A Matrix Gamma Process and Applications to Bayesian Analysis of Multivariate Time Series

Dissertation
zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Alexander Meier
geb. am 23.01.1989 in Kehl

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

Gutachter: Prof. Dr. Claudia Kirch
Prof. Dr. Renate Meyer

Eingereicht am: 22.06.2018
Verteidigung am: 06.09.2018

Abstract

While there is an increasing amount of literature about Bayesian time series analysis, only few Bayesian nonparametric approaches to multivariate time series exist. Most methods rely on Whittle's Likelihood, involving the second order structure of a stationary time series by means of its spectral density matrix. This is often modeled in terms of the Cholesky decomposition to ensure positive definiteness. However, with such nonlinear transformations, the modeling of certain prior knowledge aspects such as mean or variance often proves to be difficult. Furthermore, and most importantly, asymptotic properties such as posterior consistency or posterior contraction rates are not known.

A different idea is to model the spectral density matrix by means of random measures. This is in line with existing approaches for the univariate case, where the normalized spectral density is modeled similar to a probability density, e.g. with a Dirichlet process mixture of Beta densities. In this work we present a related approach for multivariate time series, with matrix-valued mixture weights induced by a Hermitian positive definite Gamma process. The process construction is inspired by Kingman's construction of the Gamma process and utilizes an infinitely divisible Hermitian positive definite Gamma distribution.

In conjunction with Whittle's Likelihood, our proposed Bayesian nonparametric procedure for spectral density inference is shown to perform well for both simulated and real data. Important theoretical properties such as posterior consistency and contraction rates are also established. As a preliminary result for these asymptotic considerations, we establish mutual contiguity of Whittle's Likelihood and the full Gaussian Likelihood for stationary multivariate Gaussian time series, a fact that has so far only been known in the univariate case.

We also present a semiparametric model extension, accommodating a parametric linear model in which the nonparametric time series component constitutes the error term. This model is investigated with both numerical simulations and in terms of asymptotic properties of the joint posterior and the marginal posterior of the linear model coefficients.

Zusammenfassung

Trotz zunehmendem Interesse an Bayesscher Zeitreihenanalyse in der Fachliteratur existieren bis heute nur wenige Bayessche nichtparametrische Ansätze zur multivariaten Zeitreihenanalyse. Die meisten Verfahren basieren auf der Whittle Likelihood, welche die Abhängigkeitsstruktur der Zeitreihe in Form der Spektraldichtematrix einbezieht. Letztere wird oft mithilfe der Cholesky Zerlegung modelliert, um positive Definitheit zu garantieren. Die Modellierung gewisser Aspekte des Bayesschen Vorwissens wie zum Beispiel Erwartungswert oder Varianz gestaltet sich jedoch oft schwierig bei Verwendung solcher nichtlinearer Transformationen. Darüber hinaus ist die Gültigkeit von asymptotischen Gütekriterien wie Konsistenz und Kontraktionsraten der a posteriori Verteilung nicht bekannt.

Eine andere Idee besteht darin, die Spektraldichtematrix mithilfe von zufälligen Maßen zu modellieren. Dies ist in Anlehnung an ein aus der Literatur bekanntes Verfahren für den univariaten Fall, in welchem die normalisierte Spektraldichte analog zu einer Wahrscheinlichkeitsdichte modelliert wird, zum Beispiel mit einer Dirichlet Prozess Mischung von Beta Dichten. In dieser Arbeit präsentieren wir einen verwandten Ansatz für multivariate Zeitreihen, mit matrixwertigen Mischungsgewichten, welche von einem hermitesch positiv definiten Gamma Prozess induziert werden. Die Prozesskonstruktion ist inspiriert von der Konstruktion nach Kingman und beruht auf einer unendlich teilbaren hermitesch positiv definiten Gamma Verteilung.

Zusammen mit der Whittle Likelihood liefert das hier vorgestellte Bayessche nichtparametrische Verfahren zur Spektraldichteschätzung sowohl für simulierte als auch für echte Daten gute Resultate. Wichtige theoretische Eigenschaften wie Konsistenz und Kontraktionsrate der a posteriori Verteilung werden auch hergeleitet. Als vorausgehendes Resultat für diese asymptotischen Betrachtungen leiten wir zudem die gegenseitige Kontiguität der Whittle Likelihood und der vollen Gaussschen Likelihood für multivariate stationäre Gaussche Zeitreihen her. Dieses Resultat war bisher lediglich für den univariaten Fall bekannt.

Darüber hinaus präsentieren wir eine Modellerweiterung in Form eines parametrischen linearen Modells, in welchem die nichtparametrisch modellierte Zeitreihe den Fehlerterm darstellt. Dieses wird sowohl in Form von numerischen Simulationen, als auch mittels asymptotischen Gütekriterien der gemeinsamen a posteriori Verteilung und der marginalen a posteriori Verteilung der linearen Modellkoeffizienten untersucht.

Danksagung

Mein aufrichtiger Dank gilt Prof. Dr. Claudia Kirch, die mich bereits bei der Betreuung der Masterarbeit gefördert hat. Die zahlreichen konstruktiven Gespräche und mathematischen Anregungen haben ebenso zum Gelingen dieser Arbeit beigetragen, wie der Betreuungsstil, welcher mir Freiraum für eigene Ideen und Arbeitsweisen bot und durch den Anspruch an mathematische Präzision ein solides Fundament schuf.

Ebenso danken möchte ich meiner Zweitbetreuerin Prof. Dr. Renate Meyer. Die zahlreichen Gespräche und E-Mails mit wertvollen Impulsen und Korrekturvorschlägen haben wesentlich zu dieser Arbeit beigetragen. Insbesondere möchte ich mich auch für die Unterstützung während meiner Zeit an der University of Auckland bedanken.

Vielen Dank an alle Kolleginnen und Kollegen vom IMST der OvGU für das angenehme Arbeitsumfeld, insbesondere an Kerstin Altenkirch, Kerstin Reckrühm und Christina Stöhr. I would like to thank my friend and colleague Dr. Matt Edwards for the suggestions about improving this work and for his support during my time in Auckland.

Ich möchte mich herzlich bei meiner Familie und meinen Freunden bedanken. Danke an Sebastian Ernst für die langjährige Freundschaft und Unterstützung. Danke an Tina Hund, Marianne Meier und Peter Meier für die kontinuierliche und volle Unterstützung meiner Vorhaben und den Rückhalt in allen Lebenslagen. Ebenso geht mein Dank an Matthias, Greta und Charlotte Hund. Schließlich möchte ich mich bei Katherina Hartmann bedanken für die beständige Unterstützung an meiner Seite, Liebe und Freundschaft.

Meine Arbeit wurde von der Deutschen Forschungsgesellschaft (DFG AZ KI 1443/3-1) finanziert. Mein erster Aufenthalt an der University of Auckland wurde durch den Deutschen Akademischen Austauschdienst im Rahmen eines *Kurzzeitstipendiums für Doktoranden (2016/2017)* unterstützt.

Contents

List of Figures

List of Tables

Preface

Spectral Modeling of Stationary Time Series

Spectral theory is a very popular approach to mathematically describe the dependence of a time series. The idea is to decompose the linear dependence structure as a superposition of periodic components of different frequencies. Such a *frequency-domain representation* (or *spectral representation*) lends itself naturally for inspection and quantification of periodic phenomena in the data. To elaborate, consider the weak stationary and centered d-dimensional time series $\{\underline{Z}_t \colon t \in \mathbb{Z}\}$ with autocovariance function $\boldsymbol{\Gamma}(h) = \mathrm{E}[\underline{Z}_{t+h}\underline{Z}_t^T] \in \mathbb{R}^{d\times d}$. The famous Herglotz Lemma (see Theorem 11.8.1 in Brockwell and Davis (1991)) states that $\boldsymbol{\Gamma}$ can be written in terms of a superposition of 2π-periodic trigonometric functions:

$$\boldsymbol{\Gamma}(h) = \int_{[0,2\pi]} \exp(ih\omega)\boldsymbol{F}(d\omega), \quad h \in \mathbb{Z},$$

for a Hermitian positive semidefinite matrix valued measure $\boldsymbol{F} = (F_{j,k})_{j,k=1}^d$ on $[0, 2\pi]$, that is, $F_{j,j} = F_j$ are finite measures and, for $j \neq k$, $F_{j,k} = F_{j,k}^{(r)} + iF_{j,k}^{(i)}$, where both $F_{j,k}^{(r)}$ and $F_{j,k}^{(i)}$ are differences of finite measures and, for all measurable sets B, the $d \times d$-matrix $\boldsymbol{F}(B)$ is Hermitian positive semidefinite (Hpsd). The integral in the above representation is understood componentwise. It may be noted that – despite their natural occurrence in spectral analysis of multivariate time series – Hpsd measures have also emerged as mathematical research objects in their own rights. For further details on them and their role in multivariate time series analysis, see e.g. Salehi (1968), Robertson and Rosenberg (1968), Rosenberg (1974) and the references therein.

Throughout this work, we denote by $\|\boldsymbol{A}\|$ the Frobenius norm of a $d\times d$ matrix $\boldsymbol{A}$, see (B.3) in the Appendix. If the autocovariance function of $\underline{Z}_t$ is absolutely summable, i.e. $\sum_{h\in\mathbb{Z}} \|\boldsymbol{\Gamma}(h)\| < \infty$, then $\boldsymbol{F}$ possesses a continuous Lebesgue density $\boldsymbol{f} \colon [0, 2\pi] \to \bar{\mathcal{S}}_d^+$, where $\bar{\mathcal{S}}_d^+$ denotes the set of Hpsd matrices. In this case, $\boldsymbol{f}$ is given by an inverse Fourier transform of $\boldsymbol{\Gamma}$:

$$\boldsymbol{f}(\omega) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \boldsymbol{\Gamma}(h)\exp(-ih\omega), \quad 0 \leq \omega \leq 2\pi. \tag{0.1}$$

Since $\boldsymbol{\Gamma}(h) \in \mathbb{R}^{d\times d}$ and $\boldsymbol{\Gamma}(-h) = \boldsymbol{\Gamma}(h)^T$ holds for for all $h \in \mathbb{Z}$ (with $\boldsymbol{A}^T$ denoting the transpose of a matrix $\boldsymbol{A}$), it is immediate that $\boldsymbol{f}(2\pi - \omega) = \boldsymbol{f}(\omega)^T$ holds for $0 \leq \omega \leq \pi$. Due to this symmetry property, it suffices to consider $\boldsymbol{F}$ and $\boldsymbol{f}$ on $[0, \pi]$ respectively. Since there is a one-to-one relation between $\boldsymbol{\Gamma}$ and $\boldsymbol{f}$, the spectral density matrix contains all the linear dependence structure of the time series $\{\underline{Z}_t\}$. It constitutes the object of primary interest in this work.

Bayesian Nonparametrics

Consider a statistical model with data $(Z_1, \dots, Z_n)$ obeying a *likelihood* function P_θ^n, with a parameter θ belonging to some set Θ of possible parameters. In the Bayesian framework, θ is modeled as random. Any knowledge about it (before observing any data) has to be employed in terms of the *prior* distribution $P(d\theta)$. The connection between data and parameter is described in terms of the likelihood $P_\theta^n(Z_1, \dots, Z_n) = P(Z_1, \dots, Z_n|\theta)$. Once the data $(Z_1, \dots, Z_n)$ are observed, they are considered as fixed in the Bayesian paradigm and the prior is updated with the likelihood to yield the *posterior* distribution $P(d\theta|Z_1, \dots, Z_n)$, which is given by Bayes' formula

$$P(d\theta|Z_1, \dots, Z_n) = \frac{P(Z_1, \dots, Z_n|\theta)P(d\theta)}{\int_\Theta P(Z_1, \dots, Z_n|\theta)P(d\theta)} \propto P(Z_1, \dots, Z_n|\theta)P(d\theta), \tag{0.2}$$

where the notation $p(z) \propto q(z)$ refers to the existence of a positive constant (not depending on z) such that $p(z) = cq(z)$ holds. Often the posterior distribution is not available in closed-form to analyze, but can be approximated numerically by drawing a large random sample based on (0.2), e.g. using Markov Chain Monte Carlo (MCMC) methods. There are many different aspects that make the Bayesian paradigm appealing, such as having the inference results at hand as a whole in terms of one single mathematical object, i.e. the posterior distribution. However, the discussion of the advantages and disadvantages of the Bayesian framework in comparison to the *classical* (frequentist) approach is far beyond the scope of this work and the reader is e.g. referred to Section 1.1 in Ghosal and van der Vaart (2017).

A model is called *parametric*, if the parameter space is finite dimensional, e.g. $\Theta \subset \mathbb{R}^d$. Parametric models are known to be very powerful if they are well-specified (i.e. if the data generating process belongs to the parametric class in consideration), however inference may be misleading in case of misspecification, as discussed in Kleijn and Van der Vaart (2012).

On the other hand, in a *nonparametric* model, the parameter space is of infinite dimension. Exemplary objects to think about include infinite series, measures or functions – such as the spectral density of a stationary time series. Despite being less powerful than parametric approaches, nonparametric models are typically much more flexible in the sense that they cover a much larger class of data generating processes for which "valid" (in an asymptotic sense to be specified later) inference can be conducted. However, establishing such validity results in Bayesian nonparametrics is much harder than in the finite dimensional case: In fact, an important difference is that the prior matters even asymptotically and much care has to be taken not to choose an "unfortunate" prior that ruins validity. Due to advantages in theory and due to increasing computational power and capacity for numerical posterior approximation, Bayesian nonparametrics have seen a steady growth over the last few decades and are currently a very active field of research.

A *semiparametric* model consists of both a parametric and a nonparametric part. Semiparametric models exist in numerous variants (see Chapter 12 in Ghosal and van der Vaart (2017) or Chapter 4 in Bickel et al. (1998) for examples) and in this work, we will restrict our attention to *partitioned models*, i.e. parameters of the form $(\theta, \eta) \in \Theta \times \mathcal{H}$, with Θ being finite dimensional and $\mathcal{H}$ being infinite dimensional. Often θ is considered the *parameter of interest* and η

a *nuisance parameter*. As an example from the scope of this work, consider $\theta = \underline{\mu} \in \mathbb{R}^d$ being the mean and $\eta = \boldsymbol{f}$ being the spectral density of a stationary time series.

Outline

This work is structured in three main parts and an Appendix.

Part I presents the construction of an Hpsd Gamma process. After an introduction into random processes and revisiting Kingman's process construction from the univariate case in Chapter 1, we will consider the infinitely divisible Hermitian positive definite (Hpd) Gamma distribution in Chapter 2. Based on this distribution, we will construct an Hpd Gamma process in Chapter 3. We will investigate distributional properties of the process concerning support and lower bounds for probability masses. A practical representation in terms of an almost surely convergent infinite series is derived and the Chapter is concluded with a numerical illustration of random samples from the process.

Part II discusses a new Bayesian nonparametic method for multivariate spectral density inference, based on the Hpd Gamma process from Part I. The starting point is an elaboration of Whittle's Likelihood in Chapter 4. A main result of this Chapter is the mutual contiguity of the full Gaussian Likelihood and Whittle's Likelihood for Gaussian stationary time series, which serves as a foundation for later asymptotic considerations and constitutes a result of independent interest beyond the scope of Bayesian inference. A new nonparametric prior for the spectral density matrix will be introduced in Chapter 5, where also the issue of numerical simulation of posterior samples is discussed. Chapter 6 contains an illustration of the proposed method, in terms of a real data example and a comparative simulation study. In Chapter 7, $\mathbb{L}^1$-consistency of the posterior distribution is established and posterior contraction rates in the Hellinger topology are derived.

Part III presents a semiparametric Bayesian linear model, consisting of a parametric Bayesian linear model and a time series error term that is modeled nonparametrically. In Chapter 8, the model is presented, in which the Gaussian stationary and centered error time series is modeled with the spectral density matrix prior from Part II. The performance of the proposed method is illustrated in terms of a comparative simulation study. Chapter 9 specializes to the case of a mean model, in which the linear model consists only of an intercept and mean vector. Contraction rates for the joint posterior of mean and spectral density are derived, and it is also shown that the marginal posterior of the mean converges at a parametric rate under rather mild prior assumptions. The chapter concludes with an outlook on how a Bernstein-von-Mises theorem for the mean could be established.

Appendix A summarizes all the mathematical assumptions that have been used throughout this work, whereas Appendix B contains several auxiliary definitions and results. To emphasize, Section B.1 contains all important definitions and results from matrix algebra which are used extensively throughout this work. The notation is summarized in Appendix C.

Part I.

Matrix-Valued Gamma Process

1.

Introduction

We start our considerations by a brief literature review about existing Bayesian nonparametric approaches to time series analysis in Section 1.1. One of the most famous tools in Bayesian nonparametrics is the Dirichlet process, which – along with its close connection to the Gamma process – we will revisit in Section 1.2.

1.1. Bayesian Nonparametric Time Series Analysis

The standardized spectral measure $\tilde{F} = F/F([0,\pi])$, of a univariate stationary time series is a probability measure on $[0,\pi]$. In particular, the corresponding standardized spectral density $\tilde{f} = f/F([0,\pi])$ is a probability density. This property has been used in several Bayesian nonparametric modeling approaches, where well-established methods for modeling compactly supported probability densities have been tailored to this application. As a prominent example, the *Bernstein-Dirichlet prior* from Choudhuri et al. (2004a) relies on a mixture of Bernstein polynomials with mixture weights induced by a Dirichlet process to model the normalized spectral density $\tilde{f}$ on $[0,\pi]$. To elaborate, let G_0 be a measure on $[0,1]$ and $M > 0$. Then the Bernstein-Dirichlet prior is defined as

$$\begin{aligned} \tilde{f}(\pi x) &= \sum_{j=1}^{k} G\left(\left(\frac{j-1}{k}, \frac{j}{k}\right]\right) b(x|j, k-j+1), \quad 0 \le x \le 1, \\ G &\sim \mathrm{DP}(MG_0), \\ k &\sim p(k), \end{aligned} \tag{1.1}$$

with $b(\cdot|j, k-j+1)$ denoting the density of the Beta$(j, k-j+1)$ distribution (see (B.20) in the Appendix) and $\mathrm{DP}(MG_0)$ denoting the Dirichlet process with base measure MG_0 (see the upcoming Section 1.2 for a detailed discussion thereof). Here, $G_0 = \mathrm{E}G$ is the prior mean of the random probability measure G and the prior variance is proportional to M^{-1} (thus M is also called *precision* or *concentration* parameter). We refer to measure-valued random variables as *random measures*. Accordingly, *random probability measures* are random variables that take values in the space of probability measures. Furthermore, $p(k)$ in (1.1) denotes the prior probability mass function for the polynomial degree $k \in \mathbb{N}$. The random function $\tilde{f}$ in (1.1) is

continuous and integrates to 1, because $\sum_{j=1}^{k} G(((j-1)/k, j/k]) = G((0,\pi]) = 1$ with probability 1. See Appendix B.2 (and Lorentz (2012) and Petrone (1999)) for further details on the Bernstein polynomial approximation of compactly supported functions.

It is also possible to model the normalized spectral measure $\tilde{F}$ itself with a Dirichlet process:

$$\tilde{F} \sim \mathrm{DP}(M\tilde{F}_0), \tag{1.2}$$

where $\tilde{F}_0$ is a measure on $[0,\pi]$ that reflects the prior mean. However, with probability one, the draws of a Dirichlet process are *discrete* probability measures (Blackwell, 1973), i.e. they consist of countably many random probability mass atoms. It may be emphasized that the notion of discreteness of measures is not to be confused with topological discreteness. While the former is defined as consisting of at most countably many mass atoms, the latter denotes a minimum distance between the atoms. The model (1.2) is not suited for inferring the spectral density function, since the resulting spectral measure is discrete and in particular not absolutely continuous with respect to the Lebesgue measure.

