Change Point Detection in Random Fields

Masterarbeit
im Masterstudiengang Mathematik (1-Fach)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel
vorgelegt von

Niklas von Moers (1024545)

Erstgutachter: Prof. Dr. Mathias Vetter
Zweitgutachter: Prof. Dr. Sören Christensen
Kiel im April 2023

Contents

Introduction

In this master thesis, we will be exploring detection of epidemic changes in the distribution of weakly dependent random fields, in particular through the non-parametric test procedure introduced in [BW17]. That test is designed to detect changes in the mean in rectangular-shaped subsets and can, via a translation of $\mathbb{R}^p$-valued random fields to Hilbert space-valued ones, detect any changes in the distribution. For time series data, epidemic change sets can be represented as intervals. Rectangular-shaped change sets, being characterized by two points just like intervals, are the easier of the two natural extensions, the other being connected subsets of the index set.

In order to derive asymptotic results, a functional central limit theorem (FCLT) under the null hypothesis, in particular under stationarity, is proven. As the asymptotic test statistic arising under this FCLT is difficult to calculate and depends on the, in general unknown, long-run variance of the random field, critical values are estimated via a dependent wild bootstrap.

While this thesis almost exclusively mirrors the approach of [BW17], the results are supplemented with a formulation of and theoretical results under the alternative which have not been presented previously. Additionally, some minor errors are pointed out and gaps in the proofs are filled. The original definition of strongly separated blocks turns out to be particularly problematic. As the reader is not assumed to have knowledge on Hilbert space valued random elements, concepts such as their covariance are covered relatively indepth.

We start by motivating the problem of detecting epidemic changes in Chapter 1 where we consider the simple case of independent time series. It turns out that the asymptotic test distribution is closely related to the Kolmogorov distribution and that its derivation is not much more than an application of Donsker's theorem.

In the second chapter, fundamental concepts and conventions needed to grasp the main results presented in Chapter 3 are introduced. If one wishes, one may start with Chapter 3 and go back to Chapter 2 when definitions are unclear. Additionally, a short list of symbols is found at the end.

A discussion on the limitations and possible extensions of the results is found at the end of Chapter 3.

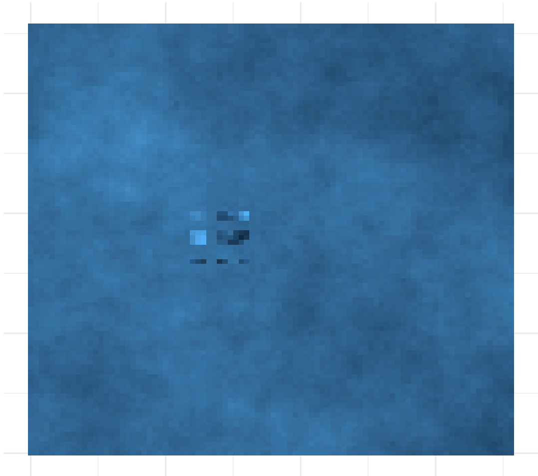

Figure 1: Simulation of a 2-parameter random field with change set.

For the sake of readability, most of the proofs have been deferred to Chapter 4. That chapter begins by introducing and proving the necessary technical lemmas. The proofs of the main results are found at the very end.

All figures found in this thesis were created by the author.

0.1 Deutsche Zusammenfassung

In dieser Masterarbeit wird das Problem der Strukturbrucherkennung in Zufallsfeldern behandelt. Dabei arbeiten wir im Wesentlichen genau die Ergebnisse aus [BW17] auf. Genauer behandeln wir das dort eingeführte statistische Testverfahren und dessen asymptotischen Eigenschaften, welches dazu ausgelegt ist, epidemische Änderungen im Erwartungswert Hilbertraum-wertiger Zufallsfelder und als Anwendung davon beliebige epidemische Änderungen in der Verteilung $\mathbb{R}^p$-wertiger Zufallsfelder zu erkennen.

In diesem Zuge werden zwei Sätze bewiesen: Der erste stellt einen funktionalen zentralen Grenzwertsatz dar und der zweite Satz erweitert den ersten um die gemeinsame Konvergenz der Partialsummenprozesse zusammen mit deren Bootstrap-Versionen. Die Anwendung dieser beiden Sätze beschreibt die asymptotischen Fehlerraten des statistischen Tests.

Ergänzend zu den Ergebnissen aus [BW17] werden die beiden Hypothesen "keine Änderung im Erwartungswert" und "epidemische Änderung im Erwartungswert" präzise formuliert und das Verhalten unter der Alternativhypothese beschrieben. Außerdem werden einige Fehler ausgebessert.

Chapter 1

Change in Mean Detection for Univariate Independent Observations

This chapter serves to give an intuition behind the topics covered. Therefore, we are for now considering the case $d = p = 1$, i.e., the random field is a stochastic process taking values in $\mathbb{R}$. Furthermore, we impose the strong condition that the observations are independent and we only test for changes in the mean.

It turns out that this special case of Theorem 3.2.1 can be thought of as an extension of Donsker's theorem 1.0.1 which itself can be thought of as a generalization of the central limit theorem.

Theorem 1.0.1 (Donsker's Theorem). *Let $X = (X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables that take values in $\mathbb{R}$ with mean μ and variance $\sigma^2 > 0$. Define the partial sum process*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

and the rescaled random walk W_n via

$$W_n(t) := \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}, \ t \in [0, 1].$$

Then, for each n, W_n is a random variable in the Skorokhod space $\mathcal{D}[0, 1]$ and the sequence $(\frac{1}{\sigma} W_n)_n$ converges in distribution to a standard Brownian motion W on $[0, 1]$.

Proof. The original proof may be found in [Don51]. □

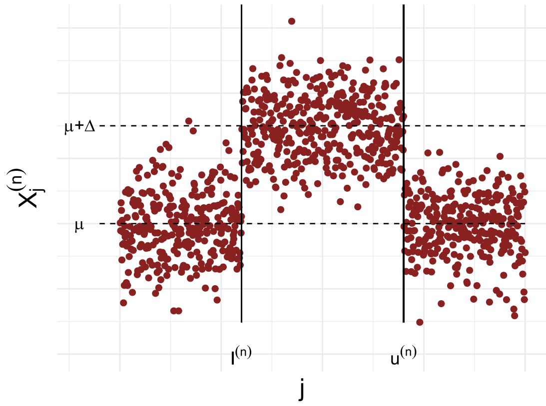

Figure 1.1: Notations under the Alternative H'_A.

1.1 Model and Results

For each $n \in \mathbb{N}$, let ${}_nX = ({}_nX_i)_{i\in\mathbb{N}}$ be a process with values in $\mathbb{R}$ and independent elements with constant variance $\mathrm{Var}({}_nX_i) = \sigma^2$. For a given $n \in \mathbb{N}$, we are interested in testing:

Hypothesis H'_0. The random variables ${}_nX_j$, $j \in \mathbb{N}$, are identically distributed and do not depend on n.

Hypothesis H'_A. There are change points ${}_nl < {}_nu \in \{0, ..., n\}$ and a change parameter $\Delta \in \mathbb{R} \setminus \{0\}$ such that the process $(\tilde{X}_j)_{j\in\mathbb{N}}$ defined by

$$\tilde{X}_j := \begin{cases} {}_nX_j - \Delta, & j \in ({}_nl, {}_nu] \\ {}_nX_j, & j \notin ({}_nl, {}_nu] \end{cases} \quad \forall j \in \mathbb{N}$$

consists of identically distributed elements and does not depend on n. Furthermore, the change set sizes ${}_nu - {}_nl =: c(n)$ are asymptotically proportional to the number of observations:

$$\lim_{n\to\infty} \frac{c(n)}{n} = \gamma \tag{1.1}$$

for a constant $\gamma \in (0, 1)$.

Remark 1.1.1. In order to derive asymptotic results as we will in Theorem 1.1.4 b) and later on in Corollary 3.2.7, one needs to model the Alternative H'_A (later H_A) using a family of stochastic processes ${}_nX$ indexed by time n because if one only considers a single process X that has a fixed change set of size $c \in \mathbb{N}$, this size vanishes as a fraction of the number of observations n. See Figure 1.2 for an illustration of this phenomenon.

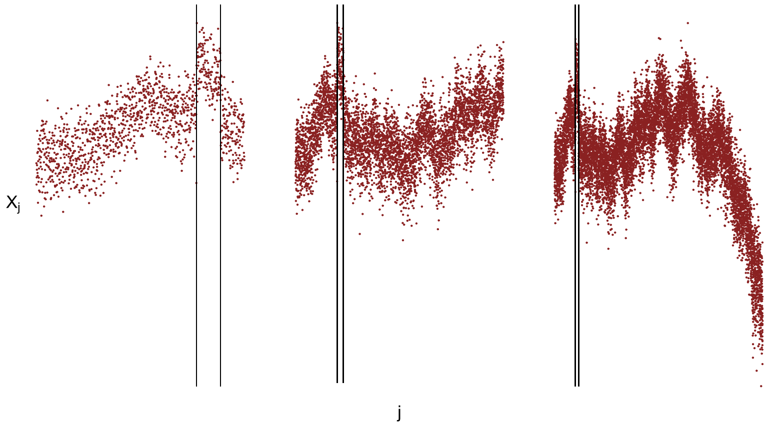

Figure 1.2: The same stochastic process with a change set (indicated by the vertical lines), sampled until three different points in time. While it seems plausible that one can detect the changes in the beginning, the change set of fixed size increasingly resembles random noise the more data is collected.

Define the test statistic

$$T_n := T_n({}_nX) := \frac{1}{\sqrt{n}} \max_{0\le k<m\le n} (m-k) \left| \frac{1}{m-k} \sum_{k<j\le m} {}_nX_j - \frac{1}{n} \sum_{1\le j\le n} {}_nX_j \right|. \tag{1.2}$$

For the exact test procedure we refer to Section 3.1.4.

Unfortunately, the asymptotic behaviour of the this test under the Alternative H'_A was not discussed in [BW17]. We have settled for this alternative in order to exploit the results we will develop under the Null Hypothesis H'_0.

Remark 1.1.2. The intuition behind this test statistic is that one tries find a subinterval in which the mean deviates from the overall mean as much as possible while weighing the difference of means by the size of that subinterval.

Lemma 1.1.3. *Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables that diverges to infinity in probability in the sense that*

$$\forall \lambda > 0 \lim_{n\to\infty} \mathbb{P}(Y_n \ge \lambda) = 1.$$

Then, for any nonnegative sequence $(U_n)_{n\in\mathbb{N}}$ of $\mathbb{R}$-valued random variables that is bounded from above by a weakly converging sequence $(U'_n)_n$, we have

$$\lim_{n\to\infty} \mathbb{P}(Y_n \ge U_n) = 1.$$

Proof. Let $\epsilon > 0$ be arbitrarily small and choose $\lambda > 0$ large enough such that

$$\lim_{n\to\infty} \mathbb{P}(U_n > \lambda) \le \lim_{n\to\infty} \mathbb{P}(U'_n > \lambda) < \epsilon.$$

Then

$$\begin{aligned}\mathbb{P}(Y_n < U_n) &= \mathbb{P}((Y_n < U_n \wedge U_n > \lambda) \vee (Y_n < U_n \wedge U_n \leq \lambda)) \\ &\leq \mathbb{P}(Y_n < U_n \wedge U_n > \lambda) + \mathbb{P}(Y_n < U_n \wedge U_n \leq \lambda) \\ &\leq \mathbb{P}(U_n > \lambda) + \mathbb{P}(Y_n < \lambda).\end{aligned}$$

By choice of λ, we have $\lim_{n\to\infty} \mathbb{P}(U_n > \lambda) < \epsilon$ and $\lim_{n\to\infty} \mathbb{P}(Y_n < \lambda) = 0$ holds by assumption. Consequently $\lim_{n\to\infty} \mathbb{P}(Y_n < U_n) < \epsilon$. As ϵ was chosen arbitrarily, $\lim_{n\to\infty} \mathbb{P}(Y_n < U_n) = 0$ is shown. □

Theorem 1.1.4. a) *Under the Null Hypothesis H_0', the test statistic T_n converges in distribution to*

$$T \coloneqq \sigma \sup_{0\leq s<t\leq 1} |W(t) - W(s) - (t-s)W(1)| \tag{1.3}$$

where W denotes a standard Brownian motion on $[0,1]$.

b) *Under the Alternative H_A', the test statistic T_n diverges in the sense that*

$$\lim_{n\to\infty} \mathbb{P}(T_n \geq Q_n) = 1 \tag{1.4}$$

for any weakly converging sequence $(Q_n)_n$ of $\mathbb{R}$-valued random variables.

Proof. a) In the following, we will apply Theorem 1.0.1, also using its notations. Without loss of generality, we can assume $\mathbb{E}\left[{}_nX_j\right] = 0$ as the test statistic T_n doesn't change when we replace ${}_nX_j$ with ${}_nX_j - \mu$. Note that by replacing m and k with $\lfloor sn \rfloor$ and $\lfloor tn \rfloor$ respectively, one has

$$\begin{aligned}T_n &= \frac{1}{\sqrt{n}} \max_{0\leq k<m\leq n} \left| \sum_{k<j\leq m} {}_nX_j - \frac{m-k}{n} \sum_{1\leq j\leq n} {}_nX_j \right| \\ &= \frac{1}{\sqrt{n}} \sup_{0\leq s<t\leq 1} \left| S_{\lfloor tn \rfloor} - S_{\lfloor sn \rfloor} - \frac{\lfloor tn \rfloor - \lfloor sn \rfloor}{n} S_n \right| \\ &= \sup_{0\leq s<t\leq 1} \left| W_n(t) - W_n(s) - \frac{\lfloor tn \rfloor - \lfloor sn \rfloor}{n} W_n(1) \right|.\end{aligned}$$

We apply Theorem 1.0.1 to get the convergence of the W_n-terms. Applying Slutsky's theorem, we get the weak convergence

$$\frac{\lfloor tn \rfloor - \lfloor sn \rfloor}{n} W_n(1) \Rightarrow (t-s)W(1).$$

Finally, apply the continuous mapping theorem to get the stated convergence.

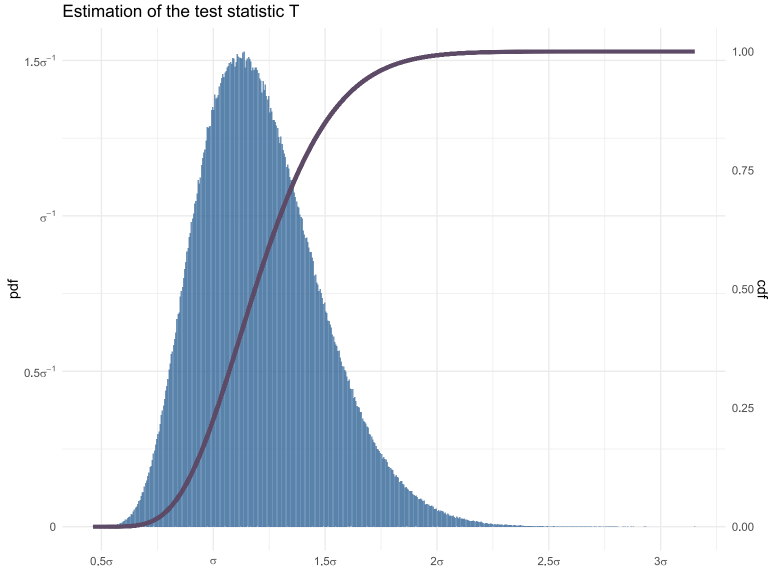

Figure 1.3: Histogram estimation of the density and empirical distribution function of a discrete approximation of the test statistic T, using approxmimately 2,250,000 samples. Note that the true test statistic is slightly larger as the supremum was only taken over 1,000 equidistant points.

b) We have

$$\begin{aligned} T_n &= \frac{1}{\sqrt{n}} \max_{0 \le k < m \le n} \left| \sum_{k < j \le m} {}_nX_j - \frac{m-k}{n} \sum_{1 \le j \le n} {}_nX_j \right| \\ &\ge \frac{1}{\sqrt{n}} \left| \sum_{{}_nl < j \le {}_nu} {}_nX_j - \frac{{}_nu - {}_nl}{n} \sum_{1 \le j \le n} {}_nX_j \right| \\ &= \frac{1}{\sqrt{n}} \left| \sum_{j={}_nl+1}^{{}_nu} \tilde{X}_j - \frac{c(n)}{n} \sum_{j=1}^{n} \tilde{X}_j + \left(1 - \frac{c(n)}{n}\right) c(n)\Delta \right|. \end{aligned}$$

By the reverse triangle inequality, the above is greater than or equal to $Y_n - V_n$ with

$$\begin{aligned} Y_n &:= \frac{1}{\sqrt{n}} \left(1 - \frac{c(n)}{n}\right) c(n)|\Delta|, \\ V_n &:= \frac{1}{\sqrt{n}} \left| \sum_{j={}_nl+1}^{{}_nu} \tilde{X}_j - \frac{c(n)}{n} \sum_{j=1}^{n} \tilde{X}_j \right|. \end{aligned}$$

Because V_n is bounded from above by $T_n(\tilde{X})$, we have for any weakly converging sequence Q_n

$$\begin{aligned}\mathbb{P}(T_n({}_nX) \geq Q_n) &\geq \mathbb{P}(Y_n - V_n \geq Q_n) \\ &\geq \mathbb{P}(Y_n \geq Q_n + T_n(\tilde{X})).\end{aligned}$$

Applying Lemma 1.1.3 to the weakly converging sequence $U_n = Q_n + T_n(\tilde{X})$, we therefore have

$$\lim_{n\to\infty} \mathbb{P}(T_n({}_nX) \geq Q_n) \geq \lim_{n\to\infty} \mathbb{P}(Y_n \geq Q_n + T_n(\tilde{X})) = 1.$$

$\square$

The following table describes the empirical quantiles from our simulations where we have replaced the supremum with a maximum over 1,000 equidistant points.

percentile	value
50	1.186597
60	1.257144
70	1.337206
80	1.436288
90	1.582738
95	1.710000
99	1.964733
99.9	2.271218
99.99	2.526544

Remark 1.1.5. As a standard Brownian bridge $B = (B(t))_{t\in[0,1]}$ is given by

$$B(t) = W(t) - tW(1)$$

in distribution, we have for the limiting test statistic

$$T = \sigma \sup_{0\leq s<t\leq 1} |W(t) - W(s) - (t-s)W(1)| = \sigma \sup_{0\leq s<t\leq 1} |B(t) - B(s)|\,.$$

This asymptotic distribution is similar to that of the Kolmogorov-Smirnov statistic which is used to test whether or not a sample comes from a given distribution. Assuming the given distribution has a continuous cdf, the Kolmogorov-Smirnov statistic has the asymptotic distribution $D = \sup_{0\leq t\leq 1} |B(t)|$, called the Kolmogorov distribution. See [Vaa98] Corollary 19.21.

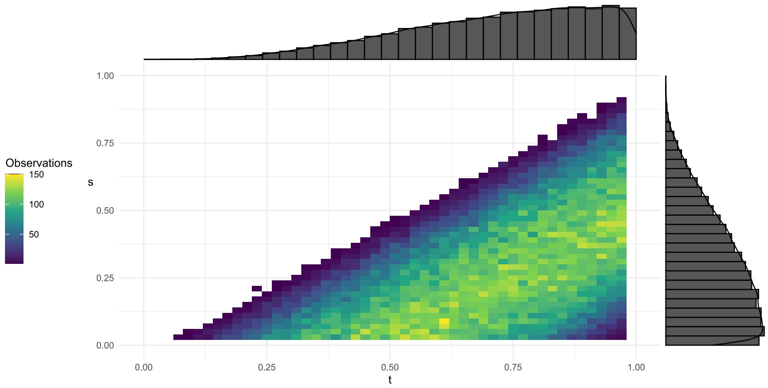

Figure 1.4: Histogram estimation of the joint distribution of the s and t maximizing the (discrete approximation) of the test statistic T, using approximately 100,000 samples.

1.2 Generalization

In the following chapters, we will generalize the previously discussed results in a variety of ways:

a) The observations are no longer assumed to be independent but rather weakly dependent, see Assumptions B, D and D' in Section 3.1.2.

b) We replace "time" with "space", i.e., the index set of our observations is $\mathbb{N}^d$ instead of $\mathbb{N}$. For technical reasons, namely to make use of the notion of stationarity, we will assume that the observations come from a random field that lives on $\mathbb{Z}^d$. This seems to be standard practice, see, e.g., [BD91] Chapter 1.3 Remark 3. From now on, if we refer to the "time" n, we are not talking about the index of the random field itself; we rather imagine that we have observed the values of X on the index set $\{1, ..., n\}^d$.

c) The observations are multivariate, taking values in some $\mathbb{R}^p$.

d) We are no longer only interested in changes in the mean but rather in all changes in the distribution - even in changes in, say, the dependence structure of the marginal distributions.

While the third generalization causes relatively few problems, the first two require a lot of machinery. The step from stochastic processes to random fields is not always as straightforward as a simple induction over the dimension d.

The last point of the above requires a "trick" which we will explain in Section 3.1.1. This "trick" will have us consider countably-infinite-dimensional Hilbert spaces. These will be approximated by their finite-dimensional subspaces.

Chapter 2

Preliminaries

This chapter mainly introduces conventions and definitions. In particular, we discuss the concepts around random fields and Hilbert space valued random elements.

The preliminary results we will be proving may be found in Chapter 4.

2.1 Conventions

We usually write vectors $\mathbf{x} \in \mathbb{R}^d$ in boldface. For a real number $x \in \mathbb{R}$, $\underline{\mathbf{x}} \in \mathbb{R}^d$ denotes the d-dimensional vector that has x in each of its components. We define $[x] \coloneqq \prod_i \mathbf{x}_i$. When we write $\mathbf{x}_i$, $i \in \{1, ..., d\}$, we mean the i-th component $\langle \mathbf{x}, e_i \rangle$ of $\mathbf{x}$. Similarly, for a matrix $A \in \mathbb{R}^{k \times l}$, $A_{i,j}$ denotes the entry in the i-th row and j-th column. If f is a function $\mathbb{R} \to \mathbb{R}$, e.g., the floor function $\lfloor \cdot \rfloor$, with $f(\mathbf{x})$ and $f(\mathbf{A})$ we mean the vector $(f(\mathbf{x}_1), ..., f(\mathbf{x}_d))$ and the matrix $(f(A_{i,j}))_{i,j}$ respectively. We denote the transpose of matrices and vectors by $\cdot^T$. However, we usually do not distinguish between row and column vectors. For two normed vector spaces V and W, $\mathcal{L}(V, W)$ is the set of bounded linear operators $V \to W$. For a finite set U, $|U|$ denotes the number of elements in U.

Let H be a separable real Hilbert space throughout unless stated otherwise. We denote the scalar product $H \times H \to \mathbb{R}$ by $\langle \cdot, \cdot \rangle$ and the associated norm by $\| \cdot \|$. We equip H with the topology induced by $\| \cdot \|$ and the associated Borel-σ-algebra. When writing H^K for some integer K, we mean the direct sum of Hilbert spaces

$$H^K \coloneqq \bigoplus_{i=1}^{K} H.$$

Interpreting H as a metric space, this is the same as equipping the cartesian product $\prod_i H$ with the 2 product metric. Hence we generally equip products of metric spaces with the 2 product metric.

As H is separable, its dimension is at most countably infinite. Therefore, it has an orthonormal basis $(e_i)_{i\in I}$, $I \subset \mathbb{N}$, i.e., $\langle e_i, e_j\rangle = \delta_{i,j}$ and $\operatorname{span}(e_i)_i = H$. We fix this orthonormal basis. In the case $H = \mathbb{R}^p$, we consider the canonical basis. We let

$$P_k : H \to H_k := \operatorname{span}(e_1, ..., e_k) \subset H, h \mapsto \sum_{i=1}^{k} \langle h, e_i\rangle e_i$$

stand for the orthogonal projections. For a (possibly random) vector $Y \in H$, we usually write $Y^{(k)} = P_k Y \in H_k$ for the projection onto H_k.

Random elements in H live on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We usually do not mention this probability space. For a random element Y, we set

$$\|Y\|_p := \mathbb{E}\left[\|Y\|^p\right]^{1/p}.$$

If a sequence of random variables Y_n converges weakly to a random variable Y, we write $Y_n \Rightarrow Y$. If we want to make clear that this convergence is to be understood with respect to n, we write $Y_n \overset{n}{\Rightarrow} Y$.

If we call $X_{\mathbf{j}}$ a random field, we mean the random field $X = (X_{\mathbf{j}})_{\mathbf{j}}$ where $\mathbf{j}$ usually indexes $\mathbb{Z}^d$. We use a similar convention for random fields of the form $S_n = (S_n(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$. We usually write observations in "space" $\mathbf{t}$ using a subscript, i.e., $X_{\mathbf{t}}$. However, in some occasions, in particular if we want to write out the components of the vector $\mathbf{t}$, we may write $X(\mathbf{t})$ instead.

2.2 Blocks

A block, also called multi-dimensional interval, is our generalization of intervals. Blocks will mainly appear in two contexts: As the change-set under our alternative hypothesis and in the context of increments of a random field. The increments of a random field will be part of the characterization of Brownian sheets, just like increments of stochastic processes may be used to characterize Brownian motion.

While the following definitions appear in [BW17], most of them seem to have been first introduced in [BS07] Section 2.1.

Definition 2.2.1 (Blocks)**.** We call a set of the form

$$(\mathbf{s}, \mathbf{t}] := \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{s}_i < \mathbf{x}_i \leq \mathbf{t}_i \ \forall i \in \{1, ..., d\}\}$$

a *block* (in $\mathbb{R}^d$). A *discrete block* is the intersection of a block (in $\mathbb{R}^d$) and $\mathbb{Z}^d$. If it is clear that a block U is discrete, we also write $U = (\mathbf{s}, \mathbf{t}] \subset \mathbb{Z}^d$ for $\mathbf{s}, \mathbf{t} \in \mathbb{Z}^d$.

Definition 2.2.2. A discrete block $W = (\mathbf{v}, \mathbf{w}]$ is said to belong standardly to another discrete block $U = (\mathbf{s}, \mathbf{t}]$ if $\mathbf{v} = \mathbf{s}$ and $\mathbf{w} \leq \mathbf{t}$. In that case, we write $W \lhd U$.

Definition 2.2.3 (Increment)**.** Let X be a random field indexed by $[0,1]^d$ ($\mathbb{Z}^d$). The *increment* $X(B)$ of X around a block $\emptyset \neq B = \prod_{i=1}^d (s_i, t_i] \subset [0,1]^d$ ($\subset \mathbb{Z}^d$) is defined by

$$X(B) = \sum_{\boldsymbol{\epsilon} \in \{0,1\}^d} (-1)^{d - \sum_{i=1}^d \epsilon_i} X(\mathbf{s} + \boldsymbol{\epsilon}(\mathbf{t} - \mathbf{s})),$$

where the multiplication of the vectors $\boldsymbol{\epsilon}$ and $\mathbf{t} - \mathbf{s}$ is to be read componentwise. We define the increment around the empty set to be 0. If B is the disjoint union of blocks $B_1, ..., B_k$, we define the increment of X around B to be the sum of the increments around the individual blocks:

$$X(B) := \sum_{j=1}^k X(B_j).$$

Remark 2.2.4. One may convince themselves that the increment around a disjoint union of blocks does not depend on the way one chooses the blocks.

Remark 2.2.5. The definition of the increment $X(\mathbf{s}, \mathbf{t}]$ can be thought of as an alternating sum of the values of X at the vertices of the block $(\mathbf{s}, \mathbf{t}]$ where the "uppermost" corner $\mathbf{t}$ receives a positive sign and all other signs are defined such that two adjacent vertices have pairwise different signs. Furthermore, the definition is designed such that the partial sum field

$$S_n(\mathbf{t}) = \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{j} \leq \lfloor \mathbf{t} n \rfloor} X_{\mathbf{j}}$$

fulfills

$$\frac{1}{n^{d/2}} \sum_{\mathbf{k} < \mathbf{j} \leq \mathbf{m}} X_{\mathbf{j}} = S_n(\mathbf{k}/n, \mathbf{m}/n] \tag{2.1}$$

for two vectors $\underline{\mathbf{0}} \leq \mathbf{k} < \mathbf{m} \leq \underline{\mathbf{n}}$. This is due to the fact that $S_n(\cdot)$, considered as a (random) function on blocks, is finitely additive as noted in [BW71].

