process:

$$\max_{i=1,\ldots,q_T} \sup_{0\leq t\leq \xi_T} \frac{1}{\sqrt{\xi_T}} \|\widetilde{\mathbf{R}}_{\theta_i}^{(\theta_i)} - \widetilde{\mathbf{R}}_{\theta_i \pm t}^{(\theta_i)}\| = O_P\left(\max\left\{\frac{T^{1/2}\nu_T}{\sqrt{\xi_T}}, \sqrt{\log(2q_T)}\right\}\right).$$

The bound in (ii) carries over if a forward and backward invariance principle as above exists starting in an arbitrary point θ_i, in this case (iii) carries over if $q_T = O(1)$ if ξ_T is large enough.

For a single change point (instead of taking the maximum over all) the bound in (a) (i) and (b) is given by $O_P(1)$.

Proof. Note that we will only show the assertions for '+' in the terms involving '$\pm$' as the proofs are analogous for '$-$'.

(a) Let $\mathbf{B}_t^{(j)} = (B_{t,1}^{(j)}, \ldots, B_{j,p}^{(j)})' = (\mathbf{\Sigma}_T^{(j)})^{-1/2}\,\mathbf{W}_t^{(j)}$ be multivariate standard Wiener processes.

(i) By Lemma B.2 (ii) it holds that

$$
\begin{aligned}
&\max_{i=1,\dots,q_T} \sup_{0\le t\le \xi_T} \frac{1}{\sqrt{\xi_T}} \left\| \mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i+t}^{(\theta_i)} \right\| \\
=\ &\max_{i=1,\dots,q_T} \sup_{0\le t\le \xi_T} \frac{1}{\sqrt{\xi_T}} \left\| \left(\mathbf{\Sigma}_T^{(\theta_i)}\right)^{1/2} \left(\mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+t}^{(\theta_i)}\right) \right\| \\
\le\ &\max_{i=1,\dots,q_T} \left\| \left(\mathbf{\Sigma}_T^{(\theta_i)}\right)^{1/2} \right\| \sup_{0\le t\le \xi_T} \frac{1}{\sqrt{\xi_T}} \left\| \mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+t}^{(\theta_i)} \right\| \\
\le\ &\max_{i=1,\dots,P} \left\| \left(\mathbf{\Sigma}_T^{(i)}\right)^{1/2} \right\| \max_{i=1,\dots,q_T} \sup_{0\le t\le \xi_T} \frac{1}{\sqrt{\xi_T}} \left\| \mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+t}^{(\theta_i)} \right\|.
\end{aligned}
$$

By Corollary B.3 (iii) it holds that

$$
\max_{i=1,\dots,P} \left\| \left(\mathbf{\Sigma}_T^{(i)}\right)^{1/2} \right\| = \max_{i=1,\dots,P} \sqrt{\left\| \mathbf{\Sigma}_T^{(i)} \right\|} = O(1). \tag{A.1}
$$

For $i = 1, \dots, q_T$ it holds due to the self-similarity and the invariance of the Wiener process under time shift for all $C > 2$ that

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{0\le t\le \xi_T} \frac{1}{\sqrt{\xi_T}} \left\| \mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+t}^{(\theta_i)} \right\| \ge \sqrt{Cp\log(2q_T)} \right) \\
=\ &\mathrm{P}\left(\sup_{0\le s\le 1} \left\| \mathbf{B}_s^{(\theta_i)} \right\| \ge \sqrt{Cp\log(2q_T)} \right) \\
\le\ &\mathrm{P}\left(\bigcup_{j=1}^{p} \left\{ \sup_{0\le s\le 1} \left| B_{s,j}^{(\theta_i)} \right| \ge \sqrt{C\log(2q_T)} \right\} \right) \\
\le\ &p\,\mathrm{P}\left(\sup_{0\le s\le 1} \left| B_{s,1}^{(\theta_i)} \right| \ge \sqrt{C\log(2q_T)} \right).
\end{aligned}
$$

By the reflection principle of the standard Wiener process, Markov's inequality and the moment-generating function of the standard normal distribution it holds with $r = \sqrt{C\log(2q_T)}$ that

$$
\begin{aligned}
&p\,\mathrm{P}\left(\sup_{0\le s\le 1} \left| B_{s,1}^{(\theta_i)} \right| \ge \sqrt{C\log(2q_T)} \right) \le 4p\,\mathrm{P}\left(B_{1,1}^{(\theta_i)} \ge \sqrt{C\log(2q_T)} \right) \\
\le\ &4p \frac{\mathbb{E}\left[e^{rB_{1,1}^{(\theta_i)}} \right]}{e^{r\sqrt{C\log(2q_T)}}} = 4p\exp\left(\frac{r^2}{2} - r\sqrt{C\log(2q_T)} \right) = 4p\exp\left(-\frac{C\log(2q_T)}{2} \right) \le 4pq_T^{-\frac{C}{2}}.
\end{aligned}
$$

Therefore by subadditivity, we obtain that

$$
\mathrm{P}\left(\max_{i=1,\dots,P} \left\| \left(\mathbf{\Sigma}_T^{(i)}\right)^{\frac{1}{2}} \right\| \max_{i=1,\dots,q_T} \sup_{0\le t\le \xi_T} \frac{1}{\sqrt{\xi_T}} \left\| \mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+t}^{(\theta_i)} \right\| \ge \sqrt{Cp\log(2q_T)} \right) \le 4pq_T^{1-\frac{C}{2}}.
$$

which in combination with (A.1) proves the assertion.

Since obviously (ii) follows from (iii), we will only prove the latter.
(iii) Similar to above it holds that

$$\max_{i=1,\dots,q_T} \sup_{\frac{D_T}{\|\mathbf{d}_i\|^2} \le s \le h_T} \frac{\sqrt{D_T} \left\| \mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i+s}^{(\theta_i)} \right\|}{s \|\mathbf{d}_i\|}$$
$$\le \max_{i=1,\dots,P} \sqrt{\left\| \mathbf{\Sigma}_T^{(i)} \right\|} \max_{i=1,\dots,q_T} \sup_{\frac{D_T}{\|\mathbf{d}_i\|^2} \le s \le h_T} \frac{\sqrt{D_T} \left\| \mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+s}^{(\theta_i)} \right\|}{s \|\mathbf{d}_i\|}.$$

By the self-similarity and the of the Wiener process we obtain with $t = s \|\mathbf{d}_i\|^2 / D_T$ that

$$\sup_{\frac{D_T}{\|\mathbf{d}_i\|^2} \le s \le h_T} \frac{\sqrt{D_T} \left\| \mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+s}^{(\theta_i)} \right\|}{s \|\mathbf{d}_i\|} \overset{\mathcal{D}}{=} \sup_{\frac{D_T}{\|\mathbf{d}_i\|^2} \le s \le h_T} \frac{\sqrt{D_T} \left\| \mathbf{B}_{s}^{(\theta_i)} \right\|}{s \|\mathbf{d}_i\|} \overset{\mathcal{D}}{=} \sup_{1 \le t \le \frac{h_T \|\mathbf{d}_i\|^2}{D_T}} \frac{\left\| \mathbf{B}_{t}^{(\theta_i)} \right\|}{t}.$$

It holds that

$$\mathrm{P}\left(\sup_{1 \le t \le \frac{h_T \|\mathbf{d}_i\|^2}{D_T}} \frac{\left\| \mathbf{B}_{t}^{(\theta_i)} \right\|}{t} \ge \sqrt{Cp \log(2q_T)} \right) \le \mathrm{P}\left(\sup_{t \ge 1} \frac{\left\| \mathbf{B}_{t}^{(\theta_i)} \right\|}{t} \ge \sqrt{Cp \log(2q_T)} \right)$$
$$\le \sum_{l \ge 0} \mathrm{P}\left(\sup_{2^l \le t \le 2^{l+1}} \frac{\left\| \mathbf{B}_{t}^{(\theta_i)} \right\|}{t} \ge \sqrt{Cp \log(2q_T)} \right).$$

With the substitution $r = t/2^{l+1}$, the reflection principle, the self-similarity of the Wiener process and Markov's inequality with $s = \sqrt{2^l C \log q_T}$ it holds that

$$\mathrm{P}\left(\sup_{2^l \le t \le 2^{l+1}} \frac{\left\| \mathbf{B}_{t}^{(\theta_i)} \right\|}{t} \ge \sqrt{Cp \log(2q_T)} \right) = \mathrm{P}\left(\sup_{\frac{1}{2} \le r \le 1} \frac{\left\| \mathbf{B}_{2^{l+1}r}^{(\theta_i)} \right\|}{2^{l+1} r} \ge \sqrt{Cp \log(2q_T)} \right)$$
$$= \mathrm{P}\left(\sup_{\frac{1}{2} \le r \le 1} \frac{\left\| \mathbf{B}_{r}^{(\theta_i)} \right\|}{\sqrt{2^{l+1} r}} \ge \sqrt{Cp \log(2q_T)} \right) \le \mathrm{P}\left(\sup_{\frac{1}{2} \le r \le 1} \left\| \mathbf{B}_{r}^{(\theta_i)} \right\| \ge \sqrt{2^l Cp \log(2q_T)} \right)$$
$$\le \mathrm{P}\left(\sup_{0 \le r \le 1} \left\| \mathbf{B}_{r}^{(\theta_i)} \right\| \ge \sqrt{2^l Cp \log(2q_T)} \right) \le p\, \mathrm{P}\left(\sup_{0 \le r \le 1} \left| B_{r,1}^{(\theta_i)} \right| \ge \sqrt{2^l C \log(2q_T)} \right)$$
$$\le 4p\, \mathrm{P}\left(B_{1,1}^{(\theta_i)} \ge \sqrt{2^l C \log(2q_T)} \right) \le 4p \frac{\mathbb{E}\left[e^{s B_{1,1}^{(\theta_i)}} \right]}{e^{s\sqrt{2^l C \log(2q_T)}}}$$
$$= 4p \exp\left(\frac{s^2}{2} - s\sqrt{2^l C \log(2q_T)} \right) = 4p \exp\left(-\frac{2^l C \log(2q_T)}{2} \right) = 4p (2q_T)^{-2^{l-1} C}.$$