Instead of modeling the normalized spectral density $\tilde{f}$ and the normalized spectral measure $\tilde{F}$ respectively with a Dirichlet process as in (1.1) and (1.2), it is also possible to model the (unnormalized) spectral density f and the (unnormalized) spectral measure F respectively with an *unnormalized process*, i.e. a random measure F on $[0,\pi]$ such that $F([0,\pi]) < \infty$ holds with probability 1. As an example, a Gamma process (that is, an independent increment process with Gamma distributed increments) is used in Section 4.4 in Hjort et al. (2010) for this purpose:

$$F(d\omega) \overset{\text{ind.}}{\sim} \mathrm{Ga}(\alpha(\omega), \beta(\omega)), \quad \alpha(\omega) = M(\omega)f_0(\omega)d\omega, \quad \beta(\omega) = M(\omega), \quad 0 \leq \omega \leq \pi, \tag{1.3}$$

where f_0 is the continuous density of the prior mean spectral measure, and $M\colon [0,\pi] \to (0,\infty)$ is a measurable *concentration function*. The notation in (1.3) is understood as the infinitesimal increments of the random measure F on $[0,\pi]$ being Gamma distributed (see the upcoming Section 3.1 for a mathematically rigorous explanation of this rather vague statement). Similarly, it is possible to formulate a prior for f as in (1.1), but with a Gamma process prior for the mixture-weight inducing random measure G instead of a Dirichlet process.

In fact, the Dirichlet process is strongly related to the Gamma process, since it can be obtained by a normalization thereof, see Chapter 9 in Kingman (1992). The Gamma process is a special case of a more general concept, namely *completely random measures*. These are widely used in Bayesian nonparametrics (see James et al. (2009), Lijoi (2010)). Before having a closer look at the Gamma process and its connection to the Dirichlet process, we consider the problem of spectral inference for multivariate time series.

In view of (1.1)–(1.3), the question arises whether there is an extension of the Gamma process or of the Dirichlet process for modeling the spectral measure and density of multivariate time series. Bayesian nonparametric approaches for related problems exist. For example, in Zhang et al. (2014) the joint prior distribution of a set of random matrices is modeled by means of Dirichlet processes, where the base measure is a matrix-variate distribution. The authors refer to this prior as *matrix-variate Dirichlet process*. Since the number of matrices in consideration is fixed there (to the observation length n), it is not clear how this approach could be embedded in a mixture setting such as in (1.1). In Dunson et al. (2008), a matrix of (dependent) probability measures

is constructed by stick-breaking, to which the authors refer as *matrix stick-breaking process*. This method offers an alternative to parametric hierarchical models when clustering studies with mutual borrowing of information is of interest. The constructed matrices of probability measures do not obey any particular structure (such as Hermitian positive definiteness). In Belitser et al. (2015), Bayesian nonparametric estimation of the intensity of an inhomogeneous Poisson process is considered. Although it does not fall into the scope of time series analysis, the work of the authors is related to this thesis in terms of the intersection between Poisson processes (see the Hpd Gamma process construction in the upcoming Section 3) and asymptotic results in the Bayesian nonparametric framework (see the upcoming Section 7). In Wolpert et al. (2011), the parameters of a Bayesian wavelet mixture model are described by a Lévy random field and in Zheng et al. (2009), the spectral density of a random field is modeled nonparametrically with a Dirichlet process in conjunction with multi-dimensional Bernstein polynomials. This constitutes a generalization of the Bernstein-Dirichlet prior (1.1) that is orthogonal to the method presented in this thesis. In fact, an interesting idea for future research would be a combination of both ideas, towards a Bayesian nonparametric method for multivariate random fields. In Cadonna et al. (2016), a hierarchical Gaussian mixture model is employed in conjunction with Whittle's Likelihood for the log spectral density of multiple time series. The proposed method of the authors yields computationally efficient inference for the spectral densities of the replicated series, whereas the cross-correlation between the series is not of interest.

Surprisingly, there exists no approach for Bayesian nonparametric modeling of matrix-valued measures, though there are several reasons that make such a modeling appealing. Most previous approaches to nonparametric Bayesian multivariate time series analysis rely on a nonparametric model of the components of the Cholesky decomposition of $\boldsymbol{f}^{-1}$ (Rosen and Stoffer, 2007; Zhang, 2016) to ensure positive definiteness, a methodology that has recently been extended to the case of locally stationary time series Li and Krafty (2018). However, due to the non-linearity of the Cholesky decomposition, it is difficult to employ prior knowledge about the mean and covariance structure of $\boldsymbol{f}$. Furthermore, the asymptotic properties of these procedures such as posterior consistency and posterior contraction rates are not known.

1.2. The Dirichlet Process and the Gamma Process

The Dirichlet process is arguably one of the most widely used tools in Bayesian nonparametric modeling. There are several equivalent characterizations of the Dirichlet process, which will be briefly discussed in the following. These characterizations form the basis for many extensions of the Dirichlet process (see Lijoi (2010) for an overview). We will also emphasize the strong connection between the Dirichlet Process and the Gamma process before revisiting the famous Kingman construction thereof, which constitutes the key technique for the construction of the matrix-valued Gamma process in the upcoming Chapter 3. For what follows, let the underlying space $\mathcal{X}$ fulfill the following assumption.

Assumption $\mathcal{X}$1. *Let $\mathcal{X}$ be a Polish space, i.e. a topological space that is homeomorphic to a complete metric space having a countable dense subset. Let $\mathcal{X}$ be equipped with a locally compact and σ-finite and non-trivial Borel measure denoted by dx.*

As suggested by the notation in $\mathcal{X}1$, we may think of dx being equal to the Lebesgue measure if $\mathcal{X}$ is a subset of a Euclidean space. This is e.g. the case for spectral inference (with $\mathcal{X} = [0, \pi]$), but the process construction discussed in this section is not limited to this case. Furthermore, we denote by $\mathcal{M}$ the set of finite measures on $\mathcal{X}$, endowed with the smallest σ-algebra such that the mapping $\mathcal{M} \ni \Phi \mapsto \Phi(A)$ is measurable for all $\Phi \in \mathcal{M}$ and all $A \subset \mathcal{X}$ measurable (with respect to the Borel σ-algebra). We denote by $\mathcal{M}^*$ the set of probability measures on $\mathcal{X}$.

The Classical Dirichlet Process

The original definition of the Dirichlet process (as given together with the proof of its existence in Ferguson (1973)) is stated as follows: Let G_0 be a probability measure on $\mathcal{X}$ and $M > 0$. A random probability measure G on $\mathcal{X}$ is a Dirichlet process if for all (measurable) finite partitions of $\mathcal{X}$, the corresponding probability vector follows a Dirichlet distribution:

$$\begin{aligned} &G \sim \mathrm{DP}(MG_0) : \Longleftrightarrow \\ &(G(A_1), \dots, G(A_k)) \sim \mathrm{Dir}(MG_0(A_1), \dots, MG_0(A_k)), \quad \forall k, \quad \forall \sum_{j=1}^{k} A_j = \mathcal{X}, \end{aligned} \tag{1.4}$$

where $\sum$ denotes the disjoint set union, and the Dirichlet distribution is as defined in (B.24) in the Appendix. From this definition, it follows readily that the expected value of G is G_0, with variance proportional to M^{-1}, see (B.25) in the Appendix (justifying the names *base measure* or *center measure* (Ghosal and van der Vaart (2017)) for G_0 and *prior precision* or *concentration parameter* for M).

Another equivalent definition of $\mathrm{DP}(MG_0)$ was given by Sethuraman (1994) and is of great importance for many practical applications, as e.g. the sampling methods presented in Ishwaran and James (2001). Indeed, in case of its existence, the Dirichlet process can be constructed by a *stick-breaking* with independent identically distributed (iid) Beta weight proportions and G_0 distributed atoms:

$$\begin{aligned} &G \sim \mathrm{DP}(MG_0) \iff \\ &G \overset{d}{=} \sum_{j \geq 1} p_j \delta_{Z_j}, \quad p_j = V_j \prod_{l=1}^{j-1} (1 - V_l), \quad V_j \overset{\text{iid}}{\sim} \mathrm{Beta}(1, M), \quad Z_j \overset{\text{iid}}{\sim} G_0, \end{aligned} \tag{1.5}$$

where the sequences $\{V_j\}$ and $\{Z_j\}$ are independent and δ_Z denotes the Dirac delta function. The name stick-breaking comes from the conception of the probability mass $G(\mathcal{X})$ of G being a stick of length 1. The first atom weight p_1 is given by breaking a (random) fraction V_1 from the stick. From the remaining fraction $(1 - V_1)$ of the stick, a fraction V_2 is broken to determine p_2. And so on. For further details on the stick-breaking, see Sethuraman (1994) or Section 4.2.5 in Ghosal and van der Vaart (2017).

Recall the definition of the Gamma distribution from (B.26) in the Appendix. The Dirichlet process can also be conceived as a normalized Gamma process. This is an infinite-dimensional extension of the well-known construction of the k-dimensional Dirichlet distribution as vector of k independent Gamma variables divided by their sum:

$$\gamma_j \overset{\text{ind.}}{\sim} \mathrm{Ga}(\alpha_j, 1), \ j = 1, \dots, k \implies \left(\frac{\gamma_1}{\sum \gamma_j}, \dots, \frac{\gamma_k}{\sum \gamma_j} \right) \sim \mathrm{Dir}(\alpha_1, \dots, \alpha_k), \tag{1.6}$$

(where this distributional equality can readily be verified by using the densities of the Gamma and Dirichlet distributions). To elaborate, let Φ be a *completely random measure* on $\mathcal{X}$, i.e. Φ is a random measure (an $\mathcal{M}$-valued random variable) such that the random variables $\Phi(A_1), \dots, \Phi(A_k)$ are independent for all disjoint subsets $A_1, \dots, A_k$ of $\mathcal{X}$ and all $k > 0$. Furthermore, let the increments of Φ be Gamma distributed:

$$\Phi(dx) \overset{\text{ind.}}{\sim} \text{Ga}(MG_0(dx), M). \tag{1.7}$$

The infinitesimal notation in (1.7) is understood as

$$\Phi(A) \sim \text{Ga}(MG_0(A), M), \quad \text{for all measurable } A \subset \mathcal{X}$$

and the notation will be made more precise for the more general multivariate setting in the upcoming Section 3.1. For convention, we define $\text{Ga}(0, M)$ to be the degenerate measure at $\{0\}$, for every $M > 0$. We shall refer to Φ also as *completely random Gamma measure*, or (in particular for $\mathcal{X} \subset \mathbb{R}$) as *Gamma process with independent increments* or – whenever the context is clear – simply as *Gamma process*. It is known (see Chapter 9 in Kingman (1992)) that

$$G \sim \text{DP}(MG_0) \iff G \overset{d}{=} \frac{\Phi}{\Phi(\mathcal{X})}, \quad \Phi \text{ as in (1.7)},$$

which is an infinite-dimensional generalization of (1.6). The Gamma process Φ from (1.7) is called *homogeneous*, because the precision parameter M is a fixed constant that does not vary along $\mathcal{X}$. In the more general case of an *inhomogeneous* process, M is a measurable mapping from $\mathcal{X}$ to $(0, \infty)$ and $\Phi(dx) \overset{\text{ind.}}{\sim} \text{Ga}(M(x)G_0(dx), M(x))$, where this notation will be made precise later in (3.5). The Dirichlet process is by far not the only process than can be constructed by normalization. In fact, the class of random probability distributions that are obtained by a normalization of completely random measures are known as *normalized random measures with independent increments (NRMIs)* in the literature Regazzini et al. (2003). Applications of NRMIs in a Bayesian nonparametric framework have been considered in James et al. (2009) (see also Lijoi (2010) for an overview). Before proceeding to the Hpd measure setting, we will highlight the Kingman construction of Φ in the following section. This technique shows how the Gamma process Φ (and hence also the Dirichlet process) can be constructed from a Poisson process Π on $\mathcal{X} \times [0, \infty)$.

Kingman's Construction of the Gamma Process

In this section, we briefly revisit the Kingman construction (see Chapter 8 and Chapter 9 in Kingman (1992)), of the Gamma process Φ from (1.7). The construction is based on Poisson processes and for a brief introduction into this topic along with the most important results, the reader may confer Section B.4 in the Appendix. Let $\mathcal{X}$ fulfill Assumption $\mathcal{X}1$ and consider the Borel space $\mathcal{Y} := \mathcal{X} \times [0, \infty)$, endowed with the product σ-algebra. Let G_0 be a probability measure on $\mathcal{X}$ and $M > 0$. Define the measure α on $\mathcal{X}$ as $\alpha(dx) = MG_0(dx)$ and consider the mapping $\beta \colon \mathcal{X} \to (0, \infty)$, given as $\beta(x) \equiv M$. Let Π be a Poisson-process on $\mathcal{Y}$ with mean measure

$$\nu(dx, dz) = \frac{\exp(-\beta(x)z)}{z} dz\alpha(dx). \tag{1.8}$$

The infinitesimal notation for ν in (1.8) is understood as follows: Let $B \subset \mathcal{Y}$ be measurable. Then $\nu(B) := \int_{\mathcal{Y}} \mathbb{1}_B(x,z)\frac{\exp(-\beta(x)z)}{z}dz\alpha(dx)$, with $\mathbb{1}_B$ denoting the indicator function of the set B. Now define

$$\begin{gathered}\Phi \sim \mathrm{CRM}(\nu) : \Longleftrightarrow \\ \Phi(A) = \sum_{(x,z)\in\Pi} \mathbb{1}_A(x)z, \quad A \in \mathbb{B}(\mathcal{X}), \quad \Pi \sim \mathrm{PP}(\nu),\end{gathered} \tag{1.9}$$

with $\mathbb{B}(\mathcal{X})$ denoting the Borel sets in $\mathcal{X}$. It follows from the independence property of Π (see Appendix B.4) that $\Phi(A_1), \ldots, \Phi(A_k)$ are independent for each disjoint partition $A_1, \ldots, A_k$ of $\mathcal{X}$. Hence Φ defines a completely random measure, motivating the notation $\Phi \sim \mathrm{CRM}(\nu)$. The measure ν is called the *Poisson mean measure* of Φ.

Remark. *An intuition of* (1.9) *is as follows: A realization of* Π *consist of a countable set of pairs* $(x,z) \in \mathcal{X}\times[0,\infty)$*. Each such pair can be interpreted as a mass atom* x *and a corresponding weight* z*. The (random) measure value* $\Phi(A)$ *of* $A \subset \mathcal{X}$ *is given by summing up all the weights* z *corresponding to atoms* $x \in A$ *of* $(x,z) \in \Pi$*. The mean measure* ν *of* Π *controls the mass/weight distribution.*

Campbell's Theorem (see Theorem B.19 in the Appendix) describes the distribution of $\Phi \sim \mathrm{CRM}(\nu)$ in terms of the Laplace transform:

$$\begin{aligned}\mathrm{E}\exp(-t\Phi(A)) &= \exp\left(-\int_{\mathcal{Y}}(1-\exp(-t\mathbb{1}_A(x)z))\,\nu(dx,dz)\right) \\ &= \exp\left(-MG_0(A)\int_0^\infty (1-\exp(-tz))\exp(-Mz)z^{-1}dz\right)\end{aligned} \tag{1.10}$$

for $t \geq 0$. It is known (see e.g. (9.6) in Kingman (1992)) that the last line in (1.10) is equal to $(1+M^{-1}t)^{-MG_0(A)}$, which shows that $\Phi(A) \sim \mathrm{Ga}(MG_0(A), M)$. In particular, the total mass $\Phi(\mathcal{X}) \sim \mathrm{Ga}(M,M)$ is finite with probability one. Accordingly, the construction (1.9) indeed yields the Gamma process (1.7). The key property of the Gamma distribution for this construction to be valid is its *infinite divisibility*. Recall that a probability distribution P is called infinitely divisible if, for every $n \in \mathbb{N}$, there exist iid random variables $X_{n,1}, \ldots, X_{n,n}$ such that $X_{n,1} + \ldots + X_{n,n} \overset{d}{=} P$. This property is needed because of the one-to-one relation between infinitely divisible laws and so called *Lévy–Khinchine representations* such as (1.10). We will elaborate this in more detail in the upcoming Section 2.1.

The Gamma process has received considerable attention in the Bayesian nonparametric literature. As an example, in Roychowdhury and Kulis (2015) a Gamma process is used for prior modeling of latent structures within a count data framework. The authors also derive a stick-breaking representation of the Gamma process similar to (1.5) and quantify truncation error bounds.

The Matrix Case: A Negative Result

The most famous extension of the Dirichlet distribution to the random Hpd matrix case is the complex matrix variate Dirichlet distribution, see Troskie (1967), Cui et al. (2005), Gupta and

Bedoya (2007) and Nagar and Gupta (2009). To elaborate, let $k > 0$ and $\alpha_1, \dots, \alpha_k > (d-1)$. The k-vector of Hermitian $d \times d$ matrices $(\boldsymbol{U}_1, \dots, \boldsymbol{U}_k)$ is said to follow the complex matrix variate Dirichlet distribution $\mathrm{Dir}_{d\times d}(\alpha_1, \dots, \alpha_k)$, if its joint probability density function is given by

$$p(\boldsymbol{U}_1, \dots, \boldsymbol{U}_k) \propto \prod_{j=1}^{k} |\boldsymbol{U}_j|^{\alpha_j - d}, \quad 0 < \boldsymbol{U}_j < \boldsymbol{I}_d, \quad \sum_{j=1}^{k} \boldsymbol{U}_j = \boldsymbol{I}_d, \tag{1.11}$$

where $|\boldsymbol{A}|$ denotes the absolute value of the determinant of the matrix $\boldsymbol{A}$ and $\boldsymbol{A} < \boldsymbol{B}$ is understood as $\boldsymbol{B} - \boldsymbol{A}$ being Hpd. The question arises whether there exists a generalization of the Dirichlet process (1.4) with the matrix variate Dirichlet distribution $\mathrm{Dir}_{d\times d}$, to a *matrix variate Dirichlet process*. A negative answer to this question has been given in Bobecka et al. (2010).

A deeper insight can be gained by the following observations. Consider the complex Wishart distribution $\mathrm{CWish}_{d\times d}(\eta, \boldsymbol{\Sigma})$, which is defined on the set $\mathcal{S}_d^+$ of Hpd matrices in terms of the probability density

$$f_W(\boldsymbol{Z}) = \frac{1}{\tilde{\Gamma}_d(\eta)|\boldsymbol{\Sigma}|^{\eta}} \operatorname{etr}(-\boldsymbol{\Sigma}^{-1}\boldsymbol{Z})|\boldsymbol{Z}|^{\eta-d}, \quad \boldsymbol{Z} \in \mathcal{S}_d^+, \tag{1.12}$$

where $\eta > d-1$ denotes the degrees of freedom and $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$ the covariance matrix. See Goodman (1963) for further details. A similar relation as (1.6) connects the complex Wishart distribution with the matrix variate Dirichlet distribution (see Theorem 1.8 in Cui et al. (2005)):

$$\begin{aligned} \boldsymbol{W}_j &\overset{\text{ind.}}{\sim} \mathrm{CWish}_{d\times d}(\eta_j, \boldsymbol{I}_d), \quad j = 1, \dots, k, \\ \implies \left(\boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1/2}\boldsymbol{W}_1\boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1/2*}, \dots, \boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1/2}\boldsymbol{W}_k\boldsymbol{\Sigma}_{\boldsymbol{W}}^{-1/2*}\right) &\sim \mathrm{Dir}_{d\times d}(\eta_1, \dots, \eta_k), \end{aligned} \tag{1.13}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{W}}^{1/2}$ denotes the Hpd square root (see (B.15) in the Appendix) of $\boldsymbol{\Sigma}_{\boldsymbol{W}} = \sum_{j=1}^{k} \boldsymbol{W}_j$. From (1.12), it can be seen that the probability density of the complex Wishart distribution is an Hpd matrix-valued generalization of the Gamma distribution. But a process construction as in Section 1.2 can *not* be conducted with the complex Wishart distribution, because it lacks the neccessary property of infinite divisibility (Lévy, 1948).