Also, increments behave well under induction, see the proof of Lemma 4.1.1.

Definition 2.2.6 (Strongly separated blocks)**.** A collection $(\mathbf{a}^1, \mathbf{b}^1], ..., (\mathbf{a}^m, \mathbf{b}^m]$ of blocks in $[0,1]^d$ is called *strongly separated* if there is a dimension $j \in \{1, ..., d\}$ such that, potentially after reordering,

$$\mathbf{a}_j^1 < \mathbf{b}_j^1 < \mathbf{a}_j^2 < ... < \mathbf{b}_j^m.$$

Remark 2.2.7. Definition 2.2.6 of strongly separated blocks does not just ensure that the distance between any two blocks is positive, but that there is a dimension in which no two blocks overlap. This is, however, not true in the original definition

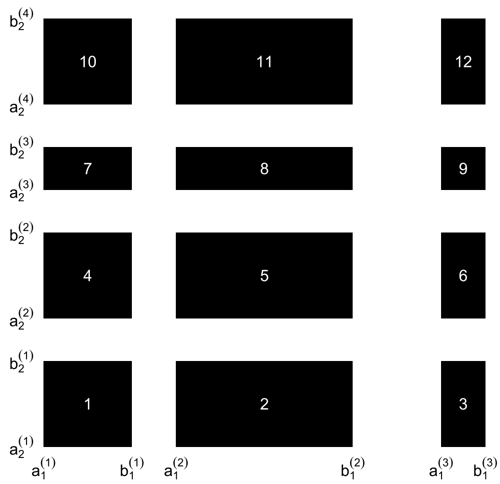

Figure 2.1: A family of blocks. The subfamily containing blocks 1, 5 and 6 is strongly separated while the one containing blocks 1, 2 and 4 is not. Using the definition of strongly separated blocks in [BW17] and [Deo75], any subfamily of the above is strongly separated.

of strongly separated blocks introduced in [Deo75] and later adopted in [BW17] which we have adjusted considerably. This can, for example, be seen by considering a collection of blocks of the form

$$\left(\begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \end{pmatrix}, \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \end{pmatrix}\right],$$
$$\left(\begin{pmatrix} a_1^{(1)} \\ a_2^{(2)} \end{pmatrix}, \begin{pmatrix} b_1^{(1)} \\ b_2^{(2)} \end{pmatrix}\right],$$
$$\left(\begin{pmatrix} a_1^{(2)} \\ a_2^{(1)} \end{pmatrix}, \begin{pmatrix} b_1^{(2)} \\ b_2^{(1)} \end{pmatrix}\right],$$

which corresponds to the Blocks 1, 2 and 4 in Figure 2.2. Therefore, the proof of Lemma 4.2.6 (Lemma 3 in [BW17]) only works with our adjusted definition.

Furthermore, the original definition implicitly demanded that the blocks do not overlap in any dimension, except when the two endpoints respectively coincide in that dimension. This is an issue as the argument at the end of the proof of Lemma 4.2.3 regarding the independence of increments around any collection of pairwise disjoint blocks would not work in that case.

Note that we prove the results in [Deo75] regarding strongly separated blocks in Section 4.1. See Remark 4.1.2 in particular.

The following definition is used in Lemma 4.2.14 for the implicit long-run variance estimator.

Definition 2.2.8. For two sets (usually blocks) $B, B' \subset \mathbb{R}^d$, we write

$$B \ominus B' := \{\mathbf{s} - \mathbf{t} \mid \mathbf{s} \in B, \mathbf{t} \in B'\} \subset \mathbb{R}^d.$$

The above product of sets is defined such that the following holds:

Lemma 2.2.9. *For a double-indexed sequence $(x_{\mathbf{a},\mathbf{b}})_{\mathbf{a}\in B, \mathbf{b}\in B'}$, with B and B' finite, one has*

$$\sum_{\mathbf{a}\in B} \sum_{\mathbf{b}\in B'} x_{\mathbf{a},\mathbf{b}} = \sum_{\mathbf{h}\in B' \ominus B} \sum_{\substack{\mathbf{a}: \\ \mathbf{a}\in B, \\ \mathbf{a}+\mathbf{h}\in B'}} x_{\mathbf{a},\mathbf{a}+\mathbf{h}}.$$

2.3 Hilbert Space Valued Random Elements

2.3.1 Expected Value and Covariance

As I have been unable to find a source in which the expected value and covariance of Hilbert space valued random elements is defined rigorously, I have decided to cover these two concepts, especially the covariance, relatively indepth. As a result of this, some of the following results and proofs are original.

Theorem 2.3.1 (Riesz-Fréchet)**.** *Let H be a (not necessarily separable) Hilbert space. For every bounded linear functional $\varphi : H \to \mathbb{R}$, there exists a unique vector $\mu_\varphi \in H$, called the Riesz representation of φ, such that*

$$\varphi(h) = \langle \mu_\varphi, h \rangle \ \forall h \in H. \tag{2.2}$$

The Riesz representation is given by

$$\mu_\varphi = \sum_i \varphi(e_i) e_i. \tag{2.3}$$

Furthermore the operator norm of φ and the norm of μ_φ are equal:

$$\|\varphi\|_{\mathrm{op}} = \|\mu_\varphi\|_H. \tag{2.4}$$

Proof. The first part and (2.4) are proven in [Con19] Theorem 3.4. To see that μ_φ defined as in (2.3) fulfills (2.2) is an easy calculation using the continuity of the inner product. □

For a random element Y in H with $\mathbb{E}\left[\|Y\|^2\right] < \infty$ define the linear functional

$$\varphi : H \to \mathbb{R}, h \mapsto \mathbb{E}\left[\langle Y, h\rangle\right]. \tag{2.5}$$

This is well-defined as Cauchy-Schwarz tells us

$$\mathbb{E}\left[|\langle Y, h\rangle|^2\right] \leq \mathbb{E}\left[\|Y\|^2\right]\|h\|^2 < \infty \tag{2.6}$$

which implies, using Jensen's inequality,

$$\mathbb{E}\left[|\langle Y, h\rangle|\right] \leq \sqrt{\mathbb{E}\left[|\langle Y, h\rangle|^2\right]} < \infty.$$

(2.6) also shows that the operator norm

$$\|\varphi\| = \sup_{\|h\|\leq 1} |\varphi(h)| \leq \sup_{\|h\|\leq 1} \mathbb{E}\left[|\langle Y, h\rangle|\right]$$

of φ is finite which means that φ is a bounded functional. We can therefore apply Theorem 2.3.1 to define the expected value of the Hilbert space valued random element Y:

Definition 2.3.2 (Expected value)**.** Let Y be a random element in H with $\mathbb{E}\left[\|Y\|^2\right] < \infty$. The *expected value* $\mathbb{E}[Y]$ of Y is the Riesz representation of φ, defined as in (2.5).

Remark 2.3.3. The expected value $\mathbb{E}[Y]$ satisfies

$$\mathbb{E}\left[\langle Y, h\rangle\right] = \langle \mathbb{E}[Y], h\rangle \ \ \forall h \in H$$

by definition and is given by

$$\mathbb{E}[Y] = \sum_i \mathbb{E}\left[Y^i\right] e_i \in H.$$

Additionally, the above definition of the expected value in H gives rise to a linear operator

$$\mathbb{E} : L^2(H) \to H, \ Y \mapsto \mathbb{E}[Y]$$

where

$$L^2(H) \coloneqq L^2((\Omega, \mathcal{F}, \mathbb{P}), H) \coloneqq \{Y : (\Omega, \mathcal{F}, \mathbb{P}) \to H \mid Y \text{ msbl.}, \ \mathbb{E}\left[\|Y\|^2\right] < \infty\}.$$

Proof. Apply Theorem 2.3.1 to obtain the first two claims. Using the linearity of the expectation in $\mathbb{R}$, a straightforward calculation shows

$$\mathbb{E}\left[aY_1 + Y_2\right] = \sum_i \mathbb{E}\left[(aY_1 + Y_2)^i\right] e_i = a\mathbb{E}[Y_1] + \mathbb{E}[Y_2]$$

for a scalar $a \in \mathbb{R}$ and random elements $Y_1, Y_2 \in L^2((\Omega, \mathcal{F}, \mathbb{P}), H)$. $\square$

Remark 2.3.4. The definition of the expected value of the H-valued random variable Y can also be read as a Bochner-integral (see [Boc33a]):

$$\mathbb{E}[Y] = \int_\Omega Y d\mathbb{P}. \tag{2.7}$$

The following proof is original.

Proof. From the Petti measurability theorem ([Pet38] Theorem 1.1.) together with the fact the we equip H with the Borel-σ-algebra and that we assume H to be separable, it follows that Y is measurable in the Bochner-sense. Now, if Y is a step function of the form

$$Y(\omega) = \sum_{i=1}^m \mathbb{1}_{A_i}(\omega) h_i$$

for a partition $(A_i)_i$ of H and vectors $h_i \in H$, then we have

$$\begin{aligned}\left\langle \int_\Omega Y d\mathbb{P}, h \right\rangle &= \left\langle \sum_{i=1}^m \mathbb{P}(A_i) h_i, h \right\rangle \\ &= \left\langle \sum_{i=1}^m \mathbb{E}\left[\mathbb{1}_{A_i}\right] h_i, h \right\rangle\end{aligned}$$

for all $h \in H$. By definition we may exchange the expectation and the inner product. Therefore, the above is equal to

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^m \mathbb{1}_{A_i} \langle h_i, h\rangle\right] &= \mathbb{E}\left[\left\langle \sum_{i=1}^m \mathbb{1}_{A_i} h_i, h \right\rangle\right] \\ &= \mathbb{E}\left[\langle Y, h\rangle\right].\end{aligned}$$

This means that 2.7 holds by Riesz-Fréchet 2.3.1 for step functions Y.

Now let Y be any element in $L^2(H)$. As Y is measurable in the Bochner-sense, there is a sequence of step functions $s_n : \Omega \to H$ that converge almost surely and in the L^2-norm to Y. Therefore, by the definition of the Bochner integral, we have for any element $h \in H$

$$\left\langle \int_\Omega Y d\mathbb{P}, h \right\rangle = \left\langle \lim_{n\to\infty} \int_\Omega s_n d\mathbb{P}, h \right\rangle.$$

By the continuity of the inner product, this is equal to

$$\lim_{n\to\infty} \left\langle \int_\Omega s_n d\mathbb{P}, h \right\rangle.$$

Using the fact that the Bochner integral and the definition of the expected value using Riesz-Fréchet 2.3.1 coincide for the step functions s_n as we have shown, the above is equal to

$$\lim_{n\to\infty} \langle \mathbb{E}[s_n], h\rangle = \lim_{n\to\infty} \mathbb{E}[\langle s_n, h\rangle].$$

Due to the convergence of $(s_n)_n$ to Y in the L^2-sense and the contiuinity of the inner product, we get

$$\lim_{n\to\infty} \mathbb{E}[\langle s_n, h\rangle] = \mathbb{E}[\langle Y, h\rangle]$$

which finally shows that the Bochner integral $\int_\Omega Y d\mathbb{P}$ and the expected value $\mathbb{E}[Y]$ are equal for general $Y \in L^2(H)$. $\square$

The following version of the Riesz-Fréchet representation theorem is [Rud91] Theorem 12.8.

Theorem 2.3.5 (Riesz-Fréchet, bilinear case). *Let $f : H \times H \to \mathbb{R}$ be a bounded bilinear form. Then there exists a unique $S \in \mathcal{L}(H, H)$ that satisfies*

$$f(x, y) = \langle x, S(y)\rangle \ \forall x, y \in H.$$

Moreover, $\|S\|_{\text{op}} = \|f\|_{\text{op}}$.

In order to define the covariance operator of two H-valued random elements Y, Z, first note that if Y and Z are random vectors in $\mathbb{R}^p$, the (cross-)covariance matrix is a nonnegative definite matrix $\text{Cov}(Y, Z) \in \mathbb{R}^{p\times p}$ defined by

$$\text{Cov}(Y, Z) := \mathbb{E}\left[(Y - \mathbb{E}[Y])(Z - \mathbb{E}[Z])^t\right]. \tag{2.8}$$

This matrix can be seen as a linear operator

$$\mathbb{R}^p \to \mathbb{R}^p,\ x \mapsto Cov(Y, Z)x$$

or as a bilinear form

$$\mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}, (x, y) \mapsto x^t\text{Cov}(Y, Z)y = \langle \text{Cov}(Y, Z)x, y\rangle$$

where the equality comes from the fact that the covariance matrix is symmetric. The generalization of the covariance of Hilbert space valued random elements as a bilinear form is straightforward. However, in order to regard the covariance as some kind of "matrix", one needs the notion of a tensor product. See Appendix A for details.

We begin the definition of the covariance operator of two Hilbert space valued random elements $Y, Z \in L^2(H)$ by the defining the bilinear covariance form

$$\operatorname{cov} := \operatorname{cov}(Y,Z) : H \times H \to \mathbb{R},\ (h_1,h_2) \mapsto \mathbb{E}\left[\langle Y - \mathbb{E}[Y], h_1\rangle \langle Z - \mathbb{E}[Z], h_2\rangle\right].$$

A simple calculation using Cauchy-Schwarz shows that cov is a bounded bilinear form. For any element $h \in \mathbb{R}$ we have, applying the Cauchy-Schwarz inequality twice and then Parseval's identity,

$$\begin{aligned}
\sum_{i\in I}\sum_{j\in I}\left|\langle \operatorname{cov}(e_i,e_j),h\rangle_{\mathbb{R}}\right|^2 &\le \sum_{i\in I}\sum_{j\in I}\left\|\mathbb{E}\left[\langle Y-\mathbb{E}[Y],e_i\rangle_H \langle Z-\mathbb{E}[Z],e_j\rangle_H\right]\right\|_{\mathbb{R}}^2 \|h\|_{\mathbb{R}}^2 \\
&\le \sum_{i\in I}\mathbb{E}\left[\langle Y-\mathbb{E}[Y],e_i\rangle_H^2\right]\sum_{j\in I}\mathbb{E}\left[\langle Z-\mathbb{E}[Z],e_i\rangle_H^2\right]\|h\|_{\mathbb{R}}^2 \\
&= \mathbb{E}\left[\sum_{i\in I}\langle Y-\mathbb{E}[Y],e_i\rangle_H^2\right]\mathbb{E}\left[\sum_{j\in I}\langle Z-\mathbb{E}[Z],e_i\rangle_H^2\right]\|h\|_{\mathbb{R}}^2 \\
&= \mathbb{E}\left[\|Y-\mathbb{E}[Y]\|_H^2\right]\mathbb{E}\left[\|Z-\mathbb{E}[Z]\|_H^2\right]\|h\|_{\mathbb{R}}^2.
\end{aligned}$$

Therefore, $\operatorname{cov}(Y,Z)$ is a weak Hilbert-Schmidt mapping (see Definition A.0.4) for

$$\alpha = \mathbb{E}\left[\|Y-\mathbb{E}[Y]\|_H^2\right]\mathbb{E}\left[\|Z-\mathbb{E}[Z]\|_H^2\right].$$

Hence we may apply the universal property A.0.5 to get a unique bounded linear functional

$$\tilde{\operatorname{cov}} := \tilde{\operatorname{cov}}(Y,Z) : H\hat{\otimes}H \to \mathbb{R}$$

such that

$$\operatorname{cov} = \tilde{\operatorname{cov}} \circ \hat{\otimes}.$$

Definition 2.3.6 (Covariance operator)**.** We define the *(cross-)covariance operator* $\operatorname{Cov}(Y,Z) \in H\hat{\otimes}H$ of Y and Z as the Riesz representation of $\tilde{\operatorname{cov}}$.

The *(auto-)covariance operator* $\operatorname{Cov}(Y)$ of Y is defined as the cross-covariance operator $\operatorname{Cov}(Y,Y)$.

Proposition 2.3.7. *The cross-covariance*

$$\operatorname{Cov}(\cdot,\cdot) : L^2(H)\times L^2(H) \to H\hat{\otimes}H, (Y,Z) \mapsto \operatorname{Cov}(Y,Z)$$

is a bilinear continuous map with operator norm $\|\operatorname{Cov}\|_{\mathrm{op}} = 1$*.*

The following proof is original.

Proof. We show that $\mathrm{Cov}(\cdot,\cdot)$ is bounded. Let $Y, Z \in L^2(H)$. Due to the equality of norms in (2.4) and (A.2), we have

$$\|\mathrm{Cov}(Y,Z)\|_{H\hat{\otimes}H} = \|\tilde{\mathrm{cov}}(Y,Z)\|_{\mathrm{op}} = \|\mathrm{cov}(Y,Z)\|_2.$$

The above is equal to (see Theorem A.0.5)

$$\sqrt{\sum_{i\in I}\sum_{j\in J}\mathrm{cov}(Y,Z)(e_i,e_j)^2} = \sqrt{\sum_{i\in I}\sum_{j\in J}\mathbb{E}\left[\langle Y-\mathbb{E}[Y], e_i\rangle \langle Z-\mathbb{E}[Z], e_j\rangle\right]^2}.$$

As shown previously, this is bounded from above by

$$\begin{aligned}\sqrt{\mathbb{E}\left[\|Y-\mathbb{E}[Y]\|_H^2\right]\mathbb{E}\left[\|Z-\mathbb{E}[Z]\|_H^2\right]} &\leq \sqrt{\mathbb{E}\left[\|Y\|_H^2\right]\mathbb{E}\left[\|Z\|_H^2\right]}\\ &= \|Y\|_{L^2(H)}\|Z\|_{L^2(H)}.\end{aligned}$$

Hence $\mathrm{Cov}(\cdot,\cdot)$ is bounded with operator norm $\|\mathrm{Cov}\|_{\mathrm{op}} \leq 1$. Cov indeed reaches this upper bound as one can see by considering $Y = Z$ with $\mathbb{E}[Y] = 0$, $\mathbb{E}[\|Y\|^2] = 1$ and $\langle Y, e_i\rangle = 0$ for all $i > 1$. $\square$

Remark 2.3.8. Another definition of the covariance operator $\mathrm{Cov}(Y,Z)$ is given by using Riesz-Fréchet 2.3.5 for bilinear forms which lets us define

$$\mathrm{Cov}_{\mathrm{op}}(Y,Z): H \to H$$

via

$$\langle h_1, \mathrm{Cov}_{\mathrm{op}}(Y,Z)(h_2)\rangle = \mathrm{cov}(Y,Z)(h_1,h_2).$$

In the following, we will identify the bilinear form

$$\mathrm{cov}(Y,Z): H\times H \to \mathbb{R},$$

the linear functional

$$\tilde{\mathrm{cov}}(Y,Z): H\hat{\otimes}H \to \mathbb{R},$$

the element of the tensor space

$$\mathrm{Cov}(Y,Z) \in H\hat{\otimes}H$$

and the linear operator

$$\mathrm{Cov}_{\mathrm{op}}(Y,Z): H \to H$$

with each other and use the one which fits our needs the best. In [BW17], the linear operator $\mathrm{Cov}_{\mathrm{op}}$ is used and defined via the bilinear form cov.

Remark 2.3.9. (i) In the finite-dimensional case $H = \mathbb{R}^p$, one has $\mathbb{R}^p \otimes \mathbb{R}^p \cong \mathbb{R}^{p\times p}$ and

$$\begin{aligned}\operatorname{cov}(Y,Z)(e_i,e_j) &= \mathbb{E}\left[(\langle Y - \mathbb{E}[Y], e_i\rangle)(\langle Z - \mathbb{E}[Z], e_j\rangle)\right] \\ &= \operatorname{Cov}(Y,Z)_{i,j}\end{aligned}$$

where $\operatorname{Cov}(Y,Z)$ denotes the usual covariance matrix. See also Example A.0.7.

(ii) The definition of the covariance operator can be extended to random elements that take values in two different Hilbert spaces.

Definition 2.3.10 (Asymptotic independence)**.** Let $k \in \mathbb{N}$ be fixed and let Y_n^i, $n \in \mathbb{N}$, $i = 1, ..., k$, be random elements in H. We say the Y_n^i, $i = 1, ..., k$, are asymptotically independent if

$$\lim_{n\to\infty}\left(\mathbb{P}(Y_n^i \in H_i\ \forall i = 1, ..., k) - \prod_{i=1}^{k}\mathbb{P}(Y_n^i \in H_i)\right) = 0$$

for all Borel sets H_i in H.

2.3.2 Hilbert Space Valued Random Fields

We define the direct extension of Brownian motion to random fields. Note that the following definition at the same time effortlessly generalizes the codomain from $\mathbb{R}$ to separable Hilbert spaces.

Definition 2.3.11 (Brownian sheet)**.** A *d-parameter Brownian sheet in H with covariance operator S* is an H-valued random field $W = (W_{\mathbf{t}})_{\mathbf{t}\in[0,1]^d}$ that satisfies the following properties:

a) W has continuous paths almost surely,

b) $W_{\mathbf{t}} = 0$ almost surely if one of the components of $\mathbf{t} \in [0,1]^d$ is 0,

c) the increments $W(B_1), \ldots, W(B_m)$ are independent Gaussian random elements in H with mean zero and covariance operator $\lambda(B_i)S$ for pairwise disjoint blocks $B_1, \ldots, B_m \subset [0,1]^d$, $m \in \mathbb{N}$.

Definition 2.3.12 (Stationarity)**.** Let $X = (X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^d}$ be a random field with values in H. X is said to be

a) *weakly stationary*, or alternatively *covariance stationary*, if

 (i) X has constant mean ($\mathbb{E}[X_{\mathbf{k}}] = \mu$ for all $\mathbf{k}$),

 (ii) X has second moments ($\mathbb{E}[\|X_k\|^2] < \infty$),

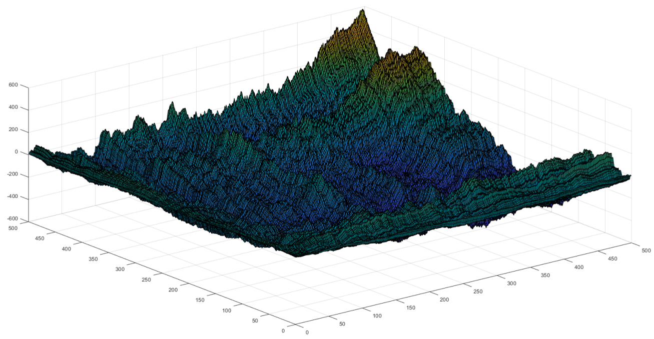

Figure 2.2: Sample path of a two-dimensional Brownian sheet in $\mathbb{R}$

(iii) The covariance of X is invariant under index shifts, i.e.,

$$\operatorname{Cov}\left(X_{\mathbf{i}+\mathbf{l}}, X_{\mathbf{j}+\mathbf{l}}\right) = \operatorname{Cov}\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right) \ \forall \mathbf{i}, \mathbf{j}, \mathbf{l}.$$

In that case we define the *auto-covariance function*, also called *covariogram*,

$$\gamma : \mathbb{Z}^d \to H \otimes H, \mathbf{v} \mapsto \gamma(\mathbf{v}) := \operatorname{Cov}(X_{\underline{\mathbf{0}}}, X_{\mathbf{v}}) \tag{2.9}$$

and the *long-run variance*

$$\Gamma := \sum_{\mathbf{v} \in \mathbb{Z}^d} \gamma(\mathbf{v})$$

in case the above series converges unconditionally (in particular if it converges absolutely). In the case $H = \mathbb{R}^k$, we also write $\gamma_{i,j}(\mathbf{v}) := (\gamma(\mathbf{v}))_{i,j}$ for the indices of the auto-covariance matrices.

b) *strictly stationary* if the finite-dimensional distributions are invariant under index shifts, i.e., the joint distributions of

$$X_{\mathbf{t}^{(1)}+\boldsymbol{\tau}}, \dots, X_{\mathbf{t}^{(k)}+\boldsymbol{\tau}}$$

do not depend on $\boldsymbol{\tau} \in \mathbb{Z}^d$.

For continuous-time random fields, weak and strict stationarity and the covariogram are defined analogously.

Example 2.3.13 (Stationarity). a) Any strictly stationary random field with finite first moment is weakly stationary. For Gaussian random fields, the converse is also true. This has been proven for Gaussian stochastic processeses in [Lap09] Proposition 25.5.1. However, the proof remains almost the same when replacing Gaussian stochastic processeses with Gaussian random fields.

b) Brownian sheets are not stationary as their covariance is not constant in space.

Definition 2.3.14 (Partial sum field)**.** Let X be an H-valued random field. The partial sum field S_n, $n \in \mathbb{N}$, is defined via

$$S_n(\mathbf{t}) := \frac{1}{n^{d/2}} \sum_{\mathbf{1} \leq \mathbf{j} \leq \lfloor n\mathbf{t} \rfloor} (X_{\mathbf{j}} - \mathbb{E}\left[X_{\mathbf{j}}\right]).$$

$\mathbf{t} \in [0,1]^d$.

2.4 Skorokhod Space

In the univariate case, Skorokhod spaces consist of càdlàg functions, that is, functions which are continuous from the right with limits from the left. As "left" and "right" may be understood in each dimension, in order to extend this definition to functions on $[0,1]^d$, we need to consider quadrants, which can be thought of as sub-intervals of $[0,1]^d$ which touch the boundary of $[0,1]^d$ in each dimension.

Skorokhod spaces are of interest as we will consider the partial sum fields as random elements in a Skorokhod space.

Definition 2.4.1 (Quadrant)**.** A *quadrant in* $[0,1]^d$ is a set of the form

$$Q(\mathbf{t}) = Q_{R_1,...,R_d}(\mathbf{t}) := \{\mathbf{s} \in [0,1]^d \mid \mathbf{s}_i R_i \mathbf{t}_i \ \forall i = 1,...,d\}$$

with relations $R_i \in \{<, \geq\}$ and corner point $\mathbf{t} \in [0,1]^d$.

Definition 2.4.2 (Càdlàg function)**.** We say that a function $f : [0,1]^d \to H$ is *càdlàg* if the limit

$$\lim_{\substack{\mathbf{s} \to \mathbf{t} \\ \mathbf{s} \in Q(\mathbf{t})}} f(\mathbf{s})$$

exists for all $\mathbf{t} \in [0,1]^d$ and quadrants $Q(\mathbf{t})$ in $[0,1]^d$ and if f is continuous from the right in the sense that

$$f(\mathbf{t}) = \lim_{\substack{\mathbf{s} \to \mathbf{t} \\ \mathbf{s} \in Q_{\geq,...,\geq}(\mathbf{t})}} f(\mathbf{s})$$

for all $\mathbf{t} \in [0,1]^d$.

Definition 2.4.3 (Skorokhod space)**.** The *Skorokhod space* $D_H[0,1]^d$ is defined as the space of càdlàg functions $[0,1]^d \to H$. We endow it with the so-called *Skorokhod metric*

$$d_S(f,g) := \inf_{\lambda \in \Lambda} \max(\|f - g \circ \lambda\|_\infty, \|\mathrm{id} - \lambda\|_\infty)$$

where Λ denotes the d-times cartesian product of the set of increasing bijections of $[0, 1]$.

We consider the space of continuous functions $C_H[0, 1]^d$ a subset of this Skorokhod space.

Remark 2.4.4. Unlike the supremum metric

$$d_\infty : (f, g) \mapsto \sup_{\mathbf{t}} \|f(\mathbf{t}) - g(\mathbf{t})\|_H,$$

which only considers "distance in space", the Skorokhod metric d_S also considers "distance in time". E.g. the distance of the indicator function of $[0.5, 1]$ and the one of $[0.5 + 1/n, 1]$ is 1 under the supremum metric while the distance converges to 0 under the Skorokhod metric, which can be seen by considering λ of the form

$$\lambda(x) = \begin{cases} (1 + 2/2n)x, & x < 0.5 \\ (1 - 2/n)x + 2/n, & x \geq 0.5 \end{cases}.$$

This implies that these two metrics are not equivalent. This is important as the weak convergences we will show later on in general do not hold when considering weak convergence with respect to the supremum metric.