Therefore, it holds that

$$\sum_{l\geq 0} \mathrm{P}\left(\sup_{2^l\leq t\leq 2^{l+1}} \frac{\left\|\mathbf{B}_t^{(\theta_i)}\right\|}{t} \geq \sqrt{Cp\log(2q_T)}\right) \leq 4p\sum_{l\geq -1} q_T^{-2^l C}$$
$$\leq 16p\int_{1/4}^{\infty}(2q_T)^{-x}\,dx = \frac{16p(2q_T)^{-C/4}}{C\log(2q_T)}.$$

By subadditivity we obtain that

$$\mathrm{P}\left(\max_{i=1,\ldots,q_T}\sup_{\frac{D_T}{\|\mathbf{d}_i\|^2}\leq s\leq h_T} \frac{\sqrt{D_T}\left\|\mathbf{B}_{\theta_i}^{(\theta_i)} - \mathbf{B}_{\theta_i+s}^{(\theta_i)}\right\|}{s\left\|\mathbf{d}_i\right\|} \geq \sqrt{Cp\log(2q_T)}\right) \leq \frac{16p(2q_T)^{1-C/4}}{C\log(2q_T)},$$

which in combination with (A.1) proves the assertion. (b) By the invariance principle from Assumption 2.1 and (a) (i) it holds that

$$\begin{aligned}
&\max_{i=1,\ldots,q_T}\sup_{0\leq t\leq \xi_T}\frac{1}{\sqrt{\xi_T}}\left\|\widetilde{\mathbf{R}}_{\theta_i}^{(\theta_i)} - \widetilde{\mathbf{R}}_{\theta_i+t}^{(\theta_i)}\right\| \\
\leq\ &\max_{i=1,\ldots,q_T}\sup_{0\leq t\leq \xi_T}\frac{1}{\sqrt{\xi_T}}\left\|\left(\widetilde{\mathbf{R}}_{\theta_i}^{(\theta_i)} - \widetilde{\mathbf{R}}_{\theta_i+t}^{(\theta_i)}\right) - \left(\mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i+t}^{(\theta_i)}\right)\right\| \\
&+ \max_{i=1,\ldots,q_T}\sup_{0\leq t\leq \xi_T}\frac{1}{\sqrt{\xi_T}}\left\|\mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i+t}^{(\theta_i)}\right\| \\
\leq\ &2\max_{i=1,\ldots,P}\sup_{0\leq t\leq T}\frac{1}{\sqrt{\xi_T}}\left\|\widetilde{\mathbf{R}}_t^{(i)} - \mathbf{W}_t^{(i)}\right\| + O_P\left(\sqrt{\log(2q_T)}\right) \\
=\ &O_P\left(\max\left\{\frac{T^{1/2}\nu_T}{\sqrt{\xi_T}}, \sqrt{\log(2q_T)}\right\}\right).
\end{aligned}$$

Note that if $q_T = O_P(1)$ and that forward or backward invariance principles starting in θ_i exist, it holds that

$$\max_{i=1,\ldots,q_T}\sup_{0\leq t\leq \xi_T}\frac{1}{\sqrt{\xi_T}}\left\|\left(\widetilde{\mathbf{R}}_{\theta_i}^{(\theta_i)} - \widetilde{\mathbf{R}}_{\theta_i+t}^{(\theta_i)}\right) - \left(\mathbf{W}_{\theta_i}^{(\theta_i)} - \mathbf{W}_{\theta_i+t}^{(\theta_i)}\right)\right\| = O_P(\nu_T) = o_P(1),$$

thus proving the assertion.

□

Remark A.1. Note that in our situation, it is typically the case that $\xi_T = h$, and therefore the order in Proposition A.4 (b) is typically given by $O_P\left(\sqrt{\log(2q_T)}\right)$ due to Assumption 3.1.

Remark A.2. In the simulation study in Section 6, we study three-dimensional renewal processes with intensity changes at $0 < c_1 < \ldots < c_q < T$. For mean inter-event times $\mu_1, \ldots, \mu_{q+1} > 0$ with $\mu_{i+1} \neq \mu_i$, the intensity changes are given by $\mathbf{d}_i = (1/\mu_{i+1} - 1/\mu_i, 1/\mu_{i+1} - 1/\mu_i, 1/\mu_{i+1} - 1/\mu_i)$. In order to explain the differences

in the detection rates between the dependent choice (B) and the dependent choice (C) for the matrix $\widehat{\boldsymbol{A}}_t^{-1}$ in (3.7), we study the behavior of the signal term of $\mathbf{M}_t'\widehat{\boldsymbol{A}}_t^{-1}\mathbf{M}_t$. By (3.5), we can write

$$\boldsymbol{m}_t'\widehat{\boldsymbol{A}}_t^{-1}\boldsymbol{m}_t = C_t\mathbf{d}_i'\boldsymbol{A}^{-1}\mathbf{d}_i$$

with some constant $C_t > 0$ depending on t, h, μ, σ, where σ is the standard deviation of the inter-event times. In (B), $\boldsymbol{A}$ is the 3×3-identity matrix. In (C) $\boldsymbol{A}$ a 3×3 matrix with 1 on the diagonal and the respective correlations between the dimensions on the off-diagonal entries. We obtain by straightforward algebraic calculus that

for type (B),

$$\boldsymbol{m}_t'\widehat{\boldsymbol{A}}_t^{-1}\boldsymbol{m}_t = 3C_t\left(\frac{1}{\mu_{i+1}} - \frac{1}{\mu_i}\right).$$

For type (C) with pairwise correlations of 0.2 between dimensions,

$$\boldsymbol{m}_t'\widehat{\boldsymbol{A}}_t^{-1}\boldsymbol{m}_t = \frac{15}{7}C_t\left(\frac{1}{\mu_{i+1}} - \frac{1}{\mu_i}\right).$$

For type (C) with pairwise correlations of -0.2 between dimensions,

$$\boldsymbol{m}_t'\widehat{\boldsymbol{A}}_t^{-1}\boldsymbol{m}_t = 5C_t\left(\frac{1}{\mu_{i+1}} - \frac{1}{\mu_i}\right).$$

A.1 $\Gamma(s, \lambda)$-distribution

Definition A.1. Let $s, \lambda > 0$. The probability density function of the $\Gamma(s, \lambda)$ distribution is given by

$$f_{\Gamma(s,\lambda)}(x) = \frac{x^{s-1}e^{-\lambda x}\lambda^s}{\Gamma(s)},$$

where $\Gamma(s)$ denotes the gamma function.

If $X \sim \Gamma(s, \lambda)$, then it holds by straightforward calculations that

$$\mathbb{E}[X] = \frac{s}{\lambda}$$
$$\mathrm{Var}[X] = \frac{s}{\lambda^2}.$$

The moment-generating function (MGF) is given by

$$M_{\Gamma(s,\lambda)}(t) = \frac{1}{\left(1 - \frac{t}{\lambda}\right)^s}.$$

for $t < 1/\lambda$.

Lemma A.5. *Let $s_1, s_2, \lambda > 0$. Let $X \sim \Gamma(s_1, \lambda)$, $Y \sim \Gamma(s_2, \lambda)$ be independent. Then $X + Y \sim \Gamma(s_1 + s_2, \lambda)$.*

Proof. This follows immediately from the MGF of the Gamma-distribution. As X and Y are independent, it holds for $t < 1/\lambda$ that

$$M_{X+Y}(t) = M_X(t)M_Y(t) = \frac{1}{\left(1 - \frac{t}{\lambda}\right)^{s_1+s_2}} = M_{\Gamma(s_1+s_2,\lambda)}(t).$$

□

Remark A.3. (a) In the simulation study in Section 6, for the results in Table 6.2, we have coupled the inter-event times by having (with a slight abuse of notation) $X_j \sim \Gamma(s, \lambda)$ and $X_4 \sim \Gamma(s^*, \lambda)$ and $Y_j = X_j + X_4$ for $j = 1, 2, 3$. By the above properties of the Γ-distribution, the correlation between Y_j and Y_k for $1 \leq j < k \leq 3$ is

$$\rho = \frac{\mathrm{Cov}\left[Y_j, Y_k\right]}{\sqrt{\mathrm{Var}\left[Y_j\right]}\sqrt{\mathrm{Var}\left[Y_k\right]}} = \frac{\mathrm{Var}\left[X_4\right]}{\sqrt{\mathrm{Var}\left[Y_j\right]}\sqrt{\mathrm{Var}\left[Y_k\right]}} = \frac{s^*}{s + s^*}.$$

Thus, for given mean $\mu > 0$, variance $0 < \sigma^2 < \infty$ of the inter-event times and given correlation $0 \leq \rho \leq 1$ between the dimensions, it holds by Lemma A.5 that

$$\mu = \frac{s + s^*}{\lambda}, \qquad \sigma^2 = \frac{s + s^*}{\lambda^2}, \qquad \rho = \frac{s^*}{s + s^*}.$$

By solving for s, s^* and λ, we obtain that

$$s = \frac{(1 - \rho)\mu^2}{\sigma^2}, \qquad \lambda = \frac{\mu}{\sigma^2}, \qquad s^* = \frac{\rho\mu^2}{\sigma^2}.$$

