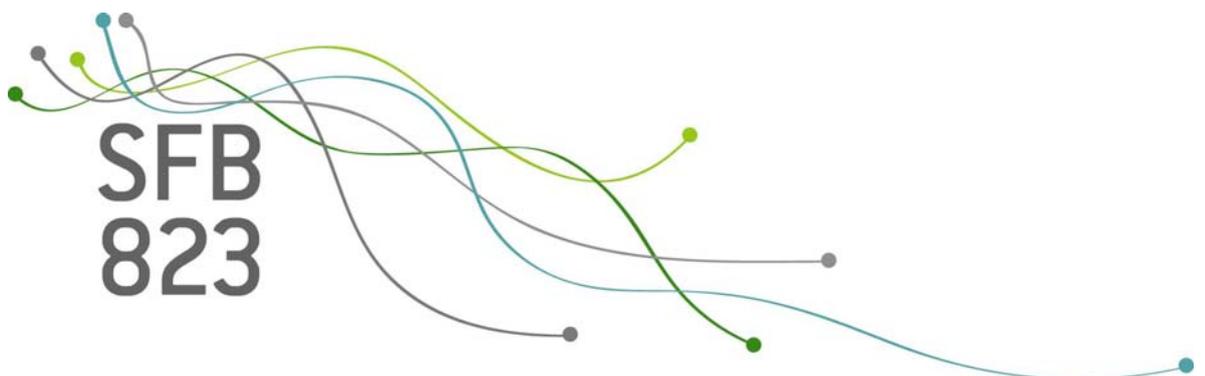


SFB
823

A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing

Anne van Delft, Holger Dette

Nr. 33/2018



Discussion Paper

A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing

Anne van Delft and Holger Dette

Ruhr-Universität Bochum,
Fakultät für Mathematik,
44780 Bochum, Germany

December 3, 2018

Abstract

Due to the surge of data storage techniques, the need for the development of appropriate techniques to identify patterns and to extract knowledge from the resulting enormous data sets, which can be viewed as collections of dependent functional data, is of increasing interest in many scientific areas. We develop a similarity measure for spectral density operators of a collection of functional time series, which is based on the aggregation of Hilbert-Schmidt differences of the individual time-varying spectral density operators. Under fairly general conditions, the asymptotic properties of the corresponding estimator are derived and asymptotic normality is established. The introduced statistic lends itself naturally to quantify (dis)-similarity between functional time series, which we subsequently exploit in order to build a spectral clustering algorithm. Our algorithm is the first of its kind in the analysis of non-stationary (functional) time series and enables to discover particular patterns by grouping together ‘similar’ series into clusters, thereby reducing the complexity of the analysis considerably. The algorithm is simple to implement and computationally feasible. As a further application we provide a simple test for the hypothesis that the second order properties of two non-stationary functional time series coincide.

Keywords: time series, functional data, clustering, spectral analysis, local stationarity

AMS Subject classification: Primary: 62M15; 62H15, Secondary: 62M10, 62M15

1 Introduction

The surge in data storage techniques over the past two decades has led to more and more data sets that are almost continuously recorded from their domain of definition. The development of tools to model these type of data is the main focus of functional data analysis. In functional data analysis, the variables of interest are perceived as random smooth functions that vary on a continuum D , i.e., $X(\tau), \tau \in D$. While the intrinsically infinite variation of such random functions can be considered a rich source of information, extracting relevant information and to identify patterns becomes more and more a challenge. Especially when the data is collected sequentially over time and the curves exhibit serial dependence, i.e., when the data set consists of a collection of d functional time series, $\{X_{i,t}(\tau) : \tau \in D\}_{t \in \mathbb{Z}, i \in 1, \dots, d}$. This type of data arises naturally in a wide

range of scientific disciplines such as astronomy, biology, finance, meteorology, medicine or yet engineering. In addition, in most real-world applications, the second order characteristics of time series change gradually over time. In meteorology, the distribution of the daily records of temperature, precipitation and cloud cover for a region, viewed as three related functional surfaces, may change over time due to global climate changes. Other relevant examples appear in the study of cognitive functions such as high-resolution recordings from local field potentials, EEG and MEG or from the financial industry where implied volatility of an option as a function of moneyness changes over time. The development of appropriate exploratory techniques that allow to discover patterns or anomalies is therefore of foremost interest for this type of data.

The most widely used technique for this preliminary step of data exploration is known as *cluster analysis*. Clustering is concerned with partitioning a data-set into a set of disjoint homogeneous groups (clusters) of realizations. Unlike supervised learning, clustering does not rely on prior knowledge of the groups or on building classifiers based upon a training set. It is therefore especially meaningful when little is known about the nature of the process and the data set is large.

A large body of literature on clustering (and related learning techniques) of Euclidean-valued time series has been published. Depending on the goal of the application, clustering algorithms can differ in a variety of aspects such as the representation of the data, how similarities are measured and the way clusters are constructed. The first two aspects are known to be crucial in terms of efficiency and accuracy of the solution and this is where most research focuses on (see section 5 of Aghabozorgi et al., 2015, for a full taxonomy of the different aspects of clustering time series). For instance, in parametric approaches, clusters are usually built based upon similarity of their estimated parameters (see e.g., Kalpakis et al., 2001; Corduas and Piccolo, 2008) some of which take a Bayesian approach (Bauwens and Rombouts, 2007; Frühwirth-Schnatter and Kaufman, 2008; Juárez and Steel, 2010). Nonparametric methods are often based upon comparing similarity of the estimated power spectra, which is a research topic on its own (see e.g., Coates and Diggle, 1986; Eichler, 2008; Dette and Paparoditis, 2009; Dette and Hildebrandt, 2012; Jentsch and Pauly, 2015). This approach is taken in, i.a., Kakizawa et al. (1998); Savvides et al. (2008); Fokianos and Promponas (2011); Holan and Ravishanker (2018). A wavelet-based approach can be found in Vlachos et al. (2003).

Clustering and classification methods have also been extended to non-stationary time series. For example, Sakiyama and Taniguchi (2004) use the framework of locally stationary time series (Dahlhaus, 1997) for clustering while Chandler and Polonik (2006) use it to develop a shape-based approach discriminant analysis. Another branch of literature focuses on piecewise stationary processes using Smooth Localized Complex EXponential (SLEX) transforms, which were introduced by Ombao et al. (2001), or variations thereof (see e.g., Huang et al., 2004; Harvill et al., 2017) for clustering approaches and Böhm et al. (2010) for classification of multivariate series.

In contrast to the Euclidean case, the literature on cluster analysis for functional data is not that rich. Some methods have been developed for clustering of i.i.d. functional data (Jacques and Preda, 2014, and references therein). A popular technique is to first reduce dimension by projecting the curves onto a basis of finite dimension and then apply a standard classical clustering algorithm such as k -means (see Peng and Müller, 2008; Abraham et al., 2003). Alternatively, non-parametric methods have been proposed that use specific distances for functional data (Ferraty and Vieu, 2006; Ieva et al., 2013) and parametric (Bayesian) approaches assuming a particular probability distribution for multivariate functional data (see Jacques and Preda, 2014; Heard et al., 2006, among others).

Despite of the vast amount of literature available on various data structures, existing methods are inappropriate to cluster possibly non-stationary functional time series. A clustering tech-

nique for functional time series requires to take into account its infinite variation and hence a clustering approach must be able to capture the complex within-curves dynamics as well as the between-curve dynamics. At the same time, it needs to be efficient to apply because of the high dimensionality of the data. In this article, we address this problem from several perspectives. We develop a new measure to compare the second order properties of non-stationary functional time series. This measure is based on the aggregation of Hilbert-Schmidt differences of the individual time-varying spectral density operators. Under fairly general conditions, the asymptotic properties of the corresponding estimator are derived and asymptotic normality is established. We then use this methodology for two purposes. Firstly, we consider this measure and its estimate to develop a new spectral clustering algorithm for functional time series, which are allowed to be non-stationary and nonlinear. Our algorithm is novel in the sense that not only is it the first of its kind for exploratory analysis of functional time series but moreover because, to our knowledge, spectral clustering has also not yet been considered for Euclidean time series. The underlying principle of spectral clustering is to reformulate the problem into a graph partitioning problem (see Figure 3.1) Geometrical properties of graphs can be conveniently described by the spectral properties of the corresponding graph Laplacian (see e.g., Chung, 1997; Chung and Radcliffe, 2011; Von Luxburg, 2007). Using these properties, we can detect clusters in non-convex regions, which classical clustering techniques may not be able to find. Furthermore, it can be solved efficiently via classical linear algebra operations. We will show that our introduced measure of similarity provides a meaningful basis for the adjacency matrix underlying the graph Laplacian. Secondly, we use this measure to develop a particularly simple level α -test for the hypothesis of equality of two time-varying spectral density operators, which uses the quantiles of the standard normal distribution.

The structure of this paper is as follows. We first introduce necessary notation and background on the type of processes considered in this paper. We then define a measure of similarity for a pair of functional time series and derive a consistent estimator to construct an empirical adjacency matrix. In Section 3, the spectral clustering algorithm is discussed in detail and it is shown that the algorithm based upon an empirical graph Laplacian, a transformation of the estimated adjacency matrix, is consistent. In Section 4, we discuss the application of hypothesis testing, whereas in Section 5 we study the properties of our algorithm in finite samples. In Section 6, the clustering method is illustrated by means of an application to high-resolution meteorological data. All proofs and technical assumptions for the main statements are relegated to the Appendix.

2 A measure of similarity

In this section we introduce a measure of similarity for functional time series, which is appropriate to use as a basis for a similarity matrix to cluster functional time series.

Example. We have generated 90 functional time series uniformly from models I, II, III (described in detail in Section 4), which should be clustered according to their second order properties. To visualize the difficulties of this task we exemplarily depict 9 series in Figure 2.1, where we took three from each distribution. As can be seen, the high dimensionality of the data makes a visual assessment of their second order properties almost impossible for 6 of them, while 3 of them appear more obvious to distinguish from the other 6.

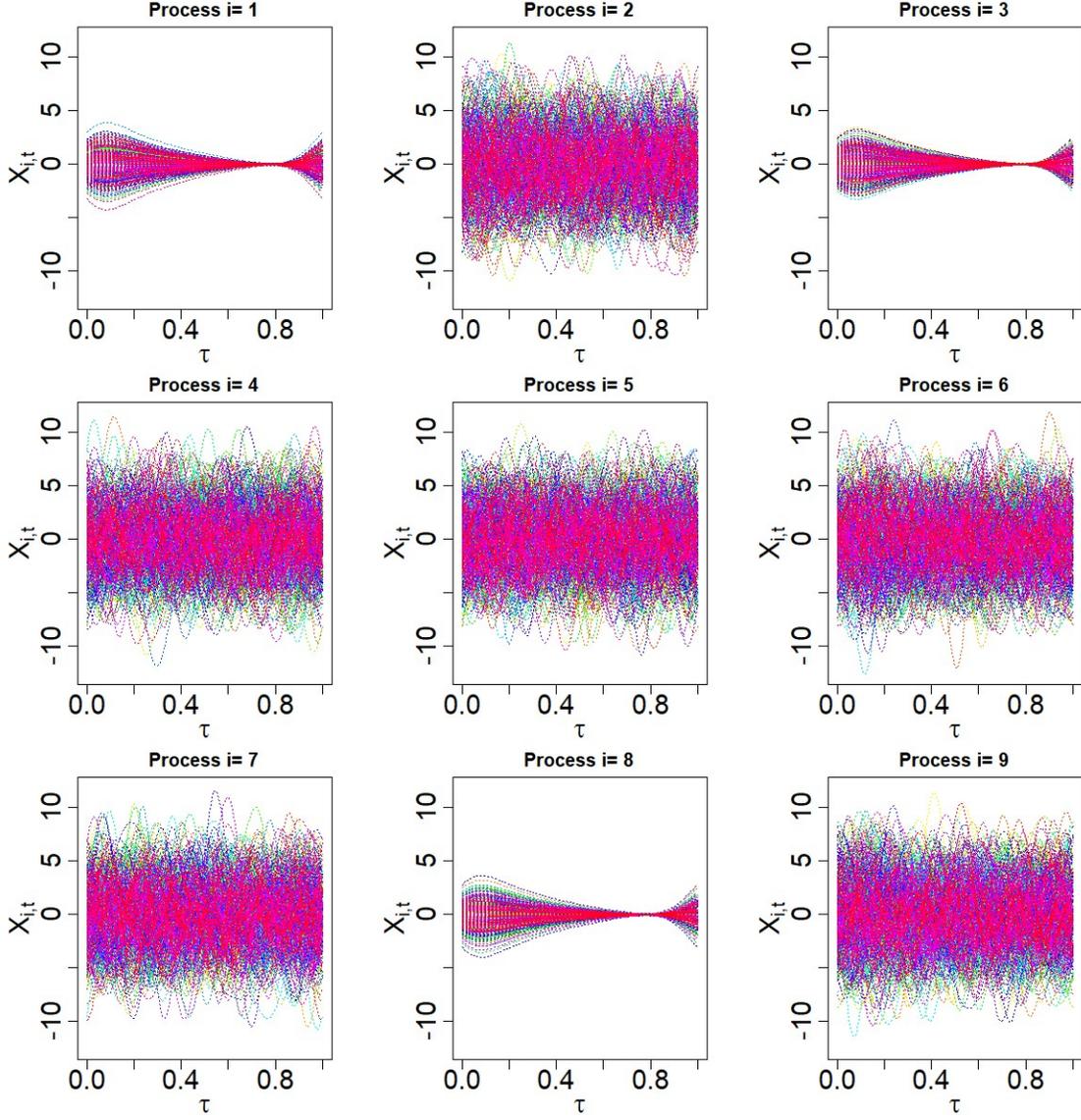


Figure 2.1: Functional time series from 3 different distributions, $T = 512$.

2.1 Notation

First, let us introduce some necessary notation. For a separable Hilbert space \mathcal{H} , we denote the inner product as $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$ and its induced norm by $\| \cdot \|$. The Banach space of bounded linear operators $A : \mathcal{H} \rightarrow \mathcal{H}$ with operator norm $\|A\|_\infty = \sup_{\|x\| \leq 1} \|Ax\|$ shall be denoted by $\mathcal{L}(\mathcal{H})$, the adjoint of $A \in \mathcal{L}(\mathcal{H})$ by A^\dagger . $A \in \mathcal{L}(\mathcal{H})$ is called self-adjoint if $A = A^\dagger$ and non-negative definite if $\langle Ax, x \rangle \geq 0$ for each $x \in \mathcal{H}$. If well-defined we denote the trace of $A \in \mathcal{L}(\mathcal{H})$ by $\text{Tr}(A)$. A compact operator $A \in \mathcal{L}(\mathcal{H})$ belongs to the Schatten class of order $1 \leq p < \infty$, denoted by $A \in \mathcal{S}_p(\mathcal{H})$, if $\|A\|_p^p = \sum_{j \geq 1} s_j^p(A) < \infty$, where $\{s_j(A) : j \geq 1\}$ are the singular values of A . Operators that belong to the Banach spaces $(\mathcal{S}_1(\mathcal{H}), \|\cdot\|_1)$ or $(\mathcal{S}_2(\mathcal{H}), \|\cdot\|_2)$ will be referred to as trace-class operators and Hilbert-schmidt operators, respectively. We remark in particular that $(\mathcal{S}_2(\mathcal{H}), \|\cdot\|_2)$ is a Hilbert space with the inner product given by $\langle A, B \rangle_{\text{HS}} = \text{Tr}(AB^\dagger) = \sum_{j \geq 1} \langle Ae_j, Be_j \rangle$ for each $A, B \in \mathcal{S}_2(\mathcal{H})$ and $\{e_j\}_{j \in \mathbb{N}}$ an orthonormal basis of \mathcal{H} . For $f, g \in \mathcal{H}$, we

define the tensor product $f \otimes g : \mathcal{H} \otimes \mathcal{H} \rightarrow \mathcal{H}$ as the bounded linear operator defined by

$$(f \otimes g)v = \langle v, g \rangle f \quad \forall v \in \mathcal{H}.$$

Without loss of generality, we consider the Hilbert space $\mathcal{H} = \mathbb{L}_{\mathbb{C}}^2([0, 1])$ of equivalence classes of square integrable measurable functions $f : [0, 1] \rightarrow \mathbb{C}$ with inner product $\int_0^1 f(\tau) \overline{g(\tau)} d\tau$, where the complex conjugate of $x \in \mathbb{C}$ is denoted as usual by \bar{x} . Since the mapping $\mathcal{T} : \mathcal{H} \otimes \mathcal{H} \rightarrow \mathcal{S}_2(\mathcal{H})$ defined by the linear extension of $\mathcal{T}(f \otimes g) = f \otimes \bar{g}$ is an isometric isomorphism, it defines a Hilbert-Schmidt operator with the kernel in \mathcal{H} given by $(f \otimes g)(\tau, \sigma) = f(\tau) \overline{g(\sigma)}$ for each $\tau, \sigma \in [0, 1]$ in an L^2 -sense. We refer to Appendix A for further details and background.

2.2 A measure of similarity for functional time series

The main object of this paper are d zero-mean stochastic processes $X_i = \{X_{i,t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$ $i = 1, \dots, d$ that take values in the Hilbert space $L_{\mathbb{R}}^2([0, 1])$. The second order dynamics are assumed to be finite but allowed to change over time. Processes of this type fit the framework of *locally stationary functional time series* as defined in van Delft and Eichler (2018) –which extends the concept of local stationarity (Dahlhaus, 1997) to the function space– and contain weakly stationary functional processes as a subclass. Asymptotic properties can be described by so-called infill-asymptotics, such that, as $T \rightarrow \infty$, we obtain more and more observations at a local level. Formally, a process $\{X_{t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$ is called *functional locally stationary* if, for all rescaled times $u \in [0, 1]$, there exists an $L_{\mathbb{R}}^2([0, 1])$ -valued strictly stationary process $\{X_t^{(u)} : t \in \mathbb{Z}\}$ such that

$$\left\| X_{t,T} - X_t^{(u)} \right\|_2 \leq \left(\left| \frac{t}{T} - u \right| + \frac{1}{T} \right) P_{t,T}^{(u)} \quad a.s. \quad (2.1)$$

for all $1 \leq t \leq T$, where $\{P_{t,T}^{(u)}\}_{t=1,\dots,T; T \in \mathbb{N}}$ is a positive real-valued process such that for some $\rho > 0$ and $C < \infty$ the process satisfies $\mathbb{E}(|P_{t,T}^{(u)}|^\rho) < C$ for all t and T and uniformly in $u \in [0, 1]$. We refer to van Delft and Eichler (2018) for further details.

What we will exploit throughout this paper is that the full second order dynamics of the the triangular array $\{X_{t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$ are completely and uniquely characterized by the *time-varying spectral density operator*

$$\mathcal{F}_{u,\omega} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \mathbb{E}(X_{t+h}^{(u)} \otimes X_t^{(u)}) e^{-i\omega h} \quad (2.2)$$

for each $u \in [0, 1]$ and $\{X_t^{(u)} : t \in \mathbb{Z}\}$. In particular, for weakly stationary functional time series we can drop the dependence on local time and thus $\mathcal{F}_{u,\omega} \equiv \mathcal{F}_\omega$. This uniquely characterizing object of a functional time series lends itself naturally as a basis for a measure of similarity.

More specifically, let, $\mathcal{F}_{u,\omega}^{(i_1)}$ and $\mathcal{F}_{u,\omega}^{(i_2)}$ denote the time-varying spectral density operator of processes $\{X_{i_1,t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$ and $\{X_{i_2,t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$, respectively. As a measure of pairwise similarity between two functional time series, we consider

$$\mathcal{A}_{i_1, i_2} = \frac{\int_0^1 \int_{-\pi}^{\pi} \|\mathcal{F}_{u,\omega}^{(i_1)} - \mathcal{F}_{u,\omega}^{(i_2)}\|_2^2 dud\omega}{\int_0^1 \int_{-\pi}^{\pi} \|\mathcal{F}_{u,\omega}^{(i_1)}\|_2^2 + \|\mathcal{F}_{u,\omega}^{(i_2)}\|_2^2 dud\omega}. \quad (2.3)$$

Clearly, if this distance is zero, then processes $\{X_{i_1,t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$ and $\{X_{i_2,t,T}\}_{t=1,\dots,T; T \in \mathbb{N}}$ must have the same second order properties and hence must belong to the same cluster. Note that (2.3) takes values in the interval $[0, 1)$. The local scaling via the denominator is an essential aspect

for its usage in a spectral clustering procedure. Differences in scales can lead spectral clustering to fail. While most similarity graphs are based upon a global scaling parameter of which the optimal value is difficult to determine and can highly affect the clustering performance [see Von Luxburg (2007)], we find that making the measure scale-invariant by accounting for local scales avoids this issue. We discuss this further in Section 5.