A closely related approach has been presented in Bru (1991), where the authors defined random processes with Wishart distributed marginals in terms of solutions to stochastic differential equations. This can be conceived as a generalization of a univariate process with χ^2_m-distributed marginals (which can be constructed as the sum of squares of m independent standard Wiener Processes), to the case of non-integer values of m and to the matrix case. Although the Wishart marginals of these processes are positive semidefinite (even positive definite for m large enough), this property does not translate to the process *increments* (think of a non-increasing process, where the increments may be negative). Accordingly, this approach is not suited for the framework of random measures, where nonnegative increments are an intrinsically necessary property for modeling the random measure increments.

Motivated from these negative results, we will present a different generalization of the Gamma distribution to the random Hpd matrix case that has the desirable property of infinite divisibility in the upcoming Chapter 2. This distribution will be used to construct an Hpd Gamma process (similar to (1.9)) in the upcoming Chapter 3.

2.

Infinitely Divisible Hpd Gamma Distributions

A class of cone-valued infinitely divisible Gamma distributions has been proposed in Pérez-Abreu and Stelzer (2014). There the authors also considered the special case of the cone of symmetric positive definite (spd) matrices, yielding the class of infinitely divisible spd Gamma distributions. In the following Section 2.2, we present a similar specialization thereof: the class of infinitely divisible complex Hermitian positive definite (Hpd) matrix Gamma distributions. Several properties of the spd matrix distributions derived in Pérez-Abreu and Stelzer (2014) will be translated to the Hpd matrix case. Distributional properties will be investigated in Section 2.3 and a special case (the Hpd $A\Gamma$ distribution) discussed in Section 2.4.

2.1. Preliminaries

We will start our considerations by a brief introduction into infinitely divisible distributions on cones and in particular the famous Lévy Khinchine representation. This is followed by a brief introduction into the basic concepts of matrix calculus and functions of matrix argument.

2.1.1. Infinitely Divisible Distributions on Cones

Recall that a *cone* $\mathcal{K}$ is a subset of a Hilbert space $(\mathcal{X}, \langle\cdot,\cdot\rangle)$, which is closed under multiplication with the positive real numbers, i.e. $\alpha\boldsymbol{Z} \in \mathcal{K}$ for all $\alpha > 0$ and $\boldsymbol{Z} \in \mathcal{K}$. A cone is called *proper*, if it is closed, not equal to the trivial cone $\{0\}$ and if it does not contain a straight line through $\mathbf{0}$. The *dual cone* of $\mathcal{K}$ is defined as $\mathcal{K}' := \{\boldsymbol{\Theta} \in \mathcal{X} \colon \langle\boldsymbol{\Theta}, \boldsymbol{Z}\rangle \geq 0 \text{ for all } \boldsymbol{Z} \in \mathcal{K}\}$. Let $\mathcal{X}$ be endowed with the Borel σ-algebra. For a probability distribution μ on $\mathcal{K}$, the *Laplace transform* of μ is defined as

$$L_\mu(\boldsymbol{\Theta}) := \int_{\mathcal{K}} \exp(-\langle\boldsymbol{\Theta}, \boldsymbol{Z}\rangle)\mu(dZ), \quad \boldsymbol{\Theta} \in \mathcal{K}'.$$

It is clear that the real half-line $[0, \infty)$ is a proper cone. Another important example is the space $\mathcal{S}_d$ of Hermitian $d \times d$ matrices, endowed with the inner product $\langle\boldsymbol{X}, \boldsymbol{Z}\rangle := \operatorname{tr}(\boldsymbol{X}\boldsymbol{Z})$ and the open cone $\mathcal{S}_d^+ \subset \mathcal{S}_d$ of Hermitian positive definite matrices and its closure $\bar{\mathcal{S}}_d^+$, the

closed cone of Hermitian positive semidefinite matrices. It is clear that $\bar{\mathcal{S}}_d^+$ is proper and it follows from Lemma B.7 in the Appendix that $(\bar{\mathcal{S}}_d^+)' = \bar{\mathcal{S}}_d^+$. The boundary $\bar{\mathcal{S}}_d^+ \setminus \mathcal{S}_d^+$ consists of all Hermitian positive semidefinite matrices that are not invertible. For a random element $\boldsymbol{X}$ in $\bar{\mathcal{S}}_d^+$ with probability distribution μ, the Laplace transform becomes $L_\mu(\boldsymbol{\Theta}) = \mathrm{E}\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{X})$ for $\boldsymbol{\Theta} \in \bar{\mathcal{S}}_d^+$, where $\operatorname{etr}(\boldsymbol{Z}) := \exp(\operatorname{tr}\boldsymbol{Z})$ denotes the exponential trace. Recall that a probability distribution μ is called *infinitely divisible* if for every $n \in \mathbb{N}$ there exist independent identically distributed random variables $X_{1,n}, \ldots, X_{n,n}$ such that the sum $\sum_{j=1}^n X_{j,n}$ has distribution μ. The following result will be of great importance.

Theorem 2.1 (Lévy Khinchine representation on proper cones)**.** *Let $\mathcal{K}$ be a proper cone on a finite-dimensional Hilbert space and let μ be a probability distribution on $\mathcal{K}$. Then μ is infinitely divisible if and only if there exists $\boldsymbol{\Gamma}_0 \in \mathcal{K}$ and a measure ν on $\mathcal{K}$ with $\nu(\{\mathbf{0}\}) = 0$ and $\int_{\mathcal{K}} \min(1, \sqrt{\langle Z, Z\rangle})\nu(d\boldsymbol{Z}) < \infty$ such that the* Lévy Khinchine representation

$$L_\mu(\boldsymbol{\Theta}) = \exp\left(-\langle \boldsymbol{\Theta}, \boldsymbol{\Gamma}_0\rangle - \int_{\mathcal{K}} (1 - \exp(-\langle \boldsymbol{\Theta}, \boldsymbol{Z}\rangle))\nu(d\boldsymbol{Z})\right), \quad \boldsymbol{\Theta} \in \mathcal{K}', \tag{2.1}$$

holds for the Laplace transform L_μ of μ.

Proof. See Theorem 1 and Remark 2 in Pérez-Abreu and Rosiński (2007). □

The measure ν in Theorem 2.1 is called the *Lévy measure* of μ. The parameter $\boldsymbol{\Gamma}_0$ can be conceived as a mere translation parameter. Indeed, for a random element $\boldsymbol{X} \sim \mu$ in $\mathcal{K}$ with μ as in (2.1), let $\tilde{\boldsymbol{X}} \sim \tilde{\mu}$ with $\tilde{\mu}$ defined in terms of the Lévy Khinchine representation

$$L_{\tilde{\mu}}(\boldsymbol{\Theta}) = \exp\left(-\int_{\mathcal{K}} (1 - \exp(-\langle \boldsymbol{\Theta}, \boldsymbol{Z}\rangle))\nu(d\boldsymbol{Z})\right), \quad \boldsymbol{\Theta} \in \mathcal{K}'.$$

Then it is not difficult to see that $\boldsymbol{X} \overset{d}{=} \tilde{\boldsymbol{X}} + \boldsymbol{\Gamma}_0$. The result from Theorem 2.1 can be extended to infinite-dimensional spaces, see Dettweiler (1976) and Pérez-Abreu and Rosiński (2007). For a formulation of the Lévy Khinchine representation for Euclidean spaces instead of cones (involving the Characteristic Function rather than the Laplace transform), see e.g. Section 8 in Sato (1999). In what follows, we will neglect the translation parameter $\boldsymbol{\Gamma}_0$ from the Lévy Khinchine representation (2.1) and assume $\boldsymbol{\Gamma}_0 = \mathbf{0}$.

A particularly important class of infinitely divisible distributions are the Compound Poisson distributions, which are characterized by the following result.

Theorem 2.2. *Let $\mathcal{K}$ be a proper cone on a finite-dimensional Hilbert space and ν^* be a probability measure on $\mathcal{K}$. Consider the random matrix $\boldsymbol{X} := \boldsymbol{S}_N$ with $\boldsymbol{S}_0 := \mathbf{0}$ and $\boldsymbol{S}_k := \sum_{j=1}^k \boldsymbol{Y}_j$ for $k \geq 1$ and $\boldsymbol{Y}_j \overset{iid}{\sim} \nu^*$ and a random variable $N \sim \mathrm{Poi}(C)$ which is independent of $\{\boldsymbol{Y}_j\}$, where C is a positive constant. Then $\boldsymbol{X}$ is called* Compound Poisson*. The Laplace transform of $\boldsymbol{X}$ is given by*

$$\mathrm{E}\exp\left(-\langle \boldsymbol{\Theta}, \boldsymbol{X}\rangle\right) = \exp\left(C\left(L_{\nu^*}(\boldsymbol{\Theta}) - 1\right)\right), \quad \boldsymbol{\Theta} \in \mathcal{K}'.$$

In this case, we write $\boldsymbol{X} \sim \mathrm{CPoi}(C, \nu^*)$.

Proof. The result is shown for $\mathbb{R}^d$ (involving the Characteristic Function rather than the Laplace Transform) in Theorem 4.3 in Sato (1999) and the proof for the cone case follows along the lines. □

Any infinitely divisible distribution with finite Lévy measure ν is Compound Poisson, as the following result summarizes.

Lemma 2.3. *Let $\mathcal{K}$ be a proper cone on a finite-dimensional Hilbert space and let $\boldsymbol{X}$ be an infinitely divisible random element in $\mathcal{K}$ with finite Lévy measure ν. Then $\boldsymbol{X} \sim \mathrm{CPoi}(C, \nu^*)$ with $C := \nu(\mathcal{K})$ and $\nu^* := \frac{\nu}{C}$.*

Proof. From Theorem 2.1, we get

$$\mathrm{E}\exp(-\langle \boldsymbol{\Theta}, \boldsymbol{X}\rangle) = \exp\left(-\int_{\mathcal{K}} (1 - \exp(-\langle \boldsymbol{\Theta}, \boldsymbol{Z}\rangle))\, \nu(d\boldsymbol{Z})\right), \quad \boldsymbol{\Theta} \in \mathcal{K}'.$$

Since ν is finite, the right hand side can be computed as

$$\exp\left(\int_{\mathcal{K}} \exp\left(-\langle \boldsymbol{\Theta}, \boldsymbol{Z}\rangle\right) \nu(d\boldsymbol{Z}) - \nu(\mathcal{K})\right) = \exp\left(C\left(\int_{\mathcal{K}} \exp\left(-\langle \boldsymbol{\Theta}, \boldsymbol{Z}\rangle\right) \nu^*(d\boldsymbol{Z}) - 1\right)\right),$$

which – in view of Theorem 2.2 – concludes the proof. □

Generally speaking, any infinitely divisible distribution μ with Lévy measure ν (not necessarily finite) can be approximated by a sequence of Compound Poisson distributions $\mathrm{CPoi}(C_n, \nu_n^*)$ with $C_n \to \nu(\mathcal{K})$ (which may be $+\infty$) as $n \to \infty$, see e.g. p.44 f. in Sato (1999). We will make use of such an approximation technique in the proof of the following results, see e.g. the proof of the upcoming Theorem 2.6.

For an infinitely divisible random element $\boldsymbol{X}$ in $\bar{\mathcal{S}}_d^+$, the Lévy Khinchine representation specializes as

$$\mathrm{E}\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{X}) = \exp\left(-\int_{\bar{\mathcal{S}}_d^+} (1 - \operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{Z}))\nu(d\boldsymbol{Z})\right), \quad \boldsymbol{\Theta} \in \bar{\mathcal{S}}_d^+, \tag{2.2}$$

with Lévy measure ν on $\bar{\mathcal{S}}_d^+$ fulfilling $\nu(\{\boldsymbol{0}\}) = 0$ and $\int_{\bar{\mathcal{S}}_d^+} \min(1, \sqrt{\operatorname{tr}(\boldsymbol{Z}^2)})\nu(d\boldsymbol{Z}) < \infty$. On $\mathcal{S}_d$, we will consider the trace norm $\|\boldsymbol{A}\|_T := \operatorname{tr}((\boldsymbol{X}^2)^{1/2})$, where $\boldsymbol{A}^{1/2}$ denotes the Hpd square root of the Hpd matrix $\boldsymbol{A}$ (see (B.7) and (B.15) in the Appendix). The trace norm is *not* the norm induced by the inner product $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B})$. However, we will use it in the following considerations because for $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$ it obeys the particularly convenient representation $\|\boldsymbol{A}\|_T = \operatorname{tr}\boldsymbol{A}$ (see (B.9) in the Appendix). Let $\mathbb{S}_d$ denote the unit sphere in $(\mathcal{S}_d, \|\cdot\|_T)$. Let $\mathbb{S}_d^+ = \bar{\mathcal{S}}_d^+ \cap \mathbb{S}_d$ be the intersection of the unit sphere with the cone, and $\bar{\mathbb{S}}_d^+$ its closure. Note that

$$\mathbb{S}_d^+ = \left\{\boldsymbol{U} \in \mathcal{S}_d^+ \colon \operatorname{tr}\boldsymbol{U} = 1\right\}. \tag{2.3}$$

Every $\boldsymbol{Z} \in \bar{\mathcal{S}}_d^+ \setminus \{\boldsymbol{0}\}$ can be decomposed into a *radial part* $r := \|\boldsymbol{Z}\|_T \in (0, \infty)$ and a *spherical part* $\boldsymbol{U} := \frac{\boldsymbol{Z}}{\|\boldsymbol{Z}\|_T} \in \bar{\mathbb{S}}_d^+$ such that $\boldsymbol{Z} = r\boldsymbol{U}$. We call this decomposition $\bar{\mathcal{S}}_d^+ \cong \bar{\mathbb{S}}_d^+ \times [0, \infty)$ the *Polar*

decomposition of the cone $\bar{\mathcal{S}}_d^+$. With respect to this Polar decomposition, the Lévy Khinchine representation from (2.2) for an infinitely divisible random element $\boldsymbol{X}$ in $\bar{\mathcal{S}}_d^+$ specializes as

$$\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{X}) = \exp\left(-\int_{\bar{\mathbb{S}}_d^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}))\nu(d\boldsymbol{U},dr)\right), \quad \boldsymbol{\Theta}\in\bar{\mathcal{S}}_d^+, \tag{2.4}$$

with the Lévy measure ν being defined on $\bar{\mathbb{S}}_d^+ \times [0,\infty)$ and fulfilling $\nu(\bar{\mathbb{S}}_d^+,\{0\}) = 0$ as well as $\int_{\bar{\mathbb{S}}_d^+}\int_0^\infty \min(1,r)\nu(d\boldsymbol{U},dr) < \infty$.

2.1.2. Matrix Calculus

For $\boldsymbol{Z}\in\mathcal{S}_d$, write $\boldsymbol{Z} = \boldsymbol{Z}_1 + i\boldsymbol{Z}_2$ with $\boldsymbol{Z}_1 = (z_{1jk}) \in \mathbb{R}^{d\times d}$ being symmetric and $\boldsymbol{Z}_2 = (z_{2jk}) \in \mathbb{R}^{d\times d}$ being skew symmetric. Then the Lebesgue measure $d\boldsymbol{Z}$ on $\mathcal{S}_d$ is defined as

$$d\boldsymbol{Z} := d\boldsymbol{Z}_1 d\boldsymbol{Z}_2, \quad d\boldsymbol{Z}_1 := \prod_{j=1}^d\prod_{k=j}^d dz_{1jk}, \quad d\boldsymbol{Z}_2 := \prod_{j=1}^d\prod_{k=j+1}^d dz_{2jk}. \tag{2.5}$$

The infinitesimal notation in (2.5) is understood as follows: For any measurable $A\subset\mathcal{S}_d$, the Lebesgue mass $\mathcal{L}_{\mathcal{S}_d}(A)$ of A in $\mathcal{S}_d \cong \mathbb{R}^{d^2}$ is

$$\mathcal{L}_{\mathcal{S}_d}(A) := \int_{\mathcal{S}_d} \mathbb{1}_A(\boldsymbol{Z})d\boldsymbol{Z} := \int_{\mathbb{R}^{d^2}} \mathbb{1}_A(\boldsymbol{Z}_1+i\boldsymbol{Z}_2)d\boldsymbol{Z}_1 d\boldsymbol{Z}_2,$$

see (1.2.10) and (3.0.2) in Mathai (1997). Similarly, for any measurable function $g\colon \mathcal{S}_d\to[0,\infty)$, the Lebesgue integral

$$\int_{\mathcal{S}_d} g(\boldsymbol{Z})d\boldsymbol{Z} := \int_{\mathbb{R}^{d^2}} g(\boldsymbol{Z}_1+i\boldsymbol{Z}_2)d\boldsymbol{Z}_1 d\boldsymbol{Z}_1$$

is defined as usual.

For $\boldsymbol{U}\in\bar{\mathbb{S}}_d^+$, write $\boldsymbol{U} = \boldsymbol{U}_1 + i\boldsymbol{U}_2$ with real matrices $\boldsymbol{U}_1,\boldsymbol{U}_2$, where we observe from (2.3) that the last diagonal entry of $\boldsymbol{U}_1$ can be written as $u_{1dd} = 1 - \sum_{j=1}^{d-1} u_{1jj}$. Thus any measurable function $h\colon \bar{\mathbb{S}}_d^+\to[0,\infty)$ can be written in the form $h(\boldsymbol{U}) = h(\underline{\tilde{u}}_1,\boldsymbol{U}_2)$, where $\underline{\tilde{u}}_1$ is a real vector of dimension $\tilde{d} := \frac{d(d+1)}{2} - 1$ containing all functionally independent entries of $\boldsymbol{U}_1$. Accordingly, we define $d\boldsymbol{U} = d\underline{\tilde{u}}_1 d\boldsymbol{U}_2$ (with $d\underline{\tilde{u}}_1$ denoting the Lebesgue measure on $\mathbb{R}^{\tilde{d}}$ and $d\boldsymbol{U}_2$ defined as in (2.5)) and

$$\int_{\bar{\mathbb{S}}_d^+} h(\boldsymbol{U})d\boldsymbol{U} := \int_{\mathbb{R}^{d^2-1}} \mathbb{1}_{\{\boldsymbol{U}_1+i\boldsymbol{U}_2\in\bar{\mathbb{S}}_d^+\}} h(\boldsymbol{U}_1+i\boldsymbol{U}_2)d\underline{\tilde{u}}_1 d\boldsymbol{U}_2.$$

In particular, for $B\subset\bar{\mathbb{S}}_d^+$ measurable, the Lebesgue mass $\mathcal{L}_{\bar{\mathbb{S}}_d^+}(B)$ of B in $\bar{\mathbb{S}}_d^+$ is

$$\mathcal{L}_{\bar{\mathbb{S}}_d^+}(B) := \int_{\bar{\mathbb{S}}_d^+} \mathbb{1}_B(\boldsymbol{U})d\boldsymbol{U} = \int_{\mathbb{R}^{d^2-1}} \mathbb{1}_B(\boldsymbol{U}_1+i\boldsymbol{U}_2)d\underline{\tilde{u}}_1 d\boldsymbol{U}_2 \tag{2.6}$$

Some further results from matrix calculus are collected in Appendix B.1 and referenced whenever needed.