(b) For the results in Table 6.3, we have coupled the inter-event times by having (with a slight abuse of notation) $Y_j = X_j + \sum_{1 \leq k < j} X_{k,j} - \sum_{j < k \leq 3} X_{j,k}$ $j = 1, 2, 3$ where the $X_j = X_j^{(i)}$ are sequences of independent in time $\Gamma(s, \lambda)$-distributed random variables. The $X_{j,k} = X_{j,k}^{(i)}$ are sequences of independent in time $\mathcal{N}(0, s_1^2)$-distributed random variables. For given mean $\mu > 0$, variance $0 < \sigma^2 < \infty$ of the inter-event times and given correlation $-1/2 < \rho \leq 0$ between the dimensions, we have by the above properties of the Γ-distribution for $i = 1, 2, 3$ and $1 \leq j < k \leq 3$ that

$$\begin{aligned}
\mu &= \mathbb{E}\left[Y_i\right] = \frac{s}{\lambda} \\
\sigma^2 &= \mathrm{Var}\left[Y_i\right] = \mathrm{Var}\left[Y_1\right] = \mathrm{Var}\left[X_1 + X_{1,2} + X_{1,3}\right] = \frac{s}{\lambda^2} + 2s_1^2 \\
\rho &= \mathrm{Cov}\left[Y_j, Y_k\right] = \mathrm{Cov}\left[Y_1, Y_2\right] = \frac{\mathrm{Cov}\left[X_1 + X_{1,2} + X_{1,3}, X_1 - X_{1,2} + X_{2,3}\right]}{\sqrt{\mathrm{Var}\left[Y_1\right]}\sqrt{\mathrm{Var}\left[Y_2\right]}}
\end{aligned}$$

$$= \frac{\mathrm{Var}\left[X_1\right] - \mathrm{Var}\left[X_{1,2}\right]}{\sigma^2} = \frac{\frac{s}{\lambda^2} - s_1^2}{\sigma^2}.$$

By solving for s, s_1^2 and λ, we obtain that

$$s = \frac{3\mu^2}{\sigma^2(2\rho+1)}, \qquad \lambda = \frac{3\mu}{\sigma^2(2\rho+1)}, \qquad s_1^2 = \frac{\sigma^2(1-\rho)}{3}.$$

A.2 The space $\mathcal{D}([0,1]^p)$

Definition A.2. (Compare Wichura (1969), Definition 1)
Let $\mathcal{C}_p$ be the set of all continuous functions from $[0,1]^p$ to $\mathbb{R}$. Let $\mathcal{D}([0,1]^p)$ be the uniform closure of the set of all step functions that are constant on sets of the form $[a_1, b_1) \times \ldots \times [a_p, b_p)$. Let $\mathcal{A}$ be the σ-algebra on D_p generated by the projection mappings $\pi_t : f \to f(t)$, $t \in [0,1]^p$.

A sequence $(P_n)_{n\geq 1}$ of probability measures on $(\mathcal{D}([0,1]^p), \mathcal{A})$ *converges weakly in the U-topology* to a probability measure P on $(\mathcal{D}([0,1]^p), \mathcal{A})$ if for every measurable function $f : \mathcal{D}([0,1]^p) \to \mathbb{R}$ that is continuous in the topology of uniform convergence on $\mathcal{D}([0,1]^p)$, it holds that

$$\int f \; dP_n \to \int f \; dP.$$

Theorem A.6. *(Wichura (1969), Theorem 2)*
A sequence $\{X_T\}$ of stochastic processes in $\mathcal{D}([0,1]^p)$ is weakly convergent to a process X which a.s. belongs to the class of uniformly continuous functions on $[0,1]^p$ if and only if

(i)

$$(X_{\boldsymbol{t}_1,T}, \ldots, X_{\boldsymbol{t}_n,T})' \xrightarrow{\mathcal{D}} (X_{\boldsymbol{t}_1}, \ldots, X_{\boldsymbol{t}_n})'$$

for all $n \in \mathbb{N}$, $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in S_0$, where S_0 is a dense subset of $[0,1]^p$.

(ii)

$$\lim_{\delta \to 0} \limsup_{T \to \infty} \mathrm{P}\left(w_{\delta,T}(X_T) \geq x\right) = 0$$

for all $x > 0$, where

$$w_{\delta,T}(X_T) = \sup_{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty < \delta} \left|X_{\boldsymbol{s},T} - X_{\boldsymbol{t},T}\right|.$$

Lemma A.7. *Let $\mathcal{D}\left([0,1]^p\right)$ be equipped with the topology induced by the maximum norm. For $f = (f_1, \ldots, f_P)'$, $g = (g_1, \ldots, g_P)' \in \left(\mathcal{D}\left([0,1]^p\right)\right)^P$ and $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)' \in \left(\mathbb{R}^P\right)^n$ let*

$$d_1(f,g) = \max_{i=1,\ldots,P} \sup_{\boldsymbol{t} \in [0,1]^p} |f_i(\boldsymbol{t}) - g_i(\boldsymbol{t})|$$

$$d_2(\boldsymbol{x},\boldsymbol{y}) = \max_{i=1,\ldots,n} \|\boldsymbol{x}_i - \boldsymbol{y}_i\|_\infty.$$

Denote for $n \in \mathbb{N}$, $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0,1]^p$ *by* $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n} : \left(\mathcal{D}\left([0,1]^p\right)\right)^P \to \left(\mathbb{R}^P\right)^n$, $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(f) = (f(\boldsymbol{t}_1), \ldots, f(\boldsymbol{t}_n))'$ *the projection mapping from* $\left(\mathcal{D}\left([0,1]^p\right)\right)^P$ *to* $\left(\mathbb{R}^P\right)^n$. *If* $\left(\mathcal{D}\left([0,1]^p\right)\right)^P$ *is equipped with the metric* d_1 *and* $\left(\mathbb{R}^P\right)^n$ *is equipped with the metric* d_2 *then* $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}$ *is continuous.*

Proof. Let $f = (f_1, \ldots, f_P)'$, $g = (g_1, \ldots, g_P)' \in \left(\mathcal{D}\left([0,1]^p\right)\right)^P$. Then it holds that

$$
\begin{aligned}
& d_2\left(\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(f), \pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(g)\right) = \max_{i=1,\ldots,P} \left\|\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(f_i) - \pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}(g_i)\right\|_\infty \\
= & \max_{i=1,\ldots,P} \left\|(f_i(\boldsymbol{t}_1), \ldots, f_i(\boldsymbol{t}_n))' - (g_i(\boldsymbol{t}_1), \ldots, g_i(\boldsymbol{t}_n))'\right\|_\infty = \max_{i=1,\ldots,P} \max_{j=1\ldots,n} \left|f_i(\boldsymbol{t}_j) - g_i(\boldsymbol{t}_j)\right| \\
\leq & \max_{i=1,\ldots,P} \sup_{\boldsymbol{t}\in[0,1]^p} \left|f_i(\boldsymbol{t}) - g_i(\boldsymbol{t})\right| = d_1(f,g),
\end{aligned}
$$

thus showing that $\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_n}$ is Lipschitz-continuous with Lipschitz-constant 1. □

Definition A.3. (Extension of Csörgö and Révész (1981), 1.11) Let $(W_{\boldsymbol{t}})_{\boldsymbol{t}\in[0,1]^p}$ be a p-parameter, real-valued stochastic process. Let $\boldsymbol{s} = (s_1, \ldots, s_p)'$, $\boldsymbol{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$ with $s_i \leq t_i$ for all i. For the hyperrectangle $R = [\boldsymbol{s}, \boldsymbol{t})$ define

$$
W_R = \sum_{\mathbf{d}=(d_1,\ldots,d_p)'\in\{0,1\}} (-1)^{p-\sum_i d_i} W_{\boldsymbol{s}+\mathbf{d}\odot(\boldsymbol{t}-\boldsymbol{s})}.
$$

(i) For any hyperrectangle $R \subset [0,\infty)^p$, it holds that

$$
W_R \sim \mathcal{N}(0, \lambda(R)).
$$

(ii) $W_{\boldsymbol{t}} = 0$ for all $\boldsymbol{t} = (t_1, \ldots, t_p)'$ with $t_i = 0$ for at least one i.

(iii) For pairwise disjoint rectangles $R_1, \ldots, R_n \subset [0,1]^p$, $W_{R_1}, \ldots, W_{R_p}$ are independent.

(iv) $(W_{\boldsymbol{t}})_{\boldsymbol{t}\in[0,1]^p}$ almost surely belongs to the class of uniformly continuous functions on $[0,1]^p$.

Note that (i)-(iii) implies that for $\boldsymbol{s} = (s_1, \ldots, s_p)'$, $\boldsymbol{t} = (t_1, \ldots, t_p)' \in \mathbb{R}^p$,

$$
\mathbb{E}\left[W_{\boldsymbol{s}} W_{\boldsymbol{t}}\right] = \prod_{i=1}^{p} \min\left\{s_i, t_i\right\}.
$$

B Norms, Eigenvalues

Definition B.1. Let $\boldsymbol{a} = (a_1, \ldots, a_n)' \in \mathbb{R}^n$, $1 \leq p < \infty$. Define

$$
\begin{aligned}
\|a\|_p &= \left(\sum_{i=1}^{n} |a_i|^p\right)^{\frac{1}{p}} \\
\|a\|_\infty &= \max_{1=1,\ldots,n} |a_i|.
\end{aligned}
$$

and write $\|\cdot\| = \|\cdot\|_2$ for the Euclidean norm.

Lemma B.1. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric, positive definite matrix. Let $\lambda_1, \ldots, \lambda_n > 0$ be the (not necessarily unique) eigenvalues of A. Then it holds that*

(i) If $\lambda > 0$ is an eigenvalue of A, then λ^2 is an eigenvalue of A^2.

(ii) If $\lambda > 0$ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}.