Since $\mathcal{F}_{u,\omega}^{(i_1)}$ and $\mathcal{F}_{u,\omega}^{(i_2)}$ are unknown, we will have to estimate (2.3). Similar to van Delft et al. (2018), we shall do this using integrated periodogram tensors. That is, we split the sample into M blocks with N elements inside each of these blocks so that $T = MN = M(T)N(T)$ for each $T \in \mathbb{N}$, where $M \in \mathbb{N}$ and N is an even number. M and N will correspond to the number of terms used in a Riemann sum approximating the integrals in (2.3) with respect to du and $d\omega$ and therefore they have to be reasonable large. The functional discrete Fourier transform (fDFT) at time point u , is a random function with values in $L^2([0, 1], \mathbb{C})$ defined by

$$D_i^{u,\omega} := \frac{1}{\sqrt{2\pi N}} \sum_{s=0}^{N-1} X_{i, \lfloor uT \rfloor - N/2 + s + 1, T} e^{-i\omega s}. \quad (2.4)$$

The local periodogram tensor for the i -th time series can be given as

$$I_i^{u,\omega} := D_i^{u,\omega} \otimes D_i^{u,\omega}, \quad (2.5)$$

for $i = 1, \dots, d$. We base our estimator upon a linear combination of the following Hilbert-Schmidt inner products

$$F_{i_1 i_2} = \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \langle I_{i_1}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}. \quad (2.6)$$

In particular, a suitable and (symmetric) estimator for the distance (2.3) is given by

$$\hat{\mathcal{A}}_{i_1, i_2} := \frac{F_{i_1 i_1} + F_{i_2 i_2} - F_{i_1 i_2} - F_{i_2 i_1}}{(F_{i_1 i_1} + F_{i_2 i_2})}, \quad (2.7)$$

To ease notation, we provide empirical quantities with $\hat{\cdot}$. The dependence of these quantities on T is therefore implicit. We find under suitable regularity conditions, which are postponed to section B,

Theorem 2.1 (consistency). *Under Assumption B.1 with $m = 8$ and Assumption B.2, we find that*

$$\hat{\mathcal{A}}_{i_1, i_2} - \mathcal{A}_{i_1, i_2} = O_p(T^{-1/2}).$$

We remark that, under suitable moment conditions, it is moreover asymptotically multivariate normal and therefore lends itself for other statistical applications such as a test for equality of time-varying spectral density operators. We shall briefly discuss this application together with more details on the statistic $\hat{\mathcal{A}}_{i_1, i_2}$ in Section 4. In the next section, we define a similarity graph based upon the measure (2.7) and introduce a spectral clustering algorithm to cluster the functional time series.

Example (continued). For 6 of the 9 functional time series depicted in Figure 2.1 it is difficult to visually distinguish the second order properties. This becomes even more difficult for all 90 series. However, we can use the measure $\hat{\mathcal{A}}$ to identify similarities. A heat map of the corresponding estimates for all 90 time series are displayed in Figure 2.2(a). As can be seen the empirical measure gives the pairs of time series different weights ranging from 0 to 1, but it is difficult to identify any structure. For the sake comparison, we ordered the times series assuming knowledge of the clusters in the right part of Figure 2.2(b). We observe that the similarity measure $\hat{\mathcal{A}}$ make the clusters visible.

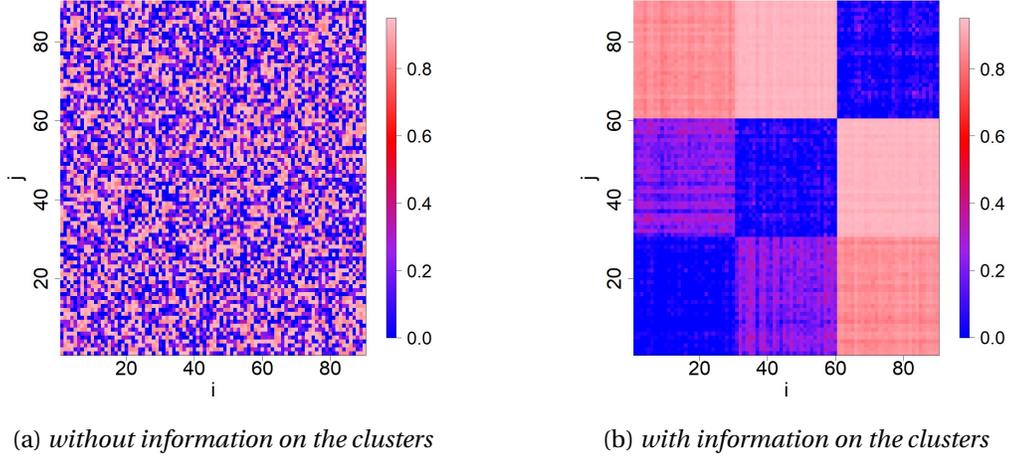


Figure 2.2: Heat map of the value of the estimates $\hat{\mathcal{A}}_{i,j}$ of $\mathcal{A}_{i,j}$ plotted for all pairs $i, j = 1, \dots, 90$ in random order (left); and ordered by cluster (right).

3 Spectral clustering of functional time series

In this section, we develop a spectral clustering algorithm, which consists of a several steps. We start by translating the problem of clustering d functional time series into k clusters into a graph partitioning problem using the previously defined similarity measure. Secondly, we construct a spectral embedding using an empirical graph Laplacian, which is shown to have spectral properties that converge to those of the population graph Laplacian. We then use the embedded points to cluster the data using k -means and show that we can bound the number of points that are clustered converges to zero as $T \rightarrow \infty$.

3.1 Construction of the graph

We construct an undirected similarity graph $G = (V, E)$, where V denotes the set of vertices and E denotes the set of edges. To each family of random curves $\{X_i\}$, $i \in [d]$ we relate a node $v_i \in V$ and describe the similarity between node v_i and v_j via non-negative weights on the edges. These weights are given by the *empirical adjacency matrix* which is defined as the following transformation of the matrix $\hat{\mathcal{A}} = \{\hat{\mathcal{A}}_{i_1, i_2} : i_1, i_2 = 1, \dots, d\}$ as defined in (2.7)

$$\hat{W} = e^{-\hat{\mathcal{A}}} \in \mathbb{R}^{d \times d}. \quad (3.1)$$

This is illustrated in Figure 3.1 for the first 6 processes depicted in Figure 2.1. The clustering problem then becomes equivalent to partitioning the similarity graph into connected components such that points with pairwise high weight belong to the same component while points with low weight are put into different components.

3.2 The spectral embedding

The main ingredient to our algorithm is the empirical graph Laplacian

$$\hat{L} = I - \hat{D}^{-1/2} \hat{W} \hat{D}^{-1/2}, \quad (3.2)$$

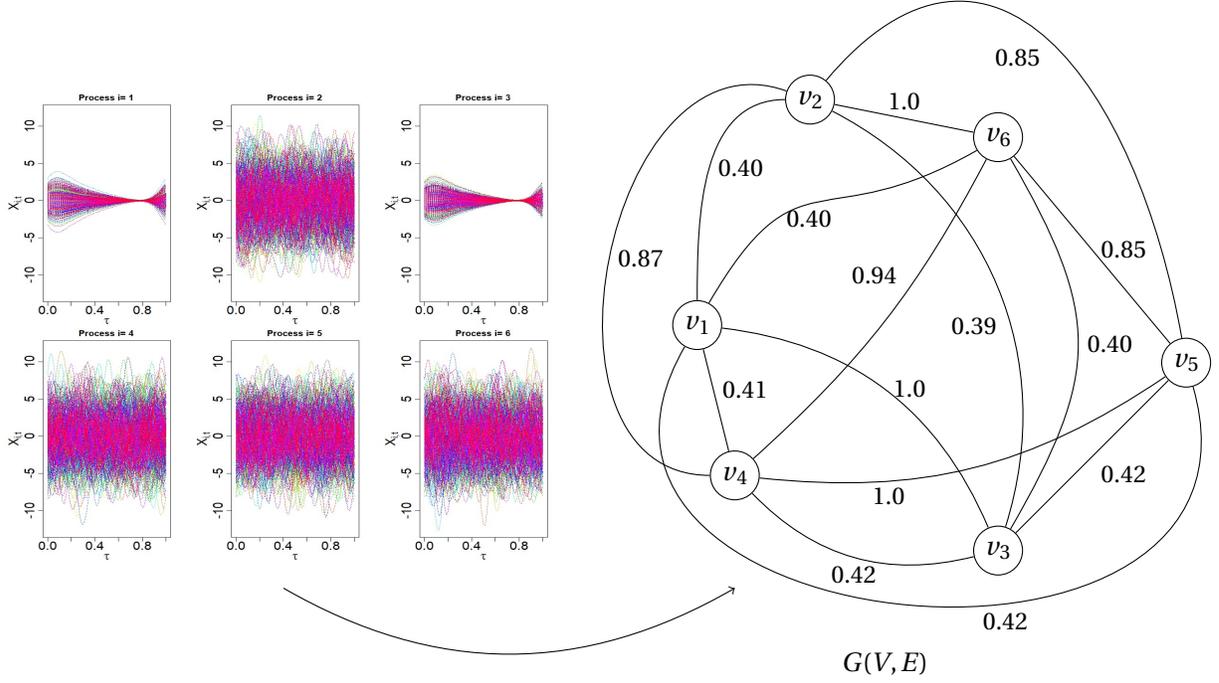


Figure 3.1: *Illustration of the map from the space of functional time series (left) into a graph partitioning problem (right).*

where $\hat{D} = \text{diag}(\{\hat{D}_i\}_{i=1}^d)$ denotes the *degree matrix* of \hat{W} which carries the degree of vertex v_i as its i -th diagonal element, i.e., $\hat{D}_i = \sum_{j=1}^d \hat{W}_{i,j}$. This Laplacian can be viewed as a perturbed version of the unknown population Laplacian

$$L = I - D^{-1/2} W D^{-1/2}, \quad (3.3)$$

where $W = e^{-\mathcal{A}}$ is the population adjacency matrix and D its corresponding degree matrix. Note that we use a Normalized Laplacian as it shows a better performance if d is large (Von Luxburg et al., 2008) and can also be applied to irregular graphs, i.e., graphs of which the vertices have different degrees.

It is straightforward to show that (3.3) is symmetric and positive-definite and therefore has an eigendecomposition, say $L = U^\top \Lambda U$ with $\Lambda = \text{diag}(\{\lambda_i\}_{i=1}^d) \in \mathbb{R}_{\geq 0}^{d \times d}$. In the case where $G = (V, E)$ is an undirected weighted graph with non-negative weights, its spectrum provides information on the connectivity of the graph. More specifically, the multiplicity k of the eigenvalue 0 of L equals the number of connected components G_1, \dots, G_k in the graph G , and the eigenspace of the eigenvalue 0 is spanned by the vectors $D^{1/2} \mathbb{1}_{G_i}$, $i = 1, \dots, k$, where $\mathbb{1}_{G_i} \in \mathbb{R}^d$ denotes the indicator vector on component G_i (see e.g., Chung, 1997).

To understand the usefulness of these properties, suppose for a moment that our graph has exactly k disconnected components where the nodes belonging to different components are infinitely far apart, i.e. have zero weight. The above implies that if we collect the k eigenvectors that belong to the k smallest eigenvalues of L and we row-normalize this matrix, we obtain a matrix of indicator vectors

$$\mathcal{U} := [\mathbb{1}_{G_1}, \dots, \mathbb{1}_{G_k}] \in \mathbb{R}^{d \times k}. \quad (3.4)$$

Per row of \mathcal{U} there will be exactly one nonzero element, indicating the component to which it belongs to. In practice, one never encounters the ideal situation that the nullspace of the em-

pirical graph Laplacian is perfectly spanned by (scaled) indicator vectors because the empirical similarity graph, by construction, consists of only one connected component. Nevertheless, the information about the structure of k clusters is still completely contained in the eigenvectors that belong to the smallest k eigenvalues of L . In particular, these eigenvectors provide a relaxation solution to the *Normalized minimal cut* problem (Ncut) (Shi and Malik, 2000), which has the objective to partition the graph into k disjoint components by ‘cutting’ it at the edges of which the total sum of normalized weights (relative to the volume of the partitioning component) is minimized. In order to exploit this information, we embed the infinite-dimensional processes X_i into the space spanned by the k eigenvectors that belong to the k smallest eigenvalues by representing the i -th process by the i -th row of \mathcal{U} . The embedded points then provide a representation of X_i in \mathbb{R}^k of which the clustering properties are enhanced. As a result, a simple algorithm such as k -means can be applied to the embedded points to identify the clusters.

To be precise, let $\hat{\mathcal{U}}_{\cdot,1}, \dots, \hat{\mathcal{U}}_{\cdot,k} \in \mathbb{R}^d$ denote the row-normalized k eigenvectors of the empirical graph Laplacian \hat{L} defined in (3.2) corresponding to the smallest k eigenvalues. Note that the eigenvectors of \hat{L} can only be identified up to an orthogonal rotation. The row normalization avoids this additional source of unidentifiability as will become clear below. It is by no means guaranteed that the embedding of our data as the rows of the matrix

$$\hat{\mathcal{U}} := [\hat{\mathcal{U}}_{\cdot,1}, \dots, \hat{\mathcal{U}}_{\cdot,k}] \in \mathbb{R}^{d \times k}, \quad (3.5)$$

provides a meaningful spectral embedding for clustering. For this purpose we need to show that \hat{L} is ‘close’ to L measured by a suitable norm such that their spectral properties converge to their population counterparts. The approach that we take to establish consistency of the empirical Laplacian as T tends to infinity is based on perturbation theory comparing the matrices \hat{L} and L (see also Ng et al., 2002; Rohe et al., 2011). The proofs are technical and rely on several auxiliary results which are relegated to the Appendix. Using consistency of $\hat{\mathcal{A}}$ and the symmetry of L , we can show that the distance in operator norm between \hat{L} and L converges to zero in probability.

Lemma 3.1. *Under the conditions of Theorem 2.1*

$$\forall \varepsilon > 0, \quad \lim_{T \rightarrow \infty} \mathbb{P}(\|\hat{L} - L\|_\infty > \varepsilon) = 0.$$

To analyze the concentration of $\hat{\mathcal{U}}$, we use a slight modification of the classical Davis-Kahan Sin Θ theorem (Stewart and Sun, 1990). While this theorem provides an upper bound on the sinus of the principal angles between the eigenspaces $\hat{U}O$, for some orthonormal rotation matrix $O \in \mathbb{R}^{k \times k}$, and U , in terms of the spectral gap δ , the dimension of the space k and on the size of the perturbation $\|H\|_\infty$, Lemma C.1 in the Appendix can be used to bound the Euclidean distance between the non-normalized empirical eigenvectors \hat{U} and their population counterparts U up to rotation. For the row-normalized matrix $\hat{\mathcal{U}}$, we avoid the additional source of unidentifiability caused by the rotation matrix. We derive the following result.

Lemma 3.2. *The matrix $\hat{\mathcal{U}}$ defined in (3.5) satisfies*

$$\|\hat{\mathcal{U}} - \mathcal{U}\|_2 \leq \frac{4\sqrt{k}}{\sqrt{\min_i \|U_{i\cdot}\|_2^2}} \frac{\|\hat{L} - L\|_\infty}{\lambda_{k+1}} \leq 4\sqrt{k} \sqrt{\frac{\mathcal{C}_{\max}}{\min_i D_i}} \frac{\|\hat{L} - L\|_\infty}{\lambda_{k+1}}$$

where λ_{k+1} is the $(k+1)$ -th smallest eigenvalue of L and where $\mathcal{C}_{\max} = \max_i \sum_{i_1, i_2 \in G_i} W_{i_1, i_2}$. Hence, Under the conditions of Lemma 3.1,

$$\forall \varepsilon > 0, \quad \lim_{T \rightarrow \infty} \mathbb{P}(\|\hat{\mathcal{U}} - \mathcal{U}\|_2 > \varepsilon) = 0.$$

These results thus justify to use the rows $\hat{\mathcal{U}}_1, \dots, \hat{\mathcal{U}}_d$ of (3.5) to embed the d functional time series. Each embedded point $\hat{\mathcal{U}}_{i,\cdot}$ then represents a process X_i (or node) in k dimensions, where these k dimensions can be seen as the features. Based on this representation we can cluster our data using k -means. This step is analyzed next. We remark that our treatment of the spectral embedding by means of row-normalized eigenvectors of the graph Laplacian is therefore similar to Ng et al. (2002) who first investigated the use of a symmetric Laplacian in the context of spectral clustering.

Example (continued). For our example, the eigenvalues in Figure 3.2(a) indicate the empirical graph Laplacian has one connected component. This is not surprising as we have a fully connected graph. The first eigenvector is therefore approximately constant, which can be seen in Figure 3.2(b), while the second and third are approximately constant on the clusters. The second eigenvector (Figure 3.2(c)) allows to separate the first 3 observations from the rest, while the third eigenvector indicates to separate the last three from the first 6 observations (Figure 3.2(d)).

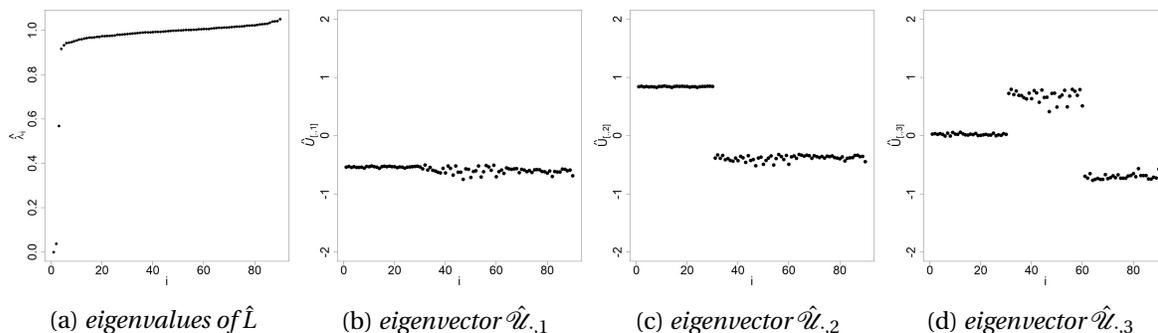


Figure 3.2: The spectrum of \hat{L} and the three row-normalized eigenvectors belonging to the eigenvalues that belong to the three smallest eigenvalues $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \hat{\lambda}_3$.

3.3 Clustering the embedded points using k -means

In this section, we analyze the final step where k -means is applied to the embedded points $\hat{\mathcal{U}}_1, \dots, \hat{\mathcal{U}}_d \in \mathbb{R}^k$. We show that the k -means algorithm, clusters, with high probability, the data correctly. More specifically, we derive a non-asymptotic bound on the number of points that are misclustered and show, under regularity conditions, that this converges to zero as $T \rightarrow \infty$.

The k -means objective aims to partition the d embedded points $\hat{\mathcal{U}}_1, \dots, \hat{\mathcal{U}}_d \in \mathbb{R}^k$ into k clusters $\{C_1, \dots, C_k\}$ in such a way that the pairwise squared deviations of points within the same cluster is minimized. The algorithm thus returns the centroids $\{c_1^*, \dots, c_k^*\}$ from the objective function,

$$\min_{\{c_1, \dots, c_k\} \subset \mathbb{R}^k} \sum_i \min_j \| \hat{\mathcal{U}}_{i,\cdot} - c_j \|^2.$$

The data points $\hat{\mathcal{U}}_{i,\cdot}$ and $\hat{\mathcal{U}}_{j,\cdot}$ are then put in the same cluster if $c_i^* = c_j^*$. More formally but equivalently, for $\hat{\mathcal{U}}$ defined in (3.5), the k means algorithm should return a matrix $C^* \in \mathbb{R}^{d \times k}$ with at most k unique rows such that

$$C^* = \underset{C \in \mathcal{M}(d,k)}{\operatorname{argmin}} \| \hat{\mathcal{U}} - C \|^2_2. \quad (3.6)$$

where $\mathcal{M}(d, k) = \{M \in \mathbb{R}^{d \times k} : M \text{ has at most } k \text{ distinct rows}\}$.