Definition 2.4.5 (Modulus of continuity)**.** We define the *modulus of continuity* as the function

$$w \coloneqq w_H : D_H([0, 1]^d) \times [0, \infty) \to [0, \infty], (f, \delta) \mapsto w_f(\delta) \coloneqq \sup_{\|\mathbf{t}-\mathbf{s}\| \leq \delta} \|f(\mathbf{t}) - f(\mathbf{s})\|.$$

Remark 2.4.6. It can be shown that $D_H[0, 1]^d$ is a complete and separable metric space (see [Neu71]). In particular, it is Polish, allowing us to use Prokhorov's theorem. One may argue as in [Bil68] Theorem 8.1 to conclude that the convergence of the finite-dimensional distributions of probability measures $\mathbb{P}_n$ on $D_H[0, 1]^d$ to those of another probability measure $\mathbb{P}$ imply the convergence of $\mathbb{P}_n$ to $\mathbb{P}$ under the assumption that the collection $(\mathbb{P}_n)_n$ is tight. Tightness of measures will be assumed in (4.9) in the form of an assumption on the modulus of continuity, see also [BS07] Chapter 5.

2.5 Mixing

Intuitively, mixing coefficients measure how much mutual information different observations of a random field carry, depending on how far apart they are in space. The question how to measure distance does not have many different answers in the case of stochastic processes, i.e., the index set being $\mathbb{Z}$. In this case, one only needs to define wheter or not index sets are allowed to be "interlaced", see [Bra01] for a demonstration of the difference between the two. However, when moving to random fields, one has further options which are expressed in the two mixing coefficients ρ and ρ^*. In addition to these two, we are also interested in the α-mixing coefficients which has been introduced in [Ros56]. For an overview on some other mixing coefficients that are used in the literature, see [Bra86] and Bradley's other work.

Definition 2.5.1 (Mixing coefficients)**.** The mixing coefficients of two σ-fields $\mathcal{A}$ and $\mathcal{B}$ are defined via

$$\alpha(\mathcal{A},\mathcal{B}) := \sup_{\substack{A\in\mathcal{A}\\ B\in\mathcal{B}}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \tag{2.10}$$

and

$$\rho_H(\mathcal{A},\mathcal{B}) := \sup\left\{\frac{\mathbb{E}\left[\langle Y, Z\rangle\right] - \langle \mathbb{E}\left[Y\right], \mathbb{E}\left[Z\right]\rangle}{\|Y\|_2\|Z\|_2} \mid {\scriptstyle Y\in L^2(\mathcal{A},H), Z\in L^2(\mathcal{B},H), \atop \mathrm{Var}(Y),\mathrm{Var}(Z)>0}\right\}. \tag{2.11}$$

Now let X be a random field and define the σ-algebra generated by the observations X for an index set $M \subset \mathbb{Z}^d$ as

$$\mathcal{A}_M := \sigma(X_{\mathbf{j}} : \mathbf{j} \in M).$$

Then the mixing coefficients of X are defined as the functions given by

$$\alpha_{k,m}(r) = \sup\{\alpha(\mathcal{A}_M, \mathcal{A}_N) \mid M, N \subset \mathbb{Z}^d, \mathrm{dist}(M,N) \geq r, |M| \leq k, |N| \leq m\}, \tag{2.12}$$

$$\rho_H(r) := \sup\{\rho_H(\mathcal{A}_M,\mathcal{A}_N) \mid M, N \subset \mathbb{Z}^d, \exists i \in \{1,...,d\} \exists A, B \subset \mathbb{Z}, \mathrm{dist}(A,B) \geq r : \forall \mathbf{j} \in M, \mathbf{k} \in N : \mathbf{j}_i \in A, \mathbf{k}_i \in B\}, \tag{2.13}$$

$$\rho^*_H(r) := \sup\{\rho_H(\mathcal{A}_M, \mathcal{A}_N) \mid M, N \subset \mathbb{Z}^d, \mathrm{dist}(M,N) \geq r\} \tag{2.14}$$

where dist denotes the distance w.r.t. the supremum norm.

A random field is said to be *c-mixing* for some given coefficient function c (here ρ, ρ^* or $\alpha_{k,m}$) if $\lim\limits_{r\to\infty} c(r) = 0$.

Remark 2.5.2. As was shown in [BB85] Theorem 4.2, the ρ_H-coefficient does not depend on what space H one chooses. For this reason, we will usually write ρ instead of ρ_H from now on. One may always think about the case $H = \mathbb{R}$, in which case the definition of the ρ-mixing coefficient measures the maximum absolute correlation.

According to [BW17], mixing conditions are very common in the literature even though they "are not easy to verify in practice". For the exact assumptions we are going to impose on the mixing coefficients, see Section 3.1.2.

Remark 2.5.3. Let X be a random field and $T(X)$ a function of it. Then the mixing coefficients of $T(X)$ are no bigger than the ones of X as the σ-algebras generated by $T(X)$ are contained in the ones generated by X and therefore the suprema in the definitions of the mixing coefficients are taken over smaller sets.

2.6 Dependent Wild Bootstrap

The dependent wild bootstrap, first introduced in [Sha10], is a generalization of the wild bootstrap introduced in [Wu86]. In the "usual" bootstrap, it is assumed that the samples are independent or at least that the sample size is large enough that the sample represents the underlying population reasonably well. The wild bootstrap was originally developed to deal with heteroscedasticity in the context of regression analysis and may be better suited in case these assumption are not fulfilled: Instead of resampling with replacement, the wild bootstrap does not resample directly, but rather introduces new independent randomness to the residuals via a multiplier field. As the name suggests, in the case of a dependent wild bootstrap, the multiplier field has a non-trivial dependence structure, in contrast to the wild bootstrap. The dependence structure is described by a kernel function.

The following definition is the multivariate generalization of the definition in [Boc33b] chapter 8.

Definition 2.6.1 (Positive-definite function)**.** A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *positive-definite* if it is bounded, continuous, even (in the sense that $f(\mathbf{x}) = f(-\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$) and for any $n \in \mathbb{N}$ and points $\mathbf{x_1}, ..., \mathbf{x_n} \in \mathbb{R}^d$, the matrix

$$(f(\mathbf{x_i} - \mathbf{x_j}))_{i,j\in\{1,\dots,n\}}$$

is positive-definite.

Definition 2.6.2 (Kernel function)**.** We call a function $\omega : \mathbb{R}^d \to \mathbb{R}$ a *kernel function* (or *lag window* in [BD91]) if it is positive-definite, $\omega(\underline{\mathbf{0}}) = 1$ and it satisfies the growth condition

$$\sum_{-\underline{\mathbf{n}}\leq\mathbf{j}\leq\underline{\mathbf{n}}} \left|\omega\left(\frac{1}{q}\mathbf{j}\right)\right| = O(q^d). \tag{2.15}$$

Remark 2.6.3. In [BH15], the stronger Assumption (W)

$$\omega\left(\mathbf{j}\right) = 0 \text{ if } \|\mathbf{j}\|_\infty \geq 1$$

instead of (2.15) is imposed on the kernel functions. In [Lav08] section 3.1, only the Bartlett-kernel is considered (see Example 2.6.7). However, the proofs in [BH15] and [Lav08] remain essentially the same with the weaker assumption (2.15).

The following theorem is a special case of the one in [Loo53] p. 142. The univariate case was originally proven in [Boc33b] Theorem 19. Note that we assume positive-definite functions to take values in $\mathbb{R}$ instead of in $\mathbb{C}$. However, characteristic functions are real-valued iff the associated probability measure is symmetric.

Theorem 2.6.4 (Bochner's theorem)**.** *A function $f : \mathbb{R}^d \to \mathbb{R}$ with $f(\underline{\mathbf{0}}) = 1$ is positive-definite iff it is the characteristic function (Fourier transform) of a symmetric Borel probability measure.*

The following definition of a dependent multiplier field will be used to bootstrap the residuals $X_{\mathbf{i}} - \hat{\mu}(\mathbf{i})$ where $\hat{\mu}$ is an estimator of the mean function, see Section 3.1.4.

Definition 2.6.5 (Dependent multiplier field)**.** Given a random field $X = (X_{\mathbf{i}})_{\mathbf{i}\in I}$, $I \subset \mathbb{Z}^d$, we call a real-valued, centered, Gaussian random field $(V_n(\mathbf{i}))_{\mathbf{i}\in I}$ a *dependent multiplier field with bandwith $q = q_n$ and kernel function ω* if it is independent of the observations X and

$$\operatorname{Cov}\left(V_n(\mathbf{i}), V_n(\mathbf{j})\right) = \omega\left(\frac{1}{q}(\mathbf{i} - \mathbf{j})\right) \tag{2.16}$$

for all $\mathbf{i}, \mathbf{j} \in I$.

Remark 2.6.6. (i) In the definition of the dependent multiplier field, we have $I = \{1, ..., n\}^d$ in mind.

(ii) Later on, we will only be interested in a bandwidth $q_n = o(\sqrt{n})$ with $q_n \to \infty$, see Theorem 3.2.5.

(iii) In both [BH15] Chapter 3.1 and [Lav08] Chapter 4.2 (in the context of testing for long memory in random fields), it is argued that the choice of a smaller bandwith q_n leads to higher rejection rates of the Null Hypothesis. As the simulations in [BW17] suggest that the here considered bootstrap test seems to overreject the Null Hypothesis for finite samples, it seems that a good choice of bandwidth is $q_n = \Omega(n^{\frac{1}{2}-\delta})$ with $\delta \in (0, \frac{1}{2})$ close to 0.

(iv) For bandwidth q_n with $O(n) \subset O(q_n)$, the Growth Condition (2.15) is always satisfied as ω, being positive-definite, is assumed to be bounded, hence

$$\sum_{-\underline{\mathbf{n}}\leq\mathbf{j}\leq\underline{\mathbf{n}}} \left|\omega\left(\frac{1}{q}\mathbf{j}\right)\right| \leq \sum_{-\underline{\mathbf{n}}\leq\mathbf{j}\leq\underline{\mathbf{n}}} C = O(n^d) \subset O(q_n^d).$$

for some constant $C > 0$.

(v) Unlike [BH15] and [BW17], we explicitly demand that a kernel function is positive-definite. This assumption ensures and is required so that such Gaussian dependent multiplier field exists, see [Sch12]. In fact, the kernel function (together with the mean which we assume to be 0) characterizes the distribution of the Gaussian random field according to [Bis06] Section 6.4.1.

(vi) The grf function in the package geoR [RDC$^+$22] of the R programming language implements sampling from such dependent multiplier field for dimensions $d = 1, 2$.

Example 2.6.7 (Kernel functions). (i) Let

$$\mathrm{si} : \mathbb{R} \to \mathbb{R}, x \mapsto \begin{cases} sin(x)/x, & x \neq 0 \\ 0, & x = 0 \end{cases}.$$

Then $x \mapsto \mathrm{si}(x/2)$ is the characteristic function of the uniform distribution on $[-1, 1]$.

(ii) The Bartlett-kernel, defined by

$$\mathbf{x} \mapsto \prod_{k=1}^{d} \mathbb{1}_{|\mathbf{x}_k|<1}(1 - |\mathbf{x}_k|),$$

is one of the most popular choices (e.g., it is one of the lag windows discussed in [BD91] p. 360 and the only one considered in [Lav08]). In the univariate case, it is the characteristic function of $x \mapsto \mathrm{si}^2(x/2)$.

(iii) Indicator functions of the form $\omega = 1_{[-\underline{\mathbf{a}},\underline{\mathbf{a}}]}$ are in general *not* kernel functions: Consider the case $d = 1$, $a = 1$, $n = 3$, $q = \sqrt[4]{n}$. Then one has

$$\left(\omega\left(\frac{1}{q}(i-j)\right)\right)_{i,j=1,\ldots,n} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

This matrix is, however, not positive-definite as it has the eigenvalue $1 - \sqrt{2}$ with corresponding eigenvector $(1, -\sqrt{2}, 1)^t$.

(iv) Let ω be the rescaled density of a multivariate normal distribution with diagonal covariance matrix $\Sigma = \lambda I_d$, $\lambda > 0$, i.e.

$$\omega(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}^t\Sigma^{-1}\mathbf{x}\right).$$

This is the Fourier transform of a multivariate normal distribution. The fact that this function is positive-definite is shown in [Sch38] (without refering to Bochner's theorem). Also $\omega(\underline{\mathbf{0}}) = 1$ and ω is symmetric. According to [BW17], it fulfills the Growth Condition (2.15).

Chapter 3

Change in Distribution Detection in Multivariate Weakly Dependent Random Fields

3.1 Model

In this section, we introduce the model for the multivariate case hinted at in Section 1.2. In order to continue the ideas of Chapter 1, we need to translate the problem of detecting changes in the distribution to a problem in which we only need to detect changes in the mean. Afterwards, we introduce some assumptions with which we will formulate the two hypotheses we will (be successfully able to) test. Lastly, the test procedure is defined.

For technical reasons, the random fields we are examining, depending on the time $n \in \mathbb{N}$, will be assumed to be indexed by the entire grid $\mathbb{Z}^d$ in each step. However, our test statistic will only use information up until time n, i.e., it will be $\sigma(X_{\mathbf{i}} \mid \underline{\mathbf{1}} \leq \mathbf{i} \leq \underline{\mathbf{n}})$-measurable if $X = (X_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ is the random field we are considering at time n.

3.1.1 Setup

For a fixed positive, bounded, integrable weight function $w : \mathbb{R}^p \to \mathbb{R}$, the weighted Hilbert space $L^2(\mathbb{R}^p, w)$ is defined as the space of measurable functions $f : \mathbb{R}^p \to \mathbb{R}$ with $\|f\|_w < \infty$, where the norm $\|\cdot\|_w$ is induced by the inner product

$$\langle f, g\rangle_w \coloneqq \int_{\mathbb{R}^p} f(\mathbf{t})g(\mathbf{t})w(\mathbf{t})d\mathbf{t}. \tag{3.1}$$

Example 4.5(b) in [Rud74] tells us that this indeed defines a Hilbert space and [Bra17] tells us that it is separable.

Let $X = (X_{\mathbf{j}})_{\mathbf{j}\in\mathbb{Z}^d}$ be a random field with observations in $\mathbb{R}^p$. If F is the distribution function of $X_{\mathbf{j}}$, then one has

$$\mathbb{E}\left[\left\langle \mathbb{1}_{X_{\mathbf{j}}\leq \cdot}, h\right\rangle\right] = \mathbb{E}\left[\int\limits_{\mathbb{R}^p} \mathbb{1}_{X_{\mathbf{j}}\leq \mathbf{t}} h(\mathbf{t}) w(\mathbf{t}) d\mathbf{t}\right] = \int\limits_{\mathbb{R}^p} F(\mathbf{t}) h(\mathbf{t}) w(\mathbf{t}) d\mathbf{t} = \langle F, h\rangle$$

by Fubini's theorem. Therefore, we have

$$F = \mathbb{E}_{L^2(\mathbb{R}^p,w)}\left[\mathbb{1}_{X_{\mathbf{j}}\leq \cdot}\right]$$

by Riesz-Fréchet 2.3.1. This lets us test for changes in the distribution of the $\mathbb{R}^p$-valued random field X by considering changes in the mean of the $L^2(\mathbb{R}^p, w)$-valued random field $\left(\mathbb{1}_{X_{\mathbf{j}}\leq \cdot}\right)_{\mathbf{j}\in\mathbb{Z}^d}$. Therefore, we will assume that the random field X (or ${}_nX$ later on) takes values in some separable Hilbert space H and we will only consider changes in the mean, making it possible to continue the ideas from Chapter 1.

Remark 3.1.1. a) We consider the weighted Hilbert space $L^2(\mathbb{R}^p, w)$ instead of $L^2(\mathbb{R}^p)$ so that the norm of bounded functions (and in particular of empirical distribution functions) is finite.

b) [BW17] only requires the weight function w to be nonnegative. However, if w vanishes on some interval I, then 3.1 does not define an inner product as one has $\langle f, f\rangle_w = 0$ for all functions f with support in I. Therefore, it is necessary that w only vanishes on a null set.

c) While in theory the test that will be described in the following works for every such weight function, the power for finite samples is greatly improved when the weight function is large at the points where the two distribution functions that arise under the alternative differ. One possible choice might be the densities of heavy-tailed distributions, e.g., of a multivariate Cauchy distribution, as these choices are less reliant on one already knowing where the two marginal distributions of the random field possibly differ, see Example 3.1.2.
One may also choose piecewise constant weight functions as these greatly simplifiy the calculations of 3.1 for empirical distribution functions since one then only has to calculate the integral of step functions.

Example 3.1.2. We illustrate how much of a difference choosing different weight functions w may make: Assume we want to detect changes in the distribution in the univariate case $d = p = 1$ where the underlying random field has the two marginal cumulative distribution functions be given by

$$F(t) = \mathbb{1}_{\{t\geq 4\}},$$
$$G(t) = \mathbb{1}_{\{t\geq 6\}}.$$

For the weight function $w_0 \sim \mathcal{N}(0,\, 1)$, we have

$$\|F - G\|_{w_0}^2 = \int_4^6 w_0(t)dt \approx 0.00003,$$

while the weight function $w_5 \sim \mathcal{N}(5,\, 1)$ gives

$$\|F - G\|_{w_5}^2 \approx 0.683.$$

The "correct" choice of the location parameter amounts to an increased performance by a factor of about $21,556$ in this case. Choosing a Cauchy distribution instead, one has

$$\|F - G\|_{w_0'}^2 \approx 0.025$$

for $w_0' \sim \text{Cauchy}(0, 1)$ and

$$\|F - G\|_{w_5'}^2 = 0.5$$

for $w_5' \sim \text{Cauchy}(5, 1)$, which gives a factor of about 20. This shows how much less sensitive the Cauchy distribution is.

3.1.2 Assumptions

In the following, we will be stating some assumptions which will be needed in order to precisely define the two hypotheses we are interested in testing. Note that we will indicate which assumptions will be used exactly in the respective results. Let $X = (X_{\mathbf{j}})_{\mathbf{j}\in\mathbb{Z}^d}$ denote an H-valued random field and $\delta > 0$ a fixed, positive constant.

Assumption A. X is strictly stationary.

Assumption A'. X is weakly stationary.

Assumption B. X is ρ-mixing.

Assumption B'. $\lim\limits_{r\to\infty} \rho(r) < 1$.

Assumption C. $\mathbb{E}\left[\left\|X_{\underline{\mathbf{1}}}\right\|^{4+2\delta}\right] < \infty$.

Assumption C'. $\mathbb{E}\left[\left\|X_{\underline{\mathbf{1}}}\right\|^{2+\delta}\right] < \infty$.

Assumption C". $\sup\limits_{\mathbf{i}} \mathbb{E}\left[\|X_{\mathbf{i}}\|^{2+\delta}\right] < \infty$.

Assumption D. $\sum\limits_{m\geq 1} m^{d-1}\alpha_{2,2}(m)^{\frac{\delta}{2+\delta}} < \infty$.

Assumption D'. $\sum\limits_{m\geq 1} m^{d-1}\alpha_{1,1}(m)^{\frac{\delta}{2+\delta}} < \infty$.

Remark 3.1.3.

- Assumption A' is needed in order to define the covariogram.
- Assumption B' regarding the ρ-mixing coefficient is used to show the Rosenthal type inequality in Lemma 4.2.1. The ρ-mixing Assumption B is used in Lemma 4.2.6 to show the asymptotic independence of increments.
- The Moment Assumption C' replaces Assumption C" in case X is strictly stationary, i.e., if Assumption A holds, and it is needed in order to show uniform integrability, see Lemma 4.2.6.
- Assumption D' is used to ensure the existence of the long-run variance, see Lemma 4.2.6, and Assumption D ensures the convergence of the implicit long-run variance estimators, see Lemma 4.2.14.

3.1.3 Hypotheses

Having stated the above assumptions, we are now able to define the exact two hypotheses we are able to test:

For every $n \in \mathbb{N}$, let ${}_nX = ({}_nX_{\mathbf{i}})_{\mathbf{i}\in\mathbb{Z}^d}$ be an H-valued random field.

Hypothesis H_0. The random field ${}_nX$ does not depend on n, i.e., ${}_nX = X$ for all n. The random field X satisfies the Assumptions A, B, C and D.

Hypothesis H_A. There are change points ${}_n\mathbf{l} < {}_n\mathbf{u} \in \{0, ..., n\}^d$ and a shift $\Delta \in H \setminus \{0\}$ such that the H-valued random fields $\tilde{X} = {}_n\tilde{X}$, $n \in \mathbb{N}$, defined by

$${}_n\tilde{X}_{\mathbf{j}} := \begin{cases} {}_nX_{\mathbf{j}} - \Delta, & \mathbf{j} \in ({}_n\mathbf{l}, {}_n\mathbf{u}] \\ {}_nX_{\mathbf{j}}, & \mathbf{j} \notin ({}_n\mathbf{l}, {}_n\mathbf{u}] \end{cases} \quad \forall \mathbf{j} \in \mathbb{Z}^d$$

satisfy the assumptions of the Null Hypothesis H_0. Additionally, the change set size

$$c(n) := [{}_n\mathbf{u} - {}_n\mathbf{l}]$$

is asymptotically proportional to the number of observations in the sense that

$$\lim_{n\to\infty} \frac{c(n)}{n^d} = \gamma \tag{3.2}$$

for a constant $0 < \gamma < 1$.

Remark 3.1.4. The translation of the change in the distribution problem to a change in the mean problem is vital in order to define the model as it is. It is currently not known whether or not the test has power if we do not assume stationarity of the random field $\tilde{X}$. Without the translation of the problem, it is not clear how one would exploit stationarity under the alternative.

Remark 3.1.5. a) The change set size restriction (3.2) may be relaxed, only requiring

$$0 < \alpha \leq \frac{c(n)}{n^d} \leq 1 - \beta < 1$$

for sufficiently large n and some constants α, β. However, under this weaker requirement, $n^{-d/2} T_n$ does in general not converge under the Alternative H_A, cf. Remark 3.2.8.

b) The model may also be changed so that the shift Δ depends on the time n, as long as it satisfies $\|\Delta_n\|_w n^{d/2} \to \infty$.

3.1.4 Statistical Test

Analogously to the univariate case in (1.2), we define the test statistic

$$T_n \coloneqq T_n({}_nX) \coloneqq \frac{1}{n^{d/2}} \max_{\underline{\mathbf{0}} \leq \mathbf{k} \leq \mathbf{m} \leq \underline{\mathbf{n}}} [\mathbf{m} - \mathbf{k}] \left\| \frac{1}{[\mathbf{m} - \mathbf{k}]} \sum_{\mathbf{k} < \mathbf{j} \leq \mathbf{m}} {}_nX_{\mathbf{j}} - \frac{1}{n^d} \sum_{\underline{\mathbf{1}} \leq \mathbf{j} \leq \underline{\mathbf{n}}} {}_nX_{\mathbf{j}} \right\| \tag{3.3}$$

In the univariate case, the asymptotic distribution of T_n depends on the variance σ^2 of X_1. However, as we no longer assume that the observations are independent, the variance is replaced with the long-run variance. We will avoid estimating this long-run variance directly and rather employ the dependent wild bootstrap.

To this end, let ${}_n\hat{\mathbf{l}}$ and ${}_n\hat{\mathbf{u}}$ be those points maximizing the expression in (3.3). Define the change set size estimator

$$\hat{c}(n) \coloneqq [{}_n\hat{\mathbf{u}} - {}_n\hat{\mathbf{l}}] \tag{3.4}$$

and the mean estimator

$$\hat{\mu}(\mathbf{i}) \coloneqq \begin{cases} \frac{1}{\hat{c}(n)} \sum\limits_{\mathbf{j} \in ({}_n\hat{\mathbf{l}}, {}_n\hat{\mathbf{u}}]} {}_nX_{\mathbf{j}}, & \mathbf{i} \in ({}_n\hat{\mathbf{l}}, {}_n\hat{\mathbf{u}}] \\ \frac{1}{n^d - \hat{c}(n)} \sum\limits_{\substack{\mathbf{j} \in \{1,\ldots,n\}^d \\ \mathbf{j} \notin ({}_n\hat{\mathbf{l}}, {}_n\hat{\mathbf{u}}]}} {}_nX_{\mathbf{j}}, & \mathbf{i} \notin ({}_n\hat{\mathbf{l}}, {}_n\hat{\mathbf{u}}] \end{cases} . \tag{3.5}$$

For a number K of bootstraps, let

$$(V_{n,j}(\mathbf{i}))_{\mathbf{i} \in \mathbb{Z}^d}, \; 1 \leq j \leq K,$$

be dependent multiplier fields. Define the bootstrapped versions of ${}_nX$

$${}_nX_j^\star \coloneqq \left({}_nX_j^\star(\mathbf{i})\right)_{\mathbf{i}} \coloneqq \left(V_{n,j}(\mathbf{i})({}_nX_{\mathbf{i}} - \hat{\mu}(\mathbf{i}))\right)_{\mathbf{i} \in \{1,\ldots,n\}^d},$$

the bootstrapped partial sum fields $S_{n,j}^\star$ by

$$S_{n,j}^\star(\mathbf{t}) = \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} V_{n,j}(\mathbf{i})({}_nX_{\mathbf{i}} - \hat{\mu}(\mathbf{i})), \tag{3.6}$$

and the bootstrapped test statistics

$$T^{\star}_{n,j} \coloneqq T_n({}_nX^{\star}_j).$$

By (2.1), we may write them as

$$T^{\star}_{n,j} = \max_{\underline{\mathbf{0}}\leq\mathbf{k}<\mathbf{m}\leq\underline{\mathbf{n}}} [\mathbf{m}-\mathbf{k}] \left\| \frac{1}{[\mathbf{m}-\mathbf{k}]} S^{\star}_{n,j}\left(\frac{\mathbf{k}}{n}, \frac{\mathbf{m}}{n}\right] - \frac{1}{n^d} S^{\star}_{n,j}(\underline{\mathbf{1}}) \right\|.$$

Let $\hat{F}_{n,K}$ be the empirical c.d.f. computed from $T^{\star}_{n,1}, ..., T^{\star}_{n,K}$ and define the sample quantile

$$q^{\star}_{n,K} \coloneqq \hat{F}^{-1}_{n,K} : (0,1) \to \mathbb{R}, y \mapsto \inf\{x \in \mathbb{R} \mid \hat{F}_{n,K}(x) \geq y\}. \tag{3.7}$$

For a given significance level $\alpha \in (0,1)$, we then reject the Null Hypothesis H_0 if $T_n \geq q^{\star}_{n,K}(1-\alpha)$.

Remark 3.1.6. The change set estimator

$$({}_n\hat{\mathbf{l}}, {}_n\hat{\mathbf{u}}]$$

need not necessarily be derived from the definition of the test statistic T_n. For some discussion see [BH15] Section 3 and 4.

3.2 Main Results

In this section, we state the theorems formulated in [BW17] and their corollaries regarding the convergence of the test statistic and the resulting performance of our test procedure. We will prove the corollaries immediately. The proofs of the two theorems will be prepared and carried out in the next chapter.