Proof. (i)+(ii) Let x be an eigenvector corresponding to λ. Then it holds by definition of an eigenvalue that

$$\begin{aligned} A^2x = A(Ax) = A(\lambda x) = \lambda Ax = \lambda^2 x \\ \lambda(A^{-1}x) = A^{-1}(\lambda x) = A^{-1}Ax = x \\ \Rightarrow A^{-1}x = \lambda^{-1}x, \end{aligned}$$

thus proving the assertion. □

Definition B.2. Let $A \in \mathbb{R}^{n\times m}$, $1 \leq p \leq \infty$. Define the *operator norm* by

$$\|A\|_p = \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

and denote by $\|\cdot\| = \|\cdot\|_2$ the operator norm induced by the vector Euclidean norm.

Lemma B.2. *Let $A, B \in \mathbb{R}^{n\times m}$.*

(i) Let $\lambda_1, \ldots, \lambda_k$ be the eigenvalues of $A'A$. Then

$$\|A\| = \sqrt{\max\{|\lambda_1|, \ldots, |\lambda_k|\}}.$$

(ii) Then it holds that

$$\|AB\| \leq \|A\| \|B\|.$$

Proof. (i) See e.g. Shores (2018), Theorem 6.14 (3).
(ii) Follows immediately from Definition B.2, see e.g. Shores (2018) Theorem 6.13. □

Corollary B.3. *Let $A \in \mathbb{R}^{n\times n}$ a symmetric positive definite matrix with eigenvalues $\lambda_{min} = \lambda_1 < \ldots < \lambda_k = \lambda_{max}$. Furthermore, let $A^{1/2}$ be a positive definite matrix such that $A = (A^{1/2})'(A^{1/2})$. Then it holds that*

(i)

$$\|A\| = \lambda_{max}$$

(ii)

$$\left\|A^{-1}\right\| = \lambda_{min}^{-1}$$

(iii)

$$\left\|A^{1/2}\right\| = \sqrt{\|A\|}.$$

Proof. (i) Since A is symmetric and positive definite, the maximum eigenvalue of $A'A = A^2$ is λ_{max}^2 by Lemma B.1 (i). Therefore the assertion follows immediately from Lemma B.2.
(ii) By Lemma B.1 (ii), it holds that if λ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}. By (i), it holds that

$$\lambda_{min}^{-1} = \max\left\{\lambda_1^{-1}, \ldots, \lambda_k^{-1}\right\} = \left\|A^{-1}\right\|.$$

(iii) By Lemma B.2 it holds that if λ is an eigenvalue of A, then $\sqrt{\lambda}$ is an eigenvalue of $A^{1/2}$. By (i), it holds that

$$\left\|A^{1/2}\right\| = \sqrt{\max\left\{\lambda_1, \ldots, \lambda_k\right\}} = \sqrt{\lambda_{max}} = \sqrt{\|A\|}.$$

□

Lemma B.4. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $\lambda_{min} = \min\{\lambda_1, \ldots, \lambda_n\}$. Then for any $x \in \mathbb{R}^n$ it holds that*

$$x'Ax \geq \lambda_{min} \|x\|^2.$$

Proof. There exists a decomposition $A = Q'DQ$ with $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and Q orthogonal, see e. g. Shores (2018) Theorem 5.13. Let $y = Qx$. Then it holds that

$$x'Ax = x'Q'DQx = y'Dy = \sum_{i=1}^{n} \lambda_i y_i^2 \geq \lambda_{min} \|y\|^2 = \lambda_{min} \|x\|^2.$$

□

C Convex sets, Algebraic properties

Lemma C.1. *(a) It holds for any $C, \gamma > 0$, $\boldsymbol{s} \in \mathbb{R}^p$ that*

$$\left|\left\{\boldsymbol{k}\gamma \middle| \boldsymbol{k} \in \mathbb{Z}^p\right\} \cap [\boldsymbol{s} - C\gamma, \boldsymbol{s} + C\gamma]\right| \leq (2\lfloor C \rfloor + 1)^p.$$

(b) Let $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^p$ be such that $s_i \leq t_i$ for all $i = 1, \ldots, p$ and let $\|\boldsymbol{s} - \boldsymbol{t}\|_\infty \leq \gamma$. If $\gamma_2 > 0$ and $\gamma_1 - \gamma_2 \geq \gamma$, then

$$[\boldsymbol{t}, \boldsymbol{t} + \gamma_2] \subset [\boldsymbol{s}, \boldsymbol{s} + \gamma_1].$$

(c) Let $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^p$ with $\|\boldsymbol{s} - \boldsymbol{t}\|_\infty \leq \gamma$ and let $\gamma_1 \geq \gamma$. Then it holds that

$$[\boldsymbol{s}, \boldsymbol{s} + \gamma_1] \cap [\boldsymbol{t}, \boldsymbol{t} + \gamma_1] \neq \emptyset.$$

Proof. (a) Let $i = 1, \dots, p$. Since for any $k_1 \neq k_2 \in \mathbb{Z}$, $|k_1\gamma - k_2\gamma| \geq \gamma$, it holds that

$$\left| \left\{ k\gamma \middle| k \in \mathbb{Z} \right\} \cap [s_i - C\gamma, s_i + C\gamma] \right| \leq 2 \lfloor C \rfloor + 1.$$

Therefore, the assertion follows by straightforward combinatorics.

(b) Let $i = 1, \dots, p$. Since $t_i - s_i \leq \gamma \leq \gamma_1 - \gamma_2$, it holds that $s_i \leq t_i \leq t_i + \gamma_2 \leq s_i + \gamma_1$, it follows that

$$[t_i, t_i + \gamma_2] \subset [s_i, s_i + \gamma_1]$$

and the assertion follows immediately.

(c) Let $i = 1, \dots, p$. W.l.o.g. let $t_i = \max\{s_i, t_i\}$. Since $t_i - s_i \leq \gamma \leq \gamma_1$, it holds that $s_i \leq t_i \leq s_i + \gamma_1$. Therefore, $\max\{s_i, t_i\} = t_i \in [s_i, s_i + \gamma_1] \cap [t_i, t_i + \gamma_1]$. It follows that

$$(\max\{s_1, t_1\}, \dots, \max\{s_p, t_p\})' \in [\boldsymbol{s}, \boldsymbol{s} + \gamma_1] \cap [\boldsymbol{t}, \boldsymbol{t} + \gamma_1],$$

thus showing the assertion. □

Lemma C.2. *(Federer (1969), Theorem 3.2.35 and Billingsley (1999), Appendix M17)*
Let $A \subset \mathbb{R}^2$ be convex. Let $\delta > 0$. It holds that

$$\lambda(A^\delta) - \lambda(A) = \ell(\partial A)\delta + \pi\delta^2.$$

Lemma C.3.
For any convex set $A \subset \mathbb{R}^2$ it holds that

$$\left| \left\{ (i,j)' \in \mathbb{Z}^2 : \left(\frac{i}{T}, \frac{j}{T}\right)' \in A \right\} \right| = T^2\lambda(A) + O(T)$$

Proof. The idea of the proof is illustrated in Figure C.1. Furthermore, this statement can be found in e. g. Gruber and Lekkerkerker (1987), p.141 and Herz (1962), p.2 who refer to the statement as the 'trivial bound'. Let

$$M_T = \bigcup_{\substack{(i,j)' \in \mathbb{Z}^2: \\ \left(\frac{i}{T}, \frac{j}{T}\right)' \in A}} \left(\left[\frac{i}{T}, \frac{i+1}{T} \right) \times \left[\frac{j}{T}, \frac{j+1}{T} \right) \right). \tag{C.1}$$

By construction,

$$\left| \left\{ (i,j)' \in \mathbb{Z}^2 : \left(\frac{i}{T}, \frac{j}{T}\right)' \in A \right\} \right| = T^2\lambda(M_T).$$

Therefore, by establishing bounds for $\lambda(M_T)$, we are able to give bounds on the number of lattice points of the form $(i/T, j/T)' \in A$. Note that by Lemma C.2, we

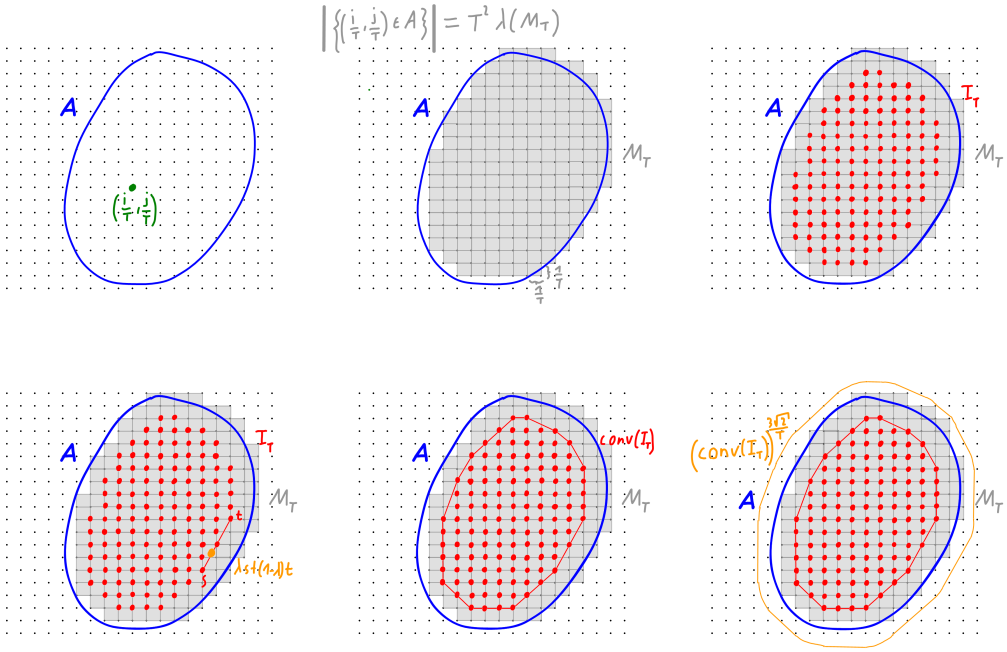

Figure C.1: Illustration of the idea of the proof of Lemma C.3: For a convex set A, the number of grid points inside A is equal to $T^2\lambda(M_T)$, with M_T defined in (C.1). With the help of the convex hull of an 'inner' set of points both in A and M_T, we are able to give an upper bound for the difference of the Lebesgue measures of A and M_T.