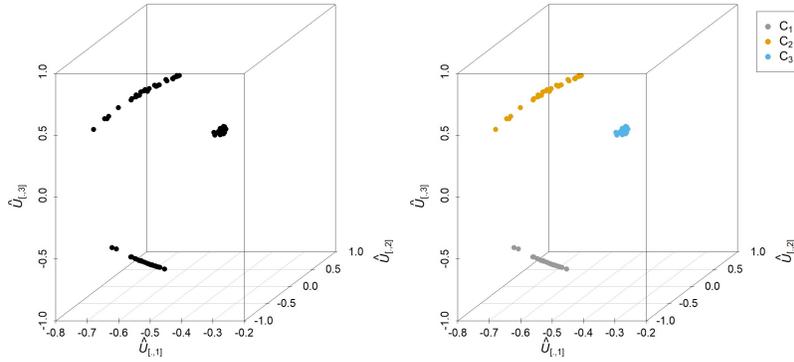


Figure 3.3: The embedded points $\hat{\mathcal{U}}_{[1,\cdot]}, \dots, \hat{\mathcal{U}}_{[90,\cdot]}$ (left) and the result of applying k -means to these points where the color indicate the cluster the point belongs to (right).

To analyze the algorithm, we need to define what we mean with a point being correctly clustered. Let C^* as in (3.6), i.e., the matrix returned from applying the k -means algorithm to $\hat{\mathcal{U}}$. Intuitively, a point $\hat{\mathcal{U}}_{i,\cdot}$ is correctly clustered if there is no other row of the population matrix \mathcal{U} defined in (3.4) that is closer to $C_{i,\cdot}^*$ than the i -th row. Using this intuition, we can provide a meaningful definition of the set of correctly clustered point (Lemma C.2) and hence of its complement set. The next theorem gives a nonasymptotic bound on the number of misclustered points.

Theorem 3.1. *Assume the graph has k components. Then the cardinality of the set of misclustered points, denoted by $|\Sigma|$, satisfies*

$$|\Sigma| \leq \iota k \frac{1}{\min_i \|U_{i,\cdot}\|_2^2} \frac{\|\hat{L} - L\|_\infty^2}{\lambda_{k+1}^2} \leq \iota k \frac{\mathcal{E}_{\max}}{\min_i D_i} \frac{\|\hat{L} - L\|_\infty^2}{\lambda_{k+1}^2}. \quad (3.7)$$

for some constant $\iota > 0$ and where $\mathcal{E}_{\max} = \max_i \sum_{i_1 \in G_i} \sum_{i_2 \in G_i} W_{i_1, i_2}$. If the conditions underlying Theorem 2.1 hold, then $|\Sigma|$ converges to zero in probability as $T \rightarrow \infty$.

This upper bound implies that the probability of misclustering is affected by various properties of the (data-induced) graph. Firstly, one can see it is an increasing function of the number of clusters k . Secondly, it is a decreasing function of the minimum degree. In particular, the second inequality implies that for an unbalanced graph as measured by the maximal sum of all entries of any of the groups relative to the minimal degree $\min_i D_i$ the probability to miscluster has a less tight upper bound. In other words, if the graph contains isolated vertices and or many points that are highly concentrated, we can expect a higher probability to miscluster. Thirdly, it is a decreasing function of the distance between the zero eigenvalues and the first nonzero eigenvalue λ_{k+1} . Finally, it is affected by how well \hat{L} estimates L .

Example (continued). For our example, the embedded points are shown in the left graph of Figure 3.3. Applying k -means identifies all clusters exactly where the colors indicate the cluster the respective data points belong to.

3.4 The choice of k

So far we have considered the case where the number of clusters k is known. In the remaining part of this section we briefly discuss a data driven choice of k . This problem has received much

attention within the clustering literature and a wide range of methods has been developed (see Gordon, 1999; Theodoridis and Koutroumbas, 2008, for overviews) and compared (see Milligan and Cooper, 1985; Tibshirani et al., 2001, and references therein). Most methods to pick k are formulated in terms of optimizing a relative criteria. For different choices of k , the quality of the clustering structure is evaluated according to some measure, such as the intra-cluster dispersion, and the number of clusters chosen that optimizes this criteria.

There is however no universally optimal method because the definition of ‘optimal’ number of clusters can highly depend on the application and on the method used to cluster to data. If the data is well-separated and there are clear pronounced clusters then there are several successful methods that will correctly detect the underlying clusters. However, in noisy datasets with overlapping of clusters, different methods will detect different number of clusters. In the case of spectral clustering, an intuitive alternative would be to pick the number of clusters k such that the first $\lambda_1, \dots, \lambda_k$ eigenvalues of \hat{L} are small but λ_{k+1} is ‘relatively’ large. This ‘eigengap’ heuristic can be justified through spectral properties of the population graph Laplacian and Lemma 3.1. In practice it is however also known to be very sensitive to the construction of the similarity graph and can quickly fail unless the data is well separated. Deemed therefore an unsolvable problem, it is not uncommon to apply multiple criteria and pick the criteria that works best for the particular problem at hand (see Charrad et al., 2014, for an implementation of various criteria).

A thorough development of a new method would be beyond the scope of this paper. In our empirical study in 5 we investigate the performance of available methods to choose k for our particular algorithm. Additionally, we consider two variations of the eigengap heuristic each with a different interpretation of a “relatively large” eigengap.

4 Testing for equality

Besides from the application of the similarity measure as a basis for spectral clustering of functional time series, a well-defined limiting distribution allows it to be meaningful in a variety of statistical applications and in particular for the construction of hypothesis tests. The problem to detect similarities or to compare time series is of interest in a wide range of scientific fields and for classical time series a variety of methods have been proposed (see references in the introduction and examples therein). In case of functional data, the literature is less well-developed. Horvath et al. (2013) proposed a procedure to test the hypothesis that two sets of functional data are identical and independently distributed using the sum of L^2 -distances of the sequence of correlation functions. Panaretos and Tavakoli (2016) instead proposed a test between two stationary functional time series based upon the Hilbert-Schmidt norm of the difference of the sample spectral density operators, restricted to a Hilbert-Schmid space of finite dimension. Bootstrap-based methods to test for equality of mean functions or covariance operators are proposed in Paparoditis and Sapatinas (2016) and, more recently, Leucht et al. (2018) discussed a test for the equality of spectral density operators.

To our knowledge, no procedure is available that allows to test for similarity of functional time series of which the second order structure is allowed to be time-dependent. In this section, we develop such a test using the previously defined similarity measure \mathcal{A}_{i_1, i_2} in (2.3). For the sake of brevity we restrict ourselves to the case of two functional series and consider the hypothesis

$$H_0 : \mathcal{F}_{u, \omega}^{(i_1)} \equiv \mathcal{F}_{\omega}^{(i_2)} \quad \text{a.e. on } [-\pi, \pi] \times [0, 1] \quad (5.1)$$

versus

$$H_a : \mathcal{F}_{u,\omega}^{(i_1)} \neq \mathcal{F}_{\omega}^{(i_2)} \text{ on a subset of } [-\pi, \pi] \times [0, 1] \text{ of positive Lebesgue measure} \quad (5.2)$$

The similarity measure in (2.3) lends itself quite naturally to test this hypothesis, i.e., we can equivalently formulate the hypothesis as

$$H_0 : \mathcal{A}_{i_1, i_2} = 0 \quad \text{versus} \quad H_a : \mathcal{A}_{i_1, i_2} > 0. \quad (5.3)$$

By Theorem 2.1, the statistic $\hat{\mathcal{A}}_{i_1, i_2}$ defined in (2.7) is a consistent estimator of the normalized distance \mathcal{A}_{i_1, i_2} . Therefore it is reasonable to reject the null hypothesis for large values of the estimator $\hat{\mathcal{A}}_{i_1, i_2}$. Using the set of regularity conditions provided in Appendix B the following result gives the asymptotic distribution of $\hat{\mathcal{A}}_{i_1, i_2}$.

Theorem 4.1. *Suppose that Assumption B.1 with $m \geq 1$ and Assumption B.2 hold. Then,*

$$\left\{ \sqrt{T} (\hat{\mathcal{A}}_{i_1, i_2} - \mathcal{A}_{i_1, i_2}) \right\}_{\{i_1, i_2 \in [d]\}} \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}) \quad \text{as } T \rightarrow \infty,$$

where $\mathbf{0} \in \mathbb{R}^d$ and $\mathbf{\Gamma}$ is a positive definite element of $\mathbb{R}^{d \times d}$.

Under the null hypothesis, the asymptotic variance reduces to a very succinct form in case the processes are moreover independent.

Theorem 4.2. *Suppose that Assumption B.1 with $m \geq 1$ and Assumption B.2 hold and suppose $\{X_{i_1, t, T}\}$ and $\{X_{i_2, t, T}\}$ are independent. Then, under the null hypothesis $H_0 : \mathcal{F}_{u,\omega}^{i_1} \equiv \mathcal{F}_{u,\omega}^{i_2}$, we have*

$$\sqrt{T} \hat{\mathcal{A}}_{i_1, i_2} \sim \mathcal{N}(0, \sigma_{H_0}^2),$$

where the asymptotic variance is given by

$$\sigma_{H_0}^2 = 4\pi \frac{\int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u,\omega}^{i_1}\|_2^4 dud\omega}{\left(\int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u,\omega}^{i_1}\|_2^2 dud\omega \right)^2}. \quad (5.4)$$

Let $I_p^{u_j, \omega_k} = (I_{i_1}^{u_j, \omega_k} + I_{i_2}^{u_j, \omega_k})/2$ be the pooled periodogram operator evaluated at u_j and ω_k . The asymptotic variance under the null can be estimated by

$$\hat{\sigma}_{H_0}^2 = \frac{2}{3T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} (\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS})^2 / \left(\frac{2}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS} \right)^2. \quad (5.5)$$

Lemma 4.1. *Under the conditions of Theorem 4.2, the estimator defined in (5.5) satisfies*

$$\hat{\sigma}_{H_0}^2 \xrightarrow{P} \sigma_{H_0}^2$$

as $T \rightarrow \infty$.

Consequently, a test for the hypotheses (5.1) and (5.2) can be based upon rejecting the null if

$$\hat{\mathcal{A}}_{i_1, i_2} > \frac{\hat{\sigma}_{H_0}^2}{\sqrt{T}} z_{1-\alpha}, \quad (5.6)$$

where $z_{1-\alpha}$ denotes the $(1 - \alpha)$ -quantile of the standard normal distribution. The finite sample performance of this test is studied in a simulation study which is provided at the end of Section 5.

Remark 4.1. *We emphasize that Theorem 4.2 still holds in case the series are dependent but the expression of the asymptotic variance is slightly more involved. It can however still be estimated similar in spirit to (5.5). See Appendix B for more details.*

5 Simulation study

To study the performance in finite samples, we apply the algorithm to a mixture of stationary and non-stationary models in a variety of settings. We vary the number of clusters k and the number of observations per cluster n as well as the included models. The algorithm is moreover considered both in the case where the number of true clusters are known and when this is unknown. In the latter case, we obtain an additional source of variability as the number of clusters needs to be estimated from the data. Finally, we consider the effect on the simulations of applying a scaling factor η to our similarity matrix. Before we discuss how to determine k , we start by introducing the simulation setting and data-generating processes.

5.1 Simulation setting

We simulate functional autoregressive and functional moving average via their basis representation as follows. A p -th order (time varying) functional autoregressive process (tvFAR(p)), $(X_t, t \in \mathbb{Z})$ can be defined as

$$X_t = \sum_{t'=1}^p A_{t,t'}(X_{t-t'}) + \epsilon_t, \quad (4.1)$$

where $A_{t,1}, \dots, A_{t,p}$ are time-varying autoregressive operators on $L^2([0, 1])$ and $\{\epsilon_t(\tau)\}_{t \in \mathbb{Z}}$ is a sequence of mean zero innovations taking values in $L^2([0, 1])$. To generate such processes, let $\{\psi_l\}_{l=1}^\infty$ be a Fourier basis of $L^2([0, 1])$. By means of a basis expansion, one can show that (see e.g., Aue and van Delft, 2017) the first L_{max} coefficients of (4.1) are generated using the p -th order vector autoregressive, VAR(p), process

$$\tilde{X}_t = \sum_{t'=1}^p \tilde{A}_{t,t'} \tilde{X}_{t-t'} + \tilde{\epsilon}_t,$$

where $\tilde{X}_t := (\langle X_t, \psi_1 \rangle, \dots, \langle X_t, \psi_{L_{max}} \rangle)^T$ is the vector of basis coefficients and the (l, l') -th entry of $\tilde{A}_{t,j}$ is given by $\langle A_{t,j}(\psi_l), \psi_{l'} \rangle$ and $\tilde{\epsilon}_t := (\langle \epsilon_t, \psi_1 \rangle, \dots, \langle \epsilon_t, \psi_{L_{max}} \rangle)^T$. The entries of the matrix $\tilde{A}_{t,j}$ are generated as $N(0, v_{l,l'}^{(t,j)})$ with $v_{l,l'}^{(t,j)}$ specified below. To ensure stationarity or existence of a causal solution the norms $\kappa_{t,j}$ of $A_{t,j}$ are required to satisfy certain conditions (see Bosq (2000) for stationary and van Delft and Eichler (2018) for local stationary time series, respectively). We also consider the following time varying functional moving-average process of order 1:

$$X_{t,T} = B_1(\epsilon_t) - \frac{1}{2} \left(1 + b \cos \left(2\pi \frac{t}{T} \right) \right) B_2(\epsilon_{t-1}) \quad (4.2)$$

where B_1 and B_2 are bounded linear operators on $L^2([0, 1])$ and $b \in \mathbb{R}$. Similarly as above, we use a basis expansion and generate data from the model

$$\tilde{X}_{t,T} = \tilde{B}_1 \tilde{\epsilon}_t - \frac{1}{2} \left(1 + b \cos \left(2\pi \frac{t}{T} \right) \right) \tilde{B}_2 \tilde{\epsilon}_{t-1}$$

where $\tilde{X}_{t,T} = (\langle X_{t,T}, \psi_1 \rangle, \dots, \langle X_{t,T}, \psi_{L_{max}} \rangle)^T$ is the vector of basis coefficients, the (l, l') -th entry of \tilde{B}_1 and \tilde{B}_2 are given by $\langle B_1(\psi_l), \psi_{l'} \rangle$ and $\langle B_2(\psi_l), \psi_{l'} \rangle$ respectively and $\tilde{\epsilon}_t$ is as above.

we consider the following data generating processes:

- (I) The functional white noise variables $\epsilon_1, \dots, \epsilon_T$ i.i.d. with coefficient variances $\text{Var}(\langle \epsilon_t, \psi_l \rangle) = \exp(-(l-1)/10)$.

(II) A FAR(2) variables X_1, \dots, X_T with operators specified by variances $v_{l,l'}^{(1)} = \exp(-l - l')$ and $v_{l,l'}^{(2)} = 1/(l + l'^{3/2})$ with norms $\kappa_1 = 0.75$ and $\kappa_2 = -0.4$ and the innovations $\epsilon_1, \dots, \epsilon_T$ are as in (I).

(III) A MA(1) with $b = 0$ and operators specified by variances $v_{l,l'}^{(1)} = \exp(-l - l')$.

(IV) A tvFAR(1) with operator specified by variances $v_{l,l'}^{(t,1)} = v_{l,l'}^{(1)} = \exp(-l - l')$ and norm $\kappa_1 = 0.8$, and innovations are as in (I) but with a multiplicative time-varying variance

$$\sigma^2(t) = \cos\left(\frac{1}{2} + \cos\left(\frac{2\pi t}{T}\right) + 0.3 \sin\left(\frac{2\pi t}{T}\right)\right).$$

(V) A tvFAR(2) with operators as in (IV), but with time-varying norm

$$\kappa_{1,t} = 1.8 \cos\left(1.5 - \cos\left(\frac{4\pi t}{T}\right)\right)$$

and constant norm $\kappa_2 = -0.81$ and innovations are as in (I).

(VI) A FAR(2) with structural break;

- for $t \leq 3T/8$, the operators are as in (II) with norms $\kappa_1 = 0.7$ and $\kappa_2 = 0.2$, with innovations as in (I).
- for $t > 3T/8$, the operators are as in (II) with norms $\kappa_1 = 0$ and $\kappa_2 = -0.2$, with innovations as in (I) but with coefficient variances $\text{Var}(\langle \epsilon_t, \psi_l \rangle) = 2 \exp((l - 1)/10)$.

Our simulation consists of the following settings:

Setting 1: $k = 3$ with models I, II and III, where the replications per cluster are taken $n = 10$ and $n = 30$

Setting 2: $k = 3$ with models IV, V and VI, where the replications per cluster are taken $n = 10$ and $n = 30$

Setting 3: $k = 6$ with models I-VI, where the replications per cluster are taken $n = 10$ and $n = 30$ and $n = 50$.

Per set-up, we run 500 simulations for both $T = 256$ with $M = 8$ and $T = 512$ with $M = 16$.

For the choice of k we investigated a subset of well-known classical criteria that have been demonstrated to work well in aforementioned comparison studies on classical clustering: the Silhouette Index (Kaufman and Rousseeuw, 1990), the CH Index (Calinski and Harabasz, 1974), the Hartigan Index (Hartigan, 1975) and the KL Index (Krzanowski and Lai, 1988). We respectively refer to these in the tables as ‘sil’, ‘ch’, ‘hartigan’ and ‘kl’. Because these cannot be applied to the spectral embedding directly, the respective criteria were constructed using the similarity graph \hat{W} .

Additionally, we considered two variations of the eigengap heuristic each with a different interpretation of ‘relatively large’ eigengap. Let $0 = \hat{\lambda}_1 \leq \dots \leq \hat{\lambda}_d$ be the estimated eigenvalues of \hat{L} in ascending order

1. ‘Relgap’: define the relative contribution of the k -th eigenvalue as

$$\rho_k = \frac{(\hat{\lambda}_k - \hat{\lambda}_{k-1})}{\hat{\lambda}_k}.$$

Then the rule is to pick k^* as the largest k for which the k -th contribution is still larger than a threshold value that is allowed to depend on the scaling parameter η of the graph (see (4.3) below).

$$k^* = \max_k \{k \in \{1, \dots, k_{\max}\} : \rho_k \leq 0.01\eta\}$$

2. ‘sd1gap’: let

$$\sigma(\hat{\lambda}_{[-1:k]}) = \frac{1}{d-k} \sum_{j=k+1}^d \left(\hat{\lambda}_j - \frac{1}{d-k} \sum_{j=k+1}^d \hat{\lambda}_j \right)^2$$

be the squared deviation from the mean excluding the first k eigenvalues. Then the rule is to pick k^* as the largest k for which the k -th gap is still larger than $\sigma(\hat{\lambda}_{[-1:k]})$, i.e.,

$$k^* = \max_k \{k \in \{1, \dots, k_{\max}\} : \hat{\lambda}_{k+1} - \hat{\lambda}_k \geq \sigma(\hat{\lambda}_{[-1:k]})\}$$

For all criteria, the maximum numbers of clusters to consider was set to $k_{\max} = 15$.

5.2 Simulation results

Table 1 provides the average k chosen according to the different criteria in each setting, while the corresponding percentage of misclustered points averaged over simulations are given in Table 2.

From the first row for $T = 256$ and $T = 512$ of Table 2, we find that our algorithm does very well if the true k is known as it has a very low percentage of misclustered points across the different settings. Based upon the percentage of misclustered points, the most difficult model is clearly setting 2 with $T = 256$. This finding is in accordance with the upper bound in (3.7), which is less tight for larger k and lower estimation precision on \hat{L} .

If the true k is not known we obtain higher percentages of misclustered points, where the percentages appear to be directly caused by the way k was selected. As can be seen from Table 1, the CH Index does best in determining the true number of clusters, while the KL index does worst when the number of true clusters increases. Both the Silhouette Index and the Hartigan Index tend to pick k more conservatively. The two rules based upon the estimated eigenvalues of the graph Laplacian –‘Relgap’ and ‘sd1gap’– appear competitive with the CH index, except in setting 2 for $n = 10$. What is in particular clear is the overall improvement as both n and T increase. It appears that in particular the eigenvalue-based methods suffer from more variation when n and T are small. This is intuitive, since the choice of k directly depends upon the estimation precision of the spectral properties, which can be expected to be more sensitive to estimation error for small T and n (see also Theorem 3.1). From the results in table Table 2, we find the CH index to work best in combination with our algorithm. It is most stable across the different settings and has the lowest percentage of misclustered points, which appears to be a direct consequence of the fact that this index estimated the true number of clusters best and that our algorithm suffers from low error conditionally upon knowing the correct number of clusters.