Theorem 3.2.1 (FCLT). *Let $X = (X_j)_{j\in\mathbb{Z}^d}$ be an H-valued random field satisfying the Assumptions A, B, C' and D'. Denote $\mathbb{E}\left[X_{\underline{\mathbf{1}}}\right] = \mu$. Then we have the functional central limit theorem*

$$S_n \coloneqq \left(\frac{1}{n^{\frac{d}{2}}} \sum_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} (X_{\mathbf{i}} - \mu) \right)_{\mathbf{t}\in[0,1]^d} \Rightarrow W$$

in $D_H[0,1]^d$, where $W = (W(\mathbf{t}))_{t\in[0,1]^d}$ is a Brownian sheet in H with covariance operator Γ, defined by

$$\langle \Gamma x, y\rangle = \sum_{\mathbf{v}\in\mathbb{Z}^d} \mathbb{E}\left[\langle X_{\underline{\mathbf{0}}} - \mu, x\rangle \langle X_{\mathbf{v}} - \mu, y\rangle\right],\ x, y \in H. \tag{3.8}$$

Furthermore, the series in (3.8) *converges absolutely.*

Corollary 3.2.2 (Asymptotic test statistic)**.** *Under the Null Hypothesis H_0, the test statistic T_n converges weakly in $\mathbb{R}$ to the statistic*

$$T := \sup_{\underline{\mathbf{0}} \leq \mathbf{s} \leq \mathbf{t} \leq \underline{\mathbf{1}}} \| W(\mathbf{s}, \mathbf{t}] - [\mathbf{t} - \mathbf{s}] W(\underline{\mathbf{1}}) \|$$

where W is the Brownian sheet defined in Theorem 3.2.1.

Proof. The Null Hypothesis H_0 implies the assumptions of Theorem 3.2.1. Without loss of generality, we may assume that the random field X is centered. By Equation (2.1), we have

$$\begin{aligned} T_n &= \frac{1}{n^{d/2}} \max_{\underline{\mathbf{0}} \leq \mathbf{k} < \mathbf{m} \leq \underline{\mathbf{n}}} \left\| \sum_{\mathbf{k} < \mathbf{i} \leq \mathbf{m}} {}_nX_{\mathbf{i}} - \frac{[\mathbf{m} - \mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \underline{\mathbf{n}}} {}_nX_{\mathbf{i}} \right\| \\ &= \max_{\underline{\mathbf{0}} \leq \mathbf{k} < \mathbf{m} \leq \underline{\mathbf{n}}} \left\| S_n(\mathbf{k}/n, \mathbf{m}/n] - \frac{[\mathbf{m} - \mathbf{k}]}{n^d} S_n(0, 1] \right\| \\ &= \sup_{\underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}}} \left\| S_n(\lfloor \mathbf{s}n \rfloor / n, \lfloor \mathbf{t}n \rfloor / n] - \frac{\lfloor \mathbf{t}n \rfloor - \lfloor \mathbf{s}n \rfloor}{n^d} S_n(1) \right\| \end{aligned}$$

To formally verify the weak convergence

$$S_n(\lfloor \mathbf{s}n \rfloor / n, \lfloor \mathbf{t}n \rfloor / n] \Rightarrow W(\mathbf{s}, \mathbf{t}] \tag{3.9}$$

in H, let $f : H \to \mathbb{R}$ be any continuous, bounded function. Note that as blocks are characterized by two points (their lower and upper corner), the block-valued random elements $S_n(\cdot)$ and $W(\cdot)$ can be seen as random fields indexed by the subset

$$I := \{ \mathbf{i} \in [0,1]^{2d} \mid \mathbf{i}_j \leq \mathbf{i}_{j+1} \text{ if } i \text{ is odd} \}$$

of $[0,1]^{2d}$ in which the odd indices represent the lower corner and the even ones represent the upper corner. Define the (deterministic) functions

$$\begin{aligned} g_n &: I \to \mathbb{R}, x \mapsto \mathbb{E}\left[f(S_n(x)) \right], \\ g &: I \to \mathbb{R}, x \mapsto \mathbb{E}\left[f(W(x)) \right]. \end{aligned}$$

As W has almost surely continuous sample paths (both the original Brownian sheet by definition and the incremental version as a finite linear combination of the original one), g is continuous. Furthermore, due to the weak convergence of S_n to W, g_n converges uniformly to g. Identifying blocks with elements of the index-set I, the sequence of blocks

$$a_n := (\lfloor \mathbf{s}n \rfloor / n, \lfloor \mathbf{t}n \rfloor / n]$$

converges to the block $a := (\mathbf{s}, \mathbf{t}]$. Combining these facts, we get the convergence

$$\lim_{n\to\infty} g_n(a_n) = g(a).$$

As f was an arbitrary continuous and bounded function, this shows the weak convergence (3.9).

The weak convergence

$$\frac{\lfloor \mathbf{t}n\rfloor - \lfloor \mathbf{s}n\rfloor}{n^d} S_n(1) \Rightarrow [\mathbf{t}-\mathbf{s}]W(\underline{\mathbf{1}})$$

is a simple application of Slutsky's theorem together with Theorem 3.2.1. The claim finally follows from the continuous mapping theorem. □

The following result applies Corollary 3.2.2 to the procedure introduced in Section 3.1.1.

Corollary 3.2.3. *Let X take values in $\mathbb{R}^p$. Under the Null Hypothesis H_0, the test statistic*

$$T_n^w := \frac{1}{n^{d/2}} \max_{\underline{\mathbf{0}}\leq\mathbf{k}\leq\mathbf{m}\leq\underline{\mathbf{n}}} \sqrt{\int_{\mathbb{R}^p} \left(\sum_{\mathbf{k}<\mathbf{i}\leq\mathbf{m}} \mathbb{1}_{\{nX_\mathbf{i}\leq\mathbf{x}\}} - \frac{[\mathbf{m}-\mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}} \mathbb{1}_{\{nX_\mathbf{i}\leq\mathbf{x}\}} \right)^2 w(\mathbf{x})d\mathbf{x}} \tag{3.10}$$

converges weakly in $\mathbb{R}$ to

$$T^w := \sup_{\underline{\mathbf{0}}\leq\mathbf{s}\leq\mathbf{t}\leq\underline{\mathbf{1}}} \|W(\mathbf{s},\mathbf{t}] - [\mathbf{t}-\mathbf{s}]W(\underline{\mathbf{1}})\|$$

where W is a Brownian sheet in $L^2(\mathbb{R}^p, w)$ with covariance operator Γ defined by

$$\langle \Gamma x, y\rangle = \sum_{\mathbf{k}\in\mathbb{Z}^d} \mathbb{E}\left[\int_{\mathbb{R}^p} \left(\mathbb{1}_{\{X_{\underline{\mathbf{0}}}\leq\mathbf{t}\}} - F(\mathbf{t})\right) x(\mathbf{t})w(\mathbf{t})d\mathbf{t} \int_{\mathbb{R}^p} \left(\mathbb{1}_{\{X_\mathbf{k}\leq\mathbf{t}\}} - F(\mathbf{t})\right) y(\mathbf{t})w(\mathbf{t})d\mathbf{t} \right]$$

for $x, y \in L^2(\mathbb{R}^p, w)$.

Proof. We need to show that the $L^2(\mathbb{R}^p, w)$-valued random field

$$\mathbb{1}_{\{X\leq\cdot\}} = (\mathbb{1}_{\{X_\mathbf{i}\leq\cdot\}})_\mathbf{i}$$

satisfies the assumptions of the Null Hypothesis H_0. Assumptions A, B, D' and D regarding stationarity, the ρ-mixing condition and the summability of the α-mixing coefficients are fulfilled as $\mathbb{1}_{\{X\leq\cdot\}}$ is a function of X and X satisfies these assumptions. The existence of moments condition C is satisfied even if it is not satisfied by X because

$$\mathbb{E}\left[\|\mathbb{1}_{\{X_{\mathbf{i}}\leq\cdot\}}\|^k\right] = \mathbb{E}\left[\left(\int\limits_{\mathbb{R}^p} {\mathbb{1}_{\{X_{\mathbf{i}}\leq\mathbf{t}\}}}^2 w(\mathbf{t})d\mathbf{t}\right)^{k/2}\right] \leq \left(\int\limits_{\mathbb{R}^p} w(\mathbf{t})d\mathbf{t}\right)^{k/2} < \infty$$

for any moment k due to the integrability assumption of the positive weight function w. $\square$

Remark 3.2.4. The test statistic T_n^w in Corollary 3.2.3 resembles both the Cramér-von Mises test statistic as an integral over a squared difference as well as the two-sided Kolgmorov-Smirnov test as the maximum of the distance between two empirical distribution functions. See Remark 1.1.5 and [Vaa98] Chapter 19.3.

The following is [BW17] Theorem 2.

Theorem 3.2.5 (Bootstrap FCLT). *Let Assumptions A, B, C and D hold. Let*

$$(V_{n,1}(\mathbf{i}))_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}, \ldots, (V_{n,K}(\mathbf{i}))_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}},\ K \in \mathbb{N},$$

be independent copies of the same dependent multiplier field. Assume that the bandwidth $q = q_n$ fulfills $q_n \to \infty$ and $q_n = o(\sqrt{n})$. Then the bootstrapped partial sum fields $S^*_{n,1}, \ldots, S^*_{n,K}$ satisfy

$$(S_n, S^*_{n,1}, \ldots S^*_{n,K}) \Rightarrow (W, W^*_1, \ldots, W^*_K)$$

in $\left(D_H[0,1]^d\right)^{K+1}$ where S_n is the partial sum field of X and $W^*_1, \ldots, W^*_K$ are independent copies of the Hilbert space valued Brownian sheet W from Theorem 3.2.1.

The following two corollaries capture the asymptotic properties of the test procedure.

Corollary 3.2.6 (Type I error rate). *For any given significance level $0 < \alpha < 1$, the type I error rate of the test approaches α in the sense that*

$$\lim_{K\to\infty}\lim_{n\to\infty}\mathbb{P}(T_n \geq q^\star_{n,K}(1-\alpha)) = \alpha$$

under the Null Hypothesis H_0.

Proof. Using Theorem 3.2.5, this will follow immediately from Proposition 4.2.15. $\square$

Corollary 3.2.7 (Statistical power). *For any number of bootstraps K, the power of the test defined in Section 3.1.4 approaches 1 as n approaches infinity.*

Proof. Assume the Alternative H_A. One can bound the test statistic from below via

$$
\begin{aligned}
T_n &= \frac{1}{n^{d/2}} \max_{\underline{\mathbf{0}} \le \mathbf{k} < \mathbf{m} \le \underline{\mathbf{n}}} \left\| \sum_{\mathbf{k} < \mathbf{i} \le \mathbf{m}} {}_nX_{\mathbf{i}} - \frac{[\mathbf{m} - \mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}} \le \mathbf{i} \le \underline{\mathbf{n}}} {}_nX_{\mathbf{i}} \right\| \\
&\ge \frac{1}{n^{d/2}} \left\| \sum_{{}_n\mathbf{l} < \mathbf{i} \le {}_n\mathbf{u}} {}_nX_{\mathbf{i}} - \frac{[{}_n\mathbf{u} - {}_n\mathbf{l}]}{n^d} \sum_{\underline{\mathbf{1}} \le \mathbf{i} \le \underline{\mathbf{n}}} {}_nX_{\mathbf{i}} \right\|.
\end{aligned}
$$

By rewriting ${}_nX$ in terms of the stationary version $\tilde{X}$ and the shift Δ and using the reverse triangle inequality, we see that this is bounded from below by

$$
\frac{1}{n^{\frac{d}{2}}} \left(1 - \frac{c(n)}{n^d}\right) c(n) \|\Delta\|_w - \frac{1}{n^{\frac{d}{2}}} \left\| \sum_{\mathbf{k} < \mathbf{i} \le \mathbf{m}} \tilde{X}_{\mathbf{i}} - \frac{[\mathbf{m} - \mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}} \le \mathbf{i} \le \underline{\mathbf{n}}} \tilde{X}_{\mathbf{i}} \right\| =: Y_n - U_n.
$$

Applying Corollary 3.2.2, $(U_n)_n$ is a sequence that is bounded from above by the weakly converging sequence $\left(T_n\left(\tilde{X}\right)\right)_n$. Rewriting the left term as

$$
Y_n = \frac{c(n)}{n^d} \left(1 - \frac{c(n)}{n^d}\right) \|\Delta\|_w n^{d/2}
$$

and applying the change set size convergence assumption (3.2), this behaves asymptotically like

$$
\gamma(1-\gamma)\|\Delta\|_w n^{d/2} \tag{3.11}
$$

which diverges to infinity. Now let $0 < \alpha < 1$ be an arbitrary significance level. Using Lemma 1.1.3, this shows

$$
\lim_{n\to\infty} \mathbb{P}(T_n \ge g^\star_{n,K}(1-\alpha)) \ge \lim_{n\to\infty} \mathbb{P}(Y_n \ge g^\star_{n,K}(1-\alpha) + U_n) = 1.
$$

□

Remark 3.2.8. Using (3.11), it is possible to see that $n^{-d/2}T_n$ has the asymptotic lower bound

$$
\gamma(1-\gamma)\|\Delta\|_w
$$

under the Alternative H_A. This suggests that the test has the greatest power if $\gamma = 0.5$, i.e., if the size of the change set is half the number of observations, as this value maximizes the function $\gamma \mapsto \gamma(1-\gamma)$.

3.3 Discussion

3.3.1 Model Limitations

While the test procedure investigated here exactly satisfies the asymptotic properties one wishes for, namely being able to detect any changes no matter how small they are while holding the significance level, in reality, the test will only be applied to samples of finite size. Therefore, in the following, we will outline some challenges one may face.

In [Cre93] Section 2.3 it is written: "[...] it is not typical to find data locations $\{s_1, ..., s_n\}$ on a regular grid nor is n often very large. Thus, most modeling of geostatistical data occurs in the spatial domain rather than the frequency domain." This hints at some problems one might face when applying the model introduced in [BW17] to real-world data:

We assume that our observations are indexed by the set $\{1, ..., n\}^d$, i.e., we need the same number of observations in each dimension. When modeling purely spatial data, this may already not apply. When considering spatial-temporal data, i.e., the dimension 1 represents time and the dimensions $2, ..., d$ represent space, this assumptions seems unnatural. However, this problem may be fixed using our approach as one can, for a given point in space, consider all observations in time to represent a single point in a function (Hilbert) space. This approach has been applied in [GK12] where temperature and precipitation data from multiple weather stations was modeled as a Hilbert space valued random field.

Even if one has the same number of observations in each dimension, the distance between two adjacent observations may not be constant. This problem is, however, not exclusive to our model, but it may rather appear whenever one deals with discrete time series or random fields.

The Alternative Hypothesis H_A assumes epidemic changes, i.e., that there are exactly two distributions which the observations may follow with a sudden leap between the two. In reality, this does often not seem to be the case: Climate change, death counts related to the COVID-19 pandemic and radioactivity related to nuclear warhead tests are some examples of changes which are (at least partially) not abrupt or constant. Hence it may be sensible to add a third region in which the observations follow neither the distribution F nor $G = F + \Delta$ but rather, say, $F + \alpha(\mathbf{i})\Delta$ with $\alpha(\mathbf{i}) \in (0, 1)$. These three regions can then, again under the assumption of rectangular change sets, be characterized by four (possibly coinciding) points.

The assumptions that the change sets form rectangles is another assumption which may rarely apply in reality. Blocks are one way to generalize one-dimensional intervals. Yet one may also see one-dimensional intervals as exactly the connected subsets of $\mathbb{R}$. However, detecting connected change sets in dimension $d > 1$ is difficult as these can no longer be parametrized by a finite number of change points.

3.3.2 Hyper Parameters

While the asymptotic performance of the test does not depend on any of the chosen parameters, its performance under finite sample sizes does greatly. The parameters include:

a) The significance level α. How to choose this level depends on how sensitive one wants the test to be.

b) The weight function w. We refer to Remark 3.1.1.

c) The kernel function ω (and thereby the dependent multiplier fields $V_{n,j}$) and the bandwidth q_n. See Remark 2.6.6. As these paramters are used to implicitly estimate the long-run variance Γ, one may consult literature dealing with long-run variance estimation on how to choose these optimally, see, e.g., [Mül07].

d) The number of bootstraps K. While less important under the Alternative H_A given a large enough sample size, a higher number may increase the odds of holding the significance level under the Null Hypothesis H_0.

We also refer to the simulation study presented in [BW17].

3.3.3 Online Test

If one is interested in a so-called "online test" where a stream of data is analyzed, the Alternative H_A may be altered such that one assumes that the upper index ${}_n\mathbf{u}$ of the change point set is $\underline{\mathbf{n}}$. In that case, one replaces the test statistic in 3.3 with the following one:

$$T_n^o := \frac{1}{n^{d/2}} \max_{\underline{\mathbf{0}} \leq \mathbf{k} \leq \underline{\mathbf{n}}} \left\| \sum_{\mathbf{k} < \mathbf{j} \leq \underline{\mathbf{n}}} {}_nX_{\mathbf{j}} - \frac{[\underline{\mathbf{n}} - \mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}} \leq \mathbf{j} \leq \underline{\mathbf{n}}} {}_nX_{\mathbf{j}} \right\|$$

This online test statistic converges weakly to the following distribution

$$T^o := \sup_{\underline{\mathbf{0}} \leq \mathbf{s} \leq \underline{\mathbf{1}}} \| W(\mathbf{s}, \underline{\mathbf{1}}] - [\underline{\mathbf{1}} - \mathbf{s}] W(\underline{\mathbf{1}}) \|$$

where W is the Brownian sheet defined in Theorem 3.2.1. In the univariate case $d = 1$ and $H = \mathbb{R}$, this is exactly the Kolmogorov distribution, provided that the long-run variance of X is 1. See also Remark 1.1.5.

As a closed form of the Kolmogorov distribution is known, one may therefore calculate critical values directly instead of using the dependent wild bootstrap approach introduced here, which may be very important if it is not feasable to bootstrap the test statistic, e.g., when observing some instrument in a machine in real-time.

One must, however, still know or approximate the long-run variance. Under a potential change set, a long-run variance estimator has been introduced in [BH15]. It seems plausible that the long-run variance may be estimated from previous runs of the instrument for which one is sure that there was no change in the mean. In that case, any estimator of the long-run variance for random fields/stochastic processes may be used, see, e.g., [Mül07].

Chapter 4

Proofs

In this chapter, we first prove the results in [Deo75] regarding strongly separated blocks. The rest almost exclusively mirrors and expands the results and proofs presented in [BW17] Section 4: We prove all preliminary results and afterwards the two main theorems.

4.1 Strongly Separated Blocks

In this section, we aim to verify that the proofs in [Deo75] still work with our adjusted definition of strongly separated blocks, see Remark 2.2.7.

The following is a slightly differently worded version of [Deo75] Lemma 2.

Lemma 4.1.1. *Let* $X = (X(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$ *be an* $\mathbb{R}$*-valued random field with sample paths in* $D_{\mathbb{R}}[0,1]^d$ *such that*

(i) $\mathbb{E}[X(\mathbf{t})] = 0$*,* $\mathrm{Var}(X(\mathbf{t})) = [\mathbf{t}]$ *for* $\mathbf{t} \in [0,1]^d$*,*

(ii) X *has continuous sample paths almost surely,*

(iii) *the increments of* X *around any collection of strongly separated blocks are independent random variables.*

then X *is a d-parameter Brownian sheet in* $\mathbb{R}$ *with covariance operator* $1 \in \mathbb{R}$.

In [Deo75], the proof only mentions that "it suffices to prove that $X_{\mathbf{t}}$ is normally distributed for each $\mathbf{t}$ and that this can be easily accomplished by induction on d in conjunction with Theorem 19.1. of [Bil68]". We carry out this proof:

Proof. We prove the statement by induction over d. The base case is the aforementioned [Bil68] Theorem 19.1. For the induction hypothesis, assume $d \geq 2$ and that $X'(\mathbf{t})$ is normally distributed for any $(d-1)$-parameter random field X' that satisfies the assumptions of this lemma. Fix $\mathbf{t} \in [0,1]^d$. Without loss of generality, we may assume $\mathbf{t}_d > 0$ since for any $\mathbf{t}$ with $\mathbf{t}_d = 0$, $X(\mathbf{t})$ has no variance by assumption and is therefore trivially normally distributed. We define the $(d-1)$-parameter random field

$$X' := \left(\frac{1}{\sqrt{\mathbf{t}_d}} X \begin{pmatrix} \mathbf{s} \\ \mathbf{t}_d \end{pmatrix}\right)_{\mathbf{s}\in[0,1]^{d-1}}.$$

As X satisfies these assumptions, we also have $\mathbb{E}\left[X'_{\mathbf{s}}\right] = 0$ for all $\mathbf{s}$ and X' has continuous sample paths almost surely. Furthermore,

$$\mathrm{Var}(X'(\mathbf{s})) = \mathrm{Var}\left(\frac{1}{\sqrt{\mathbf{t}_d}} X \begin{pmatrix} \mathbf{s} \\ \mathbf{t}_d \end{pmatrix}\right) = \frac{1}{\mathbf{t}_d}\left[\begin{pmatrix} \mathbf{s} \\ \mathbf{t}_d \end{pmatrix}\right] = \mathbf{s}.$$

The increments of X' around a block $B' = (\mathbf{a}, \mathbf{b}]$ in $[0,1]^{d-1}$ can be calculated as follows:

$$\begin{aligned}
X'(B') &= \sum_{\boldsymbol{\epsilon}\{0,1\}^{d-1}} (-1)^{d-1-\sum\limits_{i=1}^{d-1}\boldsymbol{\epsilon}_i} X'\left(\mathbf{a} + \boldsymbol{\epsilon}(\mathbf{b}-\mathbf{a})\right) \\
&= \frac{1}{\sqrt{\mathbf{t}_d}} \sum_{\boldsymbol{\epsilon}\{0,1\}^{d-1}} (-1)^{d-1-\sum\limits_{i=1}^{d-1}\boldsymbol{\epsilon}_i} X\left(\begin{pmatrix} \mathbf{a}_1 \\ \cdots \\ \mathbf{a}_{d-1} \\ 0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon} \\ 1 \end{pmatrix}\left(\begin{pmatrix} \mathbf{b}_1 \\ \cdots \\ \mathbf{b}_{d-1} \\ \mathbf{t}_d \end{pmatrix} - \begin{pmatrix} \mathbf{a}_1 \\ \cdots \\ \mathbf{a}_{d-1} \\ 0 \end{pmatrix}\right)\right) \\
&= \frac{1}{\sqrt{\mathbf{t}_d}} \sum_{\substack{\boldsymbol{\epsilon}\{0,1\}^{d} \\ \boldsymbol{\epsilon}_d=1}} (-1)^{d-\sum\limits_{i=1}^{d}\boldsymbol{\epsilon}_i} X\left(\begin{pmatrix} \mathbf{a}_1 \\ \cdots \\ \mathbf{a}_{d-1} \\ 0 \end{pmatrix} + \boldsymbol{\epsilon}\left(\begin{pmatrix} \mathbf{b}_1 \\ \cdots \\ \mathbf{b}_{d-1} \\ \mathbf{t}_d \end{pmatrix} - \begin{pmatrix} \mathbf{a}_1 \\ \cdots \\ \mathbf{a}_{d-1} \\ 0 \end{pmatrix}\right)\right).
\end{aligned}$$

This is just the increment of $\frac{1}{\mathbf{t}_d}X$ around the block in $[0,1]^d$ that we get by attaching a 0 to the bottom of $\mathbf{a}$ and $\mathbf{t}_d$ to the bottom of $\mathbf{b}$ because the part of the increment where $\boldsymbol{\epsilon}_d$ is 0 vanishes since, due to Assumption (i), $X(\mathbf{s})$ vanishes almost surely if one of the components of $\mathbf{s}$ is 0.

Now let $B'_1, ..., B'_k$ be a collection of strongly separated blocks in $[0,1]^{d-1}$. The blocks $B_1, ..., B_k$ that we get by attaching 0 and $\mathbf{t}_d$ respectively, are also strongly separated. We may therefore conclude that the increments of X around the blocks $B'_1, ..., B'_k$ are independent since they are equal to a constant multiple of the increments of X around the blocks $B_1, ..., B_k$ and they are independent by assumption. Hence $X'(\mathbf{s})$ is normally distributed by the induction hypothesis and so is

$$X(\mathbf{t}) = \sqrt{\mathbf{t}_d} X'(\mathbf{s}),$$

choosing $\mathbf{s} = (\mathbf{t}_1, ..., \mathbf{t}_{d-1})$. $\square$

Remark 4.1.2. The above proof shows why we demand that strongly separated blocks only need to not have coinciding endpoints in one of the dimensions (and therefore have positive distance in this dimension): Attaching 0 and $\mathbf{t}_d$ to the bottom of each of the blocks $B'_1, ..., B'_k$ would make them not strongly separated if we demanded that the blocks have positive distance in every dimension.

The following is [Deo75] Lemma 3.

Lemma 4.1.3. *Let $X_n = (X_n(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$, $n \in \mathbb{N}$, be a sequence of d-parameter real-valued random fields with sample paths in $D_{\mathbb{R}}[0,1]^d$ such that*

(i) $\mathbb{E}[X_n(\mathbf{t})] \to 0$, $\mathrm{Var}(X_n(\mathbf{t})) \to [\mathbf{t}]$ *for all* $\mathbf{t} \in [0,1]^d$,

(ii) *the set* $\{(X_n(\mathbf{t}))^2 \mid n \in \mathbb{N}\}$ *is uniformly integrable for all* $\mathbf{t}$,

(iii) *for any collection $B_1, ..., B_k$ of strongly separated blocks, the increments of X_n around these blocks are asymptotically independent,*

(iv) *for each $\epsilon > 0$, $\eta > 0$, we can find a $\delta > 0$ such that*

$$\mathbb{P}(w(X_n, \delta) > \epsilon) < \eta$$

for all sufficiently large n.

Then $(X_n)_n$ converges weakly to a d-parameter Brownian sheet W in $\mathbb{R}$ with covariance operator 1.

Proof. As mentioned in [Deo75], we mimic the proof of [Bil68] Theorem 19.2: Since $\mathrm{Var}(X_n(0)) \to 0$, $\{X_n(0)\}$ is tight. It follows from Condition (iv) and [Bil68] Theorem 15.5. that $\{X_n\}$ is tight and that, if X is the limit of a subsequence, then X has continuous paths almost surely. It is enough to show that any such X must be distributed as W. Conditions (i) and (ii) imply $\mathbb{E}[X(\mathbf{t})] = 0$ and $\mathrm{Var}(X(\mathbf{t})) = [\mathbf{t}]$ by [Bil68] Theorem 5.4. Now the increments of X around strongly separated blocks are independent because of Condition (iii). Thus the result follow from Lemma 4.1.1. □

4.2 Preliminary Results

This section corresponds to [BW17] Section 4.1. We state and prove (with the exception of Lemma 4.2.14 ([BW17] Lemma 7)) all results stated there with a few intermediate results added.

4.2.1 Functional Central Limit Theorem

The following result is [BW17] Lemma 1. It is a type of Rosenthal inequality, cf. [Ros70] Theorem 3. It will allow us to, in some sense, uniformly bound the expected values of rescaled averages over blocks from above. This fact will be used in Lemma 4.2.6 to show uniform integrability of higher moments of the partial sum fields. As an exception, we for now consider the random field X to be indexed by $\mathbb{N}^d$ as one does not need the technical assumption that there are negative-indexed observations. We will show a similar inequality for the bootstrapped partial sum process later on in Lemma 4.2.13.