know the difference in Lebesgue measures between a convex set and an expanded version. Therefore, the idea of the proof is to find a suitable convex set B and a suitable $\delta = \delta(T)$ such that $B \subset M_T \subset B^\delta$ such that the assertion can be shown.
In order to do that, we define a set of 'inner' points

$$I_T = \left\{ \left(\frac{i}{T}, \frac{j}{T}\right)' \middle| (i,j) \in \mathbb{Z}^2, \left(\frac{i+k}{T}, \frac{j+m}{T}\right)' \in A \text{ for all } k,m \in \{-1,0,1\} \right\}$$

of the form $(i/T, j/T)'$ such that for any point in I_T, all neighbors are inside A. Denote by $\mathrm{conv}(I_T)$ the convex hull of I_T.
Our goal is to show that $\mathrm{conv}(I_T) \subset A \cap M_T$. Note that in $\mathbb{R}^2$, the border of a convex hull of a point set I_T is a polygonal chain of a subset of points in I_T. Therefore, it is sufficient to show that for any $\boldsymbol{s}, \boldsymbol{t} \in I_T$, $\lambda \in [0,1]$, $\lambda\boldsymbol{s} + (1-\lambda)\boldsymbol{t} \in A \cap M_T$ as it implies that any polygonal chain of a subset of points in I_T is a subset of $A \cap M_T$.
Since by definition of I_T, $\boldsymbol{s}, \boldsymbol{t} \in A$, and since A is convex, $\lambda\boldsymbol{s} + (1-\lambda)\boldsymbol{t} \in A$ as well.
For $\lambda\boldsymbol{s} + (1-\lambda)\boldsymbol{t} \in M_T$, we need to show the existence of some $(i_0, j_0)' \in \mathbb{Z}^2$ such that $(i_0/T, j_0/T)' \in A$ and $\lambda\boldsymbol{s} + (1-\lambda)\boldsymbol{t} \in [i_0/T, (i_0+1)/T) \times [j_0/T, (j_0+1)/T) \subset M_T$.
By construction of I_T, it holds that

$$\left[s_1 - \frac{1}{T}, s_1 + \frac{1}{T}\right) \times \left[s_2 - \frac{1}{T}, s_2 + \frac{1}{T}\right) \subset A,$$
$$\left[t_1 - \frac{1}{T}, t_1 + \frac{1}{T}\right) \times \left[t_2 - \frac{1}{T}, t_2 + \frac{1}{T}\right) \subset A$$

and therefore, due to the convexity of A,

$$\left[\lambda s_1 + (1-\lambda)t_1 - \frac{1}{T}, \lambda s_1 + (1-\lambda)t_1 + \frac{1}{T}\right)$$
$$\times \left[\lambda s_2 + (1-\lambda)t_2 - \frac{1}{T}, \lambda s_2 + (1-\lambda)t_2 + \frac{1}{T}\right) \subset A.$$

Therefore, there exists $(i_0, j_0)' \in \mathbb{Z}^2$ with

$$\begin{aligned}\lambda \boldsymbol{s} + (1-\lambda)\boldsymbol{t} &\in \left[\frac{i_0}{T}, \frac{i_0+1}{T}\right) \times \left[\frac{j_0}{T}, \frac{j_0+1}{T}\right) \\ \subset &\left[\lambda s_1 + (1-\lambda)t_1 - \frac{1}{T}, \lambda s_1 + (1-\lambda)t_1 + \frac{1}{T}\right) \\ &\times \left[\lambda s_2 + (1-\lambda)t_2 - \frac{1}{T}, \lambda s_2 + (1-\lambda)t_2 + \frac{1}{T}\right),\end{aligned}$$

It follows that $[i_0/T, (i_0+1)/T) \times [j_0/T, (j_0+1)/T) \subset M_T$ and therefore, $\lambda \boldsymbol{s} + (1-\lambda)\boldsymbol{t} \in M_T$.
Furthermore, if T is large enough, it holds due to the convexity of A and $\lambda(A) > 0$ that for each point of the form $(i/T, j/T)' \in A$, there exist $k, m \in \{-1, 0, 1\}$ such that $((i+k)/T, (j+m)/T)' \in I_T$. Furthermore, for each $\boldsymbol{x} \in M_T \cup A$, there exist $(i_0/T, j_0/T)' \in A$ and $k_0, m_0 \in \{-1, 0, 1\}$ such that

$$\boldsymbol{x} \in \left[\frac{i_0+k_0}{T}, \frac{i_0+k_0+1}{T}\right) \times \left[\frac{j_0+m_0}{T}, \frac{j_0+m_0+1}{T}\right).$$

Therefore, we obtain that

$$\begin{aligned}&\min_{\boldsymbol{y} \in \operatorname{conv}(I_T)} \|\boldsymbol{x} - \boldsymbol{y}\| \leq \min_{\boldsymbol{y} \in I_T} \|\boldsymbol{x} - \boldsymbol{y}\| \leq \min_{\boldsymbol{z} \in M_T, \boldsymbol{y} \in I_T} (\|\boldsymbol{x} - \boldsymbol{z}\| + \|\boldsymbol{z} - \boldsymbol{y}\|) \\ \leq &\frac{2\sqrt{2}}{T} + \frac{\sqrt{2}}{T} = \frac{3\sqrt{2}}{T}\end{aligned}$$

and therefore

$$\operatorname{conv}(I_T) \subset M_T \cap A \subset M_T \cup A \subset (\operatorname{conv}(I_T))^{3\sqrt{2}/T}.$$

By Lemma C.2 (i), we obtain that

$$\begin{aligned}&|\lambda(M_T) - \lambda(A)| \leq \lambda\left((\operatorname{conv}(I_T))^{3\sqrt{2}/T}\right) - \lambda\left(\operatorname{conv}(I_T)\right) \\ \leq &\frac{3\sqrt{2}}{T}\ell\left(\partial \operatorname{conv}(I_T)\right) + \frac{18}{T^2}\pi^2,\end{aligned}$$

which shows the assertion as the boundary of $\operatorname{conv}(I_T)$ has finite length due to the convexity and boundedness of the set. □

D Miscellaneous

Lemma D.1. *(a) For $i = 1, \ldots, n$ let $f_i : (X_i, d_i) \to (X_{i+1}, d_{i+1})$ be Lipschitz-continuous functions with Lipschitz-constants $C_i > 0$. Then, the composition $f_n \circ \ldots \circ f_1 : (X_1, d_1) \to (X_{n+1}, d_{n+1})$ is Lipschitz-continuous, as well.*

(b) Addition and Subtraction in $\mathbb{R}$ are Lipschitz-continuous.

(c) The absolute value $|x|$ is a Lipschitz-continuous function from $\mathbb{R}$ to $\mathbb{R}$.

(d) $\max\{x_1, \ldots, x_n\}$ and $\min\{x_1, \ldots, x_n\}$ are Lipschitz-continuous mappings from $\mathbb{R}^n$ to $\mathbb{R}$.

Proof. (a) Let $x, y \in X_1$. We obtain iteratively that

$$\begin{aligned} & d_{n+1}\left((f_n \circ \ldots \circ f_1)(x), (f_n \circ \ldots \circ f_1)(y)\right) \\ = \; & d_{n+1}\left(f_n \circ (f_{n-1} \circ \ldots \circ f_1)(x), f_n \circ (f_{n-1} \circ \ldots \circ f_1)(y)\right) \\ \leq \; & C_n d_n\left((f_{n-1} \circ \ldots \circ f_1)(x), (f_{n-1} \circ \ldots \circ f_1)(y)\right) \leq \ldots \leq \prod_{i=1}^{n} C_i d_1(x, y). \end{aligned}$$

Therefore, $f_n \circ \ldots \circ f_1$ is Lipschitz-continuous with Lipschitz-constant $\prod_{i=1}^{n} C_i$.

(b) The addition-operator is defined as

$$\begin{array}{rccc} P: & \mathbb{R} \times \mathbb{R} & \longrightarrow & \mathbb{R} \\ & (x_1, x_2) & \mapsto & x_1 + x_2. \end{array}$$

For $\boldsymbol{x} = (x_1, x_2)'$, $\boldsymbol{y} = (y_1, y_2)'$ let w.l.o.g. $|x_1 - y_1| \geq |x_2 - y_2|$. It holds that

$$\begin{aligned} & |P(x_1, x_2) - P(y_1, y_2)|^2 = |x_1 - y_1 + x_2 - y_2|^2 \\ = \; & (x_1 - y_1)^2 + (x_2 - y_2)^2 + 2(x_1 - y_1)(x_2 - y_2) \leq 4(x_1 - y_1)^2 \\ \leq \; & 4\left((x_1 - y_1)^2 + (x_2 - y_2)^2\right) = 4\left\|\boldsymbol{x} - \boldsymbol{y}\right\|^2. \end{aligned}$$

Therefore, Addition is Lipschitz-continuous with Lipschitz-constant 2.
The proof for Subtraction is analogous.

(c) By the reverse triangle inequality,

$$||x| - |y|| \leq |x - y|$$

for all $x, y \in \mathbb{R}$. Therefore, the absolute value is Lipschitz-continuous with Lipschitz-constant 1.