As explained in Section 2, we apply a local scaling to avoid sensitivity on a global scaling parameter. To investigate the robustness with respect to scaling, we consider our clustering algorithm as explained in Section 3 but with an additional scaling parameter η in the construction of the adjacency matrix. That is, we consider the simulations but with

$$\hat{W} = e^{-\eta \hat{A}} \in \mathbb{R}^{d \times d}. \quad (4.3)$$

where $\eta = \{0.5, 2.5, 5, 10\}$. The results in Table 2 correspond to the choice $\eta = 1$ and in Table 3 we present the four alternative choices. Because of space constraints, we only report them for the

Table 1: *chosen k per method averaged over simulations (standard deviation in brackets)*

method	Setting 1		Setting 2		Setting 3		
	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 50$
$T = 256$							
true k	3	3	3	3	6	6	6
sil	2 (0)	2 (0)	3 (0.1)	3 (0)	5.34 (0.8)	5.24 (0.5)	5.17 (0.4)
ch	3 (0.5)	2.96 (0.2)	3.05 (0.2)	3 (0)	6.37 (0.6)	6.08 (0.3)	6.03 (0.2)
kl	2.19 (1.4)	2.04 (0.7)	3.09 (1)	3.22 (1.4)	10.4 (3.1)	11.13 (3.5)	10.79 (3.6)
hart	3.04 (0.2)	3 (0)	3 (0)	3 (0)	4.86 (0.9)	4.87 (0.7)	4.87 (0.6)
Relgap	3.04 (0.2)	3 (0)	3.92 (1.1)	3.09 (0.3)	5.14 (0.4)	5.15 (0.4)	5.19 (0.4)
sd1gap	3.09 (0.3)	3.03 (0.2)	3.83 (1.2)	3.55 (1)	5.65 (0.9)	6.23 (0.7)	6.41 (0.9)
$T = 512$							
true k	3	3	3	3	6	6	6
sil	2 (0)	2 (0)	3 (0)	3 (0)	5.92 (0.4)	5.95 (0.2)	5.98 (0.1)
ch	3.13 (0.4)	3 (0)	3 (0)	3 (0)	6.52 (0.7)	6.16 (0.4)	6.08 (0.3)
kl	2.44 (1.7)	2.18 (1)	3.05 (0.7)	3.43 (2.1)	9.87 (3.3)	10.45 (3.5)	9.47 (3.5)
hart	3 (0)	3 (0)	3 (0)	3 (0)	5.28 (0.7)	5.4 (0.6)	5.43 (0.5)
Relgap	3 (0)	3 (0)	3.57 (0.8)	3.03 (0.2)	5.53 (0.6)	5.88 (0.3)	5.99 (0.1)
sd1gap	3.09 (0.4)	3.01 (0.1)	3.92 (1.2)	3.4 (0.8)	6.38 (0.9)	6.3 (0.6)	6.2 (0.5)

Table 2: *percentage of misclustered points of the spectral clustering algorithm averaged over simulations (standard deviation in brackets)*

method	Setting 1		Setting 2		Setting 3		
	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 10$	$n = 30$	$n = 50$
$T = 256$							
true	0.11 (0.6)	0.05 (0.2)	0.01 (0.1)	0 (0)	3.22 (6.6)	0.3 (1.1)	0.13 (0.2)
sil	33.33 (0)	33.33 (0)	0.03 (0.4)	0 (0)	13.16 (10.2)	12.97 (8.2)	13.87 (6.9)
ch	4.65 (10.4)	1.41 (6.6)	0.49 (2.6)	0 (0)	5.11 (6.8)	0.82 (2.2)	0.31 (1.1)
kl	33.3 (6.4)	33.48 (2.4)	0.55 (6)	1.45 (9.4)	25.6 (14.7)	25.93 (15.6)	22.59 (14.4)
hart	0.61 (2.6)	0.05 (0.2)	0.01 (0.1)	0 (0)	19.76 (14.6)	18.93 (12.5)	19.02 (10.7)
Relgap	0.57 (2.4)	0.05 (0.2)	10.78 (11.9)	1.27 (4.2)	15.75 (5.6)	14.49 (6.3)	13.68 (6.7)
sd1gap	1.33 (4)	0.46 (2.3)	9.11 (12.5)	6.55 (10.9)	12.33 (7.5)	3.02 (5.5)	2.79 (5.2)
$T = 512$							
true	0 (0)	0 (0)	0 (0)	0 (0)	0.23 (1.7)	0.01 (0.1)	0 (0)
sil	33.33 (0)	33.33 (0)	0.03 (0.4)	0 (0)	2.68 (5.8)	0.77 (3.5)	0.37 (2.4)
ch	1.67 (4.6)	0 (0)	0.44 (2.1)	0 (0)	3.42 (4.4)	1.14 (2.6)	0.61 (2)
kl	27.97 (14.5)	30.63 (10.2)	1.68 (10.3)	2.59 (12.5)	20.93 (16.8)	21.37 (16.3)	15.56 (14.4)
hart	0 (0)	0 (0)	0 (0)	0 (0)	12.25 (12.1)	9.98 (9.4)	9.57 (8.7)
Relgap	0.02 (0.4)	0 (0)	6.92 (9.7)	0.39 (2.2)	8.73 (8.4)	2 (5.4)	0.17 (1.7)
sd1gap	1.13 (4.4)	0.14 (1.4)	10.16 (12.6)	4.99 (9.7)	4.15 (6)	1.92 (3.9)	1.42 (3.4)

specification $T = 512$ and $n = 30$. The results for the other cases show a very similar picture and are available upon request. It can be observed from the first row for each of the settings that the

outcomes are fairly similar when k is known. Variation again therefore seems mostly caused by the way k is chosen, where we find the Silhouette index is sensitive as well as the KL index, but also the eigengap heuristics show some sensitivity for $\eta = 10$. Overall, we may conclude that our method seems capable to detect the correct number of clusters, despite the highly complicated nature of the data. The numerical study moreover suggests that the CH index should be used to find the numbers of clusters if these are unknown.

Table 3: *Percentage of misclustered points and specified k for various values of η averaged over simulations (standard deviation in brackets) for $T = 512, n = 30$*

meth.	% of misclustered points				average chosen k			
	$\eta = 0.5$	$\eta = 2.5$	$\eta = 5$	$\eta = 10$	$\eta = 0.5$	$\eta = 2.5$	$\eta = 5$	$\eta = 10$
Setting 1: $T = 512, n = 30$								
true	0(0)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
sil	33.33(0)	33.33(0)	2.13(8.2)	0(0)	2(0)	2(0)	2(0)	2.94(0.2)
ch	0.03(0.6)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
kl	33.02(5.1)	2.8(9.3)	0(0)	0.27(4.3)	2.08(0.9)	2.18(1)	2.92(0.3)	3(0)
hart	0(0)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
Relgap	0(0)	0.02(0.5)	0.02(0.5)	0.03(0.5)	3(0)	3(0)	3(0)	3(0)
sd1gap	0.05(0.8)	0.88(3.5)	1.08(3.9)	0.6(2.7)	3(0.1)	3.01(0.1)	3.07(0.3)	3.09(0.3)
Setting 2: $T = 512, n = 30$								
true	0(0)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
sil	0(0)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
ch	0(0)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
kl	0.35(4.6)	2.44(11.2)	1.58(9.4)	0.36(4.6)	3.05(0.7)	3.43(2.1)	3.43(2)	3.29(1.7)
hart	0(0)	0(0)	0(0)	0(0)	3(0)	3(0)	3(0)	3(0)
Relgap	0.1(1.2)	2.28(5.8)	6.66(9.3)	7.14(8.5)	3.01(0.1)	3.03(0.2)	3.18(0.5)	3.59(0.9)
sd1gap	4.75(9.5)	5.48(9.7)	4.58(8.7)	2.89(5.9)	3.38(0.8)	3.4(0.8)	3.45(0.9)	3.42(0.8)
Setting 3: $T = 512, n = 30$								
true	0.01(0.1)	0(0)	0.04(0.9)	0.06(0.9)	6(0)	6(0)	6(0)	6(0)
sil	1.03(3.9)	0.03(0.7)	0.04(0.9)	0.06(0.9)	5.95(0.2)	5.95(0.2)	6(0)	6(0)
ch	1.65(3.2)	0.29(1.4)	0.04(0.9)	0.06(0.9)	6.23(0.5)	6.16(0.4)	6.04(0.2)	6(0)
kl	22.72(16.4)	13.94(15.5)	7.4(12.2)	4.56(10.5)	10.54(3.4)	10.45(3.5)	9.2(3.6)	7.84(3.1)
hart	11.48(8)	7.17(13.5)	0.04(0.9)	0.06(0.9)	5.31(0.5)	5.4(0.6)	5.57(0.8)	6(0)
Relgap	7.17(8.3)	0.57(2.2)	3.18(3.9)	3.93(4.4)	5.57(0.5)	5.88(0.3)	6.06(0.3)	6.57(0.8)
sd1gap	1.39(3.4)	4.5(5.1)	6.02(5.6)	4.18(5.1)	6.21(0.5)	6.3(0.6)	6.77(1)	7.2(1.3)

5.3 Testing for equality

We conclude this section with a small investigation of the proposed asymptotic α -level test in (5.6) for the hypothesis of equality of (possibly) time-varying spectral density operators. To investigate the finite samples properties of the test, we performed a simulation study which includes the previously defined stationary models I and II and nonstationary models V and VI with parameter specification $T = 512$ and $M = 16$. The pairwise rejection probabilities at the 5% and 10% over 1000 replications are provided in Table 4, where the diagonal elements correspond to the null hypothesis. We observe a good approximation of the nominal level with model II a bit undersized. The off-diagonal shows good power overall, with model II and model I appearing to be more difficult to distinguish than the other choices. Regardless of the second-order properties being time-varying or not, it appears therefore that the quantiles of the normal distribution

are well-captured for the various models if H_0 is true, while good power is observed under H_A .

	I	II	V	VI		I	II	V	VI
I	10.8	96.6	100.0	100.0	I	5.4	93.1	100.0	100.0
II	96.6	8.0	99.9	100.0	II	93.1	3.2	99.7	100.0
V	100.0	99.9	9.8	100.0	V	100.0	99.7	3.6	100.0
VI	100.0	100.0	100.0	10.0	VI	100.0	100.0	100.0	4.7

Table 4: Rejection probabilities of the pairwise equality test (5.6) at the 10% (left); and 5% level.

6 Data Application

We illustrate our clustering algorithm by means of an application to high resolution infrared surface temperature measurements recorded during the years 2015-2018 in the conterminous U.S., Alaska and Hawaii. The exact locations of the 126 included stations in our study are depicted in Figure 6.1.

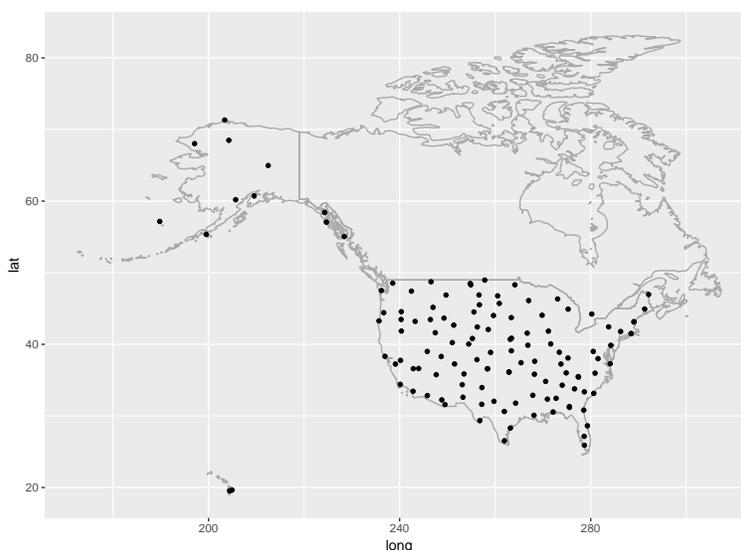


Figure 6.1: Measurement stations projected upon the map of North America according to their longitude (long) and latitude (lat) coordinates.

The measurements are part of a quality controlled dataset (Diamond et al., 2013) monitored within the United States Climate Reference Network (USCRN). These are publicly available on the website of the NOAA U.S. government agency.

Understanding the characteristics of land surface temperature is of interest in various applications as it influences for example surface energy balance and various soil hydrologic processes. Our goal is to investigate whether we can identify spatial or geographical patterns in terms of variation in the surface temperature. Variation in surface temperature can be expected to be influenced by numerous factors such as wind, elevation, proximity of major water bodies or yet land cover types. For each location $i = 1, \dots, 126$, we have of a sample of observations $\{X_{i,t}(\tau_1), \dots, X_{i,t}(\tau_{\max})\}_{t=1, \dots, 1288}$, where $X_{i,t}$ represents the sampled temperature curve on day t

for location i . The first measurement day $t = 1$ corresponds to February 28 2015 and the last measurement day $t = T = 1288$ corresponds to September 7 2018. The recordings were made at a 5 minute time-interval leading to a maximum of 288 observations per day. We remark that the measurement stations included have at most 10% of missing observations per day over the measurement period. The gridded observations were transformed into functional objects using 21 Fourier basis functions, rescaled to the unit interval, i.e., τ_1 corresponds to local time 0.00am and $\tau_{288} = \tau_{max}$ to local time 23.55pm. A few sets of realized curves are depicted in Figure 6.2.

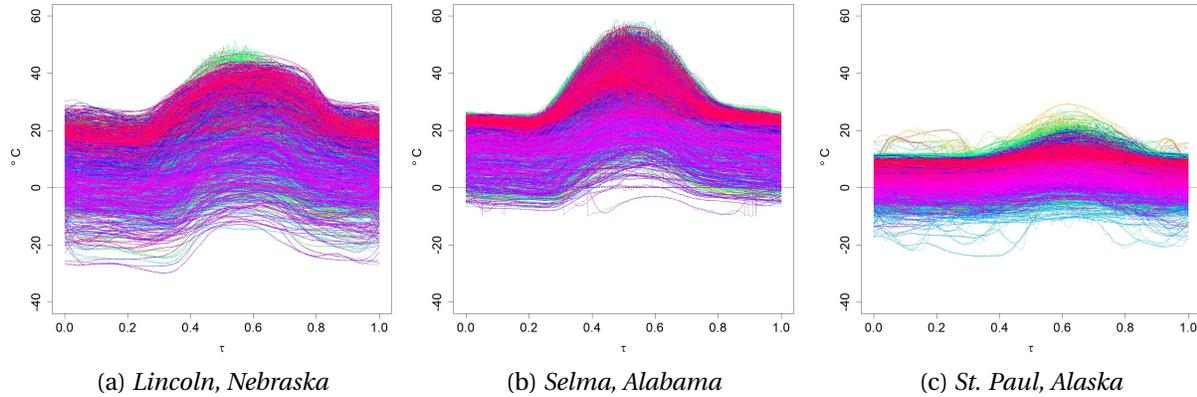


Figure 6.2: Daily surface temperature curves (solid) fitted to the observations (dotted) for the period 28-Feb. 2015-07 Sept. 2018.

It can be observed there is more between-curves variation for Lincoln, Nebraska while there is more within-curve variation for Selma, Alabama. For St. Paul, the more visible variety in colors indicate a stronger variation over longer periods of time. To cluster the locations, we factorize the 1288 daily curves into $M = 14$ complete seasons, each of length $N = 92$. We applied our algorithm described in Section 3. For the choice of k , we looked at the various criteria discussed in the previous section. These indicated 2 to 4 clusters, with the majority picking 3 clusters. To choose the number of clusters, we used the index that performed best in the simulation study, which was the CH index and which indicates 3 clusters. The results of clustering the locations in 3 clusters by means of our algorithm are given in Figure 6.3. The figure seems to indicate the seasonal variation for two clusters is related to two prevailing wind patterns; those in red to the westerlies wind, while those in green to northeasterly trade wind. We moreover observe a separate cluster consisting of mainly locations on the coast in a bay area, island or on a peninsula. In particular, these locations correspond to areas known to be more affected by different atmospheric pressures which are responsible for forming cyclones and that can vary with the El Niño-Southern Oscillations cycle. It is worth remarking that a similar pattern was observed when different time periods were considered.

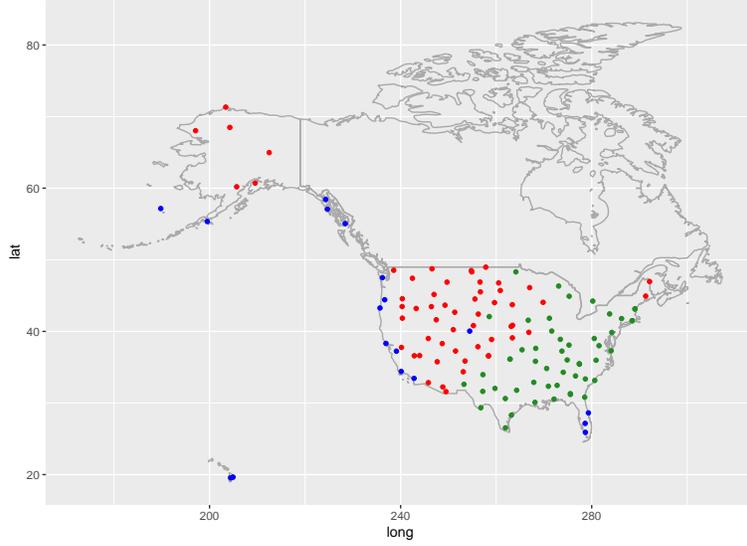


Figure 6.3: Measurement stations clustered according to variation in surface temperature

Acknowledgements This work has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823, Teilprojekt A1, A7, C1) of the German Research Foundation (DFG).

Appendix A Background material and auxiliary results

Let \mathcal{H}_i for each $i = 1, \dots, n$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. The tensor of these is denoted by

$$\mathcal{H} := \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_n = \bigotimes_{i=1}^n \mathcal{H}_i$$

If $\mathcal{H}_i = \mathcal{H} \forall i$, then this is the n -th fold tensor product of \mathcal{H} . For $A_i \in \mathcal{H}_i, 1 \leq i \leq n$ the object $\bigotimes_{i=1}^n A_i$ is a multi-antilinear functional that generates a linear manifold, the usual algebraic tensor product of vector spaces \mathcal{H}_i , to which the scalar product

$$\left\langle \bigotimes_{i=1}^n A_i, \bigotimes_{i=1}^n B_i \right\rangle = \prod_{i=1}^n \langle A_i, B_i \rangle$$

can be extended to a pre-Hilbert space. The completion of the above algebraic tensor product is $\bigotimes_{i=1}^n \mathcal{H}_i$.

For $A, B, C \in \mathcal{L}(\mathcal{H})$ we define the following bounded linear mappings. The kronecker product is defined as $(A \widehat{\otimes} B)C = ACB^\dagger$, while the transpose Kronecker product is given by $(A \widehat{\otimes}_\top B)C = (A \widehat{\otimes} \bar{B})C^\dagger$. For $A, B, C \in S_2(\mathcal{H})$, we shall denote, in analogy to elements $a, b \in \mathcal{H}$, the Hilbert tensor product as $A \otimes B$. We list the following useful properties:

Properties A.1. Let $\mathcal{H}_i = L^2_{\mathbb{C}}([0, 1]^k)$ for $i = 1, \dots, n$. Then for $a_i, b_i \in \mathcal{H}_i$ and $A_i, B_i \in S_2(\mathcal{H}_i)$, we have

1. $\langle A, B \rangle_{HS} = \text{Tr}(AB^\dagger)$

2. $\langle \otimes_{i=1}^n A_i, \otimes_{i=1}^n B_i \rangle_{HS} = \prod_{i=1}^n \langle A_i, B_i \rangle_{HS}$
3. $\langle a_1 \otimes a_2, b_1 \otimes b_2 \rangle_{HS} = \langle a_1 \otimes \bar{a}_2, b_1 \otimes \bar{b}_2 \rangle_{\mathcal{H}_1 \otimes \mathcal{H}_2} = \langle a_1, b_1 \rangle \overline{\langle a_2, b_2 \rangle}$
4. If $A_i \in S_1(H)$, then $\prod_{i=1}^n \text{Tr}(A_i) = \text{Tr}(\tilde{\otimes}_{i=1}^n A_i)$
5. $((a_1 \otimes \bar{a}_2) \otimes (a_3 \otimes \bar{a}_4)) = ((a_1 \otimes a_3) \tilde{\otimes} (\bar{a}_2 \otimes \bar{a}_4)) = ((a_1 \otimes a_4) \tilde{\otimes}_T a_2 \otimes a_3)$

Let X be a random element on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ that takes values in a separable Hilbert space H . More precisely, we endow H with the topology induced by the norm on H and assume that $X : \Omega \rightarrow H$ is Borel-measurable. The k -th order cumulant tensor is defined by (van Delft and Eichler, 2018)

$$\text{Cum}(X_1, \dots, X_k) = \sum_{l_1, \dots, l_k \in \mathbb{N}} \text{Cum}(\langle X_1, \psi_{l_1} \rangle, \dots, \langle X_k, \psi_{l_k} \rangle) (\psi_{l_1} \otimes \dots \otimes \psi_{l_k}), \quad (6.2)$$

where $\{\psi_l\}_{l \in \mathbb{N}}$ is an orthonormal basis of H and the cumulants on the right hand side are as usual given by

$$\text{Cum}(\langle X_1, \psi_{l_1} \rangle, \dots, \langle X_k, \psi_{l_k} \rangle) = \sum_{\nu=(\nu_1, \dots, \nu_p)} (-1)^{p-1} (p-1)! \prod_{r=1}^p \mathbb{E} \left[\prod_{t \in \nu_r} \langle X_t, \psi_{l_t} \rangle \right],$$

where the summation extends over all unordered partitions ν of $\{1, \dots, k\}$. The product theorem for cumulants (Brillinger, 1981, Theorem 2.3.2) can then be generalised (see e.g. van Delft and Eichler, 2018) to simple tensors of random elements of H , i.e., $X_t = \otimes_{j=1}^{J_t} X_{tj}$ with $j = 1, \dots, J_t$ and $t = 1, \dots, k$. The joint cumulant tensor is then be given by

$$\text{Cum}(X_1, \dots, X_k) = \sum_{\nu=(\nu_1, \dots, \nu_p)} S_\nu \left(\otimes_{n=1}^p \text{Cum}(X_{tj} | (t, j) \in \nu_n) \right), \quad (6.3)$$

where S_ν is the permutation that maps the components of the tensor back into the original order, that is, $S_\nu(\otimes_{r=1}^p \otimes_{(t,j) \in \nu_r} X_{tj}) = X_{11} \otimes \dots \otimes X_{kJ_t}$.