Lemma 4.2.1. *Let $(X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{N}^d}$ an H-valued centered (i.e., $\mathbb{E}[X_{\mathbf{k}}] = 0_H$ for all $\mathbf{k}$) random field satisfying Assumption B'. Then for all $r \geq 2$, there exists some constant $B_{d,r,\rho} > 0$ that does not depend on X and is monotonously increasing in the third argument such that for any finite set $S \subset \mathbb{N}^d$*

$$\mathbb{E}\left[\left\|\sum_{\mathbf{k}\in S} X_{\mathbf{k}}\right\|^r\right] \leq B_{d,r,\rho}\left(\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X_{\mathbf{k}}\|^r\right] + \left(\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X_{\mathbf{k}}\|^2\right]\right)^{\frac{r}{2}}\right). \tag{4.1}$$

If

$$\sup_{\mathbf{k}\in\mathbb{N}^d}\mathbb{E}\left[\|X_{\mathbf{k}}\|^r\right] < \infty, \tag{4.2}$$

corresponding to Assumption C”, let

$$C_{d,r,\rho} := B_{d,r,\rho}\left(\sup_{\mathbf{k}\in\mathbb{N}^d}\mathbb{E}\left[\|X_{\mathbf{k}}\|^r\right] + \left(\sup_{\mathbf{k}\in\mathbb{N}^d}\mathbb{E}\left[\|X_{\mathbf{k}}\|^2\right]\right)^{\frac{r}{2}}\right). \tag{4.3}$$

In that case

$$\mathbb{E}\left[\left\|\sum_{\mathbf{k}\in S} X_{\mathbf{k}}\right\|^r\right] \leq C_{d,r,\rho}|S|^{\frac{r}{2}} \tag{4.4}$$

and, if also $r > 2$, then

$$\mathbb{E}\left[M(U)^r\right] \leq D_{d,r}C_{d,r,\rho}|U|^{r/2} \tag{4.5}$$

for any discrete block U in $\mathbb{N}^d$ with

$$M(U) := \max_{W \lhd U}\left\|\sum_{\mathbf{j}\in W} X_{\mathbf{j}}\right\| \tag{4.6}$$

and

$$D_{d,r} := \frac{5^d}{2^d}\left(1 - \frac{1}{2^{\frac{r/2-1}{r}}}\right)^{-dr}.$$

Proof. We prove (4.1) by induction over the dimension d. In the case $d = 1$, the Inequality (4.1) follows directly from Theorem 2 in [Zha98]. To see that we can apply that theorem, note that its Condition (b) is fulfilled as any Hilbert space is a reflexive Banach space and that any Hilbert space is of type $p = 2$ which follows from the parallelogram identity, see [JN35] Theorem 1.

It can be seen from the proofs of [Zha98] Theorem 1 and Theorem 2 that $B(1, r, \cdot)$ is increasing since the constant

$$C(p, \rho) = \frac{2}{1 - 4\rho^{2/p\wedge 2/q}}$$

used in the proof of [Zha98] Theorem 1 is monotone in ρ (note $0 \le \rho^{2/p\wedge 2/q} < 1/4$ in that context) and the constant $B_{1,r\rho}$ does not introduce new dependence on ρ.

For the induction step, assume that $d \ge 2$ is fixed and that (4.1) is true for all dimensions smaller than d. For any set $M \subset \mathbb{N}^d$ and index $j \in \mathbb{N}$, write

$$M[j] \coloneqq \{\mathbf{k} \in M \mid \mathbf{k}_1 = j\},$$

$$J[M] \coloneqq \{i \in \mathbb{N} \mid M[i] \neq \emptyset\}.$$

Fix a finite, nonempty set $S \subset \mathbb{N}^d$ and define

$$Y_j \coloneqq \sum_{\mathbf{k}\in S[j]} X_{\mathbf{k}},$$

$$\zeta^{[j]} \coloneqq (\mathbb{1}_{\{\mathbf{k}\in\mathbb{N}^d[j]\}} X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{N}^d}.$$

$Y = (Y_j)_{j\in\mathbb{N}}$ and $\zeta^{[j]}$, $j \in \mathbb{N}$, can be seen as random fields on $\mathbb{N}$ and $\mathbb{N}^{d-1} \cong \mathbb{N}^d[j]$ respectively that satisfy the Mixing Condition B' as they are measurable transforms of X and X satisfies Assumption B' (see Remark 2.5.3). Applying the induction hypothesis shows

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{\mathbf{k}\in S} X_{\mathbf{k}}\right\|^r\right] &= \mathbb{E}\left[\left\|\sum_{j\in J[S]} Y_j\right\|^r\right] \\
&\le B_{1,r,\rho_Y}\left(\sum_{j\in J[S]} \mathbb{E}\left[\|Y_j\|^r\right] + \left(\sum_{j\in J[S]} \mathbb{E}\left[\|Y_j\|^2\right]\right)^{r/2}\right) \\
&\le B_{1,r,\rho_X}\left(\sum_{j\in J[S]} \mathbb{E}\left[\|Y_j\|^r\right] + \left(\sum_{j\in J[S]} \mathbb{E}\left[\|Y_j\|^2\right]\right)^{r/2}\right).
\end{aligned}$$

The above can be written as

$$B_{1,r,\rho_X} \sum_{j\in J[S]} \mathbb{E}\left[\left\|\sum_{\mathbf{k}\in S[j]} \zeta^{[j]}_{\mathbf{k}}\right\|^r\right] + B_{1,r}\left(\sum_{J[S]} \mathbb{E}\left[\left\|\sum_{\mathbf{k}\in S[j]} \zeta^{[j]}_{\mathbf{k}}\right\|^2\right]\right)^{r/2},$$

which is, due to the induction hypothesis, bounded from above by

$$\begin{aligned}&B_{1,r,\rho_X}\sum_{j\in J[S]}B_{d-1,r,\rho_{\zeta^{(j)}}}\left(\sum_{\mathbf{k}\in S[j]}\mathbb{E}\left[\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^r\right]+\left(\sum_{\mathbf{k}\in S[j]}\mathbb{E}\left[\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^2\right]\right)^{r/2}\right)\\&+B_{1,r,\rho_X}\left(\sum_{j\in J[S]}B_{d-1,2,\rho_{\zeta^{(j)}}}\left(\sum_{\mathbf{j}\in S[j]}\mathbb{E}\left[\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^2\right]+\left(\sum_{\mathbf{j}\in S[j]}\mathbb{E}\left[\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^2\right]\right)^{2/2}\right)\right)^{r/2}.\end{aligned}\tag{4.7}$$

Plugging in definitions and using the convention that sums over the empty set vanish, we get the following equality:

$$\sum_{j\in J[S]}\sum_{\mathbf{k}\in S[j]}\|\zeta_{\mathbf{k}}^{(j)}\|^r=\sum_{j\in\mathbb{N}}\sum_{\substack{\mathbf{k}\in S\\ \mathbf{k}_1=j}}\|\mathbb{1}_{\{\mathbf{k}_1=j\}}X_{\mathbf{k}}\|^r=\sum_{\mathbf{k}\in S}\|X_{\mathbf{k}}\|^r$$

Therefore we can write (4.7) as

$$\begin{aligned}&B_{1,r,\rho_X}B_{d-1,r,\rho_{\zeta^{(j)}}}\left(\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X_{\mathbf{k}}\|^r\right]+\sum_{j\in J[S]}\left(\sum_{\mathbf{k}\in S[j]}\mathbb{E}\left[\|\zeta_{\mathbf{k}}^{(j)}\|^2\right]\right)^{r/2}\right)\\&+B_{1,r,\rho_X}B_{d-1,2,\rho_{\zeta^{(j)}}}^{r/2}\left(2\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X_{\mathbf{k}}\|^2\right]\right)^{r/2}.\end{aligned}\tag{4.8}$$

Using

$$\sum_{j\in J[S]}\left(\sum_{\mathbf{k}\in S[j]}\mathbb{E}\left[\|\zeta_{\mathbf{k}}^{(j)}\|^2\right]\right)^{r/2}\le\left(\sum_{j\in J[S]}\sum_{\mathbf{k}\in S[j]}\mathbb{E}\left[\|\zeta_{\mathbf{k}}^{(j)}\|^2\right]\right)^{r/2}=\left(\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X\|^2\right]\right)^{r/2}$$

and the fact that the constants are increasing in the third argument by the induction hypothesis, (4.8) is bounded from above by

$$B_{d,r,\rho_X}\left(\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X_{\mathbf{k}}\|^r\right]+\left(\sum_{\mathbf{k}\in S}\mathbb{E}\left[\|X_{\mathbf{k}}\|^2\right]\right)^{r/2}\right)$$

with

$$B_{d,r,\rho_X}\coloneqq B_{1,r,\rho_X}B_{d-1,r,\rho_X}+2^{r/2}B_{1,r,\rho_X}B_{d-1,2,\rho_X}^{r/2}.$$

This concludes the induction and shows (4.4). Assuming

$$\sup_{\mathbf{k}\in\mathbb{N}^d}\mathbb{E}\left[\|X_{\mathbf{k}}\|^r\right]<\infty,$$

(4.6) follows directly from (4.4).

Now assume $r > 2$. Using (4.6), (4.5) follows directly from [Mór83] Corollary 1 for $\alpha = r/2$, $\gamma = r$ and the (super-)additive function of blocks

$$f(U) = C_{d,r}^{1/\alpha}|U|.$$

□

Remark 4.2.2. While this has not been made clear in [BW17], the argument that $B_{d,r,\cdot}$ is increasing is important to show that the constant does not depend on X as one can not argue that if f is a function mapping random fields to (possibly lower-dimensional) random fields, there is a function g that maps ρ-coefficients to ρ-coefficients such that the following assignments of mixing coefficients are equal:

$$\rho_{f(\cdot)} = g(\rho_\cdot).$$

The following counterexample shows that this does not hold for general f: Consider the (trivial) $\mathbb{R}^2$-valued stochastic processes

$$X = \left(\begin{pmatrix}+N\\-N\end{pmatrix}\right)_{j\in\mathbb{Z}}, \; Y = \left(\begin{pmatrix}+N\\+N\end{pmatrix}\right)_{j\in\mathbb{Z}}$$

with a non-constant, real random variable N and the function f that maps an $\mathbb{R}^2$-valued stochastic process to the $\mathbb{R}$-valued stochastic process that has the sum of the two components as its values, i.e.,

$$f(X) = (0)_{j\in\mathbb{Z}}, \; f(Y) = (2N)_{j\in\mathbb{Z}}.$$

It is easy to check that X and Y both have ρ-coefficients of 1, but $\rho_{f(X)} = -\infty$ (or $\rho_{f(X)} = 0$, depending on how one defines the supremum of the empty set) and $\rho_{f(Y)} = 1$. Therefore, there is no function that maps both ρ_X to $\rho_{f(X)}$ and ρ_Y to $\rho_{f(Y)}$.

The fact that $B_{d,r,\rho}$ does not depend on X is important as we will apply Lemma 4.2.1 in the proof of Theorem 3.2.1 to a sequence $X^{(-k)}$, $k = 1, \ldots$ of random fields whose moments we want to bound uniformly.

This lemma is [BW17] Lemma 2. The assumptions were originally based off of those of [Bil68] Theorem 19.2.

Lemma 4.2.3. *Let $\Sigma \in \mathbb{R}^{k\times k}$ be a symmetric, positive-semidefinite matrix and let $W_n = (W_n(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$ be a sequence of random fields with càdlàg sample paths in $\mathbb{R}^k$, i.e., they are elements of $D_{\mathbb{R}^k}([0,1]^d)$. Assume additionally*

a) $\lim\limits_{n\to\infty} \mathbb{E}\left[W_n(\mathbf{t})\right] = \underline{\mathbf{0}}$,

b) $\lim\limits_{n\to\infty} \operatorname{Cov}(W_n(\mathbf{t})) = [\mathbf{t}]\Sigma \; \forall \mathbf{t} \in [0,1]^d$,

c) $\{\|W_n(\mathbf{t})\|^2 \mid n \in \mathbb{N}\}$ *is uniformly integrable for all* $\mathbf{t} \in [0,1]^d$,

d) *for any collection of strongly separated blocks* $B_1, ..., B_p$, *the increments of* W_n *around these blocks are asymptotically independent.*

e)
$$\forall \epsilon > 0 \; \forall \eta > 0 \; \exists \delta > 0 \; \exists n_0 \in \mathbb{N} \; \forall n \geq n_0 : \mathbb{P}(w(W_n, \delta) > \epsilon) < \eta \tag{4.9}$$

Then W_n *converges weakly in* $D_{\mathbb{R}^k}([0,1]^d)$ *to a Brownian sheet in* $\mathbb{R}^k$ *with covariance operator (i.e., covariance matrix)* Σ.

Proof. We first show that for any vector $\boldsymbol{\gamma} \in \mathbb{R}^k$, the real-valued random fields W_n^{γ} defined by
$$W_n^{\gamma}(\mathbf{t}) := \boldsymbol{\gamma}^T W_n(\mathbf{t})$$
converge weakly to a Brownian sheet W^{γ} in $\mathbb{R}$ with covariance operator (i.e., variance) $\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma} \in \mathbb{R}$. As Σ is positive-semidefinite by assumption, we either have $\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma} = 0$ or $\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma} > 0$. The first case is trivial. For the second case, we will apply Lemma 4.1.3 to the normed random field

$$N_n^{\gamma} := \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1/2} W_n^{\gamma}.$$

(i) By linearity of the expectation, we have

$$\mathbb{E}\left[N_n^{\gamma}(\mathbf{t})\right] = \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1/2} \boldsymbol{\gamma}^T \mathbb{E}\left[W_n(\mathbf{t})\right]$$

which converges to 0 by assumption. For the variance, we have

$$\begin{aligned}
\mathrm{Var}(N_n^{\gamma}(\mathbf{t})) &= \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1} \mathrm{Var}(\boldsymbol{\gamma}^T W_n(\mathbf{t})) \\
&= \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1} \boldsymbol{\gamma}^T \mathrm{Cov}(W_n(\mathbf{t}))\boldsymbol{\gamma} \\
&\to \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1} \boldsymbol{\gamma}^T ([\mathbf{t}]\Sigma)\boldsymbol{\gamma} \\
&= [\mathbf{t}].
\end{aligned}$$

(ii) Using Cauchy-Schwarz, we get

$$\|N_n^{\gamma}(\mathbf{t})\|^2 \leq \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1} \|\boldsymbol{\gamma}\|^2 \|W_n(\mathbf{t})\|^2.$$

The uniform integrability of the set $\{\|N_n^{\gamma}(\mathbf{t})\|^2 \mid n \in \mathbb{N}\}$ for all $\mathbf{t}$ therefore follows from the uniform integrability of $\{W_n(\mathbf{t})\|^2 \mid n \in \mathbb{N}\}$ which holds by assumption.

(iii) Now let $H_1, ..., H_p$ be a collection of Borel sets in $\mathbb{R}$. Denote left multiplication with $\boldsymbol{\gamma}^T$ by $f_{\boldsymbol{\gamma}}$ and write

$$\boldsymbol{\gamma}' := \left(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}\right)^{-1/2} \boldsymbol{\gamma}.$$

Due to the continuity of $f_{\boldsymbol{\gamma}'}$, the sets $f_{\boldsymbol{\gamma}'}^{-1}(H_i)$ are Borel sets in $\mathbb{R}^k$. Let $B_1, ..., B_p$ be a collection of strongly separated blocks in $[0,1]^d$. Then we have

$$\begin{aligned}
&\mathbb{P}(N_n^{\boldsymbol{\gamma}}(B_i) \in H_i \forall i) - \prod_{i=1}^{p} \mathbb{P}(N_n^{\boldsymbol{\gamma}}(B_i) \in H_i) \\
&= \mathbb{P}(W_n(B_i) \in f_{\boldsymbol{\gamma}'}^{-1}(H_i) \forall i) - \prod_{i=1}^{p} \mathbb{P}(W_n(B_i) \in f_{\boldsymbol{\gamma}'}^{-1}(H_i)).
\end{aligned}$$

The above converges to 0 by assumption, meaning that the increments of $N_n^{\boldsymbol{\gamma}}$ around the blocks B_i are asymptotically independent.

(iv) Condition (iv) of Lemma 4.1.3 follows from Assumption (4.9) together with

$$|N_n^{\boldsymbol{\gamma}}(\mathbf{t}) - N_n^{\boldsymbol{\gamma}}(\mathbf{s})| = |\boldsymbol{\gamma}'^T (W_n(\mathbf{t}) - W_n(\mathbf{s}))| \leq \|\boldsymbol{\gamma}'\| \|W_n(\mathbf{t}) - W_n(\mathbf{s})\|$$

where we have once more used Cauchy-Schwarz.

As $f_{\boldsymbol{\gamma}'}$ is a continuous function, it maps $D_{\mathbb{R}^k}[0,1]^d$ to $D_{\mathbb{R}}[0,1]^d$. Hence we may apply Lemma 4.1.3 to see that $N_n^{\boldsymbol{\gamma}}$ converges weakly in $D_{\mathbb{R}}[0,1]^d$ to the Brownian sheet $(\boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma}) W^{\boldsymbol{\gamma}}$ with covariance operator $1 \in \mathbb{R}$.

[Bil68] Theorem 15.5 tells us that (4.9) ensures the tightness of the sequences $(W_n^{e_i})_n$, meaning that for every $\epsilon > 0$, there is a $M_\epsilon > 0$ such that

$$\mathbb{P}(\|W_n^{e_i}\|_\infty > M_\epsilon / \sqrt{k}) \leq \epsilon / k \tag{4.10}$$

for every $i = 1, ..., k$. Define the norm $\|\cdot\|_{\infty,2}$ on $D_{\mathbb{R}^k}[0,1]^d$ by

$$\|x\|_{\infty,2} := \left(\sum_{i=1}^{k} \|x_i\|_\infty^2 \right)^{1/2}.$$

The Skorokhod metric d_S on $D_{\mathbb{R}^k}[0,1]^d$ is bounded from above by the metric induced by $\|\cdot\|_{\infty,2}$ and therefore sets that are compact with respect to $\|\cdot\|_{\infty,2}$ are compact with respect to d_S. Hence if the sequence W_n is tight with respect to $\|\cdot\|_{\infty,2}$, it is also tight with respect to the Skorokhod metric d_S. And it is indeed tight with respect to $\|\cdot\|_{\infty,2}$ as we have

$$\begin{aligned}\mathbb{P}(\|W_n\|_{\infty,2} > M_\epsilon) &= \mathbb{P}\left(\left(\sum_{i=1}^k \|W_n^{e_i}\|_\infty^2\right)^{1/2} > M_\epsilon\right) \\ &\leq \mathbb{P}\left(\max_i \|W_n^{e_i}\|_\infty^2 > M_\epsilon^2/k\right) \\ &\leq \sum_{i=1}^k \mathbb{P}\left(\|W_n^{e_i}\|_\infty > M_\epsilon/\sqrt{k}\right) \\ &\leq \epsilon\end{aligned}$$

where we have used Boole's inequality and (4.10). Using Prokhorov's theorem, we may find a weakly convergent subsequence $(W_{n_m})_m$. Call the weak limit of this subsequence W. On the one hand, we know $W_{n_m}^{\gamma} \Rightarrow W^{\gamma}$. On the other hand, the continuous mapping theorem ensures $W_{n_m}^{\gamma} \Rightarrow \boldsymbol{\gamma}^T W$. Therefore, W^{γ} and $\boldsymbol{\gamma}^T W$ coincide in distribution.

We now only need to show that W is a Brownian sheet in $\mathbb{R}^k$ with covariance operator Σ because any converging subsequence of $(W_n)_n$ has the same weak limit. W is almost surely continuous because the same holds for its coordinate processes W^{e_i}. The same reasoning shows that if one of the components of $\mathbf{t} \in [0,1]^d$ is zero, then $W(\mathbf{t})$ is zero. As W has almost surely continuous sample paths, the projection maps

$$\pi_{\mathbf{t}^1,\ldots,\mathbf{t}^l} : C_{\mathbb{R}^k}[0,1]^d \to \left(\mathbb{R}^k\right)^l, x \mapsto \left(x\left(\mathbf{t}^1\right), \ldots x\left(\mathbf{t}^l\right)\right)$$

are almost surely continuous. By the continuous mapping theorem, this shows the convergence of the finite-dimensional distributions of W_{n_m} and therefore the convergence of the increments of W_{n_m} around blocks. Any increment $W(B)$ of W around a block B has a centered Gaussian distribution with covariance $\lambda(B)\Sigma$ since the coordinates of $W(B)$ are centered Gaussian and

$$\sum_{i=1}^k \boldsymbol{\gamma}_i W^{e_i}(B) = W^{\gamma}(B)$$

is centered Gaussian with variance $\lambda(B)\boldsymbol{\gamma}^T\Sigma\boldsymbol{\gamma}$. To show that W is a Brownian sheet in $\mathbb{R}^k$, it only remains to show that the increments of W around disjoint blocks $B_1, \ldots, B_l$ are independent. In the case that the blocks are strongly separated, from the weak convergence

$$(W_{n_m}(B_1), \ldots, W_{n_m}(B_l)) \overset{m}{\Rightarrow} (W(B_1), \ldots, W(B_l))$$

it follows for vectors $\mathbf{y}_1, ..., \mathbf{y}_l \in \mathbb{R}^k$ that

$$\begin{aligned}\mathbb{P}(W(B_j) \leq \mathbf{y}_j \,\forall j) &= \lim_{m\to\infty} \mathbb{P}(W_{n_m}(B_j) \leq \mathbf{y}_j \,\forall j) \\ &= \lim_{m\to\infty} \left(\mathbb{P}(W_{n_m}(B_j) \leq \mathbf{y}_j \,\forall j) - \prod_{j=1}^{l} \mathbb{P}(W_{n_m}(B_j) \leq \mathbf{y}_j) \right) \\ &+ \prod_{j=1}^{l} \lim_{m\to\infty} \mathbb{P}(W_{n_m}(B_j) \leq \mathbf{y}_j).\end{aligned}$$

Since the increments of W_n around strongly separated blocks are asymptotically independent by Assumption d), the first term is 0. The second term is

$$\prod_{j=1}^{l} \mathbb{P}(W(B_j) \leq \mathbf{y}_j).$$

Therefore, the increments of W around strongly separated blocks are independent. The almost sure continuity of W yields the independence of increments around any collection of pairwise disjoint blocks, see also the proof of [Bil68] Theorem 19.2. □

Lemma 4.2.4. *Let $j \in \mathbb{N}$ be some natural number. Then the inequality*

$$\left|\{x \in \mathbb{Z}^d \mid \|x\|_\infty = j\}\right| \leq d2^d j^{d-1} \tag{4.11}$$

holds.

Proof. Two simple inductions show

$$\begin{aligned}\left|\{x \in \mathbb{Z}^d \mid \|x\|_\infty = j\}\right| &= \left|\{x \in \mathbb{Z}^d \mid \|x\|_\infty \leq j\}\right| - \left|\{x \in \mathbb{Z}^d \mid \|x\|_\infty \leq j-1\}\right| \\ &= (2j+1)^d - (2j-1)^d \\ &\leq d2^d j^{d-1}.\end{aligned}$$

□

Lemma 4.2.5. *The limit of a sequence of symmetric, positive-semidefinite matrices is symmetric and positive-semidefinite.*

Proof. By [BV04] Example 2.15, the set of symmetric, positive-semidefinite $(d \times d)$-matrices $\mathbf{S}_+^d$ is a proper cone. In particular, it is a closed set in the complete normed space $\mathbb{R}^{d\times d}$. Hence any limit of elements in $\mathbf{S}_+^d$ lies in $\mathbf{S}_+^d$.

Alternatively, one may show the statement directly: Let $(\Sigma(n))_n$ be a sequence of symmetric, positive-semidefinite $(d \times d)$-matrices with $\Sigma(n) = (\sigma i, j(n))_{i,j=1,...,d}$ that converges to some matrix $\Sigma = (\sigma_{i,j})_{i,j=1,...,d}$. The convergence $\Sigma(n) \to \Sigma$ is equivalent of the component-wise convergence. Now let $x \in \mathbb{R}^d$ be any vector. Due to positive-semidefinitness of the matrices $\Sigma(n)$, we have

$$0 \leq x^t \Sigma(n) x = \sum_{i,j=1}^{d} \sigma_{i,j}(n) x_i x_j$$

for all n. As $(x^t \Sigma(n) x)_n$ is a sequence of non-negative numbers, its limit

$$\sum_{i,j=1}^{d} \sigma_{i,j} x_i x_j = x^t \Sigma x$$

is also non-negative. Since x was chosen arbitrarily, this shows the positive-semi-definitness of Σ. The symmetry of Σ is obvious. □

The following result is [BW17] Lemma 3.

Lemma 4.2.6. *Assume Assumptions A', B, C" and D' and let the random field X be centered and have values in $\mathbb{R}^k$ and let S_n be the partial sum field of X. Define*

$$\Gamma(n, \mathbf{t}) := \mathrm{Cov}(S_n(\mathbf{t})).$$

Then

$$\lim_{n\to\infty} \Gamma(n, \mathbf{t}) = [\mathbf{t}]\Gamma$$

for all $\mathbf{t} \in [0,1]^d$ with the long-run variance matrix

$$\Gamma = \sum_{\mathbf{v}\in\mathbb{Z}^d} \mathrm{Cov}(X_{\underline{\mathbf{0}}}, X_{\mathbf{v}})$$

of X and the above series converges absolutely. Furthermore, S_n converges weakly in $D_{\mathbb{R}^k}([0,1]^d)$ to a Brownian sheet in $\mathbb{R}^k$ with covariance matrix Γ.

Proof. We show the assumptions of Lemma 4.2.3.
Lemma 4.2.3 a) is true because X is centered and thus S_n is, too.

We now show the summability of the covariograms. It is remarked in [Guy95] p. 110 (in the context of univariate random fields) that the moment existence Assumption C" implies

$$|\mathrm{Cov}(X^i_{\underline{\mathbf{0}}}, X^i_{\mathbf{v}})| = |\gamma_{i,i}(\mathbf{v})| \leq 8 \sup_{\mathbf{l}} \|X^i_{\mathbf{l}}\|^2_{2+\delta} \alpha^{X^i}_{1,1}(\|\mathbf{v}\|_\infty)^{\frac{\delta}{2+\delta}}$$

for all $i = 1, ..., k$ and large enough $\mathbf{v}$. Since

$$\alpha^{X^i}_{1,1} \leq \alpha^{X}_{1,1},$$

according to [Guy95], Assumption D' together with the weak stationarity Assumption A' then shows the summability of the covariogram $\gamma_{i,i}$. If we want to apply this argument to $\gamma_{i,j}$, we need to consider the random field $X^{j,i}$ by which we mean the $\mathbb{R}$-valued random field X^j for which $X^j_{\underline{\mathbf{0}}}$ is replaced by $X^i_{\underline{\mathbf{0}}}$. However, this random field also satisfies the Assumption C" and

$$\alpha_{1,1}^{X^{j,i}} \leq \alpha_{1,1}^{X}$$

as the σ-fields generated by this univariate random field are included in those generated by X. By Inequality (4.11), we have

$$\begin{aligned}\sum_{\mathbf{v}\in\mathbb{Z}^d} |\gamma_{i,j}(\mathbf{v})| &\leq \sum_{\mathbf{v}\in\mathbb{Z}^d} 8 \sup_{\mathbf{l}} \|X^{j,i}_{\mathbf{l}}\|^2_{2+\delta} \alpha_{1,1}^{X^{j,i}}(\|\mathbf{v}\|_\infty)^{\frac{\delta}{2+\delta}} + C \\ &\leq 8 \sup_{\mathbf{l}} \|X_{\mathbf{l}}\|^2_{2+\delta} \sum_{m\geq 1} \left(d2^d m^{d-1}\right) \alpha_{1,1}^{X}(m)^{\frac{\delta}{2+\delta}} + C\end{aligned}$$

for a constant C that compensates for the fact that the covariance inequality only holds for large enough $\mathbf{v}$. Using Assumptions C" and D', we get

$$\sum_{\mathbf{v}\in\mathbb{Z}^d} |\gamma_{i,j}(\mathbf{v})| < \infty. \tag{4.12}$$

Using the definition of Γ, the bilinearity of the cross-covariance and the weak stationarity of X, we obtain

$$\begin{aligned}\Gamma(n,\mathbf{t})_{i,j} &= \operatorname{Cov}\left(n^{-d/2} \sum_{\underline{\mathbf{1}}\leq\mathbf{m}\leq\lfloor n\mathbf{t}\rfloor} X_{\mathbf{m}}\right)_{i,j} \\ &= \frac{1}{n^d} \sum_{\underline{\mathbf{1}}\leq\mathbf{m}\leq\lfloor n\mathbf{t}\rfloor} \sum_{\underline{\mathbf{1}}\leq\mathbf{m}'\leq\lfloor n\mathbf{t}\rfloor} \operatorname{Cov}(X^i_{\mathbf{m}}, X^j_{\mathbf{m}'}) \\ &= \frac{1}{n^d} \sum_{-\lfloor n\mathbf{t}\rfloor<\mathbf{v}<\lfloor n\mathbf{t}\rfloor} |\{(\mathbf{m},\mathbf{m}') \in \{\underline{\mathbf{1}},...,\lfloor n\mathbf{t}\rfloor\}^2 \mid \mathbf{m}-\mathbf{m}'=\mathbf{v}\}| \operatorname{Cov}(X^i_{\mathbf{v}}, X^j_{\underline{\mathbf{0}}}).\end{aligned}$$

A simple induction over d, using a second induction over $\lfloor nt\rfloor$ for the base case $d=1$, shows

$$|\{(\mathbf{m},\mathbf{m}') \in \{\underline{\mathbf{1}},...,\lfloor n\mathbf{t}\rfloor\}^2 \subset \mathbb{Z}^d\times\mathbb{Z}^d \mid \mathbf{m}-\mathbf{m}'=\mathbf{v}\}| = \prod_{l=1}^{d} (\lfloor n\mathbf{t}_l\rfloor - |\mathbf{v}_l|)$$

for $-\lfloor n\mathbf{t}\rfloor \leq \mathbf{v} \leq \lfloor n\mathbf{t}\rfloor$. Hence

$$\begin{aligned}\Gamma(n,\mathbf{t})_{i,j} &= \frac{1}{n^d} \sum_{-\lfloor n\mathbf{t}\rfloor<\mathbf{v}<\lfloor n\mathbf{t}\rfloor} \operatorname{Cov}(X^i_{\mathbf{v}}, X^j_{\underline{\mathbf{0}}}) \prod_{l=1}^{d} (\lfloor n\mathbf{t}_l\rfloor - |\mathbf{v}_l|) \\ &= \sum_{\mathbf{v}\in\mathbb{Z}^d} \mathbb{1}_{|\mathbf{v}|<\lfloor n\mathbf{t}\rfloor} \gamma_{i,j}(\mathbf{v}) \prod_{l=1}^{d} \frac{\lfloor n\mathbf{t}_l\rfloor - |\mathbf{v}_l|}{n}.\end{aligned}$$

As the sum

$$\sum_{\mathbf{v}\in\mathbb{Z}^d} \gamma_{i,j}(\mathbf{v}) = \Gamma_{i,j}$$

exists by (4.12), we can, pointwise in ω, apply the dominated convergence theorem w.r.t. the counting measure and the dominating function $[\mathbf{t}]\gamma_{i,j}$ to get

$$\begin{aligned}
\lim_{n\to\infty} \Gamma(n,\mathbf{t})_{i,j} &= \lim_{n\to\infty} \sum_{\mathbf{v}\in\mathbb{Z}^d} \mathbb{1}_{|\mathbf{v}|<\lfloor n\mathbf{t}\rfloor} \gamma_{i,j}(\mathbf{v}) \prod_{l=1}^{d} \frac{\lfloor n\mathbf{t}_l\rfloor - |\mathbf{v}_l|}{n} \\
&= \sum_{\mathbf{v}\in\mathbb{Z}^d} \gamma_{i,j}(\mathbf{v})[\mathbf{t}] \\
&= [\mathbf{t}]\Gamma_{i,j}.
\end{aligned}$$

Thus the Assumption b) of Lemma 4.2.3 is fulfilled. Applying Lemma 4.2.5, we also see that Γ is a symmetric, positive-semidefinite matrix.