(d) For $n = 2$ it holds for $x_1, x_2 \in \mathbb{R}$ that

$$\max\{x_1, x_2\} = \frac{x_1 + x_2 + |x_1 - x_2|}{2}$$

$$\min\{x_1, x_2\} = \frac{x_1 + x_2 - |x_1 - x_2|}{2}.$$

By (a)-(c), $\max\{x_1, x_2\}$ and $\min\{x_1, x_2\}$ are Lipschitz-continuous as composition of Lipschitz-continuous functions.
For general $n \in \mathbb{N}$, we obtain inductively that

$$\begin{aligned}\max\{x_1, \ldots, x_n\} &= \max\{\max\{x_1, \ldots, x_{n-1}\}, x_n\} \\ \min\{x_1, \ldots, x_n\} &= \min\{\min\{x_1, \ldots, x_{n-1}\}, x_n\}\end{aligned}$$

are compositions of Lipschitz-continuous functions. Therefore, the assertion follows by (a). □

E Additional graphics for the Simulation study in Section 12

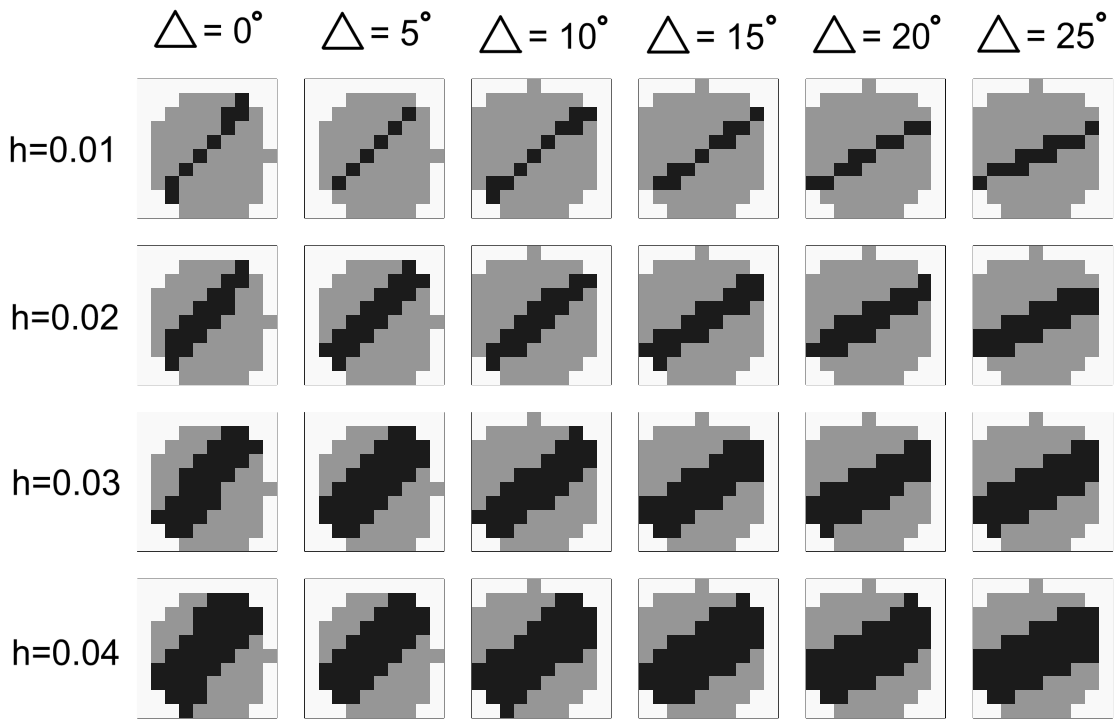

Figure E.1: The graphic shows the discretized versions of the sets $A^{(1,\alpha)}$-$A^{(3,\alpha)}$ (see Figure 11.1) in the situation of how we use them in the simulation study in Section 12. There, we have $T = 100$ and $d = 0.1$, leading to a 'circle' with a diameter of ten pixel. Δ is the difference between the angle of the 'inner strip' $A^{(1,\alpha)}$ and the true angle of the fissure in our simulation study, which is turned by $50°$ against the x-axis. $A^{(1,\alpha)}$ is displayed in black while $A^{(2,\alpha)}$ and $A^{(3,\alpha)}$ are displayed in gray.

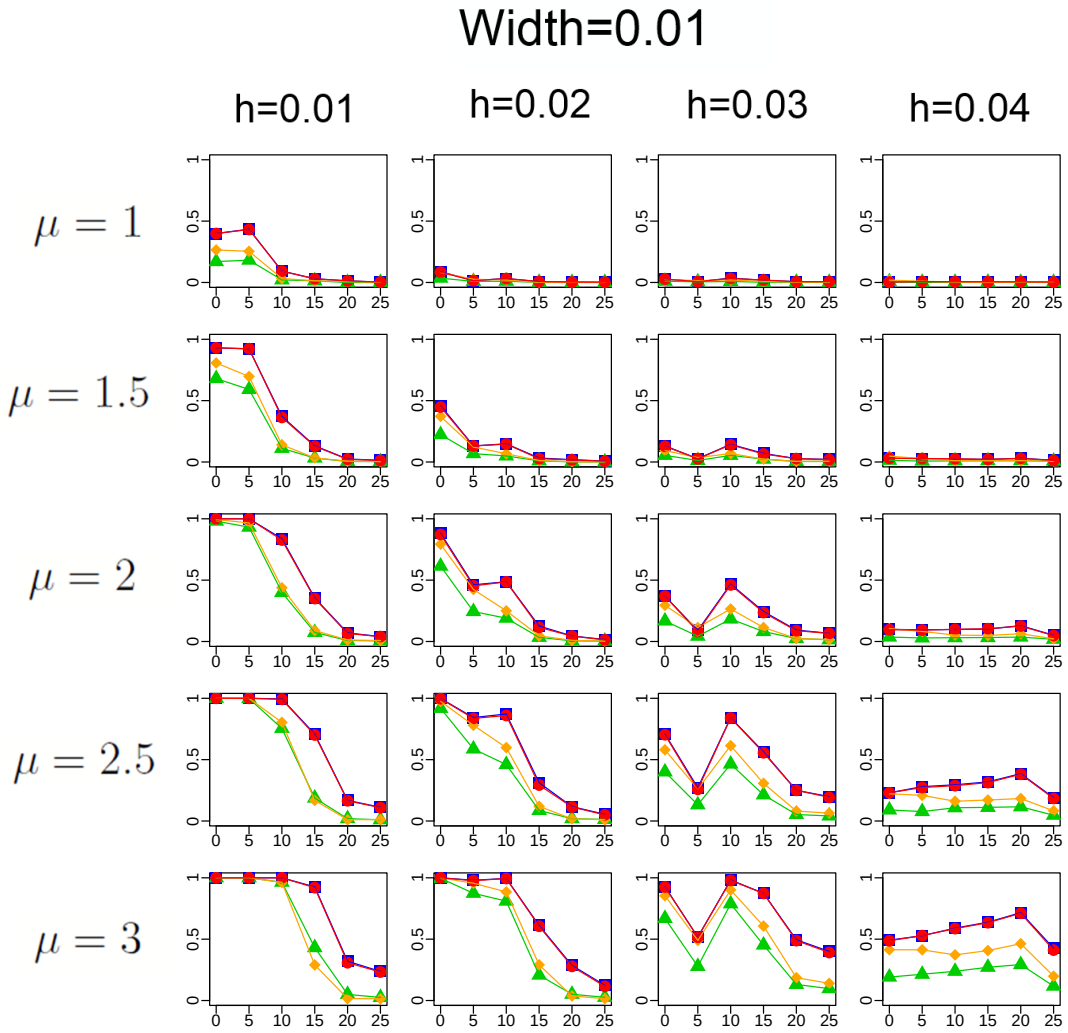

Figure E.2: Comparison of the detection rates for all combinations of signal, bandwidth, width of the fissure and variance estimators when considering the fissure to be detected if at least one of the pixel belonging to the fissure is significant. The yellow diamond-shaped points indicate the detection rates for the minimum-type estimator (A). The green triangle-shaped points indicate the detection rates for the local robust estimator (B). The red bullet-shaped points indicate the detection rates for the global robust estimator (C). The blue square-shaped points indicate the detection rates for the use of the true variance (D). On the x-axis, the misspecification of the angle of the fissure is displayed, on the y-axis the detection rates are displayed.