Appendix B Distributional properties of the similarity measure

In order to derive the distributional properties of $\hat{\mathcal{A}}_{i_1, i_2}$ defined in (2.7), we require the following assumptions on the functional processes $\{X_{i,t,T} : t \in \mathbb{Z}\}_{T \in \mathbb{N}}, i = 1, \dots, d$

Assumption B.1. Assume $\{X_{i,t,T} : t \in \mathbb{Z}\}_{T \in \mathbb{N}}$ are d locally stationary zero-mean stochastic processes taking values in $\mathcal{H}_{\mathbb{R}}$ and let $\kappa_{m; t_1, \dots, t_{m-1}} : L^2([0, 1]^{[m/2]}) \rightarrow L^2([0, 1]^{[\frac{m+1}{2}]})$ be a positive operator independent of T such that, for all $j = 1, \dots, m-1$ and some $\ell \in \mathbb{N}$,

$$\sum_{t_1, \dots, t_{k-1} \in \mathbb{Z}} (1 + |t_j|^\ell) \|\kappa_{m; t_1, \dots, t_{m-1}}\|_1 < \infty. \quad (6.4)$$

Let us denote

$$Y_{i,t}^{(T)} = X_{i,t,T} - X_{i,t}^{(t/T)} \quad \text{and} \quad Y_{i,t}^{(u,v)} = \frac{X_{i,t}^{(u)} - X_{i,t}^{(v)}}{(u-v)} \quad (6.5)$$

for $T \geq 1$, $1 \leq t \leq T$ and $u, v \in [0, 1]$ such that $u \neq v$. Suppose furthermore that m -th order joint cumulants satisfy

$$(i) \quad \|\text{Cum}(X_{i_1, t_1, T}, \dots, X_{i_{m-1}, t_{m-1}, T}, Y_{i_m, t_m}^{(T)})\|_2 \leq \frac{1}{T} \|\kappa_{m; t_1 - t_m, \dots, t_{m-1} - t_m}\|_1,$$

- (ii) $\|\text{Cum}(X_{i_1, t_1}^{(u_1)}, \dots, X_{i_{m-1}, t_{m-1}}^{(u_{m-1})}, Y_{i_m, t_m}^{(u_m, v)})\|_2 \leq \|\kappa_{m; t_1-t_m, \dots, t_{m-1}-t_m}\|_1,$
- (iii) $\sup_u \|\text{Cum}(X_{i_1, t_1}^{(u)}, \dots, X_{i_{m-1}, t_{m-1}}^{(u)}, X_{i_m, t_m}^{(u)})\|_2 \leq \|\kappa_{m; t_1-t_m, \dots, t_{m-1}-t_m}\|_1,$
- (iv) $\sup_u \|\frac{\partial^\ell}{\partial u^\ell} \text{Cum}(X_{i_1, t_1}^{(u)}, \dots, X_{i_{m-1}, t_{m-1}}^{(u)}, X_{i_m, t_m}^{(u)})\|_2 \leq \|\kappa_{m; t_1-t_m, \dots, t_{m-1}-t_m}\|_1,$

We remark that if the process is in fact stationary, the dependence on localized time u drops. As explained in Section 2, the estimator requires splitting the sample $T \in \mathbb{N}$ as $T = N(T)M(T)$, where $N(T)$ defines the resolution in frequency of the local fDFT and $M(T)$ controls the number of nonoverlapping local fDFT's in (2.6). Since they correspond to the number of terms used in a Riemann sum approximating the integrals with respect to du and $d\omega$ they have to be sufficiently large. We assume

Assumption B.2. $M \rightarrow \infty, N \rightarrow \infty$ as $T \rightarrow \infty$, such that

$$N/M \rightarrow \infty \quad \text{and} \quad N/M^3 \rightarrow 0$$

The number of elements in the blocks grows therefore must grow faster than the number of blocks, but slower than the cube number of blocks. The choice of the number of blocks is carefully discussed in van Delft et al. (2018), who used the integrated periodogram operators as a basis for a stationarity test.

In order to derive the results in Section 4, we require four auxiliary results, which can be proved using similar arguments as given in van Delft et al. (2018) and the proofs are therefore omitted. The first one expresses the cumulants of Hilbert-Schmidt inner products of local periodogram tensors, into the trace of cumulants of simple tensors of the local functional DFT's.

Theorem B.1. Let $\mathbb{E} \prod_{i_l \in \{1, \dots, d\}}^{2n} \|I_{i_l}^{u, \omega}\|_2 < \infty$ for some $n \in \mathbb{N}$ and uniformly in u and ω . Then for any $i_1, \dots, i_{2n} \in \{1, \dots, d\}$.

$$\begin{aligned} & \text{Cum}\left(\langle I_{i_1}^{u_{j_1}, \omega_{k_1}}, I_{i_2}^{u_{j_2}, \omega_{k_2}} \rangle_{HS}, \dots, \langle I_{i_{2n-1}}^{u_{j_{2n-1}}, \omega_{k_{2n-1}}}, I_{i_{2n}}^{u_{j_{2n}}, \omega_{k_{2n}}} \rangle_{HS}\right) \\ &= \text{Tr}\left(\sum_{\mathbf{P}=P_1 \cup \dots \cup P_G} \text{SP}\left(\otimes_{g=1}^G \text{cum}\left(D_{i_p}^{u_{j_p}, \omega_{k_p}} \mid p \in P_g\right)\right)\right). \end{aligned}$$

where the summation is over all indecomposable partitions $\mathbf{P} = P_1 \cup \dots \cup P_G$ of the array

$$\begin{array}{cccc} (1, 1) & (1, 2) & (1, 3) & (1, 4) \\ (2, 1) & (2, 2) & (2, 3) & (2, 4) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ (n, 1) & (n, 2) & (n, 3) & (n, 4) \end{array} \tag{6.6}$$

where $p = (l, m)$ and $i_p = i_{2l-\delta_{\{m \in \{1, 2\}\}}}$, $k_p = (-1)^m k_{2l-\delta_{\{m \in \{1, 2\}\}}}$ and $j_p = j_{2l-\delta_{\{m \in \{1, 2\}\}}}$ for $l \in \{1, \dots, n\}$ and $m \in \{1, 2, 3, 4\}$ and where $\delta_{\{A\}}$ equals 1 if event A occurs and 0 otherwise.

We additionally require auxiliary results on the properties of the local fDFTs.

Lemma B.1. Suppose Assumption B.1 is satisfied k moments and $\sum_{j=1}^k \omega_j \equiv 0 \pmod{2\pi}$ then

$$\left\| \text{Cum}\left(D_i^{u_j, \omega_1}, \dots, D_i^{u_j, \omega_k}\right) - \frac{(2\pi)^{1-k/2}}{N^{k/2-1}} \mathcal{F}_{u_i, \omega_1, \dots, \omega_{k-1}} \right\|_1 = O\left(N^{-k/2} \times \frac{N}{M^2}\right).$$

additionally, for $|i_1| = |i_2| = k/2$, we have

$$\left\| \text{Cum} \left(D_i^{u_j, \omega_1}, \dots, D_i^{u_j, \omega_k} \right) - \frac{(2\pi)^{1-k/2}}{N^{k/2-1}} \mathcal{F}_{u_i, \omega_1, \dots, \omega_{k-1}}^{i_1, i_2} \right\|_1 = O \left(N^{-k/2} \times \frac{N}{M^2} \right).$$

where $\mathcal{F}_{u_i, \omega_1, \dots, \omega_{k-1}}^{i_1, i_2}$ denotes the cross spectral density operator of series i_1 and i_2 of order k at time u_i .

When evaluated off the manifold, i.e., $\sum_{j=1}^k \omega_j \neq 0 \pmod{2\pi}$, the above cumulant is of lower order (see Corollary B.1). A direct consequence of the proof of Lemma B.1 is the following corollary

Corollary B.1. For $i_1, \dots, i_k \in \{1, \dots, d\}$ We have for any $p \geq 1$

$$\left\| \text{Cum} \left(D_{i_1}^{u_1, \omega_1}, \dots, D_{i_k}^{u_k, \omega_k} \right) \right\|_p = O \left(N^{1-k/2} \right) \quad (6.7)$$

uniformly in $\omega_1, \dots, \omega_k$ and u_1, \dots, u_k . Moreover, if $\sum_{j=1}^k \omega_j \neq 0 \pmod{2\pi}$ then

$$\left\| \text{Cum} \left(D_{i_1}^{u_1, \omega_1}, \dots, D_{i_k}^{u_k, \omega_k} \right) \right\|_p = O \left(N^{-k/2} \right). \quad (6.8)$$

Additionally, when the local fDFT's are evaluated on different midpoints then we have the following lemma.

Lemma B.2. If Assumption B.1 is satisfied and $|j_1 - j_2| > 1$ for some midpoints u_{j_1} and u_{j_2} then

$$\left\| \text{Cum} \left(D_{i_1}^{u_1, \omega_1}, \dots, D_{i_k}^{u_k, \omega_k} \right) \right\|_1 = O \left(N^{-k/2} M^{-1} \right)$$

uniformly in $\omega_1, \dots, \omega_k$.

In the following, for random variables Y_0, Y_1, Y_2, Y_3 let $\text{cum}_{m_0, m_1, m_2, m_3}(Y_0, Y_1, Y_2, Y_3)$ denote the joint cumulant

$$\text{cum} \left(\underbrace{Y_0, \dots, Y_0}_{m_0 \text{ times}}, \underbrace{Y_1, \dots, Y_1}_{m_1 \text{ times}}, \underbrace{Y_2, \dots, Y_2}_{m_2 \text{ times}}, \underbrace{Y_3, \dots, Y_3}_{m_3 \text{ times}} \right),$$

where $0 \leq m_i \leq m, i = 0, 1, 2, 3$ s.t. $\sum_{i=0}^3 m_i = m$. Using the previous statements, we can derive the order of the higher order joint cumulants of elements defined in (2.6).

Theorem B.2. If Assumption B.1 is satisfied then for finite m

$$T^{m/2} \text{cum}_{m_0, m_1, m_2, m_3}(\hat{F}_{i_1, i_2}, \hat{F}_{i_3, i_4}, \hat{F}_{i_5, i_6}, \hat{F}_{i_7, i_8}) \\ \frac{1}{T^{m/2}} \sum_{k_1, \dots, k_m=1}^{\lfloor N/2 \rfloor} \sum_{j_1, \dots, j_m=1}^M \text{Tr} \left(\sum_{P=P_1 \cup \dots \cup P_G} S_P \left(\otimes_{g=1}^G \text{cum} \left(D_{i_p}^{u_{j_p}, \omega_{k_p}} \mid p \in P_g \right) \right) \right) = O(T^{1-m/2}).$$

uniformly in $0 \leq m_i \leq m$ s.t. $\sum_{i=0}^3 m_i = m$.

Proof of Theorem B.2. For a fixed partition $P = \{P_1, \dots, P_G\}$, let the cardinality of set P_g be denoted by $|P_g| = \mathcal{C}_g$. By (6.7) of Corollary B.1 and Lemma B.2 an upperbound of (6.6) is given by

$$O \left(T^{-m/2} \sum_{k_1, \dots, k_m=1}^{\lfloor N/2 \rfloor} \sum_{j_1, \dots, j_m=1}^M \prod_{g=1}^G \frac{1}{N^{\mathcal{C}_g/2-1}} M^{-\delta_{\{\exists p_1, p_2 \in P_g: |j_{p_1} - j_{p_2}| > 1\}}} \right) \quad (6.9)$$

Similar to Lemma 4.3 of van Delft et al. (2018), we can show inductively that the indecomposability of the array (6.6) and the behavior of the joint cumulants of the local fDFT's at different midpoints imply this is at most of order

$$O(N^{m/2} M^{-m/2} E^m M N^{-2m+G}) = O(T^{1-m/2} N^{G-m-1}).$$

Thus, partitions of size $G \leq m + 1$ will vanish as $T \rightarrow \infty$. For $G \geq m + 2$, indecomposability of the array requires to stay on the frequency manifold (see equation (6.8) of Corollary B.1) and therefore imposes additional restrictions in frequency direction. It can be shown (van Delft et al., 2018, Proposition 4.1) that for a partition of size $G = m + r_1 + 1$ with $r_1 \geq 1$ of the array (6.6) only partitions with at least r_1 restrictions in frequency direction are indecomposable if $m > 2$, while if $m = 2$ there must be at least 1 restriction in frequency direction. Consequently, the joint cumulant is at most of order $O(T^{1-n/2} N^{m+r_1+1-m-1} N^{-r_1}) = O(T^{1-n/2})$. \square

Proof of Theorem 2.1. Using Theorem B.1 with $n = 1$ implies

$$\mathbb{E}F_{i_1, i_2} = \frac{1}{T} \sum_{k=1}^{\lfloor N/2 \rfloor} \sum_{j=1}^M \text{Tr} \left(\mathbb{E} [D_{i_1}^{u_{j_1}, \omega_{k_1}} \otimes D_{i_1}^{u_{j_1}, -\omega_{k_1}} \otimes D_{i_2}^{u_{j_1}, -\omega_{k_1-1}} \otimes D_{i_2}^{u_{j_1}, \omega_{k_1-1}}] \right).$$

Rewriting this expectation in cumulant tensors, we get

$$\begin{aligned} \mathbb{E}F_{i_1, i_2} &= \frac{1}{T} \sum_{k=1}^{\lfloor N/2 \rfloor} \sum_{j=1}^M \text{Tr} \left(S_{1234} \text{Cum} \left(D_{i_1}^{u_{j_1}, \omega_{k_1}}, D_{i_1}^{u_{j_1}, -\omega_{k_1}}, D_{i_2}^{u_{j_1}, -\omega_{k_1-1}}, D_{i_2}^{u_{j_1}, \omega_{k_1-1}} \right) \right) \\ &+ \frac{1}{T} \sum_{k=1}^{\lfloor N/2 \rfloor} \sum_{j=1}^M \text{Tr} \left(S_{1234} \left(\text{Cum} \left(D_{i_1}^{u_{j_1}, \omega_{k_1}}, D_{i_1}^{u_{j_1}, -\omega_{k_1}} \right) \otimes \text{Cum} \left(D_{i_2}^{u_{j_1}, -\omega_{k_1-1}}, D_{i_2}^{u_{j_1}, \omega_{k_1-1}} \right) \right) \right) \\ &+ \frac{1}{T} \sum_{k=1}^{\lfloor N/2 \rfloor} \sum_{j=1}^M \text{Tr} \left(S_{1324} \left(\text{Cum} \left(D_{i_1}^{u_{j_1}, \omega_{k_1}}, D_{i_2}^{u_{j_1}, -\omega_{k_1-1}} \right) \otimes \text{Cum} \left(D_{i_1}^{u_{j_1}, -\omega_{k_1}}, D_{i_2}^{u_{j_1}, \omega_{k_1-1}} \right) \right) \right) \\ &+ \frac{1}{T} \sum_{k=1}^{\lfloor N/2 \rfloor} \sum_{j=1}^M \text{Tr} \left(S_{1423} \left(\text{Cum} \left(D_{i_1}^{u_{j_1}, \omega_{k_1}}, D_{i_2}^{u_{j_1}, \omega_{k_1-1}} \right) \otimes \text{Cum} \left(D_{i_1}^{u_{j_1}, -\omega_{k_1}}, D_{i_2}^{u_{j_1}, -\omega_{k_1-1}} \right) \right) \right) \end{aligned}$$

By Corollary B.1 and Lemma B.1 we thus find

$$\mathbb{E}F_{i_1, i_2} = \frac{1}{T} \sum_{k=1}^{\lfloor N/2 \rfloor} \sum_{j=1}^M \langle \mathcal{F}_{u_j, \omega_k}^{i_1}, \mathcal{F}_{u_j, \omega_{k-1}}^{i_2} \rangle_{HS} + O\left(\frac{1}{M^2}\right) + O\left(\frac{1}{N}\right).$$

Hence,

$$\lim_{N, M \rightarrow \infty} \mathbb{E}F_{i_1, i_2} = \frac{1}{2\pi} \int_0^\pi \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_2} \rangle_{HS} du d\omega = \frac{1}{4\pi} \int_{-\pi}^\pi \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_2} \rangle_{HS} du d\omega.$$

Secondly, we have for any $i_1, i_2, i_3, i_4 \in \{1, \dots, d\}$

$$TCov(F_{i_1, i_2}, F_{i_3, i_4}) = TCum(F_{i_1, i_2}, \overline{F_{i_3, i_4}})$$

Hence, if Assumption B.1 is satisfied with $m = 8$, Theorem B.2 implies this term is of order $O(1)$. summarizing, we find $F_{i_1, i_1}, F_{i_2, i_2}, F_{i_1, i_2}$ and F_{i_2, i_1} are asymptotically unbiased and jointly convergence in probability. The continuous mapping theorem establishes then that $\hat{\mathcal{A}}_{i_1, i_2}$ is a \sqrt{T} -consistent estimator of \mathcal{A}_{i_1, i_2} for any $i_1, i_2, \in \{1, \dots, d\}$. \square

Proof of Theorem 4.1. If Assumption B.1 holds for all moments, then Theorem B.2, yields that for $m > 2$

$$T^{m/2} \text{cum}_{m_0, m_1, m_2, m_3}(F_{i_1, i_2}, F_{i_3, i_4}, F_{i_5, i_6}, F_{i_7, i_8}) \rightarrow 0 \quad \text{as } T \rightarrow \infty,$$

from which asymptotic joint normality of $F_{i_1, i_2}, F_{i_3, i_4}, F_{i_5, i_6}, F_{i_7, i_8}$ follows, i.e., we have

$$\sqrt{T} \begin{pmatrix} 4\pi F_{i_1, i_1} - \mathbb{E}(F_{i_1, i_1}) \\ 4\pi F_{i_2, i_2} - \mathbb{E}(F_{i_2, i_2}) \\ 4\pi F_{i_1, i_2} - \mathbb{E}(F_{i_1, i_2}) \\ 4\pi F_{i_2, i_1} - \mathbb{E}(F_{i_2, i_1}) \end{pmatrix} \rightarrow \mathcal{N}(\mathbf{0}, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \text{Var}(F_{i_1, i_1}) & \text{Cov}(F_{i_1, i_1}, F_{i_2, i_2}) & \text{Cov}(F_{i_1, i_1}, F_{i_1, i_2}) & \text{Cov}(F_{i_1, i_1}, F_{i_2, i_1}) \\ \text{Cov}(F_{i_1, i_1}, F_{i_2, i_2}) & \text{Var}(F_{i_2, i_2}) & \text{Cov}(F_{i_2, i_2}, F_{i_1, i_2}) & \text{Cov}(F_{i_2, i_2}, F_{i_2, i_1}) \\ \text{Cov}(F_{i_1, i_1}, F_{i_1, i_2}) & \text{Cov}(F_{i_2, i_2}, F_{i_1, i_2}) & \text{Var}(F_{i_1, i_2}) & \text{Cov}(F_{i_1, i_2}, F_{i_2, i_1}) \\ \text{Cov}(F_{i_1, i_1}, F_{i_2, i_1}) & \text{Cov}(F_{i_2, i_2}, F_{i_2, i_1}) & \text{Cov}(F_{i_1, i_2}, F_{i_2, i_1}) & \text{Var}(F_{i_2, i_1}) \end{pmatrix}. \quad (6.10)$$