By assumption the conditions of Lemma 4.2.1 are fulfilled. (4.4) therefore shows

$$\begin{aligned}
\mathbb{E}\left[\|S_n(\mathbf{t})\|^{2+\delta}\right] &= \left(\frac{1}{n^{d/2}}\right)^{2+\delta} \mathbb{E}\left[\left\|\sum_{\mathbf{1}\le\mathbf{j}\le\lfloor n\mathbf{t}\rfloor} X_{\mathbf{j}}\right\|^{2+\delta}\right] \\
&\le \left(\frac{1}{n^{d/2}}\right)^{2+\delta} C_{d,2+\delta} \left(\prod_{l=1}^{d} \lfloor n\mathbf{t}_l\rfloor\right)^{\frac{2+\delta}{2}} \\
&= C_{d,2+\delta} \left(\prod_{l=1}^{d} \frac{\lfloor n\mathbf{t}_l\rfloor}{n}\right)^{\frac{2+\delta}{2}} \\
&\le C_{d,2+\delta}[\mathbf{t}]^{\frac{2+\delta}{2}} \\
&\le C_{d,2+\delta}
\end{aligned}$$

for all n and $\mathbf{t}$. Therefore

$$\sup_n \mathbb{E}\left[\|S_n(\mathbf{t})\|^{2+\delta}\right] \le C_{d,2+\delta} < \infty.$$

In [Bil68] p. 32 it is remarked that this implies the uniform integrability Assumption c).

To show Condition d), let $B_j = (\mathbf{s}^j, \mathbf{t}^j]$, $j = 1, ..., q$, be a collection of strongly separated blocks. Without loss of generality (after reordering the blocks) we may assume that there is an index $i \in \{1, ..., d\}$ such the blocks are ordered in the sense that

$$0 \le \mathbf{s}_i^1 \le \mathbf{t}_i^1 < ... < \mathbf{s}_i^q \le \mathbf{t}_i^q \le 1,$$

implying $\min_{j=1,\dots,q-1}(\mathbf{s}_i^{j+1} - \mathbf{t}_i^j) > 0$ and thus

$$r_n \coloneqq \min_{j=1,\dots,q-1} \left(\lfloor n\mathbf{s}_i^{j+1} \rfloor - \lfloor n\mathbf{t}_i^j \rfloor \right) \to \infty \text{ for } n \to \infty. \tag{4.13}$$

Now let H_j, $j = 1, ..., q$ be a collection of Borel-sets in $\mathbb{R}^k$. Then

$$\left| \mathbb{P}\left(S_n(B_j) \in H_j \ \forall j\right) - \prod_{j=1}^{q} \mathbb{P}\left(S_n(B_j) \in H_j\right) \right|$$

is equal to the telescoping sum $\left| \sum_{l=1}^{q} (a_l - a_{l-1}) \right|$ with

$$a_l \coloneqq \mathbb{P}\left(S_n(B_j) \in H_j \ \forall j = 1, ..., l\right) \prod_{j=l+1}^{q} \mathbb{P}\left(S_n(B_j) \in H_j\right)$$

for $l = 0, ..., q$. Setting the events

$$A_l \coloneqq \left\{ \bigcap_{j=1}^{l-1} S_n(B_j) \in H_j \right\}, \ B_l \coloneqq \{S_n(B_l) \in H_l\}$$

for $l > 0$ and $A_0 \coloneqq B_0 \coloneqq \Omega$, we may write

$$a_l - a_{l-1} = \mathbb{P}\left(A_l \cap B_l\right) \prod_{j=l+1}^{q} \mathbb{P}\left(S_n(B_j) \in H_j\right) - \mathbb{P}(A_l)\mathbb{P}(B_l) \prod_{j=l+1}^{q} \mathbb{P}\left(S_n(B_j) \in H_j\right).$$

By definition of r_n, for each l there are, in the spirit of Definition 2.5.1 of the $\rho_{\mathbb{R}}$-mixing coefficient,

$$M_l, N_l \subset \mathbb{Z}^d : \exists A, B \subset \mathbb{Z}, \operatorname{dist}(A, B) \geq r_n : \forall \mathbf{j} \in M, \mathbf{k} \in N : \mathbf{j}_i \in A, \mathbf{k}_i \in B$$

such that $A_l \in \mathcal{A}_M \coloneqq \sigma(X_{\mathbf{k}} : \mathbf{k} \in M), B_l \in \mathcal{A}_N$. Therefore we can bound the telescoping sum as follows:

$$\begin{aligned}
\left| \sum_{l=1}^{q} (a_l - a_{l-1}) \right| &\leq \sum_{l=1}^{q} \prod_{j=l+1}^{q} \mathbb{P}\left(S_n(B_j) \in H_j\right) \left| \mathbb{P}\left(A_l \cap B_l\right) - \mathbb{P}(A_l)\mathbb{P}(B_l) \right| \\
&\leq \sum_{l=1}^{q} \left| \mathbb{P}\left(A_l \cap B_l\right) - \mathbb{P}(A_l)\mathbb{P}(B_l) \right| \\
&\leq \sum_{l=1}^{q} \alpha(\mathcal{A}_{M_l}, \mathcal{A}_{N_l}).
\end{aligned}$$

According to [Bra86] Inequality (1.8), this is bounded from above by

$$\sum_{l=1}^{q} \frac{1}{4} \rho_{\mathbb{R}}(\mathcal{A}_{M_l}, \mathcal{A}_{N_l}) \leq \frac{q}{4} \rho_{\mathbb{R}}(r_n).$$

By the ρ-mixing Assumption B in conjunction with (4.13), this tends to 0 as n goes to infinity. This means that we have shown

$$\lim_{n\to\infty} \mathbb{P}\left(S_n(B_j) \in H_j \ \forall j\right) - \prod_{j=1}^{q} \mathbb{P}\left(S_n(B_j) \in H_j\right) = 0$$

for strongly separated blocks B_j and Borel-sets H_j, which is Condition d) of Lemma 4.2.3.

To show e), note that we may, analogously to the proof of Condition c), apply Lemma 4.2.1 to get

$$\mathbb{E}\left[M(U)^{2+\delta}\right] \leq \tilde{C}_{d,2+\delta} |U|^{\frac{2+\delta}{2}}$$

from Inequality (4.5) for any discrete block $U \subset \mathbb{N}^d$. Rewriting this, we get

$$\mathbb{E}\left[\left(\frac{M(U)^2}{|U|}\right)^{\frac{2+\delta}{2}}\right] \leq \tilde{C}_{d,r} < \infty$$

where the upper bound does not depend on U. Because of $\frac{2+\delta}{2} > 1$, this shows the uniform integrability of any family

$$\left\{\left(\frac{M(U_n)^2}{|U_n|}\right)^{\frac{2+\delta}{2}} : n \in \mathbb{N}\right\}$$

of discrete blocks $U_n \subset \mathbb{N}^d$. In the proof of [BS07] Theorem 1.3. in Chapter 5 it is concluded that this ensures the condition e) regarding the modulus of continuity of the partial sum fields for univariate random fields. The proof does not change when considering multivariate ones. Note that both [BS07] Theorem 1.3. and Lemma 4.2.1 consider random fields indexed by $\mathbb{N}^d$ instead of the usual $\mathbb{Z}^d$.

We have now shown all conditions of Lemma 4.2.3. Applying that lemma concludes the proof. □

Remark 4.2.7. The constants $(2+\delta)/\delta$ and $2+\delta$ in the above proof stem from the covariance inequality

$$|Cov(Y, Z) \leq 2\alpha(Y, Z)^{1/p} \|Y\|_q \|Z\|_r$$

with positive constants p, q, r such that

$$p^{-1} + q^{-1} + r^{-1} = 1$$

where

$$\alpha(Y,Z) := 2 \sup_{y,z\in\mathbb{R}} |\mathbb{P}(Y > y, Z > z) - \mathbb{P}(Y > y)\mathbb{P}(Z > z),$$

see [Rio13] (1.12b).

The following lemma explains the rationale behind the term "long-run variance"

Lemma 4.2.8. *Let $X = (X_{\mathbf{j}})_{j\in\mathbb{R}^d}$ be an $\mathbb{R}$-valued, weakly stationary, centered random field. Assume that the long-run variance Γ of X exists. Then it is given by*

$$\Gamma = \lim_{n\to\infty} \frac{1}{n^d}\mathbb{E}\left[\left(\sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}} X_{\mathbf{j}}\right)^2\right] = \lim_{n\to\infty}\frac{1}{n^d}\mathrm{Var}\left(\sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}} X_{\mathbf{j}}\right).$$

Proof. Setting $\mathbf{t} = \underline{\mathbf{1}}$, we have essentially shown this in the proof of Lemma 4.2.6 in the part involving $\Gamma(n,\mathbf{t})$. □

The following theorem is [Bil99] Theorem 3.2. It is needed in order to prove Lemma 4.2.10.

Theorem 4.2.9. *Let (M, g) be a metric space. Suppose that $(X_{k,n}, X_n)$ are elements of $M \times M$. If*

$$X_{k,n} \Rightarrow_n Z_k \Rightarrow_k X \tag{4.14}$$

and

$$\lim_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}(g(X_{k,n}, X_n) \geq \epsilon) = 0 \tag{4.15}$$

for each $\epsilon > 0$, then $X_n \Rightarrow X$.

The following lemma, formulated in [BW17] as Lemma 4, is an adaptation of [CW98] Lemma 4.1 in which stochastic processes are replaced with vectors of random fields and the space of continuous functions is replaced with the Skorokhod space $D_H[0,1]^d$.

Lemma 4.2.10. *Let $K \in \mathbb{N}$ be some natural number and let*

$$((X_{n,1}, ..., X_{n,K}))_{n\in\mathbb{N}}$$

be a sequence of random elements with values in $\left(D_H[0,1]^d\right)^K$. For each $k \in \mathbb{N}$, let

$$X_1^k, ..., X_K^k$$

be independent Brownian sheets in

$$H_k = P_k H = \mathrm{span}(e_1, ..., e_k) \subset H$$

with covariance operators S_i^k, $i = 1, ..., K$ and $\mathbb{E}\left[X_i^k(\underline{\mathbf{1}})\right] = 0$. Assume the following:

(i)
$$\begin{pmatrix} P_k X_{n,1} \\ \dots \\ P_k X_{n,K} \end{pmatrix} \overset{n}{\Rightarrow} \begin{pmatrix} X_1^k \\ \dots \\ X_K^k \end{pmatrix} \tag{4.16}$$

in $\left(D_{H_k}[0,1]^d\right)^K$ *for all* $k \in \mathbb{N}$.

(ii)
$$\begin{pmatrix} X_1^k \\ \dots \\ X_K^k \end{pmatrix} \overset{k}{\Rightarrow} \begin{pmatrix} X_1 \\ \dots \\ X_K \end{pmatrix} \tag{4.17}$$

in $\left(D_H[0,1]^d\right)^K$ *for independent Brownian sheets* X_i, $i = 1, ..., K$, *in* H *with covariance operators* S_i.

(iii) *There is some* $r \geq 2$ *such that*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[\sup_{\mathbf{t}_1,\dots,\mathbf{t}_K \in [0,1]^d} \left\| \begin{pmatrix} X_{n,1}(\mathbf{t}_1) \\ \dots \\ X_{n,K}(\mathbf{t}_K) \end{pmatrix} - \begin{pmatrix} P_k X_{n,1}(\mathbf{t}_1) \\ \dots \\ P_k X_{n,K}(\mathbf{t}_K) \end{pmatrix} \right\|_{H^K}^r \right] = 0. \tag{4.18}$$

Then
$$\begin{pmatrix} X_{n,1} \\ \dots \\ X_{n,K} \end{pmatrix} \overset{n}{\Rightarrow} \begin{pmatrix} X_1 \\ \dots \\ X_K \end{pmatrix} \tag{4.19}$$

in $\left(D_H[0,1]^d\right)^K$.

Proof. (4.16) and (4.17) together are (4.14). Applying the fact that the Skorokhod metric is bounded from above by the supremum metric and using Markov's inequality, we see that (4.15) follows from (4.18). Hence Theorem 4.2.9 shows (4.19). □

Lemma 4.2.11. *Let* W *be a Brownian sheet in* H *with covariance operator* S. *For any* $k \in \mathbb{N}$, *the* $\mathbb{R}^k$*-valued Brownian sheet*

$$W^{(k)} = (P_k W(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$$

then has covariance operator

$$S^{(k)} := P_k S P_k : \mathbb{R}^k \to \mathbb{R}^k.$$

Proof. Let B be any block in $[0,1]^d$. We only need to show that the increment $W^{(k)}(B)$ has covariance operator $\lambda(B)S^{(k)}$, i.e.,

$$\mathbb{E}\left[\langle W^{(k)}(B), h_1\rangle \langle W^{(k)}(B), h_2\rangle\right] = \langle \lambda(B) S^{(k)} h_1, h_2\rangle$$

for all vectors $h_1, h_2 \in \mathbb{R}^k$. Using the fact P_k is linear, the increments $(P_k W)(B)$ and $P_k(W(B))$ coincide. Therefore

$$\begin{aligned}\mathbb{E}\left[\langle W^{(k)}(B), h_1\rangle\langle W^{(k)}(B), h_2\rangle\right] &= \mathbb{E}\left[\langle (P_k W)(B), h_1\rangle\langle (P_k W)(B), h_2\rangle\right] \\ &= \mathbb{E}\left[\langle P_k(W(B)), h_1\rangle\langle P_k(W(B)), h_2\rangle\right] \\ &= \mathbb{E}\left[\langle P_k(W(B)), h_1\rangle\langle P_k(W(B)), h_2\rangle\right].\end{aligned}$$

As orthonogal projections are self-adjoint, the above is equal to

$$\mathbb{E}\left[\langle W(B), P_k h_1\rangle\langle W(B), P_k h_2\rangle\right].$$

Because S is the covariance operator of W, we get

$$\begin{aligned}\mathbb{E}\left[\langle W(B), P_k h_1\rangle\langle W(B), P_k h_2\rangle\right] &= \langle \lambda(B) S(P_k h_1), P_k h_2\rangle \\ &= \langle \lambda(B)(P_k S P_k) h_1, h_2\rangle.\end{aligned}$$

□

4.2.2 Dependent Wild Bootstrap

The following result is [BW17] Lemma 5.

Lemma 4.2.12. *Let X and Y be $\mathcal{F}$- and $\mathcal{G}$-measurable random variables respectively that live in some Hilbert space H_1. Let V be another random variable that takes values in a Hilbert space H_2 and is independent of $\sigma(\mathcal{F}, \mathcal{G})$. Let $g, h : H_1 \times H_2 \to H$ be measurable functions satisfying*

$$\mathbb{E}\left[g(X, V) \mid V\right] = \mathbb{E}\left[h(Y, V) \mid V\right] = 0 \text{ a.s.}$$

Assume $\rho(\mathcal{F}, \mathcal{G}) < 1$.

Then for any $p > 1$ for which $\mathbb{E}\left[\|g(X, V)\|^p\right]$ and $\mathbb{E}\left[\|h(Y, V)\|^p\right]$ exist, there is a constant $C = C(\rho(\mathcal{F}, \mathcal{G}), p)$ such that

$$\mathbb{E}\left[\|g(X, V)\|^p\right] \leq C\mathbb{E}\left[\|g(X, V) + h(Y, V)\|^p\right].$$

Proof. For almost all $v \in H_2$, we may apply [Zha98] Theorem 1 to the H-valued random variables $g(X, v)$ and $h(Y, v)$ to get

$$\begin{aligned}\mathbb{E}\left[\|g(X, V)\|^p \mid V = v\right] &= \mathbb{E}\left[\|g(X, v)\|^p\right] \\ &\leq C\mathbb{E}\left[\|g(X, v) + h(Y, v)\|^p\right] \\ &= C\mathbb{E}\left[\|g(X, V) + h(Y, V)\|^p \mid V = v\right].\end{aligned}$$

Using iterated expectation, this shows

$$\begin{aligned}\mathbb{E}\left[\|g(X,V)\|^p\right] &= \mathbb{E}\left[\mathbb{E}\left[\|g(X,V)\|^p \mid V\right]\right] \\ &\leq \mathbb{E}\left[C\mathbb{E}\left[\|g(X,V)+h(Y,V)\|^p\right]\right] \\ &= C\mathbb{E}\left[\|g(X,V)+h(Y,V)\|^p\right].\end{aligned}$$

□

The following lemma is [BW17] Lemma 6.

Lemma 4.2.13. *Assume Assumptions A, B, C and D. Let*

$$(V_{n,1}(\mathbf{i})_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}),\ldots,(V_{n,K}(\mathbf{i})_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}),\ K\in\mathbb{N},$$

be independent copies of the same dependent multiplier field. Finally assume that the bandwidth $q = q_n$ fulfills $q_n \to \infty$ and $q_n = o(\sqrt{n})$.

Then, for any $r \geq 2$, there is a positive constant $B(d,r)$ such that for any finite subset $S \subset \mathbb{N}^d$ and any $i \in \{1, ..., K\}$

$$\frac{1}{B(d,r)}\mathbb{E}\left[\left\|\sum_{\mathbf{k}\in S^{(n)}}(X_{\mathbf{k}}-\mu)V_{n,i}(\mathbf{k})\right\|^r\right]$$
$$\leq \sum_{\mathbf{k}\in S^{(n)}}\mathbb{E}\left[\|X_{\mathbf{k}}-\mu\|^r\right]\mathbb{E}\left[\|V_{n,i}(\mathbf{k})\|^r\right]+\left(\sum_{\mathbf{k}\in S^{(n)}}\mathbb{E}\left[\|X_{\mathbf{k}}-\mu\|^2\right]\mathbb{E}\left[\|V_{n,i}(\mathbf{k})\|^2\right]\right)^{r/2}$$

where $S^{(n)} := S \cap \{1, ..., n\}^d$ and $\mu := \mathbb{E}\left[X_{\underline{\mathbf{0}}}\right]$.

If $r \in (2, 2+\delta]$, where δ is the constant from Assumptions C and D, then for any discrete block $U \subseteq \{1, ..., n\}^d$ and

$$M^\star(U) := \max_{W \lhd U}\left\|\sum_{\mathbf{j}\in W}(X_{\mathbf{j}}-\mu)V_{n,i}(\mathbf{j})\right\|,$$

it is true that

$$\mathbb{E}\left[M^\star(U)^r\right] \leq C(r)|U|^{\frac{r}{2}}$$

for some positive constant $C(r)$ that does not depend on U or n.

Proof. As noted in [BW17], this inequality follows in the same way as [Zha98] Theorem 2 and Lemma 4.2.1, using Lemma 4.2.12 instead of [Zha98] Theorem 1. □

The following result is [BW17] Lemma 7. The covariance estimator $\hat{\Sigma}_n(B)$ is the one proposed in [BH15] Chapter 3.

Lemma 4.2.14. *Assume the Assumptions A, B, C and D. Let $B \subseteq (0,1]^d$ be a finite union of disjoint blocks. For any index $\mathbf{a} \in \mathbb{Z}^d$, estimate the centered version of X in $\mathbf{a}$ via*

$$\tilde{X}(\mathbf{a}) := X_{\mathbf{a}} - \hat{\mu}(\mathbf{a}).$$

with $\hat{\mu}(\mathbf{a})$ defined as in (3.5). Let $k \in \mathbb{N}$ be fixed. Define the estimator

$$\hat{\Sigma}_n(B) := \frac{1}{n^d} \sum_{\mathbf{h} \in B_n \ominus B_n} \omega\left(\frac{\mathbf{h}}{q}\right) \sum_{\substack{\mathbf{a}: \\ \mathbf{a} \in B_n, \\ \mathbf{a}+\mathbf{h} \in B_n}} \tilde{X}(\mathbf{a})^{(k)} \left(\tilde{X}(\mathbf{a}+\mathbf{h})^{(k)}\right)^T$$

of the long-run variance matrix Σ of $X^{(k)} = P_k X$. Then

$$\hat{\Sigma}_n(B) \xrightarrow{\Pr} \lambda(B)\Sigma.$$

Proof. We refer to the proof presented in [BW17]. □

The following result is Proposition F.1 in [BK16].

Proposition 4.2.15. *For any integer $K \geq 1$, assume that there exists random variables $T_n, T^\star_{n,1}, ..., T^\star_{n,K}$ such that*

$$(T_n, T^\star_{n,1}, ..., T^\star_{n,K}) \xrightarrow{w} (T, T^\star_1, ..., T^\star_K),$$

where T, the weak limit of T_n, is a continuous random variable and $T^\star_1, ..., T^\star_K$ are independent copies of T. Then, for any significance level $\alpha \in (0,1)$,

$$\lim_{K\to\infty} \lim_{n\to\infty} \mathbb{P}(T_n \geq q^\star_{n,K}(1-\alpha)) = \alpha. \tag{4.20}$$

where $q^\star_{n,K}$ denotes the sample quantile function defined in (3.7).

Proof Idea. Firstly, one only considers the case that α is irrational as the empirical distribution functions $\hat{F}_{n,K}$ are not continuous in $\alpha = k/K$, $k = 1, ..., K$. Due to the right-continuity of empirical distribution functions, it suffices to show the weak convergence of $\hat{F}_{n,K}$ to to empirical distribution function $\hat{F}_K$ calculated from $T^\star_1, ..., T^\star_K$. This is accomplished using the continuous mapping theorem and the Portmanteau theorem. The convergence of $\hat{F}_K$ is shown using the Glivenko-Cantelli theorem. The case $\alpha \in \mathbb{Q}$ then follows by continuity. □

4.3 Proofs of the Main Results

4.3.1 Functional Central Limit Theorem

Proof of Theorem 3.2.1. We first show that the series

$$\sum_{\mathbf{v}\in\mathbb{Z}^d} \mathbb{E}\left[\left\langle X_{\underline{\mathbf{0}}} - \mu, x\right\rangle \langle X_{\mathbf{v}} - \mu, y\rangle\right], \ x, y \in H$$

defined in (3.8) converges absolutely. By replacing X with $X - \mu$, we may assume $\mu = 0$ without loss of generality. We may choose an orthonormal basis

$$E_{x,y} = E = (e_i)_{i \in I}$$

of H such that the set

$$J = \{i \in I \mid \langle x, e_i \rangle \neq 0 \vee \langle y, e_i \rangle \neq 0\}$$

is finite. By writing

$$\begin{aligned}
&\frac{1}{\|x\|\|y\|} \sum_{\mathbf{v} \in \mathbb{Z}^d} |\mathbb{E}\left[\langle X_{\underline{\mathbf{0}}}, x \rangle \langle X_{\mathbf{v}}, y \rangle\right]| \\
&\leq \frac{1}{\|x\|\|y\|} \sum_{\mathbf{v} \in \mathbb{Z}^d} \sum_{i \in J} \sum_{j \in J} |\mathbb{E}\left[\langle X_{\underline{\mathbf{0}}}, \langle x, e_i \rangle e_i \rangle \langle X_{\mathbf{v}}, \langle y, e_j \rangle e_j \rangle\right]| \\
&\leq \sum_{i \in J} \sum_{j \in J} \sum_{\mathbf{v} \in \mathbb{Z}^d} |\mathbb{E}\left[\langle X_{\underline{\mathbf{0}}}, e_i \rangle \langle X_{\mathbf{v}}, e_j \rangle\right]|,
\end{aligned}$$

we can see that it suffices to show that (3.8) converges absolutely for the case $x = e_i$, $y = e_j$. To this end, let $k \in \mathbb{N}$ be large enough such that $j \leq k$ for all $e_j \in J$ and define the $\mathbb{R}^k$-valued random field

$$X^{(k)} := (p^{(k)}(X_{\mathbf{j}}))_{\mathbf{j} \in \mathbb{Z}^d}$$

where

$$p^{(k)} : H \to \mathbb{R}^k, h \mapsto \begin{pmatrix} \langle h, e_1 \rangle \\ \dots \\ \langle h, e_k \rangle \end{pmatrix}.$$

Because of

$$\mathbb{E}\left[X_{\mathbf{j}}^{(k)}\right] = \mathbb{E}\left[\begin{pmatrix} \langle X_{\mathbf{j}}, e_1 \rangle \\ \dots \\ \langle X_{\mathbf{j}}, e_k \rangle \end{pmatrix}\right] = \begin{pmatrix} \langle \mathbb{E}\left[X_{\mathbf{j}}\right], e_1 \rangle \\ \dots \\ \langle \mathbb{E}\left[X_{\mathbf{j}}\right], e_k \rangle \end{pmatrix} = \begin{pmatrix} \langle 0, e_1 \rangle \\ \dots \\ \langle 0, e_k \rangle \end{pmatrix} = 0_{\mathbb{R}^k},$$

$X^{(k)}$ is centered. As a measurable transform of X, the random field $X^{(k)}$ also satisfies Assumptions B and D', see Remark 2.5.3. $X^{(k)}$ is also (both strictly and weakly) stationary, satisfying Assumption A': Let $\mathbf{t}_1, \dots, \mathbf{t}_k$ index time, $k \in \mathbb{N}$, let $\boldsymbol{\tau} \in \mathbb{Z}^d$ be a shift in time and let $B_1, \dots, B_k$ be arbitrary Borel-sets in $\mathbb{R}$. Then we have, due to the measurability of $p^{(k)}$ and the stationarity of X:

$$\begin{aligned}\mathbb{P}\left((X^{(k)}_{\mathbf{t}_1+\boldsymbol{\tau}}, ..., X^{(k)}_{\mathbf{t}_k+\boldsymbol{\tau}}) \in \prod_i^k B_i\right) &= \mathbb{P}\left((p^{(k)}(X_{\mathbf{t}_1+\boldsymbol{\tau}}), ..., p^{(k)}(X_{\mathbf{t}_k+\boldsymbol{\tau}})) \in \prod_i^k B_i\right) \\ &= \mathbb{P}\left((X_{\mathbf{t}_1+\boldsymbol{\tau}}, ..., X_{\mathbf{t}_k+\boldsymbol{\tau}}) \in \prod_i^k (p^{(k)})^{-1}(B_i)\right) \\ &= \mathbb{P}\left((X_{\mathbf{t}_1}, ..., X_{\mathbf{t}_k}) \in \prod_i^k (p^{(k)})^{-1}(B_i)\right) \\ &= \mathbb{P}\left((X^{(k)}_{\mathbf{t}_1}, ..., X^{(k)}_{\mathbf{t}_k}) \in \prod_i^k B_i\right).\end{aligned}$$

Lastly, $X^{(k)}$ has finite $(\delta+2)$-moments, in particular satisfying Assumption C", because of Cauchy-Schwarz and the fact that X has finite $(2+\delta)$-moments:

$$\begin{aligned}\mathbb{E}\left[\|X^{(k)}_{\mathbf{j}}\|^{2+\delta}\right] &= \mathbb{E}\left[\left(\sum_{i=1}^k \langle X_{\mathbf{j}}, e_i\rangle^2\right)^{\frac{2+\delta}{2}}\right] \\ &\leq \mathbb{E}\left[\left(\sum_{i\in I} \langle X_{\mathbf{j}}, e_i\rangle^2\right)^{\frac{2+\delta}{2}}\right] \\ &= \mathbb{E}\left[\|X_{\mathbf{j}}\|^{2+\delta}\right] \\ &< \infty.\end{aligned}$$

We have now shown all the conditions of Lemma 4.2.6. Hence we may conclude that $X^{(k)}$ satisfies the functional central limit theorem

$$\left(\frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\lfloor n\mathbf{t}\rfloor} X^{(k)}_{\mathbf{j}}\right)_{\mathbf{t}\in[0,1]^d} \Rightarrow W^{(k)}$$

in $D_{\mathbb{R}^k}[0,1]^d$ for a Brownian sheet $W^{(k)}$ in $\mathbb{R}^k$ with covariance operator

$$\begin{aligned}\Gamma^{(k)} &= \sum_{\mathbf{v}\in\mathbb{Z}^d} \operatorname{Cov}(X^{(k)}_{\underline{\mathbf{0}}}, X^{(k)}_{\mathbf{v}}) \\ &= \sum_{\mathbf{v}\in\mathbb{Z}^d} \left(\mathbb{E}\left[\langle X_{\underline{\mathbf{0}}}, e_i\rangle\langle X_{\mathbf{v}}, e_j\rangle\right]\right)_{i,j=1,...,k}\end{aligned}$$

and this series converges absolutely. As it sufficed to show the case $x = e_i$, $y = e_j$, we have shown that the series defined in (3.8) converges absolutely.