2.2. A Class of Infinitely Divisible Hpd Gamma Distributions

Let α be a finite measure on $\bar{\mathbb{S}}_d^+$ and let $\beta\colon \bar{\mathbb{S}}_d^+ \to (0,\infty)$ be measurable such that the following assumption is fulfilled:

Assumption α1. *The integral* $\int_{\bar{\mathbb{S}}_d^+} \log\left(1+\frac{1}{\beta(\boldsymbol{U})}\right)\alpha(d\boldsymbol{U})$ *is finite.*

The infinitely divisible Hpd Gamma distribution $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ with parameters α and β is defined by means of its Laplace transform in Polar decomposition (see (2.4)):

$$\boldsymbol{X}\sim \mathrm{Ga}_{d\times d}(\alpha,\beta) :\Longleftrightarrow$$
$$\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{X}) = \exp\left(-\int_{\bar{\mathbb{S}}_d^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}))\frac{\exp(-\beta(\boldsymbol{U})r)}{r}dr\alpha(d\boldsymbol{U})\right) \tag{2.7}$$

for all $\boldsymbol{\Theta}\in\bar{\mathcal{S}}_d^+$. The distribution of type (2.7) has been presented in Pérez-Abreu and Stelzer (2014) for general cones, and we will investigate the special case of Hpd matrices in more detail. It has been shown in Pérez-Abreu and Stelzer (2014) (see Proposition 3.3 there) that the $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ distribution (2.7) is well-defined if and only if Assumption α1 is fulfilled. In this case, it follows from Theorem 2.1 that

$$\int_{\bar{\mathbb{S}}_d^+}\int_0^\infty \min(1,r)\frac{\exp(-\beta(\boldsymbol{U})r)}{r}dr\alpha(d\boldsymbol{U}) < \infty. \tag{2.8}$$

If $\beta\equiv\beta_0>0$ is constant, $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ is called *homogeneous*. Conducting the integration with respect to the radial part dr in (2.7) (see Proposition 3.10 in Pérez-Abreu and Stelzer (2014) for the precise argument) yields the alternative representation of the Laplace transform

$$\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{X}) = \exp\left(-\int_{\bar{\mathbb{S}}_d^+}\log\left(1+\frac{\mathrm{tr}(\boldsymbol{\Theta}\boldsymbol{U})}{\beta(\boldsymbol{U})}\right)\alpha(d\boldsymbol{U})\right). \tag{2.9}$$

Being defined in terms of the Lévy Khinchine representation of its Laplace transform (2.7), the Hpd Gamma distribution $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ is necessarily infinitely divisible, according to the characterization of Theorem 2.1. The measure

$$\nu(d\boldsymbol{U},dr) := \frac{\exp(-\beta(\boldsymbol{U})r)}{r}dr\alpha(d\boldsymbol{U}) \tag{2.10}$$

from (2.7) is the Lévy measure of $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ on $\bar{\mathcal{S}}_d^+\cong\bar{\mathbb{S}}_d^+\times[0,\infty)$. It may be noted that (2.10) accommodates the one-dimensional Gamma distribution, since for $d=1$ we have $\bar{\mathbb{S}}_d^+=\{1\}$, reducing α to a nonnegative number and β to a positive number. Considering this, the name *Hpd* Gamma is justified since $P(\boldsymbol{X}\in\bar{\mathcal{S}}_d^+)=1$ and, under appropriate assumptions on α, even $P(\boldsymbol{X}\in\mathcal{S}_d^+)=1$, see the upcoming Theorem 2.6. In Pérez-Abreu and Stelzer (2014), the authors investigated a real spd version of the $\mathrm{Ga}_{d\times d}$ distribution in more detail and the Hpd Gamma distribution has similar properties as its spd counterpart. As an example, in the next result, the distribution under Hpd-preserving invertible linear transformations is given.

Lemma 2.4. *Let* $\boldsymbol{X}\sim\mathrm{Ga}_{d\times d}(\alpha,\beta)$ *with* α *and* β *fulfilling Assumption* α*1. Let* $\boldsymbol{C}\in\mathbb{C}^{d\times d}$ *be invertible. Let* $\boldsymbol{C}^{-*}:=(\boldsymbol{C}^*)^{-1}$ *be the inverse of the Hermitian conjugate of* $\boldsymbol{C}$*. Define the* $\boldsymbol{C}$*-weighted trace norm as* $\|\boldsymbol{B}\|_{\boldsymbol{C}}:=\|\boldsymbol{C}^{-1}\boldsymbol{B}\boldsymbol{C}^{-*}\|_T$ *and denote the unit sphere in* $\mathcal{S}_d^+$ *with respect*

to $\|\cdot\|_{\boldsymbol{C}}$ *by* $\mathbb{S}_{\boldsymbol{C}}^+ := \{\boldsymbol{V} \in \mathcal{S}_d^+ : \|\boldsymbol{V}\|_{\boldsymbol{C}} = 1\}$*, and the closure thereof as* $\bar{\mathbb{S}}_{\boldsymbol{C}}^+$*. Then* $\boldsymbol{Y} := \boldsymbol{C}\boldsymbol{X}\boldsymbol{C}^*$ *is distributed as* $\mathrm{Ga}_{d\times d}(\tilde{\alpha}_{\boldsymbol{C}}, \tilde{\beta}_{\boldsymbol{C}})$*, where*

$$\tilde{\alpha}_{\boldsymbol{C}}(E) = \int_{\bar{\mathbb{S}}_{\boldsymbol{C}}^+} \mathbb{1}_E\left(\frac{\boldsymbol{V}}{\|\boldsymbol{V}\|_T}\right)\alpha_{\boldsymbol{C}}(d\boldsymbol{V}), \quad E \subset \bar{\mathbb{S}}_d^+ \text{ measurable},$$
$$\tilde{\beta}_{\boldsymbol{C}}(\boldsymbol{U}) = \beta_{\boldsymbol{C}}\left(\frac{\boldsymbol{U}}{\|\boldsymbol{U}\|_{\boldsymbol{C}}}\right)\|\boldsymbol{U}\|_{\boldsymbol{C}}, \quad \boldsymbol{U} \in \bar{\mathbb{S}}_d^+$$

with the measure $\alpha_{\boldsymbol{C}}$ *on* $\bar{\mathbb{S}}_{\boldsymbol{C}}^+$ *and the mapping* $\beta_{\boldsymbol{C}} \colon \bar{\mathbb{S}}_{\boldsymbol{C}}^+ \to (0,\infty)$ *given by*

$$\alpha_{\boldsymbol{C}}(F) = \alpha(\boldsymbol{C}^{-1}F\boldsymbol{C}^{-*}), \quad F \subset \bar{\mathbb{S}}_{\boldsymbol{C}}^+ \text{ measurable},$$
$$\beta_{\boldsymbol{C}}(\boldsymbol{V}) = \beta(\boldsymbol{C}^{-1}\boldsymbol{V}\boldsymbol{C}^{-*}), \quad \boldsymbol{V} \in \bar{\mathbb{S}}_{\boldsymbol{C}}^+.$$

with $\boldsymbol{C}^{-1}F\boldsymbol{C}^{-*} := \{\boldsymbol{C}^{-1}\boldsymbol{V}\boldsymbol{C}^{-*} \colon \boldsymbol{V} \in F\}$.

Proof. The result is well-known in the literature (see Proposition 5.3 in Pérez-Abreu and Stelzer (2014)), however since there is no proof available, it will be presented here for the sake of completeness. For $\boldsymbol{\Theta} \in \bar{\mathcal{S}}_d^+$, we compute, using (2.7),

$$\begin{aligned}\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Y}) &= \mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{C}\boldsymbol{X}\boldsymbol{C}^*) = \mathrm{E}\,\mathrm{etr}(-\boldsymbol{C}^*\boldsymbol{\Theta}\boldsymbol{C}\boldsymbol{X})\\ &= \exp\left(-\int_{\bar{\mathbb{S}}^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{C}^*\boldsymbol{\Theta}\boldsymbol{C}\boldsymbol{U}))\frac{\exp(-r\beta(\boldsymbol{U}))}{r}dr\alpha(d\boldsymbol{U})\right)\\ &= \exp\left(-\int_{\bar{\mathbb{S}}^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\boldsymbol{C}\boldsymbol{U}\boldsymbol{C}^*))\frac{\exp(-r\beta(\boldsymbol{U}))}{r}dr\alpha(d\boldsymbol{U})\right).\end{aligned}$$

Employing the transformation $\bar{\mathbb{S}}^+ \ni \boldsymbol{U} \mapsto \boldsymbol{U}_{\boldsymbol{C}} := \boldsymbol{C}\boldsymbol{U}\boldsymbol{C}^* \in \bar{\mathbb{S}}_{\boldsymbol{C}}^+$ yields

$$\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Y}) = \exp\left(-\int_{\bar{\mathbb{S}}_{\boldsymbol{C}}^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}_{\boldsymbol{C}}))\frac{\exp(-r\beta_{\boldsymbol{C}}(\boldsymbol{U}_{\boldsymbol{C}}))}{r}dr\alpha_{\boldsymbol{C}}(d\boldsymbol{U}_{\boldsymbol{C}})\right),$$

Now consider the transformation $\bar{\mathbb{S}}_{\boldsymbol{C}}^+ \ni \boldsymbol{U}_{\boldsymbol{C}} \mapsto \boldsymbol{V} := \|\boldsymbol{U}_{\boldsymbol{C}}\|_T^{-1}\boldsymbol{U}_{\boldsymbol{C}} \in \bar{\mathbb{S}}^+$. Note that $\boldsymbol{U}_{\boldsymbol{C}} = \|\boldsymbol{V}\|_{\boldsymbol{C}}^{-1}\boldsymbol{V}$. Denote by $\tilde{\alpha}_{\boldsymbol{C}}$ the measure on $\bar{\mathbb{S}}^+$ induced by $\alpha_{\boldsymbol{C}}$ via the transformation $\boldsymbol{U}_{\boldsymbol{C}} \mapsto \boldsymbol{V}$. Then

$$\begin{aligned}\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Y}) &= \exp\left(-\int_{\bar{\mathbb{S}}^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\|\boldsymbol{V}\|_{\boldsymbol{C}}^{-1}\boldsymbol{V}))\frac{\exp(-r\beta_{\boldsymbol{C}}(\|\boldsymbol{V}\|_{\boldsymbol{C}}^{-1}\boldsymbol{V}))}{r}dr\tilde{\alpha}_{\boldsymbol{C}}(d\boldsymbol{V})\right)\\ &= \exp\left(-\int_{\bar{\mathbb{S}}^+}\int_0^\infty (1-\mathrm{etr}(-\rho\boldsymbol{\Theta}\boldsymbol{V}))\frac{\exp(-\rho\tilde{\beta}_{\boldsymbol{C}}(\boldsymbol{V}))}{\rho}d\rho\tilde{\alpha}_{\boldsymbol{C}}(d\boldsymbol{V})\right),\end{aligned}$$

where the transformation $(0,\infty) \ni r \mapsto \rho := \|\boldsymbol{V}\|_C^{-1}r \in (0,\infty)$ was used in the last step and $\tilde{\beta}_{\boldsymbol{C}}$ is as in the claim. This concludes the proof. □

2.3. Distributional Properties

In general, there is no closed form probability density available for the $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ distribution. Quantifying its support within the cone of Hpsd matrices is still possible – using different

techniques that rely on the Lévy Khinchine representation instead. Such a result will be derived in this section, see the upcoming Theorem 2.6.

Recall that for any measure μ on a Borel space $\mathcal{X}$, the *support* of μ is defined as

$$\operatorname{supp}(\mu) := \{x \in \mathcal{X} \colon \mu(U_x) > 0 \text{ for every open neighborhood } U_x \text{ of } x\}. \tag{2.11}$$

For a random variable (or matrix) X, the support $\operatorname{supp}(X)$ of X is defined as the support of the probability distribution P_X of X. Recall the notion of compound Poisson distributions from Theorem 2.2. We will start with a general result for Compound Poisson random matrices.

Lemma 2.5. *Let ν be a finite nontrivial measure on $\bar{\mathcal{S}}_d^+$ with $\nu(\{\mathbf{0}\}) = 0$ and let $C := \nu(\bar{\mathcal{S}}_d^+)$ as well as $\nu^* = \frac{\nu}{C}$. Consider the Compound Poisson random matrix $\boldsymbol{X} \sim \mathrm{CPoi}(\mathrm{C}, \nu^*)$. Then the probability distribution $P_{\boldsymbol{X}}$ of $\boldsymbol{X}$ fulfills*

$$P_{\boldsymbol{X}} = \exp(-C) \sum_{k=0}^{\infty} \frac{1}{k!} \nu^k, \tag{2.12}$$

where ν^k denotes the k-fold convolution of ν (see Section 1.3.1 in Rudin (1962)), i.e. $\nu^0 := \delta_{\mathbf{0}}$ and $\nu^k := \nu^{k-1} * \nu$ with the convolution $\mu_1 * \mu_2$ of two finite measures μ_1, μ_2 on a Borel space $\mathcal{X}$ being defined as $(\mu_1 * \mu_2)(E) = \int_{\mathcal{X}} \int_{\mathcal{X}} \mathbb{1}_E(x+y)\mu_1(dx)\mu_2(dy)$ for $E \subset \mathcal{X}$ measurable. In particular, it holds

$$\operatorname{supp}(\nu) \subset \operatorname{supp}(\boldsymbol{X}) \subset \bar{\mathcal{S}}_d^+.$$

Proof. The representation (2.12) follows from $\boldsymbol{X} \overset{d}{=} \boldsymbol{S}_N$ with $N \sim \mathrm{Poi}(C)$ and $\boldsymbol{S}_k$ for $k \geq 0$ as in Theorem 2.2. Since $P(N = 1) = C\exp(-C) > 0$, this implies $\operatorname{supp}(\nu) = \operatorname{supp}(\nu^*) \subset \operatorname{supp}(\boldsymbol{X})$. Furthermore, for every $k \in \mathbb{N}$ it holds $\operatorname{supp}(\boldsymbol{S}_k) \subset \bar{\mathcal{S}}_d^+$ by Lemma B.27 in the Appendix and since $\operatorname{supp}(\boldsymbol{X})$ is the closure of $\bigcup_{k=0}^{\infty} \operatorname{supp}(\boldsymbol{S}_k)$, this concludes the proof. □

Theorem 2.6. *Let α, β fulfill Assumption $\alpha 1$ and let $\boldsymbol{X} \sim \mathrm{Ga}_{d\times d}(\alpha, \beta)$.*

(a) Then $\operatorname{supp}(\boldsymbol{X}) \subset \bar{\mathcal{S}}_d^+$.

(b) If $\operatorname{supp}(\alpha)$ contains at least d^2 linearly independent (in $\mathcal{S}_d$) elements, then the distribution of $\boldsymbol{X}$ has a probability density with respect to the Lebesgue measure on $\mathcal{S}_d$. In particular, $P(\boldsymbol{X} \in \bar{\mathcal{S}}_d^+ \setminus \mathcal{S}_d^+) = 0$ in this case.

(c) If $\operatorname{supp}(\alpha) = \bar{\mathbb{S}}_d^+$, then $\operatorname{supp}(\boldsymbol{X}) = \bar{\mathcal{S}}_d^+$.

Proof. The proof of (a) is inspired by the one-dimensional case (see Theorem 24.7 in Sato (1999)), and relies on the idea of approximating the distribution of $\boldsymbol{X}$ by a sequence $(\boldsymbol{X}_n)$ of compound Poisson distributed variables. To elaborate, for the Lévy measure $\nu(d\boldsymbol{U}, dr)$ of the $\mathrm{Ga}_{d\times d}(\alpha, \beta)$ distribution from (2.10), define $\nu_n(d\boldsymbol{U}, dr) := \mathbb{1}_{(1/n,\infty)}(r)\nu(d\boldsymbol{U}, dr)$ and define the distribution of $\boldsymbol{X}_n$ for $n \in \mathbb{N}$ in terms of the Laplace transform

$$\mathrm{E}\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{X}_n) := \exp\left(-\int_{\mathbb{S}_d^+}\int_0^{\infty} (1 - \operatorname{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}))\nu_n(d\boldsymbol{U}, dr)\right), \quad \boldsymbol{\Theta} \in \bar{\mathcal{S}}_d^+.$$

Since

$$\int_{\mathbb{S}_d^+}\int_0^\infty \nu_n(d\boldsymbol{U},dr) = \int_{\mathbb{S}_d^+}\int_{1/n}^\infty \nu(d\boldsymbol{U},dr) \leq n\int_{\mathbb{S}_d^+}\int_0^\infty \min(r,1)\nu(d\boldsymbol{U},dr) < \infty$$

for every n by (2.8), it follows from Lemma 2.3 that $\boldsymbol{X}_n$ is Compound Poisson. Since $\operatorname{supp}(\nu_n) \subset \operatorname{supp}(\nu) \subset \bar{\mathcal{S}}_d^+$ (where the second inclusion is trivial by the definition of ν from (2.10)), it follows from Lemma 2.5 that $\operatorname{supp}(\boldsymbol{X}_n) \subset \bar{\mathcal{S}}_d^+$ for every n. As $n \to \infty$, it holds

$$\mathrm{E}\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{X}_n) \to \mathrm{E}\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{X}), \quad \boldsymbol{\Theta} \in \bar{\mathcal{S}}_d^+$$

by an application of Lebesgue's Dominated Convergence Theorem. The Lévy Continuity Theorem for Laplace transforms (see Theorem 4.3 in Kallenberg (1997)) yields that $\boldsymbol{X}_n$ converges to $\boldsymbol{X}$ in distribution. In particular, it holds $1 = P(\boldsymbol{X}_n \in \bar{\mathcal{S}}_d^+) \to P(\boldsymbol{X} \in \bar{\mathcal{S}}_d^+)$ as $n \to \infty$, concluding (a). Part (b) is Proposition 3.6 in Pérez-Abreu and Stelzer (2014) and part (c) will be shown in the upcoming Corollary 3.5. □

2.4. The Hpd AΓ Distribution

In its general parametrization from Section 2.2, the infinitely divisible Hpd Gamma distribution offers great flexibility in terms of its parameters α and β. We now consider a special case thereof, the so-called AΓ distribution which has been considered for the spd matrix case in Pérez-Abreu and Stelzer (2014). It has the desirable properties that all its cumulants are available analytically. This makes modeling of key distributional features such as mean and covariance structure easily possible, as will be shown in the upcoming Lemma 2.9.

Let $\eta > d-1$ and let $\omega > 0$ and $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$. Denote by $\boldsymbol{\Sigma}^{1/2}$ the Hpd square root of $\boldsymbol{\Sigma}$. Then the $A\Gamma$ distribution with parameters η, ω and $\boldsymbol{\Sigma}$ is defined as

$$\begin{aligned} \boldsymbol{Y} \sim A\Gamma(\eta,\omega,\boldsymbol{\Sigma}) &: \Longleftrightarrow \boldsymbol{Y} \overset{d}{=} \boldsymbol{\Sigma}^{1/2}\boldsymbol{X}\boldsymbol{\Sigma}^{1/2} \text{ with } \boldsymbol{X} \sim \boldsymbol{A}\Gamma(\eta,\omega,\boldsymbol{I}_d), \\ \boldsymbol{X} \sim A\Gamma(\eta,\omega,\boldsymbol{I}_d) &: \Longleftrightarrow \boldsymbol{X} \sim \mathrm{Ga}_{d\times d}(\omega\alpha_\eta, 1), \end{aligned} \tag{2.13}$$

where the measure α_η on $\bar{\mathbb{S}}_d^+$ is defined as

$$\alpha_\eta(d\boldsymbol{U}) = \frac{\Gamma(d\eta)}{\tilde{\Gamma}_d(\eta)}|\boldsymbol{U}|^{\eta-d}d\boldsymbol{U}, \tag{2.14}$$

with the Gamma function Γ and the complex multivariate Gamma function $\tilde{\Gamma}_d$ (see (B.14) in the Appendix). From Lemma B.10 in the Appendix, it follows that α_η is a probability measure and in particular (2.13) is well-defined, since $\omega\alpha_\eta$ is a finite measure, having total mass ω. The next Lemma shows that even for $\boldsymbol{\Sigma} \neq \boldsymbol{I}_d$, the $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$ distribution is also a member of the $\mathrm{Ga}_{d\times d}(\alpha,\beta)$ family and gives explicit formulas for the distributional parameters α and β.