To derive from this the distribution of $\hat{\mathcal{A}}_{i_1, i_2}$, consider the function $g: \mathbb{R}^4 \rightarrow \mathbb{R}$

$$g(x_1, x_2, x_3, x_4) = 1 - \frac{x_3}{(x_1 + x_2)} - \frac{x_4}{(x_1 + x_2)}$$

of which the gradient is given by

$$\nabla g^\top(\mathbf{x}) = \begin{pmatrix} x_3(x_1 + x_2)^{-2} + x_4(x_1 + x_2)^{-2} \\ x_3(x_1 + x_2)^{-2} + x_4(x_1 + x_2)^{-2} \\ -(x_1 + x_2)^{-1} \\ -(x_1 + x_2)^{-1} \end{pmatrix} = \frac{1}{(x_1 + x_2)} \begin{pmatrix} \frac{x_3 + x_4}{(x_1 + x_2)} \\ \frac{x_3 + x_4}{(x_1 + x_2)} \\ -1 \\ -1 \end{pmatrix}.$$

Then since we can write

$$\hat{\mathcal{A}}_{i_1, i_2} = g(F_{i_1, i_1}, F_{i_2, i_2}, F_{i_1, i_2}, F_{i_2, i_1}) = 1 - \frac{F_{i_1, i_2}}{(F_{i_1, i_1} + F_{i_2, i_2})} - \frac{F_{i_2, i_1}}{(F_{i_1, i_1} + F_{i_2, i_2})},$$

the Delta method implies, that as $T \rightarrow \infty$,

$$\left\{ \sqrt{T} (\hat{\mathcal{A}}_{i_1, i_2} - \mathcal{A}_{i_1, i_2}) \right\}_{\{i_1, i_2 \in [d]\}} \rightarrow \mathcal{N}(\mathbf{0}, \nabla g^\top(\mathbf{x}) \Sigma \nabla g(\mathbf{x})), \quad (6.11)$$

where for fixed $i_1, i_2 \in \{1, \dots, d\}$,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_1} \rangle_{HS} du d\omega \\ \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_2}, \mathcal{F}_{u, \omega}^{i_2} \rangle_{HS} du d\omega \\ \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_2} \rangle_{HS} du d\omega \\ \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_2}, \mathcal{F}_{u, \omega}^{i_1} \rangle_{HS} du d\omega, \end{pmatrix} \quad (6.12)$$

and Σ is defined in (6.10). Next, consider the covariance structure. We have,

$$TCov(F_{i_1, i_2}, F_{i_3, i_4}) = TCum \left(\frac{1}{T} \sum_{k_1=1}^{\lfloor N/2 \rfloor} \sum_{j_1=1}^M \langle I_{i_1}^{u_{j_1}, \omega_{k_1}}, I_{i_2}^{u_{j_1}, \omega_{k_1-1}} \rangle_{HS}, \frac{1}{T} \sum_{k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_2=1}^M \langle I_{i_3}^{u_{j_2}, \omega_{k_2}}, I_{i_4}^{u_{j_2}, \omega_{k_2-1}} \rangle_{HS} \right)$$

By Theorem B.2, this satisfies

$$TCov(F_{i_1, i_2}, F_{i_3, i_4}) = \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \text{Cum} \left(\text{Tr} (D_{i_1}^{u_{j_1}, \omega_{k_1}} \otimes D_{i_1}^{u_{j_1}, -\omega_{k_1}} \otimes D_{i_2}^{u_{j_1}, -\omega_{k_1-1}} \otimes D_{i_2}^{u_{j_1}, \omega_{k_1-1}}) \right),$$

$$\begin{aligned} & \text{Tr} \left(D_{i_3}^{u_{j_2}, -\omega_{k_2}} \otimes D_{i_3}^{u_{j_2}, \omega_{k_2}} \otimes D_{i_4}^{u_{j_2}, \omega_{k_2-1}} \otimes D_{i_4}^{u_{j_2}, -\omega_{k_2-1}} \right) \\ &= \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \text{Tr} \left(\sum_{\mathbf{P}=P_1 \cup \dots \cup P_G} S_{\mathbf{P}} \left(\otimes_{g=1}^G \text{cum} \left(D_{i_p}^{u_{j_p}, \omega_{k_p}} \mid p \in P_g \right) \right) \right) \end{aligned}$$

where $p = (l, m)$ with $k_p = (-1)^{l-m} k_l - \delta_{\{m \in \{3,4\}\}}$, $j_p = j_l$ and $i_p = i_{2l - \delta_{\{m \in \{1,2\}\}}}$ for $l \in \{1,2\}$ and $m \in \{1,2,3,4\}$. In particular, we are interested in all indecomposable partitions of the array where the summation is overall indecomposable partitions of the array

$$\begin{array}{cccc} \underbrace{D_{i_1}^{u_{j_1}, \omega_{k_1}}}_1 & \underbrace{D_{i_1}^{u_{j_1}, -\omega_{k_1}}}_2 & \underbrace{D_{i_2}^{u_{j_1}, -\omega_{k_1-1}}}_3 & \underbrace{D_{i_2}^{u_{j_1}, \omega_{k_1-1}}}_4 \\ \underbrace{D_{i_3}^{u_{j_2}, -\omega_{k_2}}}_5 & \underbrace{D_{i_3}^{u_{j_2}, \omega_{k_2}}}_6 & \underbrace{D_{i_4}^{u_{j_2}, \omega_{k_2-1}}}_7 & \underbrace{D_{i_4}^{u_{j_2}, -\omega_{k_2-1}}}_8 \end{array}$$

The significant partitions of the form Cum₄Cum₂Cum₂ are

$$\begin{aligned} & \text{Tr} \left(S_{(1256)(34)(78)} \left(\delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, \omega_{k_1}, -\omega_{k_1}, -\omega_{k_2}}^{i_1, i_3} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} + \mathcal{E}_2 \right) \right] \right) \right) \\ & \text{Tr} \left(S_{(1278)(34)(56)} \left(\delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, \omega_{k_1}, -\omega_{k_1}, \omega_{k_2-1}}^{i_1, i_4} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} + \mathcal{E}_2 \right) \right] \right) \right) \\ & \text{Tr} \left(S_{(3456)(12)(78)} \left(\delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}, \omega_{k_1-1}, -\omega_{k_2}}^{i_2, i_3} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} + \mathcal{E}_2 \right) \right] \right) \right) \\ & \text{Tr} \left(S_{(3478)(12)(56)} \left(\delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}, \omega_{k_1-1}, \omega_{k_2-1}}^{i_2, i_4} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} + \mathcal{E}_2 \right) \right] \right) \right). \end{aligned}$$

The significant partitions of the form Cum₂Cum₂Cum₂Cum₂ are

$$\begin{aligned} & \text{Tr} \left(S_{(12)(37)(56)(48)} \left(\delta_{j_1, j_2} \delta_{k_1, k_2} \left[\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2, i_4} \otimes \mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} \otimes \mathcal{F}_{u_{j_1}, \omega_{k_1-1}}^{i_2, i_4} + \mathcal{E}_2 \right] \right) \right) \\ & \text{Tr} \left(S_{(12)(36)(78)(45)} \left(\delta_{j_1, j_2} \delta_{k_1-1, k_2} \left[\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2, i_3} \otimes \mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} \otimes \mathcal{F}_{u_{j_1}, \omega_{k_1-1}}^{i_2, i_3} + \mathcal{E}_2 \right] \right) \right) \\ & \text{Tr} \left(S_{(15)(26)(37)(48)} \left(\delta_{j_1, j_2} \delta_{k_1, k_2} \left[\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1, i_3} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{i_1, i_3} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2, i_4} \otimes \mathcal{F}_{u_{j_1}, \omega_{k_1-1}}^{i_2, i_4} + \mathcal{E}_2 \right] \right) \right) \\ & \text{Tr} \left(S_{(15)(26)(34)(78)} \left(\delta_{j_1, j_2} \delta_{k_1, k_2} \left[\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1, i_3} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{i_1, i_3} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} \otimes \mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} + \mathcal{E}_2 \right] \right) \right) \\ & \text{Tr} \left(S_{(18)(27)(34)(56)} \left(\delta_{j_1, j_2} \delta_{k_1, k_2-1} \left[\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1, i_4} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{i_1, i_4} \otimes \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} \otimes \mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} + \mathcal{E}_2 \right] \right) \right). \end{aligned}$$

The covariance structure are derived in more detail in the proof of Theorem 4.2. \square

Proof of Theorem 4.2. If the series are independent, possible cross spectral terms drop out. The significant partitions of the form Cum₄Cum₂Cum₂ become therefore

$$\begin{aligned} & \text{Tr} \left(S_{(1256)(34)(78)} \left(\delta_{i_1, i_3} \delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, \omega_{k_1}, -\omega_{k_1}, -\omega_{k_2}}^{i_1} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} + \mathcal{E}_2 \right) \right] \right) \right) \\ & \text{Tr} \left(S_{(1278)(34)(56)} \left(\delta_{i_1, i_4} \delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, \omega_{k_1}, -\omega_{k_1}, \omega_{k_2-1}}^{i_1} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} + \mathcal{E}_2 \right) \right] \right) \right) \\ & \text{Tr} \left(S_{(3456)(12)(78)} \left(\delta_{i_2, i_3} \delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}, \omega_{k_1-1}, -\omega_{k_2}}^{i_2} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} + \mathcal{E}_2 \right) \right] \right) \right) \\ & \text{Tr} \left(S_{(3478)(12)(56)} \left(\delta_{i_2, i_4} \delta_{j_1, j_2} \left[\left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}, \omega_{k_1-1}, \omega_{k_2-1}}^{i_2} + \mathcal{E}_4 \right) \otimes \left(\mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} + \mathcal{E}_2 \right) \otimes \left(\mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} + \mathcal{E}_2 \right) \right] \right) \right). \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \delta_{i_2, i_4} \delta_{j_1, j_2} \left[\left\langle \left(\frac{2\pi}{N} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}, \omega_{k_1-1}, \omega_{k_2-1}}^{i_2}, \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{i_1} \otimes \mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} \right) \right\rangle_{HS} + O\left(\frac{1}{T}\right) \right] \\
& + \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \delta_{i_2, i_4} \delta_{j_1, j_2} \delta_{k_1, k_2} \left[\left\langle \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{\dagger, i_1} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2}, \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} \mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{\dagger, i_3} \right\rangle + O\left(\frac{1}{M^2}\right) \right] \\
& + \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \delta_{i_2, i_3} \delta_{j_1, j_2} \delta_{k_1-1, k_2} \left[\left\langle \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{\dagger, i_1} \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2}, \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} \mathcal{F}_{u_{j_2}, -\omega_{k_2-1}}^{i_4} \right\rangle_{HS} + O\left(\frac{1}{M^2}\right) \right] \\
& + \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \delta_{i_1, i_3} \delta_{i_2, i_4} \delta_{j_1, j_2} \delta_{k_1, k_2} \left[\left\langle \mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1}, \mathcal{F}_{u_{j_1}, \omega_{k_1-1}}^{i_2} \right\rangle_{HS} \left\langle \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{i_1}, \mathcal{F}_{u_{j_1}, -\omega_{k_1-1}}^{i_2} \right\rangle_{HS} + O\left(\frac{1}{M^2}\right) \right] \\
& + \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \delta_{i_1, i_3} \delta_{j_1, j_2} \delta_{k_1, k_2} \left[\left\langle \mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} \widetilde{\otimes} \mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1}, \mathcal{F}_{u_{j_1}, \omega_{k_1-1}}^{i_2} \otimes \mathcal{F}_{u_{j_2}, \omega_{k_2-1}}^{i_4} \right\rangle_{HS} + O\left(\frac{1}{M^2}\right) \right] \\
& + \frac{1}{T} \sum_{k_1, k_2=1}^{\lfloor N/2 \rfloor} \sum_{j_1, j_2=1}^M \delta_{i_1, i_4} \delta_{j_1, j_2} \delta_{k_1, k_2-1} \left[\left\langle \mathcal{F}_{u_{j_1}, \omega_{k_1}}^{i_1} \widetilde{\otimes}_{\top} \mathcal{F}_{u_{j_1}, -\omega_{k_1}}^{i_1}, \mathcal{F}_{u_{j_1}, \omega_{k_1-1}}^{i_2} \otimes \mathcal{F}_{u_{j_2}, -\omega_{k_2}}^{i_3} \right\rangle_{HS} + O\left(\frac{1}{M^2}\right) \right]
\end{aligned}$$

So that, as $N, M \rightarrow \infty$,

$$\begin{aligned}
\text{TCov}(F_{i_1, i_2}, F_{i_3, i_4}) & \rightarrow \delta_{i_1, i_3} \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega_1, -\omega_1, -\omega_2}^{i_1}, \mathcal{F}_{u, \omega_1}^{i_2} \otimes \mathcal{F}_{u, \omega_2}^{i_4} \right\rangle_{HS} dud\omega_1 d\omega_2 \\
& + \delta_{i_1, i_4} \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega_1, -\omega_1, \omega_2}^{i_1}, \mathcal{F}_{u, \omega_1}^{i_2} \otimes \mathcal{F}_{u, -\omega_2}^{i_3} \right\rangle_{HS} dud\omega_1 d\omega_2 \\
& + \delta_{i_2, i_3} \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, -\omega_1, \omega_1, -\omega_2}^{i_2}, \mathcal{F}_{u, -\omega_1}^{i_1} \otimes \mathcal{F}_{u, \omega_2}^{i_4} \right\rangle_{HS} dud\omega_1 d\omega_2 \\
& + \delta_{i_2, i_4} \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, -\omega_1, \omega_1, \omega_2}^{i_2}, \mathcal{F}_{u, -\omega_1}^{i_1} \otimes \mathcal{F}_{u, -\omega_2}^{i_3} \right\rangle_{HS} dud\omega_1 d\omega_2 \\
& + \delta_{i_2, i_4} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, -\omega}^{\dagger, i_1} \mathcal{F}_{u, -\omega}^{i_2}, \mathcal{F}_{u, -\omega}^{i_2} \mathcal{F}_{u, -\omega}^{\dagger, i_3} \right\rangle dud\omega \\
& + \delta_{i_2, i_3} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, -\omega}^{\dagger, i_1} \mathcal{F}_{u, -\omega}^{i_2}, \mathcal{F}_{u, -\omega}^{i_2} \mathcal{F}_{u, -\omega}^{i_4} \right\rangle_{HS} dud\omega \\
& + \delta_{i_1, i_3} \delta_{i_2, i_4} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_2} \right\rangle_{HS} \left\langle \mathcal{F}_{u, -\omega}^{i_1}, \mathcal{F}_{u, -\omega}^{i_2} \right\rangle_{HS} dud\omega \\
& + \delta_{i_1, i_3} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega}^{i_1} \widetilde{\otimes} \mathcal{F}_{u, \omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_2} \otimes \mathcal{F}_{u, \omega}^{i_4} \right\rangle_{HS} dud\omega \\
& + \delta_{i_1, i_4} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega}^{i_1} \widetilde{\otimes}_{\top} \mathcal{F}_{u, -\omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_2} \otimes \mathcal{F}_{u, -\omega}^{i_3} \right\rangle_{HS} dud\omega.
\end{aligned}$$

From this we can now derive the structures of the four distinct components of the asymptotic variance of $\mathcal{A}_{(i_1, i_2)}^T$.

1. Setting $i_1 = i_3 = i_2 = i_4$

$$\begin{aligned}
\text{TVar}(F_{i_1}) & \rightarrow \frac{2}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega_1, -\omega_1, -\omega_2}^{i_1}, \mathcal{F}_{u, \omega_1}^{i_1} \otimes \mathcal{F}_{u, \omega_2}^{i_1} \right\rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{2}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \left\langle \mathcal{F}_{u, \omega_1, -\omega_1, \omega_2}^{i_1}, \mathcal{F}_{u, \omega_1}^{i_1} \otimes \mathcal{F}_{u, -\omega_2}^{i_1} \right\rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{2}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^2 dud\omega
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u,\omega}^{i_1}\|_2^4 dud\omega \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega}^{i_1} \widetilde{\otimes} \mathcal{F}_{u,\omega}^{i_1}, \mathcal{F}_{u,\omega}^{i_1} \otimes \mathcal{F}_{u,\omega}^{i_1} \rangle_{HS} dud\omega \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega}^{i_1} \widetilde{\otimes}_{\top} \mathcal{F}_{u,-\omega}^{i_1}, \mathcal{F}_{u,\omega}^{i_1} \otimes \mathcal{F}_{u,-\omega}^{i_1} \rangle_{HS} dud\omega
\end{aligned} \tag{6.15}$$

where we used the self-adjointness of the spectral density operator and that, for any function $g : \mathbb{R} \rightarrow \mathbb{K}$, we have $\int_{-\pi}^{\pi} g(\omega) d\omega = \int_{-\pi}^{\pi} g(-\omega) d\omega$. From this it follows that the terms 1 and 4, 2 and 3 and 5 and 6 are respectively equal in the limit.

2. Setting $i_1 = i_3$ and $i_2 = i_4$, $i_1 \neq i_2 \rightarrow i_3 \neq i_4$, $i_1 \neq i_4$, $i_3 \neq i_2$

$$\begin{aligned}
T\text{Var}(F_{i_1 i_2}) & \rightarrow \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega_1,-\omega_1,-\omega_2}^{i_1}, \mathcal{F}_{u,\omega_1}^{i_2} \otimes \mathcal{F}_{u,\omega_2}^{i_2} \rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,-\omega_1,\omega_1,\omega_2}^{i_2}, \mathcal{F}_{u,-\omega_1}^{i_1} \otimes \mathcal{F}_{u,-\omega_2}^{i_1} \rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,-\omega}^{i_1} \mathcal{F}_{u,-\omega}^{i_2}, \mathcal{F}_{u,-\omega}^{i_2} \mathcal{F}_{u,-\omega}^{i_1} \rangle dud\omega \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega}^{i_1}, \mathcal{F}_{u,\omega}^{i_2} \rangle_{HS} \langle \mathcal{F}_{u,-\omega}^{i_1}, \mathcal{F}_{u,-\omega}^{i_2} \rangle_{HS} dud\omega \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega}^{i_1} \widetilde{\otimes} \mathcal{F}_{u,\omega}^{i_1}, \mathcal{F}_{u,\omega}^{i_2} \otimes \mathcal{F}_{u,\omega}^{i_2} \rangle_{HS} dud\omega.
\end{aligned} \tag{6.16}$$

3. Setting $i_1 = i_2$ and $i_3 = i_4$, $i_1 \neq i_3$, we have to due independence

$$TCov(F_{i_1, i_1}, F_{i_3, i_3}) = 0. \tag{6.17}$$

4. Setting $i_3 = i_2$, $i_4 = i_1$, $i_1 \neq i_2 \rightarrow i_3 \neq i_4$, $i_1 \neq i_3$, $i_2 \neq i_4$

$$\begin{aligned}
TCov(F_{i_1, i_2}, F_{i_2, i_1}) & \rightarrow \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega_1,-\omega_1,\omega_2}^{i_1}, \mathcal{F}_{u,\omega_1}^{i_2} \otimes \mathcal{F}_{u,-\omega_2}^{i_2} \rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,-\omega_1,\omega_1,-\omega_2}^{i_2}, \mathcal{F}_{u,-\omega_1}^{i_1} \otimes \mathcal{F}_{u,\omega_2}^{i_1} \rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,-\omega}^{i_1} \mathcal{F}_{u,-\omega}^{i_2}, \mathcal{F}_{u,-\omega}^{i_2} \mathcal{F}_{u,-\omega}^{i_1} \rangle_{HS} dud\omega \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega}^{i_1} \widetilde{\otimes}_{\top} \mathcal{F}_{u,-\omega}^{i_1}, \mathcal{F}_{u,\omega}^{i_2} \otimes \mathcal{F}_{u,-\omega}^{i_2} \rangle_{HS} dud\omega.
\end{aligned} \tag{6.18}$$

5. Setting $i_2 = i_3 = i_1$ with $i_1 \neq i_4$,

$$\begin{aligned}
TCov(F_{i_1 i_1}, F_{i_1 i_4}) & \rightarrow \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega_1,-\omega_1,-\omega_2}^{i_1}, \mathcal{F}_{u,\omega_1}^{i_1} \otimes \mathcal{F}_{u,\omega_2}^{i_4} \rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,-\omega_1,\omega_1,-\omega_2}^{i_1}, \mathcal{F}_{u,-\omega_1}^{i_1} \otimes \mathcal{F}_{u,\omega_2}^{i_4} \rangle_{HS} dud\omega_1 d\omega_2 \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,-\omega}^{\dagger, i_1} \mathcal{F}_{u,-\omega}^{i_1}, \mathcal{F}_{u,-\omega}^{i_1} \mathcal{F}_{u,-\omega}^{i_4} \rangle_{HS} dud\omega \\
& + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u,\omega}^{i_1} \widetilde{\otimes} \mathcal{F}_{u,\omega}^{i_1}, \mathcal{F}_{u,\omega}^{i_1} \otimes \mathcal{F}_{u,\omega}^{i_4} \rangle_{HS} dud\omega.
\end{aligned} \tag{6.19}$$