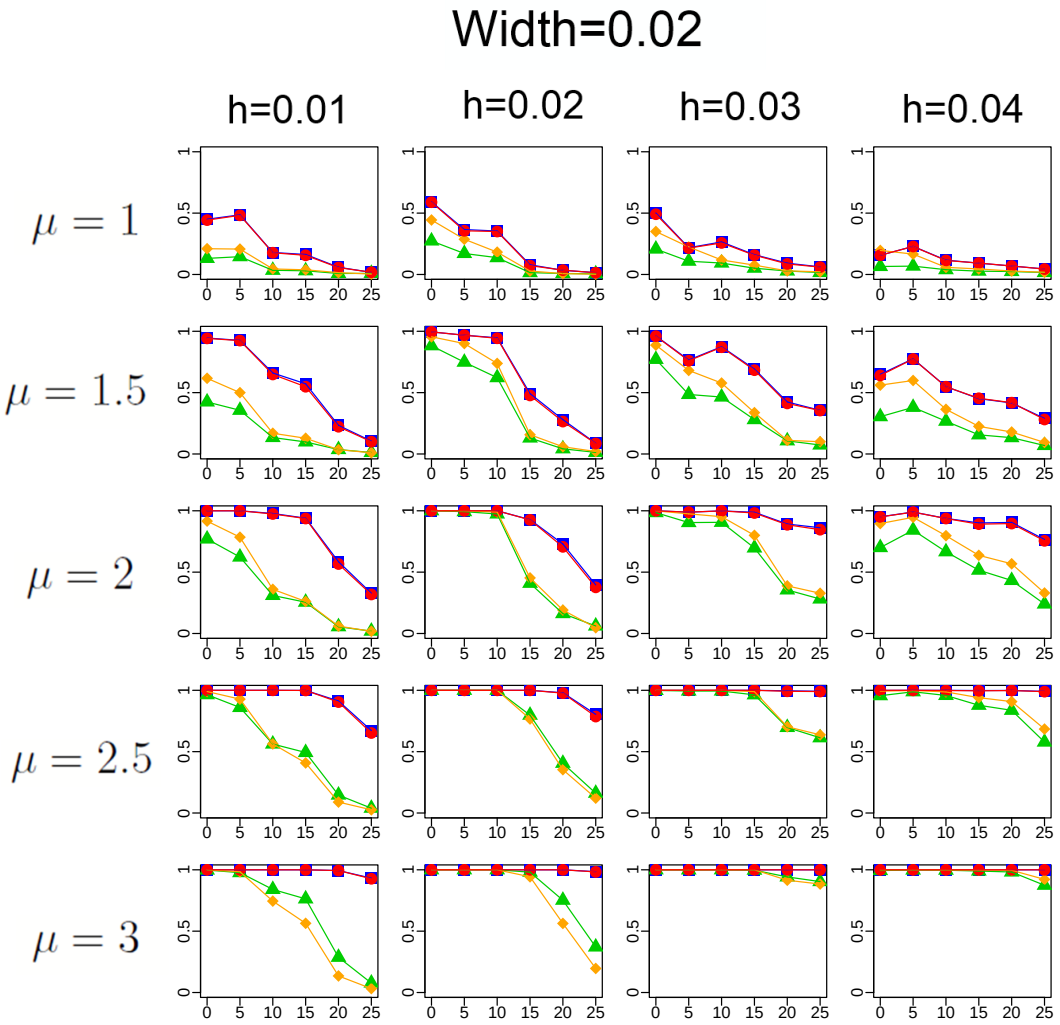

Figure E.3: Comparison of the detection rates for all combinations of signal, bandwidth, width of the fissure and variance estimators when considering the fissure to be detected if at least one of the pixel belonging to the fissure is significant. The yellow diamond-shaped points indicate the detection rates for the minimum-type estimator (A). The green triangle-shaped points indicate the detection rates for the local robust estimator (B). The red bullet-shaped points indicate the detection rates for the global robust estimator (C). The blue square-shaped points indicate the detection rates for the use of the true variance (D). On the x-axis, the misspecification of the angle of the fissure is displayed, on the y-axis the detection rates are displayed.

References

Acosta, J., Figueroa, J., and Mullen, R. (1992). "Low-cost video image processing system for evaluating pavement surface distress." In: *Transp. Res. Record*, pp. 63–72.

Aggarwal, R., Inclan, C., and Leal, R. (1999). "Volatility in emerging stock markets." In: *Journal of Financial and Quantitative Analysis* 34(1), pp. 33–55.

Amhaz, R., Chambon, S., Idier, J., and Baltazart, V. (2015). "A new minimal path selection algorithm for automatic crack detection on pavement images." In: *IEEE Image Proc.,* pp. 788–792.

Antoch, J. and Hušková, M. (1999). "Asymptotics, Nonparametrics, and Time Series". In: Marcel Dekker. Chap. Estimators of changes, pp. 533–578.

Arias-Castro, E., Candes, E. J., and Durand, A. (2011). "Detection of an anomalous cluster in a network." In: *The Annals of Statistics* 39, pp. 278–304.

Arias-Castro, E., Donoho, D., and Huo, X. (2005). "Near-optimal detection of geometric objects by fast multiscale methods". In: *EEE Trans. Inform. Theory* 51(7), 2402–2425.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2016). "A deep convolutional encoder-decoder architecture for image segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, pp. 2481–2495.

Baranowski, T., Dobrovolskij, D., Dremel, K., Hölzing, A., Lohfink, G., Schladitz, K, and Zabler, S. (2019). "Local fiber orientation from X-ray region-of-interest computed tomography of large fiber reinforced composite components". In: *Composites Science and Technology* 183, p. 107786.

Barisin, T., Jung, C., Müsebeck, F., Redenbach, C., and Schladitz, K. (2021). "Methods for segmenting cracks in 3d images of concrete: A comparison based on semi-synthetic images". In: *arXiv:2112.09493*.

Bauer, P. and Hackl, P. (1980). "An extension of the MOSUM technique for quality control." In: *Technometrics* 22(1), pp. 1–7.

Billingsley, P. (1995). *Probability and Measure*. Third Edition. John Wiley & Sons.

— (1999). *Convergence of Probability Measures*. Second edition. Wiley, New York.

Braun, J. V., Braun, RK., and Müller, H.-G. (2000). "Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation." In: *Biometrika* 87(2), pp. 301–314.

Brown E. N. an Kass, R. E. and Mitra, P. P. (2004). "Multiple neural spike train data analysis: state-of-the-art and future challenges". In: *Nature Neuroscience* 7(5), 456–461.

Bucchia, B. (2014). "Testing for epidemic changes in the mean of a multiparameter stochastic process." In: *J. Stat. Plan. Infer.* 150, pp. 124–142.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S., Brox, T., and Ronneberger, O (2016). "3D U-net: Learning dense volumetric segmentation from sparse annotation." In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* LNCS 9901, pp. 424–432.

Chan, H. P. and Chen, H. (2017). "Multi-sequence segmentation via score and higher-criticism tests." In: *arXiv:1706.07586*.

Chen, H., Chen, S., and Ding, X. (2019). "A Universal Nonparametric Event Detection Framework for Neuropixels Data". In: *bioRxiv:650671*.

Cho, H. and Fryzlewicz, P. (2012). "Multiscale and multilevel technique for consistent segmentation of nonstationary time series." In: *Statistica Sinica* 22, pp. 207–229.

Cho, H. and Kirch, C. (2020). "Data segmentation algorithms: Univariate mean change and beyond". In: *arXiv:2012.12814v2*.

— (2021+). "Two-stage data segmentation permitting multiscale changepoints, heavy tails and dependence". In: *Annals of the Institute of Statistical Mathematics, to appear*.

Csenki, A. (1979). "An Invariance Principle in k-dimensional Extended Renewal Theory." In: *J. Appl. Prob.* 16, pp. 567–574.

Csörgö, M. and Horvàth, L. (1997). *Limit theorems in change-point analysis*. vol. 18. John Wiley & Sons Inc.

Csörgö, M., Horváth, L., and Steinebach, J. (1987). "Invariance principles for renewal processes." In: *Ann. Probab.* 14, pp. 1441–1460.

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press.

Ehrig, K., Goebbels, Jue., Meinel, D., Paetsch, O., Prohaska, S., and Zobel, V. (2011). "Comparison of Crack Detection Methods for Analyzing Damage Processes in Concrete with Computed Tomography." In: *Int. Symp. Dig. Ind. Radiol. Comp. Tomogr.*

Eichinger, B. and Kirch, C. (2018). "A MOSUM procedure for the estimation of multiple random change points." In: *Bernoulli* 24, pp. 526–564.

Einmahl, U. (1987). "Strong invariance principles for partial sums of independent random vectors." In: *Ann. Probab.* 15, 1419–1440.

Fearnhead, P. and Rigaill, G. (2020). "Relating and Comparing Methods for Detecting Changes in Mean". In: *Stat*, e291.

Federer, H. (1969). *Geometric Measure Theory*. Springer Verlag.

Fisch, A. T. M., Eckley, I. A., and Fearnhead, P. (2018). "A linear time method for the detection of point and collective anomalies." In: *arXiv:1806.01947*.

Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (2000). "Multiscale Vessel Enhancement Filtering". In: *Lect. Notes Comput. Sc.* 1496, pp. 130–137.

Fryzlewicz, P. (2014). "Wild Binary Segmentation for multiple change-point detection." In: *The Annals of Statistics* 42, pp. 2243–2281.

Furat, O., Wang, M., Neumann, M., Petrich, L., Weber, M., Krill, C. E., and Schmidt, V. (2019). "Machine Learning Techniques for the Segmentation of Tomographic Image Data of Functional Materials". In: *Frontiers in Materials* 6, p. 145.

Glaz, J., Naus, J., and Wallenstein, S. (2001). *Scan statistics*. Springer.

Gruber, P. M. and Lekkerkerker, C. G. (1987). *Geometry of Numbers*. 2. Edition. ELSEVIER SCIENCE PUBLISHERS B.V.

Grün, S., Diesmann, M., and Aertsen, A. (2002). "Unitary events' in multiple single-neuron activity. II. Non-stationary data". In: *Neural Computation* 14(1), 81–119.

Grün, S. and Rotter, S. (2010). *Analysis of parallel spike trains*. Springer.

Gut, A. and Steinebach, J. (2002). "Truncated Sequential Change-point Detection based on Renewal Counting Processes." In: *Scandinavian Journal of Statistics* 29, pp. 693–719.

— (2009). "Truncated Sequential Change-point Detection based on Renewal Counting Processes II." In: *Journal of Statistical Planningand Inference* 139, pp. 1921–1936.

Haiman, G. and Preda, C. (2006). "Estimation for the distribution of two-dimensional discrete scan statistics." In: *Methodology and Computing in Applied Probability* 8(3), pp. 373–382.

Herz, C.S. (1962). "On the Number of Lattice Points in a convex set". In: *American Journal of Mathematics.* 84(1), pp. 126–133.

Heunis, A. J. (2003). "Strong invariance principle for singular diffusions." In: *Stochastic Processes and their Applications.* 104, pp. 57–80.

Horváth, L. and Rice, G. (2014). "Extensions of some classical methods in change point analysis". In: *Test* 23.2, pp. 219–255.

Horváth, L. and Steinebach, J. (2000). "Testing for changes in the mean or variance of a stochastic process under weak invariance." In: *J. Statist. Plann. Infer.* 91, pp. 365–376.

Hušková, M. and Slabỳ, A. (2001). "Permutation tests for multiple changes." In: *Kybernetika* 37(5), pp. 605–622.

Hušková, M. and Steinebach, J. (2000). "Limit theorems for a class of tests of gradual changes." In: *Journal of Statistical Planning and Inference* 89, pp. 57–77.