Lemma 2.7. *The $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$ distribution is the $\mathrm{Ga}_{d\times d}(\omega\alpha_{\eta,\boldsymbol{\Sigma}},\beta_{\boldsymbol{\Sigma}})$ distribution with*

$$\alpha_{\eta,\boldsymbol{\Sigma}}(d\boldsymbol{U}) = |\boldsymbol{\Sigma}|^{-\eta}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})^{-d\eta}\alpha_\eta(d\boldsymbol{U})$$

and

$$\beta_{\boldsymbol{\Sigma}}(\boldsymbol{U}) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$$

and α_η as in (2.14).

Proof. By Lemma 2.4 with $\boldsymbol{C} = \boldsymbol{\Sigma}^{1/2}$ and $\alpha = \alpha_\eta$, it holds $\beta_{\boldsymbol{\Sigma}}(\boldsymbol{U}) = \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})$ and $\tilde{\alpha}_{\boldsymbol{C}}(E) = \omega \int \mathbb{1}_E(\boldsymbol{V}/\|\boldsymbol{V}\|_T)\alpha_{\boldsymbol{C}}(d\boldsymbol{V}) = \omega\alpha_{\eta,\boldsymbol{\Sigma}}(E)$. Without loss of generality, we take $\omega = 1$ in the following. This yields

$$\begin{aligned}\alpha_{\eta,\boldsymbol{\Sigma}}(E) &= \int_{\mathbb{S}^+_{\boldsymbol{\Sigma}^{1/2}}} \mathbb{1}_E\left(\frac{\boldsymbol{V}}{\|\boldsymbol{V}\|_T}\right)\alpha_{\boldsymbol{\Sigma}^{1/2}}(d\boldsymbol{V}) \\ &= \int_{\mathbb{S}^+_d} \mathbb{1}_E\left(\frac{\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}\|_T}\right)\alpha_\eta(d\boldsymbol{U}) \\ &= \frac{\Gamma(\eta d)}{\tilde{\Gamma}_d(\eta)}\int_{\mathbb{S}^+_d} \mathbb{1}_E\left(\frac{\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}\|_T}\right)|\boldsymbol{U}|^{\eta-d}d\boldsymbol{U},\end{aligned} \tag{2.15}$$

where the transformation $\mathbb{S}^+_{\Sigma^{1/2}} \ni \boldsymbol{V} \mapsto \boldsymbol{U} := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{V}\boldsymbol{\Sigma}^{-1/2} \in \mathbb{S}^+_d$ was employed. We will introduce an auxiliary radial Gamma component, to lift the integral from the sphere $\mathbb{S}^+_d$ to the cone $\mathcal{S}^+_d$, where in the latter we can apply transformations whose Jacobians are known. Indeed, using the transformation transformation

$$\boldsymbol{Z} = r\boldsymbol{U}, \quad d\boldsymbol{Z} = r^{d^2-1}drd\boldsymbol{U} \tag{2.16}$$

from Lemma B.8 in the Appendix, we obtain

$$\begin{aligned}\alpha_{\eta,\boldsymbol{\Sigma}}(E) &= \frac{1}{\tilde{\Gamma}_d(\eta)}\int_{\mathbb{S}^+_d}\int_0^\infty \mathbb{1}_E\left(\frac{\boldsymbol{\Sigma}^{1/2}r\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}}{\|\boldsymbol{\Sigma}^{1/2}r\boldsymbol{U}\boldsymbol{\Sigma}^{1/2}\|_T}\right)r^{d^2-1}\exp(-r)|r\boldsymbol{U}|^{\eta-d}drd\boldsymbol{U} \\ &= \frac{1}{\tilde{\Gamma}_d(\eta)}\int_{\mathcal{S}^+_d} \mathbb{1}_E\left(\frac{\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}}{\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}\|_T}\right)\operatorname{etr}(-\boldsymbol{Z})|\boldsymbol{Z}|^{\eta-d}d\boldsymbol{Z}.\end{aligned}$$

Employing successively the transformation

$$\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}\boldsymbol{\Sigma}^{1/2} \in \mathcal{S}^+_d, \quad d\boldsymbol{Z} = |\boldsymbol{\Sigma}|^{-d}d\boldsymbol{X}$$

from Lemma B.9 in the Appendix and (2.16) yields

$$\begin{aligned}\alpha_{\eta,\boldsymbol{\Sigma}}(E) &= \frac{1}{\tilde{\Gamma}_d(\eta)|\boldsymbol{\Sigma}|^\eta}\int_{\mathcal{S}^+_d} \mathbb{1}_E\left(\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|_T}\right)\operatorname{etr}(-\boldsymbol{\Sigma}^{-1}\boldsymbol{X})|\boldsymbol{X}|^{\eta-d}d\boldsymbol{X} \\ &= \frac{1}{\tilde{\Gamma}_d(\eta)|\boldsymbol{\Sigma}|^\eta}\int_{\mathbb{S}^+_d}\int_0^\infty \mathbb{1}_E(\boldsymbol{U})\operatorname{etr}(-r\boldsymbol{\Sigma}^{-1}\boldsymbol{U})r^{\eta d-1}|\boldsymbol{U}|^{\eta-d}drd\boldsymbol{U} \\ &= \frac{1}{\tilde{\Gamma}_d(\eta)|\boldsymbol{\Sigma}|^\eta}\int_{\mathbb{S}^+_d} \mathbb{1}_E(\boldsymbol{U})|\boldsymbol{U}|^{\eta-d}\rho(\boldsymbol{U})d\boldsymbol{U},\end{aligned}$$

where

$$\rho(\boldsymbol{U}) = \int_0^\infty \operatorname{etr}(-r\boldsymbol{\Sigma}^{-1}\boldsymbol{U})r^{\eta d-1}dr = \frac{\Gamma(\eta d)}{\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})^{\eta d}}.$$

This shows $\alpha_{\eta,\boldsymbol{\Sigma}}(d\boldsymbol{U}) = |\boldsymbol{\Sigma}|^{-\eta}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})^{-d\eta}\alpha_\eta(d\boldsymbol{U})$, completing the proof. □

From (2.10), we find that the Lévy measure $\nu_{\eta,\boldsymbol{\Sigma}}$ of $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$ has the Lebesgue density

$$g_{\eta,\boldsymbol{\Sigma}}(\boldsymbol{U},r) = \frac{\omega\Gamma(\eta d)}{\Gamma_d(\eta)|\boldsymbol{\Sigma}|^\eta}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{U})^{-\eta d}|\boldsymbol{U}|^{\eta-d}\frac{\operatorname{etr}(-r\boldsymbol{\Sigma}^{-1}\boldsymbol{U})}{r}, \quad \boldsymbol{U} \in \mathbb{S}^+_d, r > 0.$$

We call $g_{\eta,\boldsymbol{\Sigma}}$ the *Lévy density* of $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$. Applying the transformation from Lemma B.8 yields the alternative form

$$g_{\eta,\boldsymbol{\Sigma}}(\boldsymbol{Z}) = \frac{\omega\Gamma(d\eta)}{\tilde{\Gamma}_d(\eta)|\boldsymbol{\Sigma}|^\eta}\frac{\operatorname{etr}(-\boldsymbol{\Sigma}^{-1}\boldsymbol{Z})}{(\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}))^{d\eta}}|\boldsymbol{Z}|^{\eta-d}, \quad \boldsymbol{Z}\in\mathcal{S}_d^+. \tag{2.17}$$

The next Lemma is in line with Proposition 5.9 in Pérez-Abreu and Stelzer (2014). It is a general result that relates the cumulants of the $A\Gamma$ distribution to moments of the complex Wishart distribution from (1.12).

Lemma 2.8. *Let $\eta > d-1$ and $q>0$ and let the set $\mathcal{H}$ be given either by $\mathcal{H}=\mathcal{S}_d^+$ or $\mathcal{H}=(\mathcal{S}_d^+)^{\otimes p}$ for some $p>0$ or $\mathcal{H}=[0,\infty)$. Let $h\colon \mathcal{S}_d^+\to\mathcal{H}$ be a q-homogeneous function, i.e. $h(r\boldsymbol{Z}) = r^q h(\boldsymbol{Z})$ for all $r>0$. Then for the Lévy density $g_{\eta,\boldsymbol{\Sigma}}$ from (2.17) of the $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$ distribution it holds*

$$\int_{\mathcal{S}_d^+} h(\boldsymbol{Z})g_{\eta,\boldsymbol{\Sigma}}(\boldsymbol{Z})d\boldsymbol{Z} = \omega B(\eta d,q)\mathrm{E}h(\boldsymbol{W}), \tag{2.18}$$

where $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ denotes the beta function and $\boldsymbol{W}\sim\mathrm{CWish}_{d\times d}(\eta,\boldsymbol{\Sigma})$.

Proof. The proof is analogous to the proof of Proposition 5.9 in Pérez-Abreu and Stelzer (2014). □

Lemma 2.8 is applicable e.g. for the matrix power $h(\boldsymbol{Z})=\boldsymbol{Z}^p$ or the tensor product power $h(\boldsymbol{Z})=\boldsymbol{Z}^{\otimes p}$ with $q=p$ (i.e. $\boldsymbol{Z}^{\otimes 1}:=\boldsymbol{Z}\in\mathbb{C}^{d\times d}$ and $\boldsymbol{Z}^{\otimes m}:=\boldsymbol{Z}\otimes\boldsymbol{Z}^{\otimes(m-1)}\in\mathbb{C}^{d^m\times d^m}$ for $m>1$) or the determinant power $h(\boldsymbol{Z})=|\boldsymbol{Z}|^p$ (with $q=dp$). When applied to $h(\boldsymbol{Z})=\boldsymbol{Z}^{\otimes p}$, (2.18) relates the cumulants of the $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$ distribution to the moments of the $\mathrm{CWish}_{d\times d}(\eta,\boldsymbol{\Sigma})$ distribution. In particular, we obtain the following result for the mean and covariance structure.

Lemma 2.9. *Let $\boldsymbol{X}\sim A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$. Then, with $\mathrm{Cov}\boldsymbol{X} := \mathrm{E}[\boldsymbol{X}^{\otimes 2}]-(\mathrm{E}\boldsymbol{X})^{\otimes 2}\in\mathbb{C}^{d^2\times d^2}$ it holds*

$$\mathrm{E}\boldsymbol{X} = \frac{\omega}{d}\boldsymbol{\Sigma}, \quad \mathrm{Cov}\boldsymbol{X} = \frac{\omega}{d(\eta d+1)}\left(\eta\boldsymbol{I}_{d^2}+\boldsymbol{H}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}).$$

where $\boldsymbol{I}_{d^2}$ is the $d^2\times d^2$ identity matrix and $\boldsymbol{H}=\sum_{i,j=1}^d \boldsymbol{H}_{i,j}\otimes\boldsymbol{H}_{j,i}$ with the matrices $\boldsymbol{H}_{i,j}$ having a one at (i,j) and zeros elsewhere.

Proof. Using Lemma B.29 in the Appendix and the result from Lemma 2.8, we obtain

$$\mathrm{E}\boldsymbol{X} = \int_{\bar{\mathcal{S}}_d^+}\boldsymbol{Z}g_{\eta,\boldsymbol{\Sigma}}(\boldsymbol{Z})d\boldsymbol{Z} = \omega B(\eta d,1)\mathrm{E}\boldsymbol{W}$$

for $\boldsymbol{W}\sim\mathrm{CWish}_{d\times d}(\eta,\boldsymbol{\Sigma})$. It is known (see e.g. Maiwald and Kraus (2000) or Graczyk et al. (2003)) that $\mathrm{E}\boldsymbol{W}=\eta\boldsymbol{\Sigma}$ and $\mathrm{Cov}\boldsymbol{W}=\eta\boldsymbol{H}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})$. Using the properties $\Gamma(z+1)=z\Gamma(z)$ for $z\geq 1$ and $\Gamma(n)=(n-1)!$ for $n\in\mathbb{N}$ (with $0!:=1$) of the Gamma function, we also obtain $B(\eta d,1)=\Gamma(\eta d)\Gamma(1)/\Gamma(\eta d+1)=1/(\eta d)$, concluding $\mathrm{E}\boldsymbol{X}=\omega/d\boldsymbol{\Sigma}$. Similarly, we obtain

$$\mathrm{Cov}\boldsymbol{X} = \int_{\bar{\mathcal{S}}_d^+}\boldsymbol{Z}\otimes\boldsymbol{Z}g_{\eta,\boldsymbol{\Sigma}}(\boldsymbol{Z})d\boldsymbol{Z} = \omega B(\eta d,2)\mathrm{E}[\boldsymbol{W}\otimes\boldsymbol{W}].$$

From $\mathrm{E}[\boldsymbol{W}\otimes\boldsymbol{W}] = \mathrm{Cov}\boldsymbol{W}+\mathrm{E}\boldsymbol{W}\otimes\mathrm{E}\boldsymbol{W} = \eta\boldsymbol{H}(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})+\eta^2\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}$ and $B(\eta d, 2) = \Gamma(\eta d)\Gamma(2)/\Gamma(\eta d+2) = 1/(\eta d(\eta d+1))$, this leads to

$$\mathrm{Cov}\boldsymbol{X} = \frac{\omega}{d(\eta d+1)}\left(\eta\boldsymbol{I}_{d^2}+\boldsymbol{H}\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}).$$

□

Remark 2.10. *The parameter η of the $A\Gamma(\eta,\omega,\boldsymbol{\Sigma})$ distribution can be interpreted as a regularity parameter. Decreasing η results in more mass put towards the boundary (i.e. the non-invertible elements) of $\bar{\mathcal{S}}_d^+$. This is similar to the spd case, see the discussion in Remark 5.7 in Pérez-Abreu and Stelzer (2014) for further details.*

3.

The Hpd Gamma Process

In this section, we utilize the Hpd Gamma distribution from Section 2 to construct an Hpd Gamma process. The construction in the upcoming Section 3.1 will be similar to the Kingman construction (1.9) of the Gamma process from Section 1.2. Support properties and lower bounds for probability mass of the process will be investigated in Section 3.2 and an almost surely convergent series representation will be derived in Section 3.3 whereas in Section 3.4 it is explained and illustrated how random samples from the process can be generated numerically.

3.1. Kingman's Construction Revisited

Denote by $\mathbb{B}(\bar{\mathbb{S}}_d^+)$ the Borel sets in $\bar{\mathbb{S}}_d^+$. Let $\mathcal{X}$ fulfill Assumption $\mathcal{X}1$. Throughout this section, we will make the following assumptions on the process parameters $\alpha(x,\cdot)$ and $\beta(x,\cdot)$ on $\mathcal{X}$:

Assumption GP1. *(a) Let $\alpha\colon \mathcal{X} \times \mathbb{B}(\bar{\mathbb{S}}_d^+) \to [0,\infty)$ such that $\{\alpha(x,\cdot)\}_{x\in\mathcal{X}}$ is a family of finite measures on $\bar{\mathcal{S}}_d^+$ and such that for all $B \in \mathbb{B}(\bar{\mathbb{S}}_d^+)$ the mapping $\mathcal{X} \ni x \mapsto \alpha(x,B)$ is measurable.*

(b) Let $\beta\colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \to (0,\infty)$ be measurable.

We start by defining the mean measure ν of the underlying Poisson process on $\mathcal{X} \times \bar{\mathcal{S}}_d^+ \cong \mathcal{X} \times \bar{\mathbb{S}}_d^+ \times [0,\infty)$ as

$$\nu(dx, d\boldsymbol{U}, dr) = \frac{\exp\left(-\beta(x,\boldsymbol{U})r\right)}{r} dr\alpha(x, d\boldsymbol{U})dx. \tag{3.1}$$

The notation in (3.1) is understood as follows: Let $A \subset \mathcal{X} \times \bar{\mathbb{S}}_d^+ \times [0,\infty)$ be measurable (with respect to the product Borel sigma algebra). Then

$$\nu(A) := \int_{\mathcal{X}} \int_{\bar{\mathbb{S}}_d^+} \int_0^\infty \mathbb{1}_A(x,\boldsymbol{U},r) \frac{\exp\left(-\beta(x,\boldsymbol{U})r\right)}{r} dr\alpha(x, d\boldsymbol{U})dx.$$

In addition to Assumption GP1, let the process parameters $\alpha(x,\cdot)$ and $\beta(x,\cdot)$ be such that ν fulfills the following assumption:

Assumption GP2. *The integral $\int_{\mathcal{X}\times\bar{\mathbb{S}}_d^+\times[0,\infty)} \min(1,r)\nu(dx, d\boldsymbol{U}, dr)$ is finite for ν as in (3.1).*

For $x \in \mathcal{X}$, the measure

$$\nu(x, d\boldsymbol{U}, dr) = \frac{\exp\left(-\beta(x, \boldsymbol{U})r\right)}{r} dr\alpha(x, d\boldsymbol{U})$$

on $\bar{\mathbb{S}}_d^+ \times [0, \infty)$ corresponds to the Lévy measure from (2.10) of the $\mathrm{Ga}_{d\times d}(\alpha(x,\cdot), \beta(x,\cdot))$ distribution (provided that $\alpha(x,\cdot)$ and $\beta(x,\cdot)$ fulfill Assumption α1). The dependence on x models changing parameters of $\mathrm{Ga}_{d\times d}$ along $\mathcal{X}$. This is similar to the one-dimensional Gamma process, see (1.8) in Section 1.2.

Remark. *For our later application of modeling the spectral measure of a multivariate time series, we will use $\mathcal{X} = [0, \pi]$ and the process parameters $\alpha(x,\cdot), \beta(x,\cdot)$ may be chosen to represent the prior knowledge about the spectrum at $0 \leq x \leq \pi$. In this case (and generally, whenever $\mathcal{X}$ is a bounded measurable subset of $\mathbb{R}^k$ for some $k > 0$), Assumption GP2 is fulfilled if α is a family of* uniformly *bounded measures (i.e. there exists $\alpha_0 > 0$ such that $\alpha(x, \bar{\mathbb{S}}_d^+) \leq \alpha_0$ for all $x \in \mathcal{X}$) and β is uniformly bounded away from 0 (i.e. there exists $\beta_0 > 0$ such that $\beta(x, \boldsymbol{U}) \geq \beta_0$ for all $x \in \mathcal{X}$ and all $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$). This follows readily from*

$$\int_{\mathcal{X}\times\bar{\mathbb{S}}_d^+\times[0,\infty)} \min(1, r)\nu(dx, d\boldsymbol{U}, dr) \leq \alpha_0 \int_{\mathcal{X}} dx \int_0^\infty \frac{\min(1, r)\exp(-\beta_0 r)}{r} dr < \infty.$$

Denote by $\Pi \sim \mathrm{PP}(\nu)$ the Poisson process on $\mathcal{X} \times \bar{\mathbb{S}}_d^+ \times [0, \infty)$ having mean measure ν from (3.1). By Assumption GP2, the assumptions of the existence theorem for Poisson processes are fulfilled (see Theorem B.17 in the Appendix) and hence Π is well-defined. Indeed, ν has no mass atoms due to the absolutely continuous radial part dr and ν is σ-finite because the measure with density $[r \mapsto \exp(-\beta r)/r]$ on $[0, \infty)$ is σ-finite for all $\beta \geq 0$. Based on Π, we define

$$\begin{aligned} &\boldsymbol{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu) :\Longleftrightarrow \\ &\boldsymbol{\Phi}(A) = \sum_{(x,\boldsymbol{U},r)\in\Pi} \mathbb{1}_A(x) r\boldsymbol{U}, \quad A \in \mathbb{B}(\mathcal{X}), \quad \Pi \sim \mathrm{PP}(\nu), \end{aligned} \tag{3.2}$$

where $\mathbb{B}(\mathcal{X})$ denote the Borel sets in $\mathcal{X}$. We call $\boldsymbol{\Phi}$ in (3.2) an *Hpd Gamma process* on $\mathcal{X}$ with parameters $\alpha = \alpha(x,\cdot)$ and $\beta = \beta(x,\cdot)$, or *completely random Hpd measure*, as already indicated by the notation $\mathrm{CRM}_{d\times d}(\nu)$.