For the second part of the proof let $k \in \mathbb{N}$ be arbitrary. We will apply Lemma 4.2.10 in order to show that

$$P_k S_n \overset{n}{\Rightarrow} P_k W \overset{k}{\Rightarrow} W$$

implies our claim

$$S_n \overset{n}{\Rightarrow} W.$$

The following table shows the correspondence between the variable names of that lemma and the ones we will use in this proof:

Lemma 4.2.10	Theorem 3.2.1
K	1
$X_n = (X_{n,1}, ..., X_{n,K})$	$S_n(\mathbf{t})$
r	$2+\delta$
$X_i^k,\ i = 1, ..., K$	$W^{(k)}$
$X_i,\ i = 1, ..., K$	W
$S_i^k,\ i = 1, ..., K$	$\Gamma^{(k)}$
$S_i,\ i = 1, ..., K$	Γ

By identifying the two Hilbert spaces $H_k := P_k H$ and $\mathbb{R}^k$, we can reuse the FCLT of $X^{(k)}$ that we have just proven, showing Condition (4.16) of 4.2.10

$$S_n^{(k)} := P_k S_n \overset{n}{\Rightarrow} W^{(k)} = P_k W$$

in $D_{H_k}[0,1]^d$.

We now show Condition (4.17). Define $W^i := (\langle W_{\mathbf{t}}, e_i\rangle)_{\mathbf{t}\in[0,1]^d}$. Using Parseval's identity, we have

$$\sup_{\mathbf{t}\in[0,1]^d} \|W_{\mathbf{t}} - W_{\mathbf{t}}^{(k)}\|^2 = \sup_{\mathbf{t}\in[0,1]^d} \sum_{i\geq k+1} \langle W_{\mathbf{t}}, e_i\rangle^2 = \sup_{\mathbf{t}\in[0,1]^d} \sum_{i\geq k+1} W_{\mathbf{t}}^i$$

and therefore

$$\begin{aligned}\mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \|W_{\mathbf{t}} - W_{\mathbf{t}}^{(k)}\|^2\right] &= \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \sum_{i\geq k+1} (W_{\mathbf{t}}^i)^2\right] \\ &\leq \sum_{i\geq k+1} \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} (W_{\mathbf{t}}^i)^2\right].\end{aligned}$$

Since W^i is a Brownian sheet in $\mathbb{R}$, we may apply Cairoli's Strong $(2,2)$ Inequality, found in [Kho02] Theorem 2.3.2 in Chapter 7, (using Theorem 2.4.1 in that same chapter to see that W^i is a martingale with respect to its history) to the nonnegative submartingale $|W^i|$ to get

$$\begin{aligned}\sum_{i\geq k+1} \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} (W_{\mathbf{t}}^i)^2\right] &\leq \sum_{i\geq k+1} \left(\frac{2}{2-1}\right)^{2d} \mathbb{E}\left[(W_{\underline{\mathbf{1}}}^i)^2\right] \\ &= 4^d \sum_{i\geq k+1} \mathbb{E}\left[(W_{\underline{\mathbf{1}}}^i)^2\right].\end{aligned}$$

The fact that Γ is the covariance operator of W implies

$$\mathbb{E}\left[\langle W_{\underline{\mathbf{1}}}, e_i\rangle\langle W_{\underline{\mathbf{1}}}, e_i\rangle\right] = \langle \Gamma e_i, e_i\rangle.$$

Combining the above, we get

$$\mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \|W_{\mathbf{t}} - W_{\mathbf{t}}^{(k)}\|^2\right] \le 4^d \sum_{i\ge k+1} \langle \Gamma e_i, e_i\rangle.$$

As the series in (3.8) converges absolutely, the above right-hand side is finite and thus

$$\lim_{k\to\infty} \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \|W_{\mathbf{t}} - W_{\mathbf{t}}^{(k)}\|^2\right] = 0,$$

i.e.,

$$\sup_{\mathbf{t}\in[0,1]^d} \|W_{\mathbf{t}} - W_{\mathbf{t}}^{(k)}\|^2 \to 0$$

in probability and in particular $W^{(k)} \Rightarrow W$ in $D_H[0,1]^d$, showing (4.17).

Let $X_{\mathbf{j}}^{(-k)} := X_{\mathbf{j}} - X_{\mathbf{j}}^{(k)}$. We have

$$\begin{aligned}
\|X_{\mathbf{j}}^{(-k)}\| &= \|\sum_{i\ge k+1} \langle X_{\mathbf{j}}, e_i\rangle e_i\| \\
&= \sqrt{\sum_{i\ge k+1} \|\langle X_{\mathbf{j}}, e_i\rangle e_i\|^2} \\
&= \sqrt{\sum_{i\ge k+1} \langle X_{\mathbf{j}}, e_i\rangle^2} \\
&\to 0
\end{aligned}$$

almost surely and therefore, using the dominated convergence theorem with the dominating random variable $\|X_{\underline{\mathbf{1}}}\|$ which is integrable by Assumption C',

$$\max(\mathbb{E}\left[\|X_{\underline{\mathbf{1}}}^{(-k)}\|^{2+\delta}\right], \mathbb{E}\left[\|X_{\underline{\mathbf{1}}}^{(-k)}\|^{2}\right]) \to 0. \tag{4.21}$$

We can apply (4.5) in Lemma 4.2.1 to the random field $X^{(-k)}$ and the block $U = (\underline{\mathbf{0}}, \underline{\mathbf{n}}]$ to obtain that

$$\begin{aligned}
\mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \|S_n(\mathbf{t}) - S_n^{(k)}(\mathbf{t})\|^{2+\delta}\right] &= \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left\|\frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\le\mathbf{j}\le\lfloor n\mathbf{t}\rfloor} X_{\mathbf{j}}^{(-k)}\right\|^{2+\delta}\right] \\
&= \left(\frac{1}{n^{d/2}}\right)^{2+\delta} \mathbb{E}\left[\max_{W\lhd U} \left\|\sum_{\mathbf{j}\in W} X_{\mathbf{j}}^{(-k)}\right\|^{2+\delta}\right]
\end{aligned}$$

is bounded from above by

$$D_{d,2+\delta} B_{d,2+\delta,\rho_{X^{(-k)}}} \left(\sup_{\mathbf{j}\in\mathbb{N}^d} \mathbb{E}\left[\|X_{\mathbf{j}}^{(-k)}\|^{2+\delta} \right] + \left(\sup_{\mathbf{j}\in\mathbb{N}^d} \mathbb{E}\left[\|X_{\mathbf{j}}^{(-k)}\|^{2} \right] \right)^{\frac{2+\delta}{2}} \right).$$

As X, and therefore also $X^{(-k)}$, is stationary, we can leave out the suprema and see that the above equals

$$\begin{aligned} &D_{d,2+\delta} B_{d,2+\delta,\rho_{X^{(-k)}}} \left(\mathbb{E}\left[\|X_{\underline{\mathbf{1}}}^{(-k)}\|^{2+\delta} \right] + \left(\mathbb{E}\left[\|X_{\underline{\mathbf{1}}}^{(-k)}\|^{2} \right] \right)^{\frac{2+\delta}{2}} \right) \\ &\leq D_{d,2+\delta} B_{d,2+\delta,\rho_{X}} \left(\mathbb{E}\left[\|X_{\underline{\mathbf{1}}}^{(-k)}\|^{2+\delta} \right] + \left(\mathbb{E}\left[\|X_{\underline{\mathbf{1}}}^{(-k)}\|^{2} \right] \right)^{\frac{2+\delta}{2}} \right), \end{aligned}$$

which converges to 0 according to (4.21). Having now shown all conditions of Lemma 4.2.10, the proof is finished. □

4.3.2 Dependent Wild Bootstrap

Proof of Theorem 3.2.5. We will once more use Lemma 4.2.10 with the following correspondence of variables:

Lemma 4.2.10	Theorem 3.2.5
K	$K+1$
$X_n = (X_{n,1}, ..., X_{n,K})$	$(S_n, S_{n,1}^{\star}, ..., S_{n,K}^{\star})$
$P_k X_{n,1}, ..., P_k X_{n,K}$	$S_n^{(k)}, S_{n,1}^{\star,(k)}, ..., S_{n,K}^{\star,(k)}$
r	$2+\delta$
X_i^k, $i = 1, ..., K$	$W^{(k)}, W_1^{\star,(k)}, ..., W_K^{\star,(k)}$
X_i, $i = 1, ..., K$	$W, W_1^{\star}, ..., W_K^{\star}$
S_i, $i = 1, ..., K$	$\Gamma, ..., \Gamma$

To show Condition (i), i.e.,

$$(S_n^{(k)}, S_{n,1}^{\star,(k)}, ..., S_{n,K}^{\star,(k)}) \overset{n}{\Rightarrow} (W^{(k)}, W_1^{\star,(k)}, ..., W_K^{\star,(k)})$$

in $(D_{H_k}[0,1]^d)^{K+1}$ for all k, we will proceed as described in [BS07] Chapter 5 after Definition 1.2.: We show the tightness of the above left-hand sequence and the weak convergence of the finite-dimensional distributions. Note that the notions of sequential tightness as defined in [BS07] and tightness are equivalent as the process we are considering is not multiindexed.

To show the tightness of the process

$$(S_n^{(k)}, S_{n,1}^{\star,(k)}, ..., S_{n,K}^{\star,(k)})_n,$$

note that the tightness of $S_n^{(k)}$ was established as part of the proof of Theorem 3.2.1. To show the tightness of the above sequence, it therefore suffices to show the tightness of the process

$$(S_{n,1}^{\star,(k)}, ..., S_{n,K}^{\star,(k)})_n.$$

We will show this by proving that it satisfies Assumption (e) of Lemma 4.2.3 regarding its modulus of continuity. For any $j = 1, ..., K$ and $\mathbf{t} \in [0,1]^d$, we may write

$$\begin{aligned} S_{n,j}^{\star,(k)}(\mathbf{t}) &= \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} \left(X_{\mathbf{i}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{i}) \right) V_{n,j}(\mathbf{i}) \\ &= \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} \left(X_{\mathbf{i}}^{(k)} - \mu^{(k)} \right) V_{n,j}(\mathbf{i}) + \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} \left(\mu^{(k)} - \hat{\mu}^{(k)}(\mathbf{i}) \right) V_{n,j}(\mathbf{i}). \end{aligned}$$

Analogously to the end of the proof of Lemma 4.2.6, using Lemma 4.2.13 and [BS07] Theorem 1.3. in Chapter 5, it can be seen that the first summand

$$\left(\frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} \left(X_{\mathbf{i}}^{(k)} - \mu^{(k)} \right) V_{n,j}(\mathbf{i}) \right)_{n \in \mathbb{N}}$$

is tight. For the second summand

$$Y_n(\cdot) := \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \lfloor n\cdot \rfloor} \left(\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)} \right) V_{n,j}(\mathbf{i}),$$

we bound the modulus of continuity via

$$\begin{aligned} \mathbb{P}\left(\omega_{Y_n}(\eta) \geq \epsilon\right) &= \mathbb{P}\left(\sup_{\|\mathbf{t}-\mathbf{s}\|_2 < \eta} \|Y_n(\mathbf{t}) - Y_n(\mathbf{s})\| \geq \epsilon \right) \\ &\leq \mathbb{P}\left(\sup_{\|\mathbf{t}-\mathbf{s}\|_\infty < \eta} \|Y_n(\mathbf{t}) - Y_n(\mathbf{s})\| \geq \epsilon \right). \end{aligned}$$

Letting only one component vary between $\mathbf{t}$ and $\mathbf{s}$ but compensating with the factor d, this may be bounded from above by

$$\mathbb{P}\left(d \max_{h=1,...,d} \sup_{\substack{\mathbf{t} \in [0,1]^d \\ \mathbf{t}_h \leq 1-\eta \\ \gamma \in (0,\eta)}} \|Y_n(\mathbf{t}_1, ..., \mathbf{t}_{h-1}, \mathbf{t}_h + \gamma, \mathbf{t}_{h+1}, ..., \mathbf{t}_d) - Y_n(\mathbf{t})\| \geq \epsilon \right).$$

Using Boole's inequality, this probability is dominated by

$$\sum_{h=1}^{d} \mathbb{P}\left(d \sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} \|Y_n(\mathbf{t}_1,...,\mathbf{t}_{h-1},\mathbf{t}_h+\gamma,\mathbf{t}_{h+1},...,\mathbf{t}_d) - Y_n(\mathbf{t})\| \ge \epsilon \right)$$

$$= \sum_{h=1}^{d} \mathbb{P}\left(d \sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} \frac{1}{n^{d/2}} \left\| \sum_{\substack{\underline{\mathbf{1}}\le\mathbf{i}\le\lfloor n\mathbf{t}\rfloor \\ \lfloor n\mathbf{t}_h < \mathbf{i}_h \le \lfloor n(\mathbf{t}_h+\gamma)\rfloor\rfloor}} \left(\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right) V_{n,j}(\mathbf{i}) \right\| \ge \epsilon \right)$$

$$\le \sum_{h=1}^{d} \mathbb{P}\left(d \sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} \frac{1}{n^{d/2}} \sum_{\substack{\underline{\mathbf{1}}\le\mathbf{i}\le\lfloor n\mathbf{t}\rfloor \\ \lfloor n\mathbf{t}_h < \mathbf{i}_h \le \lfloor n(\mathbf{t}_h+\gamma)\rfloor\rfloor}} \left\|\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right\| |V_{n,j}(\mathbf{i})| \ge \epsilon \right)$$

$$\le \sum_{h=1}^{d} \mathbb{P}\left(\max_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}} n^{d/2} \left\|\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right\| \sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} \frac{1}{n^{d}} \sum_{\substack{\underline{\mathbf{1}}\le\mathbf{i}\le\lfloor n\mathbf{t}\rfloor \\ \lfloor n\mathbf{t}_h < \mathbf{i}_h \le \lfloor n(\mathbf{t}_h+\gamma)\rfloor\rfloor}} |V_{n,j}(\mathbf{i})| \ge \frac{\epsilon}{d} \right).$$

Due to the independence of the dependent multiplier field and the mean estimator $\hat{\mu}$, introducing a new variable $C > 0$, we may bound this from above by

$$d \cdot \mathbb{P}\left(\max_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}} n^{d/2} \left\|\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right\| > C \right)$$
$$+ \sum_{h=1}^{d} \mathbb{P}\left(\sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} \frac{1}{n^{d}} \sum_{\substack{\underline{\mathbf{1}}\le\mathbf{i}\le\lfloor n\mathbf{t}\rfloor \\ \lfloor n\mathbf{t}_h < \mathbf{i}_h \le \lfloor n(\mathbf{t}_h+\gamma)\rfloor\rfloor}} |V_{n,j}(\mathbf{i})| \ge \frac{\epsilon}{dC} \right).$$

Note that the term

$$\left\|\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right\|$$

only takes two values for each n. Applying Theorem 3.2.1 to each of the two subsets, we see that the term

$$d \cdot \mathbb{P}\left(\max_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}} n^{d/2} \left\|\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right\| > C \right)$$

approaches 0 uniformly in n as C grows to infinity. To show the convergence of the second term to 0, partition $(0,1]^d$ into the blocks

$$A_m(h,\eta) := (0,1]^{h-1} \times ((m-1)\eta, m\eta \wedge 1] \times (0,1]^{d-h}$$

for $m = 1, ..., p$ with $p \coloneqq p(\eta) \coloneqq \lfloor \eta^{-1} \rfloor + 1$ and let

$$Z_n(\cdot) \coloneqq \frac{1}{n^d} \sum_{\underline{\mathbf{1}} \le \mathbf{i} \le \lfloor n \cdot \rfloor} |V_{n,j}(\mathbf{i})|.$$

Then

$$\begin{aligned}
&\sum_{h=1}^{d} \mathbb{P}\left(\sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} \frac{1}{n^d} \sum_{\substack{\underline{\mathbf{1}} \le \mathbf{i} \le \lfloor n\mathbf{t} \rfloor \\ \lfloor n\mathbf{t}_h < \mathbf{i}_h \le \lfloor n(\mathbf{t}_h+\gamma) \rfloor \rfloor}} |V_{n,j}(\mathbf{i})| \ge \frac{\epsilon}{dC} \right) \\
&= \sum_{h=1}^{d} \mathbb{P}\left(\sup_{\substack{\mathbf{t}\in[0,1]^d \\ \mathbf{t}_h \le 1-\eta \\ \gamma\in(0,\eta)}} |Z_n(\mathbf{t}_1, ..., \mathbf{t}_{h-1}, \mathbf{t}_h + \gamma, \mathbf{t}_{h+1}, ..., \mathbf{t}_d) - Z_n(\mathbf{t})| \ge \frac{\epsilon}{dC} \right) \\
&\le \sum_{h=1}^{d} \sum_{m=1}^{p} \mathbb{P}\left(\sup_{\substack{\mathbf{s},\mathbf{t}\in A_m(h,\eta) \\ \mathbf{s}_a=\mathbf{t}_a \text{ for } a\neq h}} |Z_n(\mathbf{s}) - Z_n(\mathbf{t})| \ge \frac{\epsilon}{2dC} \right)
\end{aligned}$$

where the 2 in the denominator compensates for the fact that we are no longer allowing the supremum to go over intervals that contain some $m\eta$. Let

$$A_{m,n}(h,\eta) \coloneqq nA_m(h,\eta) \cap \mathbb{Z}^d \subset \{1, ..., n\}^d.$$

We continue to estimate from above via

$$\begin{aligned}
&\sum_{h=1}^{d} \sum_{m=1}^{p} \mathbb{P}\left(\sup_{\substack{\mathbf{s},\mathbf{t}\in A_m(h,\eta) \\ \mathbf{s}_a=\mathbf{t}_a \text{ for } a\neq h}} |Z_n(\mathbf{s}) - Z_n(\mathbf{t})| \ge \frac{\epsilon}{2dC} \right) \\
&\le \sum_{h=1}^{d} \sum_{m=1}^{p} \mathbb{P}\left(\frac{1}{n^d} \sum_{\mathbf{i}\in A_{m,n}(h,\eta)} |V_{n,j}(\mathbf{i})| \ge \frac{\epsilon}{4dC} \right) \\
&= \sum_{h=1}^{d} \sum_{m=1}^{p} \mathbb{P}\left(\left| \frac{1}{n^d} \sum_{\mathbf{i}\in A_{m,n}(h,\eta)} |V_{n,j}(\mathbf{i})| \right|^r \ge \left(\frac{\epsilon}{4dC} \right)^r \right).
\end{aligned}$$

with $r \ge 2$. Using Markov's inequality, this is bounded from above by

$$\sum_{h=1}^{d} \sum_{m=1}^{p} \mathbb{E}\left[\left| \frac{1}{n^d} \sum_{\mathbf{i}\in A_{m,n}(h,\eta)} |V_{n,j}(\mathbf{i})| \right|^r \right] \frac{4^r d^r C^r}{\epsilon^r}.$$

Due to the Gaussian distribution of the dependent multiplier field, there is a constant $C_r > 0$ such that this is bounded from above by

$$
\begin{aligned}
\sum_{h=1}^{d} \frac{1}{n^{dr}} \sum_{m=1}^{p} C_r |A_{m,n}(h,\eta)|^r \frac{4^r d^r C^r}{\epsilon^r} &\leq \sum_{h=1}^{d} \frac{1}{n^{dr}} \sum_{m=1}^{p} C_r (n^d \eta)^r \frac{4^r d^r C^r}{\epsilon^r} \\
&= d(1 + \lfloor \eta^{-1} \rfloor) \frac{1}{n^{dr}} C_r n^{dr} \eta^r \frac{4^r d^r C^r}{\epsilon^r} \\
&\leq d(\delta + 1) C_r \eta^{r-1} \frac{4^r d^r C^r}{\epsilon^r} \\
&\xrightarrow{\eta \to 0} 0
\end{aligned}
$$

where we have used $|A_{m,n}(h,\eta)| \leq n^d \eta$. Hence we have finally shown that the processes $\left(S_{n,j}^{\star,(k)}\right)_n$, $j = 1, ..., K$, are tight. Therefore, the processes

$$
\left(S_{n,1}^{\star,(k)}, ..., S_{n,K}^{\star,(k)}\right)_n
$$

is tight as well.

Next, we show the convergence of the finite-dimensional distributions. Using Prokhorov's theorem, the tightness of the sequence

$$
\left(S_{n,1}^{\star,(k)}, ..., S_{n,K}^{\star,(k)}\right)_n
$$

implies that it is sequentially compact with respect to the topology of weak convergence, i.e., any subsequence has a further subsequence that converges weakly. We only need to show that this further subsequence converges to the correct limit. Let $(n_m)_{m\in\mathbb{N}}$ index some subsequence and let $(n_m)_{m\in M}$ be a further subsequence with an infinite subset $M \subset \mathbb{N}$ such that all finite-dimensional distributions of the form

$$
\mathbf{W}_{m,j}^{\star,(k)} := \left(S_{n_m,j}^{\star,(k)}(B_1), ..., S_{n_m,j}^{\star,(k)}(B_l)\right)
$$

converge weakly to some limit for all indices $j = 1, ..., K$ and collections of disjoint blocks $B_1, ..., B_l$ with rational corners (i.e., with corners in $[0,1]^d \cap \mathbb{Q}^d$). We may choose such subsequence by a diagonal sequence argument as we only consider countably many blocks. We need to show that the above converges to the following "correct" limit:

$$
\mathbf{W}_{j}^{\star,(k)} := \left(W_{j}^{\star,(k)}(B_1), ..., W_{j}^{\star,(k)}(B_l)\right)
$$

Note that the $\mathbf{W}_{m,j}^{\star,(k)}$, $j = 1, ..., K$, are centered since the dependent multiplier field is centered and it is independent of the random field X. Furthermore, conditioned on $X_{\mathbf{i}}$, $\underline{\mathbf{1}} \leq \mathbf{i} \leq \underline{\mathbf{n}}$, the $\mathbf{W}_{m,j}^{\star,(k)}$ are independent and have a Gaussian distribution.