— (2002). "Asymptotic tests for gradual changes." In: *Statistics & Decisions* 20, pp. 137–151.

Ito, A., Aoki, Y., and Hashimoto, S. (2002). "Accurate extraction and measurement of fine cracks from concrete block surface image." In: *IEEE Ind. Elec.* 3, pp. 2202–2207.

Jaruskova, D. and Piterbarg, V. I. (2011). "Log-likelihood ratio test for detecting transient change." In: *Statist. Probab. Lett.* 81(5), pp. 552–559.

Jiang, T. (2002). "Maxima of partial sums indexed by geometrical structures." In: *Ann. Probab.* 30(4), 1854–1892.

Kabluchko, K. (2011). "Extremes of the standardized Gaussian noise". In: *Stochastic Processes and their Applications* 121(3), pp. 515–533.

Killick, R., Eckley, I. A., Ewans, K., and Jonathan, P (2010). "Detection of changes in variance of oceanographic time-series using changepoint analysis." In: *Ocean Engineering* 37, pp. 1120–1126.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). "Optimal detection of change-points with a linear computational cost." In: *Journal of the American Statistical Association* 102(500), pp. 1590–1598.

Kirch, C. and Klein, P. (2021). "Moving sum data segmentation for stochastic processes based on invariance". In: *Statistica Sinica, to appear*. DOI: 10.5705/ss.202021.0048.

Kirch, C. and Steinebach, J. (2006). "Permutation principles for the change analysis of stochastic processes under strong invariance." In: *Journal of Computational and Applied Mathematics* 184, pp. 64–88.

Kuelbs, J. and Philipp, W. (1980). "Almost sure invariance principles for partial sums of mixing B-valued random variables". In: *Ann. Probab.* 8, pp. 1003–1036.

Kulldorff, M. (1997). "A spatial scan statistic." In: *Comm. Statist. Theory Methods* 26(6), 1481–1496.

— (2016). "Exact asymptotics for the scan statistic and fast alternatives". In: *Electronic Journal of Statistics* 10, pp. 2641–2684.

Kühn, C. (2001). "An estimator of the number of change points based on a weak invariance principle". In: *Statistics & Probability Letters* 51, pp. 189–196.

Kühn, C. and Steinebach, J. (2002). "On the estimation of change parameters based on weak invariance principles." In: *Limit Theorems in Probability and Statistics II, I. Berkes, E. Csáki, M. Csörgö, eds., János Bolyai Math. Soc. Budapest*, pp. 237–260.

Li, H., Munk, A., and Sieling, H. (2016). "FDR-control in multiscale change-point segmentation." In: *Electronic Journal of Statistics* 10, pp. 918–959.

Lin, Z. and Bai, Z. (2010). *Probability Inequalities*. Springer.

Maidstone, R., Hocking, T., Rigaill, G., and Fearnhead, P. (2017). "On optimal multiple changepoint algorithms for large data." In: *Statistics & Computing* 27, pp. 519–533.

McInerney, T. and Terzopoulos, D. (1996). "Deformable models in medical image analysis: a survey." In: *Medical Image Analysis* 1(2), pp. 91–108.

Meier, A., Cho, H., and Kirch, C. (2021). "mosum: A package for moving sums in change point analysis." In: *Journal of Statistical Software* 97(8), pp. 1–42.

Messer, M., Costa, K. M., Roeper, J., and Schneider, G. (2017). "Multi-scale detection of rate changes in spike trains with weak dependencies." In: *Journal of Computational Neuroscience* 42, pp. 187–201.

Messer, M., Kirchner, M., Schiemann, J., Roeper, J., Neininger, R., and Schneider, G. (2014). "A multiple filter test for the detection of rate changes in renewal processes with varying variance." In: *The Annals of Applied Statistics* 8(4), pp. 2027–2067.

Messer, M. and Schneider, G. (2017). "The shark fin function: asymptotic behavior of the filtered derivative for point processes in case of change points." In: *Statistical Inference for Stochastic Processes* 20, pp. 253–272.

Mihalache, S.-R. (2011). "Sequential Change-Point Detection for Diffusion Processes". dissertation. Universität zu Köln.

Moon, N., Bullitt, K., Leemput, K. van, and Gerig, G. (2006). "Automatic brain and tumor segmentation." In: *MICCAI '02: Proceedings of the 5th International Conference on Medical Image Computing and Computer- Assisted Intervention-Part I* 8(3), pp. 373–382.

Munk, A., Proksch, K., Li, H., and Werner, F. (2020). "Photonic Imaging with Statistical Guarantees: From Multiscale Testing to Multiscale Estimation". In: *Nanoscale Photonic Imaging*. Vol. 134. Springer Verlag. Chap. 11, pp. 283–312.

Müsebeck, F., Moghiseh, A., Redenbach, C., and Schladitz, K. (2020). "Minimal paths for 3d crack detection in concrete." In: *Forum Bild-verarb.*, pp. 143–155.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Second Edition. Springer.

Niu, Y. S. and Zhang, H. (2012). "The screening and ranking algorithm to detect DNA copy number variations." In: *The Annals of Applied Statistics* 6, pp. 1306–1326.

Olshen, A. B., Venkatraman, E., Lucito, R., and Wigler, M. (2004). "Circular binary segmentation for the analysis of array-based DNA copy number data." In: *Biostatistics* 5, pp. 557–572.

Paetsch, O., Baum, D., Ehrig, K., Meinel, D., and Prohaska, S. (2012). "Automated 3D Crack Detection for Analyzing Damage Processes in Concrete with Computed Tomography." In: *Proc. Conf. Ind. Comp. Tomogr.*, pp. 321–330.

Page, E. S. (1954). "Continuous inspection schemes." In: *Biometrika* 41, pp. 100–115.

Perkel, D. H., Gernstein, G. L., and Moore, G. P. (1967). "Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains." In: *Biophysical Journal* 7(4), pp. 419–440.

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

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* LNCS 9351, pp. 234–241.

Roseman, A. M. (2003). "Particle finding in electron micrographs using a fast local correlation algorithm." In: *Ultramicroscopy* 94(3), pp. 225–236.

Sato, Y., Westin, C., Bhalerao, A., Nakajima, S., Shiraga, N., Tamura, S., and Kikinis, R. (2000). "Tissue classification based on 3d local intensity structures for volume rendering." In: *IEEE T. Vis. Comput. Gr.* 6(2), pp. 160–180.

Schneider, G. (2008). "Messages of oscillatory correlograms - a spike-train model." In: *Neural Computation* 20(5), 1211–1238.

Sharpnack, J. and Arias-Castro, E. (2016). "Exact asymptotics for the scan statistic and fast alternatives". In: *Electronic Journal of Statistics* 10, pp. 2641–2684.

Shi, Y., Cui, L., Qi, Z., Meng, F., and Chen, Z. (2016). "Automatic road crack detection using random structured forests." In: *IEEE T. Intell. Transp.* 17(12), pp. 3434–3445.

Shores, T. S. (2018). *Applied Linear Algebra and Matrix Analysis*. Second Edition. Springer.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & HallCRC.

Steinebach, J. (2000). "Some remarks on the testing of smooth changes in the linear drift of a stochastic process". In: *Theory of Probability and Mathematical Statistics* 61, pp. 173–185.

Steinebach, J. and Eastwood, V. R. (1996). "Extreme Value Asymptotics for Multivariate Renewal Processes." In: *Journal of multivariate analysis* 56(2), pp. 284–302.

Tang, J. and Gu, Y. (2013). "Automatic crack detection and segmentation using a hybrid algorithm for road distress analysis." In: *IEEE Sys. Man. Cybern.* 3, pp. 2202–2207.

Verzelen, N., Fromont, N., Lerasle, M., and Reynaud-Bouret, P. (2020). "Optimal change point detection and localization." In: *arXiv:2010.11470*.

Vostrikova, L. (1981). "Detecting disorder in multidimensional random processes." In: *Soviet Mathematics Doklady* 24, pp. 55–59.

Wang, D., Yu, Y., and Rinaldo, A. (2020). "Univariate mean change point detection: Penalization, CUSUM and optimality". In: *Electronic Journal of Statistics* 14.1, pp. 1917–1961.

Weise, F., Wiedmann, A., Voland, K., Kotan, E., Ehrig, K., and Müller, H. S. (2015). "Auswirkungen von Ermüdungsbeanspruchungen auf Struktur und Eigenschaften von Fahrbahndeckenbeton". In: *Beton- und Stahlbeton* 110(1), pp. 22–33.

Wichura, M. J. (1969). "Inequalities with Applications to the Weak Convergence of Random Processes with Multi-Dimensional Time Parameters". In: *The Annals of Mathematical Statistics* 40(2), pp. 681–687.

Wirjadi, O., Rack, A., Liebscher, A., Meinhardt, J., Schladitz, K., Shafei, B., and Steidl, G. (2014). "Microstructural Analysis of a C/SiC Ceramic Based on the Segmentation of X-ray Phase Contrast Tomographic Data." In: *Int. J. Mater. Res.* 105, pp. 702–708.

Yamaguchi, T. and Hashimoto, S. (2010). "Fast crack detection method for large-size concrete surface images using percolation-based image processing." In: *Mach. Vision Appl.* 21, pp. 797–809.

Yao, Y.-C. (1988). "Estimating the number of change-points via Schwarz' criterion." In: *Statistics & Probability Letters* 6(3), pp. 181–189.

Yao, Y.-C. and Au, S. T. (1989). "Least-squares estimation of a step function." In: *Sankhya: The Indian Journal of Statistics*, pp. 370–381.

Yau, C. Y. and Zhao, Z. (2016). "Inference for multiple change points in time series via likelihood ratio scan statistics". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78.4, pp. 895–916.

Zemlys, V. (2008). "Invariance principle for multiparameter summation processes and applications." dissertation. Vilnius University.