Remark 3.1. *Let $A_1, \ldots, A_k$ be a disjoint partition of $\mathcal{X}$. For $j = 1, \ldots, k$, consider the measure $\nu_j(dx, d\boldsymbol{U}, dr) := \mathbb{1}_{A_j}(x)\nu(dx, d\boldsymbol{U}, dr)$. It is clear that $\nu = \sum_{j=1}^k \nu_j$. Let $\Pi_1, \ldots, \Pi_k$ be independent Poisson processes, with respective mean measure ν_j. By the Superposition Theorem (see Theorem B.18 in the Appendix), it follows that Π is equal in distribution to $\cup_{j=1}^k \Pi_j$. Since*

$$\boldsymbol{\Phi}(A_j) \overset{d}{=} \sum_{(x,\boldsymbol{U},r)\in\cup_{j=1}^k\Pi_j} \mathbb{1}_{A_j}(x) r\boldsymbol{U} = \sum_{(x,\boldsymbol{U},r)\in\Pi_j} r\boldsymbol{U},$$

it follows from the independence of the Π_j's that $\boldsymbol{\Phi}(A_1), \ldots, \boldsymbol{\Phi}(A_k)$ are independent.

The name *completely random* Hpd measure is justified by the independence property from Remark 3.1. The following result shows that the increments of $\boldsymbol{\Phi}$ are Hpd Gamma distributed.

Theorem 3.2. *Let $\mathcal{X}$ fulfill Assumption $\mathcal{X}$1 and let α, β and ν fulfill Assumptions GP1 and GP2. Let $\mathbf{\Phi}$ be defined as in (3.2) and let $A \subset \mathcal{X}$ be measurable. Then $\mathbf{\Phi}(A) \in \bar{\mathcal{S}}_d^+$ with probability 1. Furthermore, the distribution of $\mathbf{\Phi}(A)$ is given in the Lévy-Khinchine representation of its Laplace transformation*

$$\mathrm{E}\operatorname{etr}(-\mathbf{\Theta}\mathbf{\Phi}(A)) = \exp\left(-\int_{\bar{\mathbb{S}}_d^+}\int_0^\infty (1-\operatorname{etr}(-r\mathbf{\Theta}\boldsymbol{U}))\,\nu_A(d\boldsymbol{U},dr)\right), \quad \mathbf{\Theta} \in \bar{\mathcal{S}}_d^+,$$

with Lévy measure on $\bar{\mathbb{S}}_d^+ \times [0,\infty)$ given by

$$\nu_A(d\boldsymbol{U},dr) = \int_A \alpha(x,d\boldsymbol{U})\frac{\exp(-\beta(x,\boldsymbol{U})r)}{r}drdx. \tag{3.3}$$

Proof. Fix $A \subset \mathcal{X}$ and consider the measurable mapping $\phi\colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \times [0,\infty) \ni (x,\boldsymbol{U},r) \mapsto \phi(x,\boldsymbol{U},r) = \|\mathbb{1}_A(x)r\boldsymbol{U}\|_T = \mathbb{1}_A(x)r \in [0,\infty)$. Let $\Pi \sim \mathrm{PP}(\nu)$ be the Poisson process with mean measure ν from (3.1). Per definition, Π is a random subset of $\mathcal{X} \times \bar{\mathbb{S}}_d^+ \times [0,\infty)$ consisting of at most countably many elements. Put these elements in some order, such that the distributional representation

$$\Pi \overset{d}{=} \{(x_j,\boldsymbol{U}_j,r_j)\colon j \geq 1\}$$

holds. By assumption GP2, ϕ and Π fulfill the assumptions of Campbell's Theorem (see Theorem B.19 in the Appendix). An application thereof yields the almost sure convergence of the infinite series $\sum_{(x,\boldsymbol{U},r)\in\Pi}\phi(x,\boldsymbol{U},r) = \sum_{j=1}^\infty \phi(x_j,\boldsymbol{U}_j,r_j)$. The definition (3.2) of $\mathbf{\Phi}(A)$ can equivalently be written as

$$\mathbf{\Phi}(A) = \sum_{j=1}^\infty \mathbb{1}_A(x_j)r_j\boldsymbol{U}_j. \tag{3.4}$$

Since $\|\mathbf{\Phi}(A)\|_T \leq \sum_{j=1}^\infty \phi(x_j,\boldsymbol{U}_j,r_j) < \infty$ with probability 1, it follows that any partial sum sequence of (3.4) is a Cauchy sequence in the Banach space $\mathcal{S}_d$ of Hermitian matrices endowed with the trace norm $\|\cdot\|_T$, hence $\mathbf{\Phi}(A) \in \mathcal{S}_d$ with probability 1. Since all summands are from the closed cone $\bar{\mathcal{S}}_d^+$, it also holds $\mathbf{\Phi}(A) \in \bar{\mathcal{S}}_d^+$ with probability 1.

Now let $\mathbf{\Theta} \in \bar{\mathcal{S}}_d^+$ and consider the measurable mapping

$$\phi_{\mathbf{\Theta}}\colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \times [0,\infty) \ni (x,\boldsymbol{U},r) \mapsto \phi_{\mathbf{\Theta}}(x,\boldsymbol{U},r) = \operatorname{tr}(\mathbf{\Theta}\mathbb{1}_A(x)r\boldsymbol{U}) \in [0,\infty).$$

By (B.9) and Lemma B.4 (b) and (d) in the Appendix, it holds $|\phi_{\mathbf{\Theta}}(x,\boldsymbol{U},r)| \leq \|\mathbf{\Theta}\|_2\phi(x,\boldsymbol{U},r)$. Thus, by similar argumentation as above, we can apply Campbell's theorem to the random variable

$$\operatorname{tr}(\mathbf{\Theta}\mathbf{\Phi}(A)) = \sum_{(x,\boldsymbol{U},r)\in\Pi}\phi_{\mathbf{\Theta}}(x,\boldsymbol{U},r)$$

to obtain for any $t \geq 0$

$$\begin{aligned}\mathrm{E}\operatorname{etr}(-t\mathbf{\Theta}\mathbf{\Phi}(A)) &= \mathrm{E}\exp\left(-t\operatorname{tr}(\mathbf{\Theta}\mathbf{\Phi}(A))\right)\\ &= \exp\left(-\int_{\mathcal{X}\times\bar{\mathbb{S}}_d^+\times[0,\infty)}(1-\exp(-t\phi_{\mathbf{\Theta}}(x,\boldsymbol{U},r)))\nu(dx,d\boldsymbol{U},dr)\right).\end{aligned}$$

In particular, for $t = 1$, the right hand side is equal to

$$\exp\left(-\int_A \int_{\bar{\mathbb{S}}_d^+} \int_0^\infty (1 - \operatorname{etr}(-r\boldsymbol{\Theta}\boldsymbol{U})) \frac{\exp(-\beta(x,\boldsymbol{U})r)}{r} dr\alpha(x, d\boldsymbol{U})dx\right)$$
$$= \exp\left(-\int_{\bar{\mathbb{S}}_d^+} \int_0^\infty (1 - \operatorname{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}))\, \nu_A(d\boldsymbol{U}, dr)\right),$$

by Tonelli's theorem (which is applicable by assumption $\mathcal{X}1$). □

Remark 3.3. *Let $\boldsymbol{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$ with ν from (3.1). If $\beta(x,\boldsymbol{U}) = \beta(\boldsymbol{U})$ does not depend on $x \in \mathcal{X}$, then $\boldsymbol{\Phi}$ is called* homogenous. *In this case, Theorem 3.2 shows that $\boldsymbol{\Phi}(A)$ is a random matrix with Lévy measure*

$$\nu_A(d\boldsymbol{U}, dr) = \alpha_A(d\boldsymbol{U})\frac{\exp(-\beta(\boldsymbol{U})r)}{r}dr,$$

where the measure α_A on $\bar{\mathbb{S}}_d^+$ is defined as the mixture of the $\alpha(x,\cdot)$'s with respect to the Borel measure dx on A, i.e. $\alpha_A(E) = \int_A \alpha(x,E)dx$ for any measurable $E \subset \bar{\mathbb{S}}_d^+$. With a view on (2.7), this shows that $\boldsymbol{\Phi}(A) \sim \mathrm{Ga}_{d\times d}(\alpha_A, \beta)$.

In the general (non-homogeneous) case (3.1), the distribution of $\boldsymbol{\Phi}(A)$ is not Hpd Gamma. But according to Theorem 3.2, its Lévy measure is a mixture of Hpd Gamma Lévy measures with parameters $\alpha(x,\cdot)$ and $\beta(x,\cdot)$ for $x \in A$ with respect to the Borel measure dx on A. In this sense, the (infinitesimal) increments $\boldsymbol{\Phi}(dx)$ among $x \in \mathcal{X}$ are independent and locally $\mathrm{Ga}_{d\times d}$ distributed, which we will denote as

$$\boldsymbol{\Phi}(dx) \overset{\text{ind.}}{\sim} \mathrm{Ga}_{d\times d}(\alpha(x,\cdot), \beta(x,\cdot)). \tag{3.5}$$

3.2. Distributional Properties

In this section, we will investigate the support and probability mass of the Hpd Gamma Process.

3.2.1. Support

We will start by giving sufficient conditions for full support with the upcoming Theorem 3.4. To formulate these results, we will need to strengthen the assumptions on the process parameters:

Assumption GP3. *(a) There exists $N \subset \mathcal{X}$ with $\int_N dx = 0$ such that $\operatorname{supp}(\alpha(x,\cdot)) = \bar{\mathbb{S}}_d^+$ holds for all $x \in \mathcal{X} \setminus N$.*

(b) The function β is locally bounded, i.e. for every $x_0 \in \mathcal{X}$ there exists a neighborhood $\mathcal{U}_0$ of x_0 such that $\sup_{x\in\mathcal{U}_0, \boldsymbol{U}\in\bar{\mathbb{S}}_d^+} \beta(x,\boldsymbol{U}) < \infty$.

Theorem 3.4. *Let $\mathcal{X}$ fulfill Assumption $\mathcal{X}1$ and let α, β and ν fulfill Assumptions GP1 and GP2 as well as Assumption GP3. Let $\boldsymbol{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$. Let $A \subset \mathcal{X}$ be measurable with $\int_A dx > 0$. Then $\operatorname{supp}(\boldsymbol{\Phi}(A)) = \bar{\mathcal{S}}_d^+$.*

Proof. First recall that the Lévy measure ν_A of $\boldsymbol{\Phi}(A)$ is given as in (3.3). The proof is a slight modification of the one one-dimensional case (see Theorem 24.10 (iii) in Sato (1999)). The idea is to split up $\boldsymbol{\Phi}(A)$ into a *large jumps* part $\boldsymbol{Y}$ and a *small jumps* part $\boldsymbol{Z}$. To elaborate, for $\varepsilon > 0$, define $\nu_{\boldsymbol{Y}}(d\boldsymbol{U}, dr) := \mathbb{1}_{(\varepsilon/2,\infty)}(r)\nu_A(d\boldsymbol{U}, dr)$, and $\nu_{\boldsymbol{Z}}(d\boldsymbol{U}, dr) := \mathbb{1}_{[0,\varepsilon/2]}(r)\nu_A(d\boldsymbol{U}, dr)$ and let $\boldsymbol{Y}, \boldsymbol{Z}$ be independent defined in terms of the Laplace transforms by

$$\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Y}) := \exp\left(-\int_{\mathbb{S}_d^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}))\nu_{\boldsymbol{Y}}(d\boldsymbol{U}, dr)\right), \tag{3.6}$$

$$\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Z}) := \exp\left(-\int_{\mathbb{S}_d^+}\int_0^\infty (1-\mathrm{etr}(-r\boldsymbol{\Theta}\boldsymbol{U}))\nu_{\boldsymbol{Z}}(d\boldsymbol{U}, dr)\right), \tag{3.7}$$

for $\boldsymbol{\Theta} \in \bar{\mathcal{S}}_d^+$. Then it holds $\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{\Phi}(A)) = \mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Y})\mathrm{E}\,\mathrm{etr}(-\boldsymbol{\Theta}\boldsymbol{Z})$ for all $\boldsymbol{\Theta}$, and hence (since $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are independent) $\boldsymbol{\Phi}(A) \overset{d}{=} \boldsymbol{Y} + \boldsymbol{Z}$. Let $\boldsymbol{X} \in \mathcal{S}_d^+$ with $\|\boldsymbol{X}\|_T > \varepsilon$, i.e. $\boldsymbol{X} = r_0\boldsymbol{U}_0$ with $r_0 > \varepsilon$ and $\boldsymbol{U}_0 \in \mathbb{S}_d^+$. We will show $\boldsymbol{X} \in \mathrm{supp}(\boldsymbol{\Phi}(A))$ and start by showing $\boldsymbol{X} \in \mathrm{supp}(\nu_{\boldsymbol{Y}})$. To do so, let $\mathcal{U}_{\mathcal{S}_d^+}$ be an open neighborhood of $\boldsymbol{X}$, where the openness refers to the topology in $\mathcal{S}_d^+$ (as indicated by the subscript). Since $\mathcal{S}_d^+ \cong (0,\infty) \times \mathbb{S}_d^+$, there exist r_1, r_2 with $\varepsilon/2 < r_1 < r_0 < r_2 < \infty$ and an open neighborhood $\mathcal{U}_{\mathbb{S}_d^+}$ (with respect to the topology in $\mathbb{S}_d^+$) such that $[r_1, r_2] \times \mathcal{U}_{\mathbb{S}_d^+} \subseteq \mathcal{U}_{\mathcal{S}_d^+}$. This yields

$$\nu_{\boldsymbol{Y}}(\mathcal{U}_{\mathcal{S}_d^+}) \geq \int_{r_1}^{r_2}\int_{\mathcal{U}_{\mathbb{S}_d^+}} \nu_{\boldsymbol{Y}}(d\boldsymbol{U}, dr) = \int_{r_1}^{r_2}\int_{\mathcal{U}_{\mathbb{S}_d^+}} \nu_A(d\boldsymbol{U}, dr).$$

Using that the function $r \mapsto \exp(-\beta r)/r$ is monotonically decreasing on $(0,\infty)$ for every $\beta \geq 0$, we get

$$\begin{aligned}\int_{r_1}^{r_2}\int_{\mathcal{U}_{\mathbb{S}_d^+}} \nu_A(d\boldsymbol{U}, dr) &= \int_{r_1}^{r_2}\int_{\mathcal{U}_{\mathbb{S}_d^+}}\int_A \frac{\exp(-\beta(x,\boldsymbol{U})r)}{r}\alpha(x, d\boldsymbol{U})dxdr \\ &\geq \frac{r_2 - r_1}{r_2}\int_{\mathcal{U}_{\mathbb{S}_d^+}}\int_A \exp(-\beta(x,\boldsymbol{U})r_2)\alpha(x, d\boldsymbol{U})dx.\end{aligned}$$

Now let $\tilde{A} \subset A$ such that $\int_{\tilde{A}} dx > 0$ and $\sup_{x\in\tilde{A},\boldsymbol{U}\in\bar{\mathbb{S}}_d^+} \beta(x,\boldsymbol{U}) =: \beta_1 < \infty$. Note that $\tilde{A}$ and β_1 exist by Assumption GP3. Then

$$\begin{aligned}\int_{\mathcal{U}_{\mathbb{S}_d^+}}\int_A \exp(-\beta(x,\boldsymbol{U})r_2)\alpha(x, d\boldsymbol{U})dx &\geq \int_{\mathcal{U}_{\mathbb{S}_d^+}}\int_{\tilde{A}} \exp(-\beta(x,\boldsymbol{U})r_2)\alpha(x, d\boldsymbol{U})dx \\ &\geq \exp(-\beta_1 r_2)\int_{\mathcal{U}_{\mathbb{S}_d^+}}\int_{\tilde{A}} \alpha(x, d\boldsymbol{U})dx.\end{aligned}$$

By Assumption GP3, it holds $\alpha(x, \mathcal{U}_{\mathbb{S}_d^+}) > 0$ for every $x \in \tilde{A} \setminus N$ (with N being a Lebesgue null set), yielding

$$\int_{\mathcal{U}_{\mathbb{S}_d^+}}\int_{\tilde{A}} \alpha(x, d\boldsymbol{U})dx = \int_{\tilde{A}} \alpha(x, \mathcal{U}_{\mathbb{S}_d^+})dx > 0,$$

which concludes $\nu_{\boldsymbol{Y}}(\mathcal{U}_{\mathcal{S}_d^+}) > 0$ and hence (since $\mathcal{U}_{\mathcal{S}_d^+}$ was an arbitrary neighborhood of $\boldsymbol{X}$), $\boldsymbol{X} \in \mathrm{supp}(\nu_{\boldsymbol{Y}})$. Since $\nu_{\boldsymbol{Y}}([0,\infty), \bar{\mathbb{S}}_d^+) < \infty$ by Assumption GP2, we find from Lemma 2.3

that $\boldsymbol{Y}$ is compound Poisson and an application of Lemma 2.5 yields $\operatorname{supp}(\nu_{\boldsymbol{Y}}) \subset \operatorname{supp}(\boldsymbol{Y})$, which concludes $\boldsymbol{X} \in \operatorname{supp}(\boldsymbol{Y})$.