6. Setting $i_2 = i_4 = i_1$ with $i_1 \neq i_3$,

$$\begin{aligned}
TCov(F_{i_1 i_1}, F_{i_3 i_1}) &\rightarrow + \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, \omega_1, -\omega_1, \omega_2}^{i_1}, \mathcal{F}_{u, \omega_1}^{i_1} \otimes \mathcal{F}_{u, -\omega_2}^{i_3} \rangle_{HS} dud\omega_1 d\omega_2 \\
&+ \frac{1}{8\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, -\omega_1, \omega_1, \omega_2}^{i_1}, \mathcal{F}_{u, -\omega_1}^{i_1} \otimes \mathcal{F}_{u, -\omega_2}^{i_3} \rangle_{HS} dud\omega_1 d\omega_2 \\
&+ \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, -\omega}^{i_1}, \mathcal{F}_{u, -\omega}^{i_2}, \mathcal{F}_{u, -\omega}^{i_1}, \mathcal{F}_{u, -\omega}^{i_3} \rangle dud\omega \\
&+ \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \langle \mathcal{F}_{u, \omega}^{i_1} \widetilde{\otimes}_{\top} \mathcal{F}_{u, -\omega}^{i_1}, \mathcal{F}_{u, \omega}^{i_1} \otimes \mathcal{F}_{u, -\omega}^{i_3} \rangle_{HS} dud\omega \quad (6.20)
\end{aligned}$$

Let \mathbf{e} be the identity vector in \mathbb{R}^4 . We note that we the variance structure can be written in the form

$$\mathbf{e}^{\top} [\Sigma \nabla g(\mathbf{x}) \nabla g^{\top}(\mathbf{x})] \mathbf{e} = \mathbf{e}^{\top} \frac{\Sigma}{(x_1 + x_2)^2} \begin{pmatrix} \frac{(x_3 + x_4)^2}{(x_1 + x_2)^2} & \frac{(x_3 + x_4)^2}{(x_1 + x_2)^2} & -\frac{x_3 + x_4}{(x_1 + x_2)} & -\frac{x_3 + x_4}{(x_1 + x_2)} \\ \frac{(x_3 + x_4)^2}{(x_1 + x_2)^2} & \frac{(x_3 + x_4)^2}{(x_1 + x_2)^2} & -\frac{x_3 + x_4}{(x_1 + x_2)} & -\frac{x_3 + x_4}{(x_1 + x_2)} \\ -\frac{x_3 + x_4}{(x_1 + x_2)} & -\frac{x_3 + x_4}{(x_1 + x_2)} & 1 & 1 \\ -\frac{x_3 + x_4}{(x_1 + x_2)} & -\frac{x_3 + x_4}{(x_1 + x_2)} & 1 & 1 \end{pmatrix} \mathbf{e}$$

where Σ is defined in (6.10) where the expression of the individual entries are given by (6.15)-(6.20) and where the vector is evaluated as in (6.12). Under the null hypothesis H_0 , the matrix becomes

$$\begin{aligned}
\Sigma \nabla g(\mathbf{x}) \nabla g^{\top}(\mathbf{x}) &= \frac{1}{\left(\frac{2}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^2 dud\omega\right)^2} \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix} \\
&\times \begin{pmatrix} \text{Var}(F_{i_1, i_1}) & 0 & \text{Cov}(F_{i_1, i_1}, F_{i_1, i_2}) & \text{Cov}(F_{i_1, i_1}, F_{i_2, i_1}) \\ 0 & \text{Var}(F_{i_2, i_2}) & \text{Cov}(F_{i_2, i_2}, F_{i_1, i_2}) & \text{Cov}(F_{i_2, i_2}, F_{i_2, i_1}) \\ \text{Cov}(F_{i_1, i_1}, F_{i_1, i_2}) & \text{Cov}(F_{i_2, i_2}, F_{i_1, i_2}) & \text{Var}(F_{i_1, i_2}) & \text{Cov}(F_{i_1, i_2}, F_{i_2, i_1}) \\ \text{Cov}(F_{i_1, i_1}, F_{i_2, i_1}) & \text{Cov}(F_{i_2, i_2}, F_{i_2, i_1}) & \text{Cov}(F_{i_1, i_2}, F_{i_2, i_1}) & \text{Var}(F_{i_2, i_1}) \end{pmatrix} \quad (6.21)
\end{aligned}$$

where we used that $\text{Cov}(F_{i_1, i_1}, F_{i_2, i_2}) = 0$ because of the independence assumption. Using the expression of the entries of the asymptotic variance (6.15)-(6.20) we exploit that all of these components are restricted forms of (6.15) and that this equation consists of 6 distinct terms. For the structure of (6.21), a tedious derivation yields

1. the first and second term of (the fourth order terms) (6.15): 2 times in $\text{Var}(F_{i_1, i_1})$, $\text{Var}(F_{i_2, i_2})$, once in $\text{Var}(F_{i_2, i_1})$ and $\text{Var}(F_{i_1, i_1})$, $\text{Cov}(F_{i_1, i_2}, F_{i_2, i_1})$, and once in $\text{Cov}(F_{i_1, i_1}, F_{i_1, i_2})$, $\text{Cov}(F_{i_1, i_1}, F_{i_2, i_1})$. Hence using the number of occurrence in the covariance matrix of these entries, we find these terms to arise $4 + 2 + 2 - 4 - 4 = 0$ times and hence cancels.
2. the third term of (6.15): two times in $\text{Var}(F_{i_1, i_1})$, $\text{Var}(F_{i_2, i_2})$, once in $\text{Var}(F_{i_2, i_1})$ and $\text{Var}(F_{i_2, i_1})$, $\text{Cov}(F_{i_1, i_2}, F_{i_2, i_1})$, and once in $\text{Cov}(F_{i_1, i_1}, F_{i_1, i_2})$, $\text{Cov}(F_{i_1, i_1}, F_{i_2, i_1})$. Hence using the number of occurrence in the matrix, we find this term to arise $4 + 2 + 2 - 4 - 4 = 0$ times and hence cancels.
3. the fourth term of (6.15): once in $\text{Var}(F_{i_1, i_1})$, $\text{Var}(F_{i_2, i_2})$ and in once in $\text{Var}(F_{i_2, i_1})$ and $\text{Var}(F_{i_2, i_1})$ and does not arise in the other components. Therefore, we find a term $\frac{4}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^2 dud\omega$ to remain.

4. the fifth term of (6.15): once in $\text{Var}(F_{i_1, i_1})$, $\text{Var}(F_{i_2, i_2})$ and in once in $\text{Var}(F_{i_2, i_1})$ and $\text{Var}(F_{i_1, i_2})$ and off diagonal it occurs in the structures of the form $\text{Cov}(F_{i_1, i_1}, F_{i_1, i_2})$, which arises 4 times. Hence $4 - 4 = 0$ and cancels.
5. the sixth term of (6.15): once in $\text{Var}(F_{i_1, i_1})$, $\text{Var}(F_{i_2, i_2})$, and from the off-diagonal terms once in $\text{Cov}(F_{i_1, i_2}, F_{i_2, i_1})$ and once in $\text{Cov}(F_{i_1, i_1}, F_{i_2, i_1})$. As $\text{Cov}(F_{i_1, i_2}, F_{i_2, i_1})$ occurs twice off-diagonal in (6.21) and $\text{Cov}(F_{i_1, i_1}, F_{i_2, i_1})$ four times, we find the sixth term of (6.15) to occur $2 + 2 - 4 = 0$ and cancels as well.

Thus, under the null hypothesis H_0

$$\text{Var}(\hat{\mathcal{A}}_{(i_1, i_2)}) \xrightarrow{p} \frac{\frac{4}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^4 dud\omega}{\left(\frac{2}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^2 dud\omega\right)^2} = 4\pi \frac{\int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^4 dud\omega}{\left(\int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^2 dud\omega\right)^2}.$$

□

Proof of Lemma 4.1. We have,

$$\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS} = \left\langle \frac{1}{2} (I_{i_1}^{u_j, \omega_k} + I_{i_2}^{u_j, \omega_k}), \frac{1}{2} (I_{i_1}^{u_j, \omega_{k-1}} + I_{i_2}^{u_j, \omega_{k-1}}) \right\rangle_{HS}.$$

Hence the proof of theorem Theorem 2.1 shows that

$$\left(\frac{2}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS} \right)^2 \xrightarrow{p} \left(\frac{2}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^2 dud\omega \right)^2,$$

and it remains to show that under H_0

$$\frac{2}{3T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} (\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS})^2 \xrightarrow{p} \frac{1}{\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^4 dud\omega.$$

To this purpose note that

$$\begin{aligned} \mathbb{E} \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} (\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS})^2 &= \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \text{Cov}(\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS}) \\ &\quad + \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} (\mathbb{E} \langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS})^2, \end{aligned} \quad (6.22)$$

and that the second term converges to $\frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u, \omega}^{i_1}\|_2^4 dud\omega$. For the first term we write

$$\begin{aligned} &\text{Cov}(\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS}) \\ &= \frac{1}{4} \left(\text{Cov}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS} + \langle I_{i_2}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS} + \langle I_{i_1}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS} + \langle I_{i_2}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) \right. \\ &= \frac{1}{4} \left[\text{Var}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) + \text{Var}(\langle I_{i_2}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}) + \text{Var}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}) + \text{Var}(\langle I_{i_2}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) \right. \\ &\quad + 2\text{Cov}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}, \langle I_{i_1}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}) + 2\text{Cov}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}, \langle I_{i_2}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) \\ &\quad + 2\text{Cov}(\langle I_{i_2}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}, \langle I_{i_1}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}) + 2\text{Cov}(\langle I_{i_2}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}, \langle I_{i_2}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) \\ &\quad \left. + 2\text{Cov}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}, \langle I_{i_2}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) + 2\text{Cov}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}, \langle I_{i_2}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}) \right] \end{aligned} \quad (6.23)$$

Consider the structure of the first two terms of (6.23). By Theorem B.2, we are interested in all indecomposable partitions of the array where the summation is overall indecomposable partitions of the array

$$\begin{array}{cccc} \underbrace{D_{i_1}^{u_j, \omega_k}}_1 & \underbrace{D_{i_1}^{u_j, -\omega_k}}_2 & \underbrace{D_{i_1}^{u_j, -\omega_{k-1}}}_3 & \underbrace{D_{i_1}^{u_j, \omega_{k-1}}}_4 \\ \underbrace{D_{i_1}^{u_j, -\omega_k}}_5 & \underbrace{D_{i_1}^{u_j, \omega_k}}_6 & \underbrace{D_{i_1}^{u_j, \omega_{k-1}}}_7 & \underbrace{D_{i_1}^{u_j, -\omega_{k-1}}}_8 \end{array}$$

It is immediate from Corollary B.1 that all terms not consisting of second order cumulants will be of lower order. Additionally, certain partitions will be of lower order when it involves a cumulant component that is off the frequency manifold. Indecomposability moreover requires the first row to hook with the second. Those partitions that remain are

$$\begin{aligned} & \text{Tr} \left(S_{(15)(26)(34)(78)} \left(\mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} + \mathcal{E}_2 \right) \right) \\ & \text{Tr} \left(S_{(15)(26)(37)(48)} \left(\mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} + \mathcal{E}_2 \right) \right) \\ & \text{Tr} \left(S_{(12)(56)(37)(48)} \left(\mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} + \mathcal{E}_2 \right) \right) \end{aligned}$$

Hence, a similar argument as provided in Proposition B.1 demonstrates that the original order has correspondence $1 \rightarrow 3, 2 \rightarrow 4, 5 \rightarrow 7, 6 \rightarrow 8$, if we plug these in, we find

$$\begin{aligned} & \text{Tr} \left(S_{(15)(26)(12)(56)} \left(\mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} + \mathcal{E}_2 \right) \right) \\ & \text{Tr} \left(S_{(15)(26)(15)(26)} \left(\mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} + \mathcal{E}_2 \right) \right) \\ & \text{Tr} \left(S_{(12)(56)(15)(26)} \left(\mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_k}^{i_1} \otimes \mathcal{F}_{u_j, -\omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} + \mathcal{E}_2 \right) \right) \end{aligned}$$

But because switching the tensors at position 2 and 5 has no effect (being the same object), we find using Properties A.1

$$\begin{aligned} \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \text{Var}(\langle I_{i_1}^{u_j, \omega_k}, I_{i_1}^{u_j, \omega_{k-1}} \rangle_{HS}) &= \frac{3}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \langle \mathcal{F}_{u_j, \omega_k}^{i_1} \otimes \mathcal{F}_{u_j, \omega_k}^{i_1}, \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} \otimes \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} \rangle_{HS} + O\left(\frac{1}{M^2}\right) \\ &= \frac{3}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \langle \mathcal{F}_{u_j, \omega_k}^{i_1}, \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} \rangle_{HS} \langle \mathcal{F}_{u_j, \omega_k}^{i_1}, \mathcal{F}_{u_j, \omega_{k-1}}^{i_1} \rangle_{HS} + O\left(\frac{1}{M^2}\right) \\ &\rightarrow \frac{3}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}^{i_1}\|_2^2 \|\mathcal{F}_{u_j, \omega_k}^{i_1}\|_2^2 dud\omega \\ &= \frac{3}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}^{i_1}\|_2^4 dud\omega, \end{aligned} \tag{6.24}$$

Similarly

$$\frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \text{Var}(\langle I_{i_2}^{u_j, \omega_k}, I_{i_2}^{u_j, \omega_{k-1}} \rangle_{HS}) = \frac{3}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}^{i_2}\|_2^4 dud\omega, \tag{6.25}$$

under H_0 . Consider then the structure of the third and fourth term of (6.23). We are again interested in all indecomposable partitions of the array where the summation is overall indecompos-

able partitions of the array

$$\begin{array}{cccc} \underbrace{D_{i_1}^{u_j, \omega_k}}_1 & \underbrace{D_{i_1}^{u_j, -\omega_k}}_2 & \underbrace{D_{i_2}^{u_j, -\omega_{k-1}}}_3 & \underbrace{D_{i_2}^{u_j, \omega_{k-1}}}_4 \\ \underbrace{D_{i_1}^{u_j, -\omega_k}}_5 & \underbrace{D_{i_1}^{u_j, \omega_k}}_6 & \underbrace{D_{i_2}^{u_j, \omega_{k-1}}}_7 & \underbrace{D_{i_2}^{u_j, -\omega_{k-1}}}_8 \end{array}$$

Because of uncorrelatedness under H_0 between series i_1 and i_2 and that the rows of the array must hook results in three permutations remaining; $S_{(15)(26)(34)(78)}$, $S_{(15)(26)(37)(48)}$ and $S_{(12)(56)(37)(48)}$. A similar reasoning shows that, for the fifth to eight terms of (6.23) which involve an array of the form

$$\begin{array}{cccc} \underbrace{D_{i_1}^{u_j, \omega_k}}_1 & \underbrace{D_{i_1}^{u_j, -\omega_k}}_2 & \underbrace{D_{i_1}^{u_j, -\omega_{k-1}}}_3 & \underbrace{D_{i_1}^{u_j, \omega_{k-1}}}_4 \\ \underbrace{D_{i_1}^{u_j, -\omega_k}}_5 & \underbrace{D_{i_1}^{u_j, \omega_k}}_6 & \underbrace{D_{i_2}^{u_j, \omega_{k-1}}}_7 & \underbrace{D_{i_2}^{u_j, -\omega_{k-1}}}_8, \end{array}$$

will only have the permutation $S_{(15)(26)(34)(78)}$ remaining. Uncorrelatedness under H_0 between series i_1 and i_2 and the lag Fourier imply that the last two terms of (6.23) will be of lower order. We thus have that the total sum in (6.23) consists of $3 \times 2 + 3 \times 2 + 4 \times 2 = 20$ terms and similar to the proof of (6.24) it can be shown these terms all converge to the same limit under H_0 . Therefore, we find under H_0

$$\mathbb{E} \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \text{Cov}(\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS}) = \frac{20}{4} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}\|_2^4 dud\omega$$

which together with the second term of (6.22), implies

$$\begin{aligned} \mathbb{E} \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} (\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS})^2 &= \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} \text{Cov}(\langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS}) \\ &+ \frac{1}{T} \sum_{j=1}^M \sum_{k=1}^{\lfloor N/2 \rfloor} (\mathbb{E} \langle I_p^{u_j, \omega_k}, I_p^{u_j, \omega_{k-1}} \rangle_{HS})^2 \\ &\rightarrow \frac{20}{4} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}\|_2^4 dud\omega + \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}\|_2^4 dud\omega \\ &= \frac{24}{4} \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_0^1 \|\mathcal{F}_{u_j, \omega_k}\|_2^4 dud\omega \end{aligned}$$

It can be shown along the lines of the proof of Theorem 2.1 that this is in fact a \sqrt{T} consistent estimator. The joint convergence in probability therefore immediately follows and the result follows from an application of the continuous mapping theorem. \square

Appendix C Analysis of the spectral clustering algorithm

C.1 Consistency of \hat{L} for L

Proof of Lemma 3.1. From Theorem 2.1 we have that $\hat{\mathcal{A}} \in \mathbb{R}^{d \times d}$ is a \sqrt{T} -consistent estimator of the distance measure \mathcal{A} . The continuous mapping theorem therefore implies that \hat{W}^T is consistent, i.e., A simple calculation shows that, as $T \rightarrow \infty$,

$$\mathbb{P}(\|\hat{W}^T - W\|_{\infty} \geq \varepsilon) \leq \mathbb{P}(d \max_{i,j} |\hat{W}_{i,j}^T - W_{i,j}| \geq \varepsilon) \rightarrow 0. \quad (6.26)$$

Similarly,

$$\mathbb{P}(\max_i |D_i - \hat{D}_i| \geq \varepsilon) = \mathbb{P}(\max_i |\sum_j \hat{W}_{i,j}^T - \sum_j \hat{W}_{i,j}^T| \geq \varepsilon) \leq \mathbb{P}(d \max_{i,j} |\hat{W}_{i,j}^T - W_{i,j}| \geq \varepsilon) \rightarrow 0. \quad (6.27)$$

Similar to Chung and Radcliffe (2011), we use the decomposition

$$\begin{aligned} \hat{L} - L &= \hat{D}^{-1/2} \hat{W}^T \hat{D}^{-1/2} - D^{-1/2} \hat{W}^T D^{-1/2} + D^{-1/2} \hat{W}^T D^{-1/2} - D^{-1/2} W D^{-1/2} \\ &= (\hat{D}^{-1/2} - D^{-1/2}) \hat{W}^T \hat{D}^{-1/2} + D^{-1/2} \hat{W}^T (\hat{D}^{-1/2} - D^{-1/2}) + D^{-1/2} (\hat{W}^T - W) D^{-1/2} \\ &= (I - D^{-1/2} \hat{D}^{1/2}) \hat{D}^{-1/2} \hat{W}^T \hat{D}^{-1/2} + (D^{-1/2} \hat{D}^{1/2}) \hat{D}^{-1/2} \hat{W}^T \hat{D}^{-1/2} (I - \hat{D}^{1/2} D^{-1/2}) \\ &\quad + D^{-1/2} (\hat{W}^T - W) D^{-1/2} \end{aligned}$$

and bound these terms separately. Note that as D and \hat{D} are degree matrices, they are diagonal with nonnegative entries. We therefore have

$$\|I - D^{-1/2} \hat{D}^{1/2}\|_\infty = \max_i \left| 1 - \sqrt{\frac{\hat{D}_i}{D_i}} \right| \leq \max_i \left| 1 - \frac{\hat{D}_i}{D_i} \right| \leq \max_i \frac{|D_i - \hat{D}_i|}{\min_i D_i}.$$

The triangle inequality gives

$$\|D^{-1/2} \hat{D}^{1/2}\|_\infty = \|I - (I - D^{-1/2} \hat{D}^{1/2})\|_\infty \leq 1 + \max_i \frac{|D_i - \hat{D}_i|}{\min_i D_i}$$