Hence the weak limit has independent components and is Gaussian as well. Therefore, to characterize the distribution of the limit, it suffices to show the convergence of the covariance operators

$$\lim_{n\to\infty} \operatorname{Cov}\left(S_{n,j}^{\star,(k)}(B_{l_1}), S_{n,j}^{\star,(k)}(B_{l_2})\right) = \begin{cases} \lambda(B_{l_1})\Gamma & l_1 = l_2 \\ 0 & l_1 \neq l_2 \end{cases}, \tag{4.22}$$

$l_1, l_2, = 1, ..., l$, where Γ is the long-run variance of X. To do this, we will apply the law of total covariance to the (p, q)-components, $p, q = 1, ..., k$. We simplify the notation and write

$$U_s := S_{n,j}^{\star,(k)}(B_{l_s}),\ s = 1, 2,$$

for the increments of the bootstrapped, projected partial sum fields and write

$$\mathcal{A}_n := \sigma\left(X_{\mathbf{i}}, \underline{\mathbf{1}} \leq \mathbf{i} \leq \underline{\mathbf{n}}\right)$$

for the σ-algebra that contains all information on X until time n. Then we have by the law of total covariance

$$\operatorname{Cov}\left(U_1, U_2\right)_{p,q} = \mathbb{E}\left[\operatorname{Cov}(U_1, U_2 \mid \mathcal{A}_n)_{p,q} + \operatorname{Cov}(\mathbb{E}\left[U_1 \mid \mathcal{A}_n\right], \mathbb{E}\left[U_2 \mid \mathcal{A}_n\right])_{p,q}\right].$$

The U_s are linear combinations of elements of the form

$$\left(X_{\mathbf{i}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{i})\right) V_{n,j}(\mathbf{i})$$

of which the first factor is $\mathcal{A}_n$-measurable and the second factor is independent of $\mathcal{A}_n$ by definition of the dependent multiplier field. Hence the conditional expectations $\mathbb{E}\left[U_s \mid \mathcal{A}_n\right]$ as linear combinations of elements of the form

$$\begin{aligned}
\mathbb{E}\left[\left(X_{\mathbf{i}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{i})\right) V_{n,j}(\mathbf{i}) \mid \mathcal{A}_n\right] &= \left(X_{\mathbf{i}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{i})\right) \mathbb{E}\left[V_{n,j}(\mathbf{i}) \mid \mathcal{A}_n\right] \\
&= \left(X_{\mathbf{i}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{i})\right) \mathbb{E}\left[V_{n,j}(\mathbf{i})\right] \\
&= 0
\end{aligned}$$

vanish since dependent multiplier fields are centered. Therefore we can write the covariance operators as

$$\operatorname{Cov}\left(U_1, U_2\right) = \mathbb{E}\left[\operatorname{Cov}(U_1, U_2 \mid \mathcal{A}_n)\right].$$

We will thus examine the asymptotic behaviour of

$$\begin{aligned}
\operatorname{Cov}(U_1, U_2 \mid \mathcal{A}_n) &= \mathbb{E}\left[(U_1 - \mathbb{E}\left[U_1 \mid \mathcal{A}_n\right])(U_2 - \mathbb{E}\left[U_2 \mid \mathcal{A}_n\right])^T \mid \mathcal{A}_n\right] \\
&= \mathbb{E}\left[U_1 U_2^T \mid \mathcal{A}_n\right].
\end{aligned}$$

We remember the definition

$$B_n := nB \cap \mathbb{Z}^d$$

for a block $B \subset [0,1]^d$. Using Equation (2.1), we may write

$$\mathbb{E}\left[U_1 U_2^T \mid \mathcal{A}_n\right] = \mathbb{E}\left[S_{n,j}^{\star,(k)}(B_{l_1}) S_{n,j}^{\star,(k)}(B_{l_2})^T \mid \mathcal{A}_n\right]$$

as

$$\frac{1}{n^d} \sum_{\mathbf{a} \in B_{l_1,n}} \sum_{\mathbf{b} \in B_{l_2,n}} \mathbb{E}\left[\left(\left(X_{\mathbf{a}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{a})\right) V_{n,j}(\mathbf{a})\right) \left(\left(X_{\mathbf{b}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{b})\right) V_{n,j}(\mathbf{b})\right)^T \mid \mathcal{A}_n\right].$$

Once more using the fact that the factors $X_{\mathbf{i}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{i})$ are $\mathcal{A}_n$-measurable and that $V_{n,j}(\mathbf{b})$ is independent of $\mathcal{A}_n$, the above equals

$$\frac{1}{n^d} \sum_{\mathbf{a} \in B_{l_1,n}} \sum_{\mathbf{b} \in B_{l_2,n}} \left(X_{\mathbf{a}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{a})\right) \left(X_{\mathbf{b}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{b})\right)^T \mathbb{E}\left[V_{n,j}(\mathbf{a}) V_{n,j}(\mathbf{b})\right].$$

By Assumption (2.16) of the dependent multiplier field, we have

$$\mathbb{E}\left[V_{n,j}(\mathbf{a}) V_{n,j}(\mathbf{b})\right] = \omega\left(\frac{1}{q_n}(\mathbf{a} - \mathbf{b})\right).$$

Using this together with Lemma 2.2.9, we deduce that $\mathbb{E}\left[U_1 U_2^T \mid \mathcal{A}_n\right]$ equals

$$\sum_{\mathbf{h} \in B_{l_2,n} \ominus B_{l_1,n}} \omega\left(\frac{\mathbf{h}}{q_n}\right) \frac{1}{n^d} \sum_{\substack{\mathbf{a}: \\ \mathbf{a} \in B_{l_1,n}, \\ \mathbf{a}+\mathbf{h} \in B_{l_2,n}}} \left(X_{\mathbf{a}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{a})\right) \left(X_{\mathbf{a}+\mathbf{h}}^{(k)} - \hat{\mu}^{(k)}(\mathbf{a}+\mathbf{h})\right)^T.$$

For $l_1 = l_2$, we may apply Lemma 4.2.14 to conclude that the above converges to $\lambda(B_{l_1})\Gamma$ in probability. We conclude

$$\begin{aligned}
\lim_{n\to\infty} \operatorname{Cov}\left(S_{n,j}^{\star,(k)}(B_{l_1})\right) &= \lim_{n\to\infty} \mathbb{E}\left[\operatorname{Cov}\left(S_{n,j}^{\star,(k)}(B_{l_1}) \mid \mathcal{A}_n\right)\right] \\
&= \mathbb{E}\left[\lim_{n\to\infty} \operatorname{Cov}\left(S_{n,j}^{\star,(k)}(B_{l_1}) \mid \mathcal{A}_n\right)\right] \\
&= \mathbb{E}\left[\lambda(B_{l_1})\Gamma\right] \\
&= \lambda(B_{l_1})\Gamma.
\end{aligned}$$

For $l_1 \neq l_2$, from

$$\operatorname{Cov}(U_1 + U_2 \mid \mathcal{A}_n) = \operatorname{Cov}\left(U_1 \mid \mathcal{A}_n\right) + 2\operatorname{Cov}\left(U_1, U_2 \mid \mathcal{A}_n\right) + \operatorname{Cov}\left(U_2 \mid \mathcal{A}_n\right)$$

it follows that

$$\operatorname{Cov}\left(U_1, U_2 \mid \mathcal{A}_n\right) = \frac{1}{2}\left(\operatorname{Cov}\left(U_1 \mid \mathcal{A}_n\right) + \operatorname{Cov}\left(U_2 \mid \mathcal{A}_n\right) - \operatorname{Cov}(U_1 + U_2 \mid \mathcal{A}_n)\right).$$

We have calculated that the first two terms respectively converge to $\lambda(B_{l_1})\Gamma$ and $\lambda(B_{l_2})\Gamma$ in probability. Applying Lemma 4.2.14, it follows that the third term converges to

$$\lambda(B_{l_1}\dot\cup B_{l_2})\Gamma = (\lambda(B_{l_1}) + \lambda(B_{l_2}))\Gamma$$

in probability. Hence

$$\operatorname{Cov}\left(U_1, U_2 \mid \mathcal{A}_n\right) = \operatorname{Cov}\left(S_{n,j}^{\star,(k)}(B_{l_1}), S_{n,j}^{\star,(k)}(B_{l_2}) \mid \mathcal{A}_n\right)$$

converges to 0 in probability for $l_1 \neq l_2$. With the same argument as before, it follows that

$$\lim_{n\to\infty} \operatorname{Cov}\left(S_{n,j}^{\star,(k)}(B_{l_1}), S_{n,j}^{\star,(k)}(B_{l_2})\right) = 0.$$

This means that we have shown (4.22) and thus finished to proof of Condition (i) of Lemma 4.2.10.

Lemma 4.2.10 Condition (ii)

$$(W^{(k)}, W_1^{\star,(k)}, ..., W_K^{\star,(k)}) \overset{k}{\Rightarrow} (W, W_1^\star, ..., W_K^\star)$$

in $(D_H[0,1]^d)^{K+1}$ follows from the proof of Theorem 3.2.1: The weak convergence $W^{(k)} \Rightarrow W$ was shown directly and $W_1^{\star,(k)}, ..., W_K^{\star,(k)}$ and $W_1^\star, ..., W_K^\star$ are independent copies of $W^{(k)}$ and W respectively.

It only remains to prove Condition (iii). Let $r \coloneqq 2 + \delta$. For two positive numbers a and b, applying the triangle inequality and Jensen's inequality to the convex function $\cdot^r$ (with domain $\mathbb{R}_{\geq 0}$) gives

$$\sqrt{a^2 + b^2}^r \leq (a+b)^r = 2^r\left(\frac{a+b}{2}\right)^r \leq 2^r\frac{a^r + b^r}{2} = 2^{r-1}(a^r + b^r).$$

Applying the above K times, we get

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathbf{s},\mathbf{t}_1,...,\mathbf{t}_K\in[0,1]^d}\left\|\begin{pmatrix} S_n(\mathbf{s}) \\ S_{n,1}^\star(\mathbf{t}_1) \\ \cdots \\ S_{n,K}^\star(\mathbf{t}_K)\end{pmatrix} - \begin{pmatrix} S_n^{(k)}(\mathbf{s}) \\ S_{n,1}^{\star,(k)}(\mathbf{t}_1) \\ \cdots \\ S_{n,K}^{\star,(k)}(\mathbf{t}_K)\end{pmatrix}\right\|^r\right] \\
&\leq 2^{r-1}\mathbb{E}\left[\sup_{\mathbf{s},\mathbf{t}_1,...,\mathbf{t}_K\in[0,1]^d}\left\|S_n(\mathbf{s}) - S_n^{(k)}(\mathbf{s})\right\|^r + \left\|\begin{pmatrix} S_{n,1}^\star(\mathbf{t}_1) \\ \cdots \\ S_{n,K}^\star(\mathbf{t}_K)\end{pmatrix} - \begin{pmatrix} S_{n,1}^{\star,(k)}(\mathbf{t}_1) \\ \cdots \\ S_{n,K}^{\star,(k)}(\mathbf{t}_K)\end{pmatrix}\right\|^r\right] \\
&\leq 2^{r-1}\mathbb{E}\left[\sup_{\mathbf{s}\in[0,1]^d}\left\|S_n(\mathbf{s}) - S_n^{(k)}(\mathbf{s})\right\|^r\right] + 2^{K(r-1)}\sum_{j=1}^K \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d}\left\|S_{n,j}^\star(\mathbf{t}) - S_{n,j}^{\star,(k)}(\mathbf{t})\right\|^r\right].
\end{aligned}$$

We have shown in the proof of Theorem 3.2.1 that the first time converges to 0 for $k \to \infty$. Continuing the notation $h^{(-k)} := h - h^{(k)}$ for a vector $h \in H$ from the proof of Theorem 3.2.1, the second term is equal to

$$
\begin{aligned}
&2^{K(r-1)} \sum_{j=1}^{K} \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left\| \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} V_{n,j}(\mathbf{i}) \left(X_{\mathbf{i}}^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right)\right\|^r\right] \\
&\leq 2^{(K+1)(r-1)} \sum_{j=1}^{K} \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left\| \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} V_{n,j}(\mathbf{i}) \left(X_{\mathbf{i}}^{(-k)} - \mu^{(-k)}\right)\right\|^r\right] \\
&+ 2^{(K+1)(r-1)} \sum_{j=1}^{K} \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left\| \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} V_{n,j}(\mathbf{i}) \left(\mu^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right)\right\|^r\right].
\end{aligned}
$$

The convergence to 0 of the above first term on the right-hand side of the inequality follows in the same way as the one of the non-bootstrapped version when one replaced Lemma 4.2.1 with Lemma 4.2.13 in the proof of Theorem 3.2.1. To show the convergence of the second term, consider

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left\| \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} V_{n,j}(\mathbf{i}) \left(\mu^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right)\right\|^r\right] \\
&\leq \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left(\frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} \left\| V_{n,j}(\mathbf{i}) \left(\mu^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right)\right\| \right)^r\right] \\
&\leq \mathbb{E}\left[\sup_{\mathbf{t}\in[0,1]^d} \left(\max_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} \left\| \mu^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right\| \right)^r \left(\frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor} |V_{n,j}(\mathbf{i})| \right)^r\right] \\
&\leq \mathbb{E}\left[n^{rd/2} \max_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \underline{\mathbf{n}}} \left\| \mu^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right\|^r \left(\frac{1}{n^{d}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \underline{\mathbf{n}}} |V_{n,j}(\mathbf{i})| \right)^r\right].
\end{aligned}
$$

Due to independence, this is equal to

$$
\mathbb{E}\left[n^{rd/2} \max_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \underline{\mathbf{n}}} \left\| \mu^{(-k)} - \hat{\mu}^{(-k)}(\mathbf{i})\right\|^r \right] \mathbb{E}\left[\left(\frac{1}{n^{d}} \sum_{\underline{\mathbf{1}}\leq \mathbf{i} \leq \underline{\mathbf{n}}} |V_{n,j}(\mathbf{i})| \right)^r\right].
$$

As the second factor is dominated by

$$
\frac{1}{n^{dr}} C_r n^{dr} = C_r
$$

for some constant C_r, it is bounded uniformly in n and k. Since the sizes of the two sets on which the mean estimator $\hat{\mu}$ takes different values are both asymptotically proportional to n^d (and we find a subblock in both sets whose size is proportional to n^d as well), we see that for sufficiently large n, the first factor is dominated by

$$
\begin{aligned}
&C\mathbb{E}\left[n^{rd/2} \max_{\underline{\mathbf{1}}\leq \mathbf{l}<\mathbf{u}\leq \underline{\mathbf{n}}} \left\| \frac{1}{n^d} \sum_{\mathbf{l}<\mathbf{i}\leq \mathbf{u}} \left(X_{\mathbf{i}}^{(-k)} - \mu^{(-k)}\right) \right\|^r\right] \\
&= C\mathbb{E}\left[\frac{1}{n^{rd/2}} \max_{\underline{\mathbf{1}}\leq \mathbf{l}<\mathbf{u}\leq \underline{\mathbf{n}}} \left\| \sum_{\mathbf{l}<\mathbf{i}\leq \mathbf{u}} \left(X_{\mathbf{i}}^{(-k)} - \mu^{(-k)}\right) \right\|^r\right]
\end{aligned}
$$

for some constant C. By Lemma 4.2.1, the above is uniformly bounded in n and approaches 0 as $k \to \infty$, see also the proof of Theorem 3.2.1. Condition (iii) is now shown and the proof is completed. $\square$

Appendix A

Tensor Products

This chapter serves to give a brief introduction to tensor products which are needed in order to define the covariance of Hilbert space valued random elements. It is mainly based on [Rya02], [KR97] and [Wei80].

Definition A.0.1 (Tensor product of vector spaces)**.** Let V and W be two (possibly infinite-dimensional) $\mathbb{R}$-vector spaces. Let $F(V, W)$ be the $\mathbb{R}$-vector space spanned by $V \times W$, i.e., every pair $(v, w) \in V \times W$ is a basis element of $F(V, W)$. Furthermore, consider the linear subspace N of $F(V, W)$ spanned by elements of the form

(i) $(v_1 + v_2, w) - (v_1, w) - (v_2, w)$,

(ii) $(v, w_1 + w_2) - (v, w_1) - (v, w_2)$,

(iii) $(\lambda v, w) - \lambda(v, w)$,

(iv) $(v, \lambda w) - \lambda(v, w)$

with vectors $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$ and a scalar $\lambda \in \mathbb{R}$. The *(algebraic) tensor product* $V \otimes W$ is constructed as the quotient vector space $F(V, W)/N$. The image of the projection of (v, w) onto $V \otimes W$ is denoted $v \otimes w$. We call elements of the form $v \otimes w$ *simple tensors*.

Remark A.0.2. a) Every element of a tensor product space is a linear combination of simple tensors.

b) The tensor product is constructed so that $\otimes : V \times W \to V \otimes W$ is bilinear in the sense that

(i) $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$,

(ii) $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$,

(iii) $(\lambda v) \otimes w = \lambda(v \otimes w) = v \otimes (\lambda w)$

for $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, $\lambda \in \mathbb{R}$.

c) The dimension of the tensor product $V \otimes W$ is the product of the dimensions of V and W as cardinal numbers. In particular, if V and W are separable Hilbert spaces, then $V \otimes W$ is also separable.

Tensor products of Hilbert spaces themselves carry a Hilbert space structure. The following construction is taken from [Wei80] Chapter 3.4.

Definition A.0.3 (Tensor product of Hilbert spaces)**.** Let H_1 and H_2 be two Hilbert spaces with respective inner products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$. Then the inner product $\langle\cdot,\cdot\rangle_{H_1\otimes H_2} : H_1 \otimes H_2 \times H_1 \otimes H_2 \to \mathbb{R}$ defined via

$$\left\langle \sum_{i=1}^{n} c_i f_i \otimes g_i, \sum_{j=1}^{m} c_j' f_j' \otimes g_j' \right\rangle_{H_1\otimes H_2} := \sum_{i=1}^{n}\sum_{j=1}^{m} c_i c_j' \langle f_i, f_j'\rangle_1 \langle g_i, g_j'\rangle_2.$$

turns the algebraic tensor product $H_1 \otimes H_2$ into a pre-Hilbert space. The completion of $H_1 \otimes H_2$ under the metric induced by $\langle\cdot,\cdot\rangle_{H_1\otimes H_2}$, denoted $H_1 \hat{\otimes} H_2$, is called the *(complete) tensor product* of the Hilbert spaces H_1 and H_2.

Definition A.0.4 (Weak Hilbert-Schmidt mapping)**.** Let H_1, H_2 and H be Hilbert spaces. A *weak Hilbert-Schmidt mapping* is a bounded bilinear map

$$L : H_1 \times H_2 \to H$$

for which a real number $\alpha \in \mathbb{R}$ exists such that

$$\sum_{i\in I}\sum_{j\in J} |\langle L(e_i, f_j), h\rangle|^2 \leq \alpha^2 \|h\|^2 \tag{A.1}$$

for all $h \in H$ and some orthonormal basis $(e_i)_{i\in I}$ of H_1 and $(f_j)_{j\in J}$ of H_2.

The following theorem is due to [KR97] Theorem 2.6.4.

Theorem A.0.5 (Universal property of the tensor product)**.** *There is a weak Hilbert-Schmidt mapping*

$$\hat{\otimes} : H_1 \times H_2 \to H_1 \hat{\otimes} H_2$$

such that for any weak Hilbert-Schmidt mapping

$$L : H_1 \times H_2 \to H$$

to a Hilbert space H, there is a unique bounded linear operator

$$T : H_1 \hat{\otimes} H_2 \to H$$

such that the following diagram commutes:

$$\begin{tikzcd} H_1 \times H_2 \arrow[r, "\hat{\otimes}"] \arrow[dr, "L"'] & H_1 \hat{\otimes} H_2 \arrow[d, "T"] \\ & H \end{tikzcd}$$

Moreover,

$$\|T\|_{\text{op}} = \|L\|_2 \tag{A.2}$$

with

$$\|L\|_2 := \inf\{\alpha \geq 0 \mid \sqrt{\sum_{i\in I}\sum_{j\in J} |\langle L(e_i, f_j), u\rangle|^2} \leq \alpha \|u\| \ \forall u \in H\}$$

for orthonormal basis $(e_i)_{i\in I}$ of H_1 and $(f_j)_{f\in J}$ of H_2. If $H = \mathbb{R}$, this simplifies to

$$\|L\|_2 = \sqrt{\sum_{i\in I}\sum_{j\in J} L(e_i, f_j)^2}.$$

Remark A.0.6. The construction of the tensor product $H_1\hat{\otimes}H_2$ in the proof of [KR97] Theorem 2.6.4 is not the one we presented in Definition A.0.3. However, the equivalence of the two constructions is remarked on page 125 in [KR97]. See also [Wei80] Exercise 7.15.(b) or [Tre06] Theorem 48.3, which state that one may identify $H_1'\hat{\otimes}H_2$ with the space $B_2(H_1, H_2)$ of Hilbert-Schmidt operators from H_1 into H_2. H_1' here denotes the dual space of H_1.

Example A.0.7. The algebraic tensor product $\mathbb{R}^k \otimes \mathbb{R}^l$ is isomorphic to the vector space $\mathbb{R}^{k\times l}$ of $k \times l$-matrices. Isomorphisms are given via (the linear continuation of)

$$\varphi : \mathbb{R}^k \otimes \mathbb{R}^l \to \mathbb{R}^{k\times l}, a \otimes b \mapsto (a_i b_j)_{i,j}$$

and

$$\varphi^- : \mathbb{R}^{k\times l} \to \mathbb{R}^k \otimes \mathbb{R}^l, (x_{ij})_{i,j} \mapsto \sum_{i=1}^{k}\sum_{j=1}^{l} x_{ij}(e_i \otimes e_j),$$

see [Rya02] chapter 1.5.

As $\mathbb{R}^{k\times l}$ is complete, the algebraic tensor product $\mathbb{R}^k \otimes \mathbb{R}^l$ and complete tensor product $\mathbb{R}^k\hat{\otimes}\mathbb{R}^l$ are equal up to isomorphism.

By vectorizing matrices, we can calculate their inner product: Let $n = m = kl$, $f_i = f_i' = e_{(i \bmod k)+1}$ and $g_i = g_i' = e_{(i \bmod l)+1}$. Then we have

$$\{(f_i, g_i) \mid i = 1, ..., n\} = \{(e_i, e_j) \mid i = 1, ..., k, j = 1, ..., l\}.$$

If we identify matrices $A = (a_{i,j}), B = (b_{i,j}) \in \mathbb{R}^{k\times l}$ with their images under φ^{-}, their inner product $\langle A, B\rangle$ is

$$\begin{aligned}\langle A, B\rangle &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{(i \bmod k)+1,(i \bmod l)+1} b_{(j \bmod k)+1,(j \bmod l)+1} \langle e_i, e_j\rangle\langle e_i, e_j\rangle \\ &= \sum_{i=1}^{n} a_{(i \bmod k)+1,(i \bmod l)+1} b_{(j \bmod k)+1,(j \bmod l)+1} \\ &= \sum_{i=1}^{k}\sum_{j=1}^{l} a_{i,j}b_{i,j}.\end{aligned}$$

This inner product therefore coincides with the Frobenius inner product, which is an example of a so-called Hilbert-Schmidt inner product.

List of Symbols

α	α-mixing coefficient 22
$\ominus$	dash-product of sets 12
d	dimension of the index set 7
$\delta_{i,j}$	Kronecker delta 9
$\lfloor \cdot \rfloor$	Gaussian brackets $\lfloor x \rfloor = \max\{j \in \mathbb{Z} \mid j \leq x\}$ 8
Γ	Long-run variance 19
γ	auto-covariance function 19
H	separable Hilbert space 8
$\hat{\otimes}$	(complete) tensor product II
λ	Lebesgue measure 18
$\lhd$	belonging standardly to 9
$\mathcal{L}(V, W)$	set of bounded linear operators $V \to W$ 8
$[x]$	$\prod_i \mathbf{x}_i$ 8
$\|\varphi\|_{\mathrm{op}}$	operator norm of φ 12
$\otimes$	(algebraic) tensor product I
ρ	ρ-mixing coefficient 22
$\underline{\mathbf{x}}$	vector containing x in each of its components 8
w	modulus of continuity 21
$\Rightarrow$	convergence in distribution, weak convergence 9

Bibliography

[BB85] Richard C Bradley and Wlodzimierz Bryc. Multilinear forms and measures of dependence between random variables. *Journal of Multivariate Analysis*, 16(3):335–367, 1985.

[BD91] Peter J Brockwell and Richard A Davis. *Time Series: Theory and Methods*. Springer science & business media, second edition, 1991.

[BH15] Béatrice Bucchia and Christoph Heuser. Long-run variance estimation for spatial data under change-point alternatives. *Journal of Statistical Planning and Inference*, 165:104–126, 2015.

[Bil68] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, first edition, 1968.

[Bil99] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, second edition, 1999.

[Bis06] Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[BK16] Axel Bücher and Ivan Kojadinovic. A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*, 22(2):927–968, 2016.

[Boc33a] Salomon Bochner. Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind. *Fundamenta Mathematicae*, 20(1):262–176, 1933.

[Boc33b] Salomon Bochner. Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. *Mathematische Annalen*, 108(1):378–410, 1933.

[Bra86] Richard C. Bradley. Basic Properties of Strong Mixing Conditions. In *Dependence in Probability and Statistics*, pages 165–192. Springer, 1986.

[Bra01] Richard C Bradley. A Stationary Rho-Mixing Markov Chain Which Is Not "Interlaced" Rho-Mixing. *Journal of Theoretical Probability*, 14:717–727, 2001.

[Bra17] Vladislav V. Branishti. On some properties of weighted Hilbert spaces. *Journal of Siberian Federal University. Mathematics & Physics.*, 10(4):410 – 421, 2017.

[BS07] Alexander Bulinski and Alexey Shashkin. *Limit Theorems for Associated Random Fields and Related Systems*. WORLD SCIENTIFIC, 2007.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[BW71] P. J. Bickel and M. J. Wichura. Convergence Criteria for Multiparameter Stochastic Processes and Some Applications. *The Annals of Mathematical Statistics*, 42(5):1656 – 1670, 1971.

[BW17] Béatrice Bucchia and Martin Wendler. Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis*, 155:344–368, 2017.

[Con19] John B Conway. *A course in functional analysis*, volume 96. Springer, second edition, 2019.

[Cre93] Noel Cressie. *Statistics for spatial data*. John Wiley & Sons, 1993.

[CW98] Xiaohong Chen and Halbert White. Central Limit and Functional Central Limit Theorems for Hilbert-valued Dependent Heterogenous Arrays with Applications. *Econometric Theory*, 14(2):260–284, 1998.

[Deo75] Chandrakant M Deo. A Functional Central Limit Theorem for Stationary Random Fields. *The Annals of Probability*, pages 708–715, 1975.

[Don51] Monroe David Donsker. *An Invariance Principle for Certain Probability Limit Theorems*. American Mathematical Society. Memoirs. American Mathematical Society, 1951.

[GK12] Oleksandr Gromenko and Piotr Kokoszka. Testing the equality of mean functions of ionospheric critical frequency curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(5):715–731, 2012.

[Guy95] Xavier Guyon. *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer Science & Business Media, 1995.

[JN35] P. Jordan and J. V. Neumann. On Inner Products in Linear, Metric Spaces. *Annals of Mathematics*, 36(3):719–723, 1935.

[Kho02] Davar Khoshnevisan. *Multiparameter Processes: An Introduction to Random Fields*. Springer Science & Business Media, 2002.

[KR97] Richard V. Kadison and John R. Ringrose. *Fundamentals of the Theory of Operator Algebras. Volume I: Elementary Theory*, volume 15. American Mathematical Society, 1997.

[Lap09] Amos Lapidoth. *A Foundation in Digital Communication*. Cambridge University Press, Cambridge, UK, first edition, July 2009.

[Lav08] Frédéric Lavancier. The V/S test of long-range dependence in random fields. *Electronic Journal of Statistics*, 2(none):1373 – 1390, 2008.

[Loo53] Lynn H Loomis. *Introduction to Abstract Harmonic Analysis*. D. Van Nostrand Company, 1953.

[Mór83] F Móricz. A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Mathematica Hungarica*, 41(3-4):337–346, 1983.

[Mül07] Ulrich K Müller. A theory of robust long-run variance estimation. *Journal of Econometrics*, 141(2):1331–1352, 2007.

[Neu71] Georg Neuhaus. On Weak Convergence of Stochastic Processes with Multidimensional Time Parameter. *The Annals of Mathematical Statistics*, 42(4):1285–1295, 1971.

[Pet38] Billy James Pettis. On integration in vector spaces. *Transactions of the American Mathematical Society*, 44(2):277–304, 1938.

[RDC+22] Paulo Justiniano Ribeiro Jr, Peter Diggle, Ole Christensen, Martin Schlather, Roger Bivand, and Brian Ripley. *geoR: Analysis of Geostatistical Data*, 2022. R package version 1.9-2.

[Rio13] Emmanuel Rio. Inequalities and limit theorems for weakly dependent sequences. Lecture, September 2013.

[Ros56] Murray Rosenblatt. A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the national Academy of Sciences*, 42(1):43–47, 1956.

[Ros70] Haskell P. Rosenthal. On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8:273 – 303, 1970.

[Rud74] Walter Rudin. *Real and Complex Analysis*, volume 3. Mcgraw hill International Book Company, 1974.

[Rud91] Walter Rudin. *Functional Analysis.* McGraw-Hill, Inc., second edition, 1991.

[Rya02] Raymond A Ryan. *Introduction to Tensor Products of Banach spaces*, volume 73. Springer, 2002.

[Sch38] I. J. Schoenberg. Metric Spaces and Positive Definite Functions. *Transactions of the American Mathematical Society*, 44(3):522–536, 1938.

[Sch12] Martin Schlather. Construction of Covariance Functions and Unconditional Simulation of Random Fields. In Emilio Porcu, José-María Montero, and Martin Schlather, editors, *Advances and Challenges in Space-time Modelling of Natural Events*, pages 25–54, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[Sha10] Xiaofeng Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235, 2010.

[Tre06] François Treves. *Topological Vector Spaces, Distributions and Kernels.* Dover Publications, Inc., 2006.

[Vaa98] A. W. van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[Wei80] Joachim Weidmann. *Linear Operators in Hilbert Spaces*, volume 68. Springer Science & Business Media, 1980.

[Wu86] C. F. J. Wu. Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, 14(4):1261 – 1295, 1986.

[Zha98] Lixin Zhang. Rosenthal type inequalities for B-valued strong mixing random fields and their applications. *Science in China Series A: Mathematics*, 41(7):736–745, 1998.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Die eingereichte schriftliche Fassung der Arbeit entspricht der PDF Fassung.

Weiterhin versichere ich, dass diese Arbeit noch nicht als Abschlussarbeit an anderer Stelle vorgelegen hat.

Datum, Unterschrift