By Lemma B.27 in the Appendix, it suffices to show $\mathbf{0} \in \operatorname{supp}(\boldsymbol{Z})$, to conclude $\boldsymbol{X} \in \operatorname{supp}(\boldsymbol{\Phi}(A))$. Since $\tilde{\boldsymbol{Z}} = \mathbf{0}$ is equivalent to $\operatorname{tr} \tilde{\boldsymbol{Z}} = \|\tilde{\boldsymbol{Z}}\|_T = 0$ for $\tilde{\boldsymbol{Z}} \in \bar{\mathcal{S}}_d^+$ (see (B.9) in the Appendix), we will show $0 \in \operatorname{supp}(\operatorname{tr}(\boldsymbol{Z}))$. Recalling $\operatorname{tr} \boldsymbol{U} = 1$ for $\boldsymbol{U} \in \mathbb{S}_d^+$, the (one-dimensional) Laplace transform of $\operatorname{tr}(\boldsymbol{Z})$ is given as, for $t \geq 0$,

$$\begin{aligned}\operatorname{E} \exp(-t \operatorname{tr}(\boldsymbol{Z})) = \operatorname{E} \operatorname{etr}(-(t\boldsymbol{I}_d)\boldsymbol{Z}) &= \exp\left(-\int_{\bar{\mathbb{S}}_d^+} \int_0^\infty (1 - \operatorname{etr}(-rt\boldsymbol{U}))\nu_{\boldsymbol{Z}}(d\boldsymbol{U}, dr)\right) \\ &= \exp\left(-\int_0^\infty (1 - \exp(-rt))\tilde{\nu}_{\boldsymbol{Z}}(dr)\right)\end{aligned}$$

with the one-dimensional Lévy measure

$$\tilde{\nu}_{\boldsymbol{Z}}(dr) = \int_{\bar{\mathbb{S}}_d^+} \nu_{\boldsymbol{Z}}(d\boldsymbol{U}, dr) = \mathbb{1}_{[0,\varepsilon/2]}(r) \int_A \int_{\mathbb{S}_d^+} \frac{\exp(-\beta(x, \boldsymbol{U})r)}{r} \alpha(x, d\boldsymbol{U}) dx dr. \tag{3.8}$$

For any $0 < \delta < \varepsilon/2$ it holds

$$\tilde{\nu}_{\boldsymbol{Z}}([0, \infty)) \geq \tilde{\nu}_{\boldsymbol{Z}}([0, \delta]) \geq \exp(-\beta_1 \delta) \int_A \alpha(x, \mathbb{S}_d^+) dx \int_0^\delta \frac{1}{r} dr = \infty.$$

Furthermore,

$$\int_0^1 r\tilde{\nu}_{\boldsymbol{Z}}(dr) = \int_0^{\varepsilon/2} \int_A \int_{\mathbb{S}_d^+} r\nu(dx, d\boldsymbol{U}, dr) \leq \int_0^1 \int_{\mathcal{X}} \int_{\mathbb{S}_d^+} r\nu(dx, d\boldsymbol{U}, dr) < \infty$$

by Assumption GP2. Thus the assumptions of Lemma B.28 in the Appendix are fulfilled and this yields $0 = \min \operatorname{supp}(\operatorname{tr}(\boldsymbol{Z}))$. Since the support is closed by definition, this concludes the proof. □

The full support property for the $\operatorname{Ga}_{d\times d}(\alpha, \beta)$ distribution can be obtained as a special case of the previous Lemma. The result has already been stated in Theorem 2.6, and is restated and proven in the following Corollary.

Corollary 3.5. *Let α, β fulfill Assumption $\alpha 1$ and let $\boldsymbol{X} \sim \operatorname{Ga}_{d\times d}(\alpha, \beta)$. Assume that α is of full support, i.e. $\operatorname{supp}(\alpha) = \bar{\mathbb{S}}_d^+$. Then $\boldsymbol{X}$ is of full support, i.e. $\operatorname{supp}(\boldsymbol{X}) = \bar{\mathcal{S}}_d^+$.*

Proof. Let $\mathcal{X} = [0, 1]$ and consider $\tilde{\alpha} \colon \mathcal{X} \times \mathbb{B}(\bar{\mathbb{S}}_d^+)$ with $\tilde{\alpha}(x, d\boldsymbol{U}) = \alpha(d\boldsymbol{U})$ and $\tilde{\beta} \colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \to (0, \infty)$ with $\tilde{\beta}(x, \boldsymbol{U}) = \beta(\boldsymbol{U})$ for $x \in \mathcal{X}$ and $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$. Denote by $\tilde{\boldsymbol{\Phi}}$ the Hpd Gamma Process with parameters $\tilde{\alpha}$ and $\tilde{\beta}$. Since $\mathcal{X}, \tilde{\alpha}$ and $\tilde{\beta}$ fulfill the assumptions of Theorem 3.4, we conclude for $A = \mathcal{X} = [0, 1]$ that $\operatorname{supp}(\tilde{\boldsymbol{\Phi}}(A)) = \bar{\mathcal{S}}_d^+$. On the other hand, from Remark 3.3 it follows that $\tilde{\boldsymbol{\Phi}}(A) \sim \operatorname{Ga}_{d\times d}(\alpha, \beta)$, which concludes the proof. □

3.2.2. Probability Mass Bounds

In this Section, we will derive lower bounds for the probability mass of the Hpd Gamma process. The main results are Theorem 3.7 and Corollary 3.8. We will need the following stronger assumptions on the underlying space and the process parameters:

Assumption GP4. *(a) It holds* $\alpha(x, d\boldsymbol{U}) = g(x, \boldsymbol{U})d\boldsymbol{U}$ *for a measurable function* $g\colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \to (0, \infty)$*. Furthermore, there exist positive constants* g_0, g_1 *such that* $g_0 \leq g(x, \boldsymbol{U}) \leq g_1$ *for all* $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$ *and all* $x \in \mathcal{X} \setminus N$*, where* $N \subset \mathcal{X}$ *is a null set, i.e.* $\int_N dx = 0$*.*

(b) There exist constants $0 < \beta_0 \leq \beta_1 < \infty$ *and a null set* $N \subset \mathcal{X}$ *such that* $\beta_0 \leq \beta(x, \boldsymbol{U}) \leq \beta_1$ *holds for all* $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$ *and all* $x \in \mathcal{X} \setminus N$*.*

Assumption $\mathcal{X}2$. *It holds* $\int_{\mathcal{X}} dx < \infty$*.*

First we observe that Assumption GP4 and Assumption $\mathcal{X}2$ imply Assumption GP2. Furthermore, it may be noted that Assumption $\mathcal{X}2$ is implied by Assumptions GP2 and GP4, since in this case it holds

$$\infty > \int_{\mathcal{X}} \int_{\bar{\mathbb{S}}_d^+} \int_1^\infty \frac{\exp(-\beta(x, \boldsymbol{U})r}{r} dr \alpha(x, d\boldsymbol{U}) dx \geq g_0 \int_{\mathcal{X}} dx \int_{\mathbb{S}_d^+} d\boldsymbol{U} \int_1^\infty \frac{\exp(-\beta_1 r)}{r} dr > 0.$$

The assumption $g(x, \boldsymbol{U}) \leq g_1$ is actually not needed for the results in this Section but included for the sake of later reference to avoid notational overhead. In fact, this particular assumption will be needed for the derivation of posterior contraction rates (see the upcoming Section 7.2 and Section 9.1). We start our considerations with the observation that the total weight of α is finite under the assumptions of this Section.

Lemma 3.6. *Let* $\mathcal{X}$ *fulfill Assumptions* $\mathcal{X}1$*-*$\mathcal{X}2$ *and let* α, β *and* ν *fulfill Assumption GP4. Then* $C_\alpha := \int_{\mathcal{X}} \alpha(x, \bar{\mathbb{S}}_d^+) dx$ *is finite.*

Proof. By Assumption GP2 and Assumption GP4 it holds

$$\infty > \int_{\mathcal{X}} \int_{\bar{\mathbb{S}}_d^+} \int_1^\infty \alpha(x, d\boldsymbol{U}) \frac{\exp(-\beta(x, \boldsymbol{U})r)}{r} dr dx \geq \int_{\mathcal{X}} \alpha(x, \bar{\mathbb{S}}_d^+) dx \int_1^\infty \frac{\exp(-\beta_1 r)}{r} dr$$

and since $\int_1^\infty \frac{\exp(-\beta_1 r)}{r} dr > 0$ it follows that $C_\alpha = \int_{\mathcal{X}} \alpha(x, \bar{\mathbb{S}}_d^+) dx$ is finite. □

The following result gives a lower bound for the probability mass that the distribution of $\boldsymbol{\Phi}(A)$ puts in balls of radius ε.

Theorem 3.7. *Let* $\mathcal{X}$ *fulfill Assumptions* $\mathcal{X}1$ *and* $\mathcal{X}2$ *and let* α, β *and* ν *fulfill Assumption GP4. Let* $A \subset \mathcal{X}$ *with* $\mathcal{L}(A) := \int_A dx > 0$*. Let* $\boldsymbol{X}_0 \in \mathcal{S}_d^+$ *with* $\|\boldsymbol{X}_0\|_T \leq \tau$ *for some* $\tau > 1$*. Then, with* $C_\alpha(A) := \int_A \alpha(x, \bar{\mathbb{S}}_d^+) dx$ *it holds*

$$P\big(\|\boldsymbol{\Phi}(A) - \boldsymbol{X}_0\|_T < \varepsilon\big) \geq C \kappa_\alpha(A) \exp\big(\big(d^2 + C_\alpha(A) + 1\big) \log \varepsilon\big),$$

for all $\varepsilon > 0$ *small enough, where* C *is a positive constant only depending on* β_0, β_1 *(as in Assumption GP4),* d *and* τ*. Furthermore,* $\kappa_\alpha(A)$ *is defined as*

$$\kappa_\alpha(A) = \exp(-cC_\alpha(A)) \left(1 - \exp\left(-\frac{1}{2} W\left(\frac{2}{C_\alpha(A)}\right)\right)\right) g_0 \mathcal{L}(A),$$

with g_0 *from Assumption GP4 and* c *being a positive constant only depending on* β_0, β_1 *and* τ*, and* W *denoting the Lambert W function (see* (B.27) *in Section B.5 in the Appendix).*

From Theorem 3.7, we obtain the following immediate corollary.

Corollary 3.8. *Under the assumptions of Theorem 3.7, it holds*

$$P\big(\|\boldsymbol{\Phi}(A) - \boldsymbol{X}_0\| < \varepsilon\big) \geq C\kappa_\alpha(A) \exp\big(\big(d^2 + C_\alpha(A) + 1\big)\log\varepsilon\big),$$

for all $\varepsilon > 0$ small enough, where C is a positive constant only depending on β_0, β_1, (as in Assumption GP4), C_α, d and τ. Furthermore, $\kappa_\alpha(A)$ is defined in Theorem 3.7.

Proof. The result follows from Theorem 3.7, because the matrix norms $\|\cdot\|_T$ and $\|\cdot\|$ are equivalent. □

Before proceeding to the proof of Theorem 3.7, we will consider the concrete example of the Hpd $A\Gamma(\eta, \omega\boldsymbol{\Sigma})$ process and establish conditions under which the result is applicable.

Lemma 3.9. *Let $\mathcal{X}$ fulfill Assumptions $\mathcal{X}1$ and $\mathcal{X}2$ and consider the Hpd $\mathrm{A}\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ distribution from Section 2.4. Let $\eta \equiv \eta_0$ be fixed, and let $\omega\colon \mathcal{X} \to (0, \infty)$ measurable and $\boldsymbol{\Sigma}\colon \mathcal{X} \to \mathcal{S}_d^+$ measurable. Let $\boldsymbol{\Phi}$ be an $\mathrm{A}\Gamma(\eta_0, \omega, \boldsymbol{\Sigma})$ process on $\mathcal{X}$, i.e. an Hpd Gamma process with parameters $\alpha(x, d\boldsymbol{U}) = \omega(x)\alpha_{\eta_0, \boldsymbol{\Sigma}(x)}(d\boldsymbol{U})$ and $\beta(x, \boldsymbol{U}) = \beta_{\boldsymbol{\Sigma}(x)}(\boldsymbol{U})$ (see Lemma 2.7). Let the process parameters fulfill the following assumptions:*

(a) $\eta_0 = d$.

(b) It holds

$$\inf_{x\in\mathcal{X}} \lambda_{\min}(\boldsymbol{\Sigma}(x)) = \tau_0 > 0, \quad \sup_{x\in\mathcal{X}} \lambda_{\max}(\boldsymbol{\Sigma}(x)) = \tau_1 < \infty,$$

with $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ denoting the smallest and the largest eigenvalue of $\boldsymbol{A} \in \mathcal{S}_d^+$.

(c) It holds

$$\inf_{x\in\mathcal{X}} \omega(x) = \omega_0 > 0, \quad \sup_{x\in\mathcal{X}} \omega(x) = \omega_1 < \infty.$$

Then $\boldsymbol{\Phi}$ fulfills Assumption GP4. In particular, Theorem 3.7 is applicable to $\boldsymbol{\Phi}$.

Proof. From Lemma 2.7, we know that

$$\alpha(x, d\boldsymbol{U}) = g(x, \boldsymbol{U})d\boldsymbol{U}, \quad g(x, \boldsymbol{U}) = C\omega(x)|\boldsymbol{\Sigma}(x)|^{-d} \operatorname{tr}\left(\boldsymbol{\Sigma}(x)^{-1}\boldsymbol{U}\right)^{-d^2}$$

with a positive constant C, as well as

$$\beta(x, \boldsymbol{U}) = \operatorname{tr}\left(\boldsymbol{\Sigma}(x)^{-1}\boldsymbol{U}\right).$$

An application of Lemma B.4 (d) in the Appendix yields

$$0 < \tau_1^{-1} \leq \beta(x, \boldsymbol{U}) \leq \tau_0^{-1} < \infty, \quad x \in \mathcal{X}, \boldsymbol{U} \in \mathbb{S}_d^+,$$

and similarly, also using Lemma B.5 in the Appendix,

$$0 < C\omega_0\tau_0^{d^2}\tau_1^{-d^2} \leq g(x, \boldsymbol{U}) \leq C\omega_1\tau_0^{-d^2}\tau_1^{d^2} < \infty, \quad x \in \mathcal{X}, \boldsymbol{U} \in \mathbb{S}_d^+.$$

concluding the proof. □

To prove Theorem 3.7, we will use the same idea of splitting up $\boldsymbol{\Phi}(A) \stackrel{d}{=} \boldsymbol{Y} + \boldsymbol{Z}$ into a *large jumps* compound Poisson component $\boldsymbol{Y}$ and a *small jumps* component $\boldsymbol{Z}$, as in the proof of Theorem 3.4. To do so, we need the following two auxiliary results. The first Lemma 3.10 shows that the Lévy measure of $\boldsymbol{\Phi}(A)$ puts enough mass in a neighborhood of $\boldsymbol{X}_0$. It will be used later to show that the distribution of $\boldsymbol{Y}$ also puts enough mass in a neighborhood of $\boldsymbol{X}_0$. The second Lemma 3.11 will be used later to show that the distribution of $\boldsymbol{Z}$ puts enough mass in a neighborhood of $\mathbf{0}$.

Lemma 3.10. *Let the assumptions of Theorem 3.7 be fulfilled. Then with*

$$B_\varepsilon(\boldsymbol{X}_0) = \{\boldsymbol{X} \in \mathcal{S}_d^+ \colon \|\boldsymbol{X} - \boldsymbol{X}_0\|_T < \varepsilon\},$$

the Lévy measure ν_A from (3.3) of $\boldsymbol{\Phi}(A)$ fulfills

$$\nu_A\big(B_\varepsilon(\boldsymbol{X}_0)\big) \geq C g_0 \mathcal{L}(A) \exp\big((d^2+1)\log\varepsilon\big),$$

for all $\varepsilon > 0$ small enough, where C is a positive constant only depending on β_1, d and τ.

Proof. Write $\boldsymbol{X}_0 = r_0 \boldsymbol{U}_0$ with $\boldsymbol{U}_0 \in \mathbb{S}_d^+$ and $0 < r_0 \leq \tau$. For any $\boldsymbol{X} = r\boldsymbol{U} \in \mathcal{S}_d^+$ we have

$$\|\boldsymbol{X} - \boldsymbol{X}_0\|_T \leq r_0 \|\boldsymbol{U} - \boldsymbol{U}_0\|_T + |r - r_0| \leq \tau \|\boldsymbol{U} - \boldsymbol{U}_0\|_T + |r - r_0|.$$

Hence, recalling $\mathcal{S}_d^+ \cong (0,\infty) \times \mathbb{S}_d^+$, we find that $B_\varepsilon(\boldsymbol{X}_0) \supset B_{\varepsilon/2}(r_0) \times \tilde{B}(\boldsymbol{U}_0)$, with $\tilde{B}(\boldsymbol{U}_0) = B_{\varepsilon/(2\tau)}(\boldsymbol{U}_0) \cap \mathbb{S}_d^+$, where $B_{\varepsilon/(2\tau)}(\boldsymbol{U}_0)$ denotes the ball in $\mathcal{S}_d$ around $\boldsymbol{U}_0$ of radius $\varepsilon/(2\tau)$ (with respect to $\|\cdot\|_T$). Recall that $\beta(x, \boldsymbol{U}) \leq \beta_1$ for all x outside a null set and all $\boldsymbol{U}$ by Assumption GP4. Thus, using that $r \mapsto \frac{\exp(-\beta_1 r)}{r}$ is monotonically decreasing, and that $\varepsilon, r_0 \leq \tau$, we compute

$$\begin{aligned}
\nu_A(B_\varepsilon(\boldsymbol{X}_0)) &= \int_A \int_{\bar{\mathbb{S}}_d^+} \int_0^\infty \mathbb{1}_{B_\varepsilon(\boldsymbol{X}_0)}(r\boldsymbol{U}) \frac{\exp(-\beta(x,\boldsymbol{U})r)}{r} dr \alpha(x, d\boldsymbol{U}) dx \\
&\geq \int_A \int_{\tilde{B}(\boldsymbol{U}_0)} \int_{r_0-\varepsilon/2}^{r_0+\varepsilon/2} \frac{\exp(-\beta(x,\boldsymbol{U})r)}{r} dr \alpha(x, d\boldsymbol{U}) dx \\
&\geq \alpha_A(\tilde{B}(\boldsymbol{U}_0)) \int_{r_0-\varepsilon/2}^{r_0+\varepsilon/2} \frac{\exp(-\beta_1 r)}{r} dr \\
&\geq \alpha_A(\tilde{B}(\boldsymbol{U}_0)) \frac{2\varepsilon \exp(-3/2\beta_1\tau)}{3\tau} = C\alpha_A(\tilde{B}(\boldsymbol{U}_0))\varepsilon,
\end{aligned}$$

with the measure α_A on $\bar{\mathbb{S}}_d^+$ being defined as $\alpha_A(d\boldsymbol{U}) =: \int_A \alpha(x, d\boldsymbol{U}) dx$. For this, we compute

$$\alpha_A(\tilde{B}(\boldsymbol{U}_0)) = \int_A \int_{\tilde{B}(\boldsymbol{U}_0)} \alpha(x, d\boldsymbol{U}) dx \geq g_0 \mathcal{L}(A) \int_{\tilde{B}(\boldsymbol{U}_0)} d\boldsymbol{U}$$

by Assumption GP4. Now we use $\Gamma(a) = \int_0^\infty x^{a-1}\exp(-x)dx$ for $a > 0$ and the transformation $\mathcal{S}_d^+ \ni \boldsymbol{Z} = r\boldsymbol{U}$ with $r = \operatorname{tr}\boldsymbol{Z} > 0$ and $\boldsymbol{U} \in \mathbb{S}_d^+$, $d\boldsymbol{Z} = r^{d^2-1} dr d\boldsymbol{U}$ from Lemma B.8 in the appendix to obtain

$$\begin{aligned}
\int_{\tilde{B}(\boldsymbol{U}_0)} d\boldsymbol{U} &= \frac{1}{\Gamma(d^2)} \int_{\tilde{B}(\boldsymbol{U}_0)} \int_0^\infty r^{d^2-1} \exp(-r) dr d\boldsymbol{U} \\
&= \frac{1}{\Gamma(d^2)} \int_{\mathbb{S}_d^+} \int_0^\infty \mathbb{1}_{\tilde{B}(\boldsymbol{U}_0)}(\boldsymbol{U}) \operatorname{etr}(-r\boldsymbol{U}) r^{d^2-1} dr d\boldsymbol{U} \\
&= \frac{1}{\Gamma(d^2)} \int_{\mathcal{S}_d^+} \mathbb{1}_{\tilde{B}(\boldsymbol{U}_0)}(\boldsymbol{Z}/\operatorname{tr}(\boldsymbol{Z})) \operatorname{etr}(-\boldsymbol{Z}) d\boldsymbol{Z}
\end{aligned}$$