Additionally, since $\hat{D}_i = \sum_j \hat{W}_{i,j}^T$ it follows that $\|\hat{D}^{-1/2} \hat{W}^T \hat{D}^{-1/2}\|_\infty = 1$. Furthermore, $\|D^{-1/2} (\hat{W}^T - W) D^{-1/2}\|_\infty \leq \frac{1}{\min_i D_i} \|\hat{W}^T - W\|_\infty$. Therefore,

$$\|\hat{L} - L\|_\infty \leq \frac{\max_i |D_i - \hat{D}_i|}{\min_i D_i} \left(2 + \frac{\max_i |D_i - \hat{D}_i|}{\min_i D_i} \right) + \frac{1}{\min_i D_i} \|\hat{W}^T - W\|_\infty.$$

Consequently, (6.26) and (6.27) imply

$$\forall \varepsilon > 0, \quad \lim_{T \rightarrow \infty} \mathbb{P}(\|\hat{L} - L\|_\infty > \varepsilon) = 0.$$

□

C.2 Concentration of $\hat{\mathcal{U}}$

We shall use Lemma 3.1 to analyze the concentration of $\hat{\mathcal{U}}$. We first need the following auxiliary lemma, which follows from the Davis-Kahan theorem (Davis and Kahan, 1970)

Lemma C.1. *Let $S \subset \mathbb{R}$ an interval. Let $A, H \in \mathbb{R}^{d \times d}$ be two symmetric matrices and let $\hat{A} = A + H$ denote a perturbed version of A . Denote \hat{Q} and Q be orthonormal matrices of dimension $\mathbb{R}^{d \times k}$ whose column spaces equal the eigenspace of \hat{A} and A respectively. Then there exists an orthonormal rotation matrix $O \in \mathbb{R}^{k \times k}$ such that*

$$\|\hat{Q} - QO\|_2 \leq \frac{\sqrt{2k} \|H\|_\infty}{\delta}$$

where

$$\delta = \min\{|\lambda - s| : \lambda \text{ eigenvalue of } A, \lambda \notin S, s \in S\}$$

Proof of Lemma C.1. Using the singular value decomposition, we can find orthonormal matrices P_1 and P_2 such that the singular values of $Q^\top \hat{Q}$ are exactly the cosines of the principal angles Θ , i.e., we can find P_1 and P_2 such that $Q^\top \hat{Q} = P_1 \Sigma P_2^\top$ where the diagonal of Σ contains the principal angles between the column space of \hat{Q} and Q . Define the rotation matrix O as $O = P_1 P_2^\top$. Then, by definition of the Frobenius norm, the orthonormality of \hat{Q} and Q

$$\begin{aligned} \|\hat{Q} - QO\|_2^2 &= \text{Tr}((\hat{Q} - QO)^\top (\hat{Q} - QO)) \\ &= 2k - 2\text{Tr}(OQ^\top \hat{Q}) \\ &= 2k - 2\text{Tr}(\cos \Theta) = 2k - 2 \sum_{i=1}^k \cos \theta_i \\ &\leq 2k - 2 \sum_{i=1}^k \cos(\theta)_i^2 = 2k - 2k + 2 \sum_{i=1}^k \sin(\theta)_i^2 = 2\|\sin \Theta\|_2^2. \end{aligned}$$

The classical Davis-Kahan theorem then yields

$$\|\hat{Q} - QO\|_2^2 \leq 2\|\sin \Theta\|_2^2 \leq 2 \frac{\|H\|_2^2}{\delta^2}.$$

Finally, since $\|H\|_2^2 \leq k \max_j |\lambda_j^H|^2 = k\|H\|_\infty^2$, we obtain

$$\|\hat{Q} - QO\|_2 \leq \sqrt{2k} \frac{\|H\|_\infty}{\delta}.$$

□

Corollary C.1. *There exists an orthonormal rotation matrix $O \in \mathbb{R}^{k \times k}$ such that*

$$\|\hat{U} - UO\|_2 \leq \frac{2\sqrt{k}\|\hat{L} - L\|_\infty}{\lambda_{k+1}}$$

where λ_{k+1} is the $(k+1)$ -th smallest eigenvalue of L .

Proof of Corollary C.1. By construction, \hat{L} and L are symmetric and it is clear that we can view \hat{L} as a perturbed version of L . Additionally, the columns of \hat{U} and U contain the eigenvectors that correspond to the k smallest eigenvalues of \hat{L} and L , respectively. It follows therefore directly from Lemma C.1 that

$$\|\hat{U} - UO\|_2 \leq \frac{2\sqrt{k}\|\hat{L} - L\|_\infty}{\delta} \leq \frac{2\sqrt{k}\|\hat{L} - L\|_\infty}{\lambda_{k+1}}.$$

The last inequality is a consequence of the following observation. The matrix L has exactly k zero eigenvalues. Hence if we take $S = [0, \epsilon)$ for arbitray small $\epsilon > 0$ or actually the singleton $S = \{0\}$, then the first k eigenvalues of L all belong to S . The smallest distance between eigenvalues that belong to S and that do not belong to S is thus given by $|0 - \lambda_{k+1}|$. Hence $\delta = \lambda_{k+1}$. □

Proof of Lemma 3.2. We note that by definition we have $\hat{\mathcal{U}}_{i,\cdot} = \frac{\hat{U}_{i,\cdot}}{\|\hat{U}_{i,\cdot}\|_2}$ and $\mathcal{U}_{i,\cdot} = \frac{(UO)_{i,\cdot}}{\|U_{i,\cdot}\|_2}$. Therefore standard linear algebra shows

$$\|\hat{\mathcal{U}} - \mathcal{U}\|_2^2 = \sum_{i=1}^d \left\| \frac{\hat{U}_{i,\cdot}}{\|\hat{U}_{i,\cdot}\|_2} - \frac{(UO)_{i,\cdot}}{\|U_{i,\cdot}\|_2} \right\|_2^2$$

$$\begin{aligned}
&\leq 2 \sum_{i=1}^d \left\| \frac{\hat{U}_{i,\cdot} (\|U_{i,\cdot}\|_2 - \|\hat{U}_{i,\cdot}\|_2)}{\|\hat{U}_{i,\cdot}\|_2 \|U_{i,\cdot}\|_2} \right\|_2^2 + \left\| \frac{\hat{U}_{i,\cdot} - (UO)_{i,\cdot}}{\|U_{i,\cdot}\|_2} \right\|_2^2 \\
&= 2 \sum_{i=1}^d \frac{\|U_{i,\cdot}\|_2 - \|\hat{U}_{i,\cdot}\|_2}{\|U_{i,\cdot}\|_2^2} + \frac{\|\hat{U}_{i,\cdot} - (UO)_{i,\cdot}\|_2^2}{\|U_{i,\cdot}\|_2^2} \\
&\leq 4 \sum_{i=1}^d \frac{\|\hat{U}_{i,\cdot} - (UO)_{i,\cdot}\|_2^2}{\|U_{i,\cdot}\|_2^2} \\
&\leq \frac{4}{\min_i \|U_{i,\cdot}\|_2^2} \|\hat{U} - (UO)\|_2^2 = \frac{4}{\min_i D_i} \|\hat{U} - (UO)\|_2^2
\end{aligned}$$

The last equality follows since U collects eigenvectors of the form $\sqrt{D} \mathbb{1}_{C_l}$ for $l = 1, \dots, k$, where $\mathbb{1}_{C_l} \in \mathbb{R}^d$ denotes the indicator vector that equals 1 if point i belongs to component C_l . This means in particular that U has exactly one nonzero entry per row. A trivial lower bound on $\min_i D_i$ can be given by

$$\min_i \|U_{i,\cdot}\|_2^2 \geq \frac{\min_i D_i}{\mathcal{C}_{\max}}$$

where $\mathcal{C}_{\max} = \max_i \sum_{i_1 \in G_i} \sum_{i_2 \in G_i} W_{i_1, i_2}$. Hence, using Corollary C.1 and Lemma 3.1

$$\|\hat{\mathcal{U}} - \mathcal{U}\|_2 \leq 4\sqrt{k} \sqrt{\frac{\mathcal{C}_{\max}}{\min_i D_i}} \frac{\|\hat{L} - L\|_\infty}{\lambda_{k+1}} \rightarrow 0 \text{ as } T \rightarrow \infty.$$

□

C.3 Analyzing the k -means step

Using the properties of the row-normalized eigenvectors of L , we proceed by providing a definition of the set of points that are clustered correctly and then derive a bound on the complement set. The technique is therefore similar to, among others, Rohe et al. (2011) and Lei and Rinaldo (2015).

Lemma C.2. *Assume the graph has k components. Let C^* defined in (3.6) and \mathcal{U} defined in (3.4). Then, the set of correctly clustered points is defined as the complement of the set*

$$\Sigma = \{i : \|C_{i,\cdot}^* - \mathcal{U}_{i,\cdot}\|_2 \geq \frac{1}{\sqrt{2}}\} \quad (6.28)$$

Proof of Lemma C.2. By construction and using the properties of the Laplacian, \mathcal{U} has exactly one 1 per row. All other entries in that row are zero. In total, there are k distinct rows which are orthonormal. Therefore, $\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 = 0$ if the embedded points i and j belong to the same component and $\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 = \sqrt{2}$ if they belong to different components. At the same time, $\|C_{i,\cdot}^* - C_{j,\cdot}^*\|_2 = 0$ if and only if the algorithm has clustered i, j in the same cluster. So let, i and j belong to Σ^c . Minkowski's inequality yields

$$\|\mathcal{U}_{i,\cdot} - \mathcal{U}_{j,\cdot}\|_2 \leq \|\mathcal{U}_{i,\cdot} - C_{i,\cdot}^*\|_2 + \|C_{i,\cdot}^* - C_{j,\cdot}^*\|_2 + \|C_{j,\cdot}^* - \mathcal{U}_{j,\cdot}\|_2 \leq 2 \frac{1}{\sqrt{2}} = \sqrt{2}$$

if and only if i and j are clustered in the same cluster. Otherwise, we have a contradiction. Additionally, since $C^* \in \mathcal{M}(d, k)$ clusters cannot be split. Hence, points in Σ^c must be correctly clustered □

Proof of Theorem 3.1. First note that $\mathcal{U} \in \mathcal{M}(d, k)$ since it has exactly k distinct rows. Consequently,

$$\operatorname{argmin}_{C \in \mathcal{M}(d, k)} \|\hat{\mathcal{U}} - C\|_2^2 = \|\hat{\mathcal{U}} - C^*\|_2^2 \leq \|\hat{\mathcal{U}} - \mathcal{U}\|_2^2$$

$$\begin{aligned} |\Sigma| &= \sum_{i \in \Sigma} 1 \leq \sum_{i \in \Sigma} 2\|C_{i,\cdot}^* - \mathcal{U}_{i,\cdot}\|_2^2 \\ &\leq 2\|C^* - \mathcal{U}\|_2^2 \\ &\leq 4(\|C^* - \hat{\mathcal{U}}\|_2^2 + \|\hat{\mathcal{U}} - \mathcal{U}\|_2^2) \\ &= 8\|\hat{\mathcal{U}} - \mathcal{U}\|_2^2 \end{aligned}$$

The result now follows from Lemma 3.2. □

References

- Abraham, C., Cornillon, P.A., Matzner-Løber, E. and Molinari, N. “Unsupervised curve clustering using B-splines.” *Scandinavian Journal of Statistics. Theory and Applications*, 30(3):581–595 (2003).
- Aghabozorgi, S. Sirkhorshidi, A.S., and Wah, Y.W. “Time-series Clustering – A decade review.” *Information Systems* 53:16–38 (2015).
- Aue, A. and van Delft, A. “Testing for stationarity of functional time series in the frequency domain.” *arXiv:1701.0174* (2017).
- Bauwens, L. and Rombouts, J.V.K. “Bayesian Clustering of Many Garch Models.” *Econometric Reviews*, 26(2-4):365–386 (2007).
- Böhm, H-Y., Ombao, H., von Sachs, R., and Sanes, J. “Discrimination and Classification of non-stationary Time Series Using the SLEX Model.” *Journal of Statistical Inference and Planning*, 140(12):3754–3763 (2010).
- Bosq, D. *Linear Processes in Function Spaces*. New York: Springer-Verlag (2000).
- Brillinger, D. *Time Series: Data Analysis and Theory*. McGraw Hill, New York. (1981).
- Calinski R. B. and Harabasz, J. “A dendrite method for cluster analysis.” *Communications in Statistics - Theory and Methods*, 3(1):1–27 (1974).
- Chandler, G. and Polonik, W. “Discrimination of Locally Stationary Time Series Based on the Excess Mass Functional.” *Journal of the American Statistical Association*, 101(73):240–253 (2006).
- Charrad, M., Ghazzali N., Boiteau, V. and Niknafs, A. “NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set.” *Journal of Statistical Software, Articles.*, 61(6):1–36 (2014).
- Chung, F. *Spectral graph theory*. Washington: Convergence Board of the Mathematical Sciences. (1997).
- Chung, F. and Radcliffe, M. “On the spectra of general random graphs.” *Electronic Journal of Combinatorics*, 18(1):215–229 (2011).

- Coates, D.S. and Diggle, P.J. “Tests for comparing two estimated spectral densities.” *Journal of Time Series Analysis*, 7:7–20 (1986).
- Corduas, M. and Piccolo, D. “Time series clustering and classification by the autoregressive metric.” *Computational Statistics & Data Analysis*, 52(4):1860–1872 (2008).
- Dahlhaus, R. “Fitting time series models to non-stationary processes.” *Annals of Statistics*, 25(1):1–37 (1997).
- Davis, C. and Kahan, W.M. “The rotation of eigenvectors by a perturbation. III.” *Journal of Numerical Analysis*, 7:1–46 (1970).
- van Delft, A., Characiejus V., and Dette H. “A nonparametric test for stationarity in functional time series.” *arXiv:1708.05248*, (2018).
- van Delft, A. and Eichler, M. “Locally stationary functional time series.” *Electronic Journal of Statistics*, 12:107–170 (2018).
- Dette, H. and Hildebrandt, T. “A note on testing hypotheses for stationary processes in the frequency domain.” *Journal of Multivariate Analysis*, 104:101–114 (2012).
- Dette, H. and Paparoditis, E. “Bootstrapping frequency domain tests in multi-variate time series with an application to comparing spectral densities.” *Journal of the Royal Statistical Society. Series B*, 71(4):831–857 (2009).
- Diamond, H.J., Karl, T.R., Palecki, A., Baker, C.B., Bell, J.E., Leeper, R.D., Easterling, D.R., Lawrimore, J.H., Meyers T.P., Helfert, M.R., Goodge, G. and Thorne, P.W. “U.S. Climate Reference Network after one decade of operations: status and assessment.” *Bull. Amer. Meteor. Soc.*, 94: 489–498.
- Eichler, M. “Testing nonparametric and semiparametric hypotheses in vector stationary processes.” *Journal of Multivariate Analysis*, 99(5):986–1009 (2008).
- Ferraty, F. and Vieu, P. *Nonparametric functional data analysis*. Springer, New York. (2006).
- Fokianos, K. and Promponas, V.J. “Biological applications of time series frequency domain clustering.” *Journal of Time Series Analysis*, 33(5):–1009 (2011).
- Frühwirth-Schnatter, S. and Kaufman, S. “Model-Based Clustering of Multiple Time Series.” *Journal of Business & Economic Statistics*, 26(1):78–89 (2008).
- Gordon, A. *Classification, 2nd edn*. London: Chapman and Hall–CRC. (1999).
- Hartigan, J. *Clustering Algorithms*. New York: Wiley. (1975).
- Harvill, J.L., Kohli, P. and Ravishanker, N. “Clustering Nonlinear, non-stationary Time Series Using BSLEX.” *Methodology and Computing in Applied Probability*, 19(3):935–955 (2017).
- Heard, N.A., Holmes, C.C., and Stephens, D.A. “A Quantitative Study of Gene Regulation Involved in the Immune Response of Anopheline Mosquitoes.” *Journal of the American Statistical Association*, 101(473):18–29 (2006).
- Holan, S.H. and Ravishanker, N. “Time series clustering and classification via frequency domain methods.” *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1444 (2018).

- Huang, H-Y., Ombao, H. and Stoffer, D.S. "Discrimination and Classification of non-stationary Time Series Using the SLEX Model." *Journal of the American Statistical Association*, 99(467):763–774 (2004).
- Ieva, F., Paganoni, A.M., Pigoli, D. and Vitelli, V. "Multivariate functional clustering for the analysis of ecg curves morphology." *Journal of the Royal Statistical Society. Series C*, 62(3):401–418 (2013).
- Jacques, J. and Preda, C. "Model-based clustering for multivariate functional data." *Computational Statistics & Data Analysis*, 71:92–106 (2014).
- Jentsch, C. and Pauly, M. "Testing equality of spectral densities using randomization techniques." *Bernoulli*, 21(2):697–739 (2015).
- Juárez, M.A. and Steel, M.F.J. "Model-Based Clustering of Non-Gaussian Panel Data Based on Skew- t Distributions." *Journal of Business & Economic Statistics*, 28(1):52–66 (2010).
- Kakizawa, Y., Shumway, R.H, and Taniguchi, M. "Discrimination and Clustering for Multivariate Time Series." *Journal of the American Statistical Association*, 93(441):328–340 (1998).
- Kalpakis, K. Gada, D. and Puttagunta, V. "Distance measures for effective clustering of arima time-series." In: *Proceedings of the 2001 IEEE international conference on data mining. San Jose*, 273–280 (2001).
- Kaufman, L. and Rousseeuw, P. *Finding Groups in Data: an Introduction to Cluster Analysis*. New York: Wiley. (1990).
- Krzanowski, W. J. and Lai, Y. T. "A Criterion for Determining the Number of Groups in a Data set: Using Sum-of-Squares Clustering." *Biometrics*, 44(1):23–34 (1988).
- Lei, J. and Rinaldo, A. "Consistency of spectral clustering in stochastic block models." *Annals of Statistics*, 43(1):215–237 (2015).
- Leucht, A., Paporoditis, E. and Sapatinas, T. "Testing equality of spectral density operators for functional linear processes." *arXiv:1804.03366* (2018).
- Milligan, G. W. and Cooper, M. C. "An examination of procedures for determining the number of clusters in a data set." *Psychometrika*, 50:159–179 (1985).
- Ombao, H.C., Raz, J.A., von Sachs, R. and Malow, B.A. "Automatic Statistical Analysis of Bivariate non-stationary Time Series." *Journal of the American Statistical Association*, 96(454):543–560 (2001).
- Panaretos, V. and Tavakoli, S. "Detecting and Localizing Differences in Functional Time Series Dynamics: A Case Study in Molecular Biophysics" *Journal of the American Statistical Association*, 111:1020–1035 (2016).
- Paporoditis, E. and Sapatinas, T. "Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data." *Biometrika*, (103):727–733 (2016).
- Peng, J., and Müller, H-G. "Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions." *The Annals of Applied Statistics*, 2(3):1056–1077 (2008).

- Ng, A., Jordan, S. & Weiss, Y. (2002). "On spectral clustering: analysis and an algorithm." In T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14*. MIT Press.
- Rohe, K., Chatterjee, S. & Yu, B. "Spectral clustering and the high-dimensional stochastic block-model." *Annals of Statistics*, 39:1878–1915 (2011).
- Horvath, L., Hušková, M. & Rice, G. "Test of independence for functional data." *Journal of Multivariate Analysis*, 117:100–119 (2013).
- Jacques, J. and Preda, C. Functional data clustering: a survey. *Advances in Data Analysis and Classification*, 8(3):231–255 (2014).
- Sakiyama, K., and Taniguchi, M. "Discriminant analysis for locally stationary processes." *Journal of Multivariate Analysis*, 90:282–300 (2004).
- Savvides, A., Promponas, V.J. and Fokianos, K. "Clustering of biological time series by cepstral coefficients based distances." *Pattern Recognition*, 41(7):2398–2412(2008).
- Shi, J., and Malik, J. "Normalized cuts and image segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905 (2002).
- Stewart, G. and Sun, J. *Matrix Perturbation Theory*. New York: Academic Press. (1990).
- Theodoridis, S. and Koutroumbas, K. *Pattern Recognition. 4th edition*. New York: Academic Press. (2008).
- Tibshirani, R., Walther, G. & Hastie, T. "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society B*, 63:411–423 (2001).
- Vlachos, M., Lin, J. Keogh, E., and Gunopulos, D. "A wavelet-based anytime algorithm for k-means clustering of time series." In: *Proceedings of the Workshop on Clustering High Dimensionality Data and Its Applications*, San Francisco, (2003).
- Von Luxburg, U. "A tutorial on spectral clustering." *Statistics and Computing*, 17:395–416 (2007).
- Von Luxburg, U., Belkin, M. & Bousquet O. "Consistency of spectral clustering." *The Annals of Statistics*, 36:555–586 (2008).