We will now show that $B_{\varepsilon/(4\tau)}(\boldsymbol{U}_0) \subset \{\boldsymbol{Z} \in \mathcal{S}_d^+ \colon \boldsymbol{Z}/\operatorname{tr}(\boldsymbol{Z}) \in \tilde{B}(\boldsymbol{U}_0)\}$. To do so, let $\boldsymbol{Z} \in B_{\varepsilon/(4\tau)}(\boldsymbol{U}_0)$ and compute

$$\left\| \frac{\boldsymbol{Z}}{\operatorname{tr} \boldsymbol{Z}} - \boldsymbol{U}_0 \right\|_T \leq \|\boldsymbol{Z} - \boldsymbol{U}_0\|_T + \left\| \boldsymbol{Z} - \frac{\boldsymbol{Z}}{\operatorname{tr} \boldsymbol{Z}} \right\|_T \leq \frac{\varepsilon}{4\tau} + \left| 1 - \frac{1}{\operatorname{tr} \boldsymbol{Z}} \right| \|\boldsymbol{Z}\|_T. \tag{3.9}$$

Noting $\|\boldsymbol{Z}\|_T = \operatorname{tr} \boldsymbol{Z}$ and $|1 - \frac{1}{\operatorname{tr} \boldsymbol{Z}}| = \frac{1}{\operatorname{tr} \boldsymbol{Z}} |1 - \operatorname{tr} \boldsymbol{Z}|$ yields

$$\left| 1 - \frac{1}{\operatorname{tr} \boldsymbol{Z}} \right| \|\boldsymbol{Z}\|_T = |1 - \operatorname{tr} \boldsymbol{Z}| = |\operatorname{tr}(\boldsymbol{U}_0) - \operatorname{tr}(\boldsymbol{Z})| = |\operatorname{tr}(\boldsymbol{U}_0 - \boldsymbol{Z})|.$$

For $\boldsymbol{A} \in \mathcal{S}_d$, denote by $\lambda_1(\boldsymbol{A}) \leq \ldots \leq \lambda_d(\boldsymbol{A})$ the eigenvalues of $\boldsymbol{A}$ in nondecreasing order. With this notation, we recall the characterization (B.7) of the trace norm and continue to compute

$$|\operatorname{tr}(\boldsymbol{U}_0 - \boldsymbol{Z})| = \left| \sum_{i=1}^d \lambda_i(\boldsymbol{U}_0 - \boldsymbol{Z}) \right| \leq \sum_{i=1}^d |\lambda_i(\boldsymbol{U}_0 - \boldsymbol{Z})| = \|\boldsymbol{U}_0 - \boldsymbol{Z}\|_T \leq \frac{\varepsilon}{4\tau},$$

which concludes $\|1 - \frac{1}{\operatorname{tr} \boldsymbol{Z}}\|_T \|\boldsymbol{Z}\|_T \leq \frac{\varepsilon}{4\tau}$ and thus $\boldsymbol{Z}/\operatorname{tr}(\boldsymbol{Z}) \in \tilde{B}(\boldsymbol{U}_0)$ by (3.9). For ε small enough, it holds $\operatorname{tr} \boldsymbol{Z} \leq \tau$. We use this result to obtain

$$\begin{aligned} \int_{\tilde{B}(\boldsymbol{U}_0)} d\boldsymbol{U} &= \frac{1}{\Gamma(d^2)} \int_{\mathcal{S}_d^+} \mathbb{1}_{\tilde{B}(\boldsymbol{U}_0)}(\boldsymbol{Z}/\operatorname{tr}(\boldsymbol{Z})) \operatorname{etr}(-\boldsymbol{Z}) d\boldsymbol{Z} \\ &\geq \frac{1}{\Gamma(d^2)} \int_{B_{\varepsilon/(4\tau)}(\boldsymbol{U}_0)} \operatorname{etr}(-\boldsymbol{Z}) d\boldsymbol{Z} \geq \frac{\exp(-\tau)}{\Gamma(d^2)} \int_{B_{\varepsilon/(4\tau)}(\boldsymbol{U}_0)} d\boldsymbol{Z}. \end{aligned}$$

By the equivalence of matrix norms, there exists a positive constant C (depending only on d and τ), such that with $|a|_\infty := \max\{|\Re a|, |\Im a|\}$ for $z \in \mathbb{C}$ it holds

$$B_{C\varepsilon,\max}(\boldsymbol{U}_0) := \left\{ \boldsymbol{Z} \in \mathcal{S}_d \colon \max_{i,j=1,\ldots,d} |U_{0ij} - Z_{ij}|_\infty < C\varepsilon \right\} \subset B_{\varepsilon/(4\tau)}(\boldsymbol{U}_0),$$

thus, recalling the definition of the Lebesgue measure $d\boldsymbol{Z}$ on $\mathcal{S}_d$ from (2.5),

$$\int_{B_{\varepsilon/(4\tau)}(\boldsymbol{U}_0)} d\boldsymbol{Z} \geq \int_{B_{C\varepsilon,\max}(\boldsymbol{U}_0)} d\boldsymbol{Z} = (C\varepsilon)^{d^2},$$

yielding the claim. □

Lemma 3.11. *Let ν be a measure on $[0,1]$ such that $\int_{[0,1]} r\nu(dr) < \infty$. Let Z be a nonnegative random variable distributed with Lévy measure ν, i.e.*

$$\mathrm{E} \exp(-tZ) = \exp\left(- \int_{[0,1]} (1 - \exp(-rt)) \nu(dr) \right), \quad t \geq 0.$$

Assume that the distribution of Z is non-trivial, i.e. not equal to a Dirac Delta distribution. Let

$$\psi(u) := \int_{[0,1]} (\exp(ur) - 1 - ur) \nu(dr), \quad u \in \mathbb{R}.$$

Then the derivative $\psi'(u)$ is continuous and strictly monotonically increasing on $\mathbb{R}$. Denote by $u = u(\xi)$ the inverse function of $\xi = \psi'(u)$ for $\xi \in (0, \infty)$. Then for every $\delta > 0$ it holds

$$P(Z \geq \delta) \leq \exp\left(- \int_0^\delta u(\xi) d\xi \right).$$

Proof. See Lemma 26.4 in Sato (1999). □

Now we can present the proof of the main result.

Proof of Theorem 3.7. Let $\nu_A(d\boldsymbol{U}, dr)$ denote the Lévy measure from (3.3) of $\boldsymbol{\Phi}(A)$. We will split up $\boldsymbol{\Phi}(A)$ into a large jump compound Poisson component $\boldsymbol{Y}$ and a small jump component $\boldsymbol{Z}$. The idea is to bound (from below) the probability of $\boldsymbol{Y}$ being close to $\boldsymbol{X}_0$ and the probability of $\boldsymbol{Z}$ being close to $\mathbf{0}$. To elaborate, let $\nu_{\boldsymbol{Y}}(d\boldsymbol{U}, dr) := \mathbb{1}_{(\varepsilon/2,\infty)}(r)\nu_A(d\boldsymbol{U}, dr)$ and $\nu_{\boldsymbol{Z}}(d\boldsymbol{U}, dr) := \mathbb{1}_{[0,\varepsilon/2]}(r)\nu_A(d\boldsymbol{U}, dr)$ and let $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be independent with distribution as in (3.6) and (3.7). Then $\boldsymbol{\Phi}(A) \overset{d}{=} \boldsymbol{Y} + \boldsymbol{Z}$. From $\|\boldsymbol{\Phi}(A) - \boldsymbol{X}_0\|_T \leq \|\boldsymbol{Y} - \boldsymbol{X}_0\|_T + \|\boldsymbol{Z}\|_T$ it follows

$$P(\boldsymbol{\Phi}(A) \in B_\varepsilon(\boldsymbol{X}_0)) \geq P(\boldsymbol{Y} \in B_{\varepsilon/2}(\boldsymbol{X}_0))P(\boldsymbol{Z} \in B_{\varepsilon/2}(\mathbf{0})). \tag{3.10}$$

In view of (3.10), it is sufficient to show the following two results:

(a) There exist positive constants c, C only depending on β_0, β_1, d and τ such that

$$P(\boldsymbol{Y} \in B_{\varepsilon/2}(\boldsymbol{X}_0)) \geq C g_0 \mathcal{L}(A) \exp(-cC_\alpha(A)) \exp\left((d^2 + C_\alpha(A) + 1)\log\varepsilon\right).$$

(b) It holds

$$P(\boldsymbol{Z} \in B_{\varepsilon/2}(\mathbf{0})) \geq 1 - \exp\left(-\frac{1}{2}W\left(\frac{2}{C_\alpha(A)}\right)\right).$$

First recall from Lemma 2.3 that $\boldsymbol{Y}$ is compound Poisson. From Lemma 2.5 we obtain

$$P(\boldsymbol{Y} \in B_{\varepsilon/2}(\boldsymbol{X}_0)) = \exp(-C_{\boldsymbol{Y}}) \sum_{k=0}^{\infty} \frac{1}{k!}\nu_{\boldsymbol{Y}}^k(B_{\varepsilon/2}(\boldsymbol{X}_0)) \geq \exp(-C_{\boldsymbol{Y}})\nu_{\boldsymbol{Y}}(B_{\varepsilon/2}(\boldsymbol{X}_0)),$$

with $C_{\boldsymbol{Y}} := \nu_{\boldsymbol{Y}}(\mathcal{S}_d^+) < \infty$ and $\nu_{\boldsymbol{Y}}^k$ being the k-fold convolution of $\nu_{\boldsymbol{Y}}$. We get

$$C_{\boldsymbol{Y}} = \int_A \int_{\mathbb{S}_d^+} \int_{\varepsilon/2}^{\infty} \frac{\exp(-\beta(x, \boldsymbol{U})r)}{r} dr\alpha(x, d\boldsymbol{U})dx \leq \int_A \alpha(x, \mathbb{S}_d^+)dx \int_{\varepsilon/2}^{\infty} \frac{\exp(-\beta_0 r)}{r} dr.$$

The right hand side is equal to $C_\alpha(A)E_1(\beta_0\varepsilon/2)$, with

$$E_1(x) = \int_x^{\infty} \frac{\exp(-r)}{r} dr, \quad x > 0, \tag{3.11}$$

denoting the exponential integral function. Using the bound (see 5.1.20 in Abramowitz and Stegun (1964))

$$E_1(x) \leq \exp(-x)\log\left(1 + \frac{1}{x}\right) \leq \log\left(1 + \frac{1}{x}\right), \quad x > 0,$$

we arrive at

$$C_{\boldsymbol{Y}} \leq C_\alpha(A)\log\left(1 + \frac{2}{\beta_0\varepsilon}\right) \leq C_\alpha(A)\log\frac{3}{\beta_0\varepsilon} = C_\alpha(A)\left(\log\frac{3}{\beta_0} - \log\varepsilon\right)$$

for $\varepsilon < \beta_0$. Letting $c := \max\{1, \log(3/\beta_0)\} > 0$, this yields

$$\begin{aligned} P(\boldsymbol{Y} \in B_{\varepsilon/2}(\boldsymbol{X}_0)) &\geq \exp(-cC_\alpha(A)) \exp(C_\alpha(A) \log \varepsilon) \nu_{\boldsymbol{Y}}(B_{\varepsilon/2}(\boldsymbol{X}_0)) \\ &= \exp(-cC_\alpha(A)) \exp\left(C_\alpha(A) \log \varepsilon\right) \nu(B_{\varepsilon/2}(\boldsymbol{X}_0)), \end{aligned}$$

where $B_{\varepsilon/2}(\boldsymbol{X}_0) \subset \operatorname{supp}(\nu_{\boldsymbol{Y}})$ (and hence $\nu_{\boldsymbol{Y}}(B_{\varepsilon/2}(\boldsymbol{X}_0)) = \nu(B_{\varepsilon/2}(\boldsymbol{X}_0))$) for ε small enough was used in the last step. An application of Lemma 3.10 for $\nu(B_{\varepsilon/2}(\boldsymbol{X}_0))$ yields

$$P(\boldsymbol{Y} \in B_{\varepsilon/2}(\boldsymbol{X}_0)) \geq C g_0 \mathcal{L}(A) \exp(-cC_\alpha(A)) \exp(C_\alpha(A) \log \varepsilon) \exp((d^2+1) \log \varepsilon),$$

concluding (a).

To show (b), we will apply Lemma 3.11 to $\operatorname{tr} \boldsymbol{Z}$ in order to bound $P(\boldsymbol{Z} \in B_{\varepsilon/2}(\boldsymbol{0})) = P(\operatorname{tr} \boldsymbol{Z} < \varepsilon/2)$ from below. Recall from (3.8) that the Lévy measure of $\operatorname{tr} \boldsymbol{Z}$ is given by

$$\tilde{\nu}_{\boldsymbol{Z}}(dr) = \mathbb{1}_{[0,\varepsilon/2]}(r) \int_A \int_{\mathbb{S}_d^+} \frac{\exp(-\beta(x, \boldsymbol{U})r)}{r} \alpha(x, d\boldsymbol{U}) dx dr.$$

For ε small enough (i.e. $\varepsilon < 2$), the support of $\tilde{\nu}_{\boldsymbol{Z}}$ is contained in $[0,1]$. Since

$$\int_{[0,1]} r \tilde{\nu}_{\boldsymbol{Z}}(dr) = \int_0^{\varepsilon/2} \int_A \int_{\mathbb{S}_d^+} \exp(-\beta(x, \boldsymbol{U})r) \alpha(x, d\boldsymbol{U}) dx dr \leq \frac{\varepsilon}{2} C_\alpha(A) < \infty,$$

it follows that $\tilde{\nu}_{\boldsymbol{Z}}$ fulfills the assumptions of Lemma 3.11. Consider the function

$$\psi(u) := \int_{[0,1]} \left(\exp(ur) - 1 - ur\right) \tilde{\nu}_{\boldsymbol{Z}}(dr), \quad u \in \mathbb{R}.$$

The derivative of ψ can be computed by an application of Lebesgue's Dominated Convergence Theorem as

$$\begin{aligned} \psi'(u) &= \int_{[0,1]} \left(\exp(ur) - 1\right) r \tilde{\nu}_{\boldsymbol{Z}}(dr) \\ &= \int_0^{\varepsilon/2} \int_A \int_{\mathbb{S}_d^+} \left(\exp(ur) - 1\right) \exp(-\beta(x, \boldsymbol{U})r) \alpha(x, d\boldsymbol{U}) dx dr \\ &= \int_A \int_{\mathbb{S}_d^+} \left(\int_0^{\varepsilon/2} \exp\big((u - \beta(x, \boldsymbol{U}))r\big) dr - \int_0^{\varepsilon/2} \exp\big((-\beta(x, \boldsymbol{U}))r\big) dr \right) \alpha(dx, d\boldsymbol{U}) dx \\ &= \int_A \int_{\mathbb{S}_d^+} L\big(u - \beta(x, \boldsymbol{U})\big) - L\big(-\beta(x, \boldsymbol{U})\big) \alpha(x, \boldsymbol{U}) dx \end{aligned}$$

with the function $L \colon \mathbb{R} \to (0, \infty)$ being defined as $L(x) = \frac{\exp(\varepsilon/2x) - 1}{x}$ for $x \neq 0$. Since L is continuously differentiable with derivative $L'(x) = \frac{(\varepsilon/2x - 1)\exp(\varepsilon/2x) + 1}{x^2}$, it follows from the Mean Value Theorem that for $u > 0$, there exists $\tau = \tau(x, u, \boldsymbol{U}) \in (-\beta(x, \boldsymbol{U}), -\beta(x, \boldsymbol{U}) + u)$ such that

$$L\big(u - \beta(x, \boldsymbol{U})\big) - L\big(-\beta(x, \boldsymbol{U})\big) = L'(\tau(x, u, \boldsymbol{U}))u.$$

This yields

$$\psi'(u) = \rho(u)u, \quad u > 0, \tag{3.12}$$

with

$$\rho(u) = \int_A \int_{\mathbb{S}_d^+} L'(\tau(x,u,\boldsymbol{U}))\alpha(x,d\boldsymbol{U})dx.$$

By Assumption GP4 it holds $\tau(x,u,\boldsymbol{U}) \in (-\beta_1, -\beta_0+u) \subset (-\beta_1, u)$ for all $x, \boldsymbol{U}$ and all $u > 0$. Let $\delta := \varepsilon/2$. From Lemma B.25 in the Appendix, we find that $L'(t) > 0$ for all $t \in \mathbb{R}$ and $L'(t) \leq \delta^2/2\exp(\delta u)$ for $t \in [0,u]$. Furthermore, since L' is monotonically increasing on $\mathbb{R}$ (this can be seen by computing the derivative $L''(t) = \frac{[(\delta t-1)^2+1]\exp(\delta t)-2}{t^3}$ and by observing that $L''(t) \geq 0$ for all $t \in \mathbb{R}$ is equivalent to $\log(z^2-2z+2)+z \geq \log 2$ for all $z \geq 0$ and $\log(z^2+2z+2)-z \leq \log 2$ for all $z > 0$ and these properties *are* fulfilled, since the function $h(z) := \log(z^2-2z+2)+z$ is monotonically increasing on $[0,\infty)$ with $h(0) = \log 2$ and the function $\tilde{h}(z) := \log(z^2+2z+2)-z$ is monotonically decreasing on $[0,\infty)$ with $\tilde{h}(0) = \log 2$), it even holds $L'(t) \leq \delta^2/2\exp(\delta u)$ for all $t \leq u$. This concludes $L'(\tau(x,u,\boldsymbol{U})) \in (0, \delta^2/2\exp(\delta u))$, which leads to $\rho(u) \in (0, \frac{1}{2}C_\alpha(A)\delta^2\exp(\delta u))$, and from (3.12) we arrive at

$$0 < \psi'(u) \leq \frac{1}{2}C_\alpha(A)\delta^2\exp(\delta u)u, \quad u > 0. \tag{3.13}$$

Denote by $u = u(\xi) \geq 0$ the inverse function of $\xi = \psi'(u)$ for $\xi \in (0,\infty)$. Then (3.13) is equivalent to

$$\delta u(\xi)\exp(\delta u(\xi)) \geq \frac{2\xi}{C_\alpha(A)\delta}, \quad \xi > 0. \tag{3.14}$$

Applying the Lambert W function (see (B.27) in Section B.5 in the Appendix) to the left hand side of (3.14) yields

$$W\left(\delta u(\xi)\exp(\delta u(\xi))\right) = \delta u(\xi).$$

Since W is strictly monotonically increasing on $(0,\infty)$, this yields $\delta u(\xi) \geq W(\frac{2\xi}{C_\alpha(A)\delta})$ or

$$u(\xi) \geq \frac{1}{\delta}W\left(\frac{2\xi}{C_\alpha(A)\delta}\right), \quad \xi > 0.$$

An application of Lemma 3.11 yields

$$P(\operatorname{tr}\boldsymbol{Z} \geq \delta) \leq \exp\left(-\int_0^\delta u(\xi)d\xi\right) \leq \exp\left(-\frac{1}{\delta}\int_0^\delta W\left(\frac{2\xi}{C_\alpha(A)\delta}\right)d\xi\right).$$

Using that W is concave on $[0,\infty)$ and that $W(0) = 0$ yields $W(\frac{2\xi}{C_\alpha(A)\delta}) \geq \frac{\xi}{\delta}W(\frac{2}{C_\alpha(A)})$ for $0 \leq \xi \leq \delta$, hence

$$P(\operatorname{tr}\boldsymbol{Z} \geq \delta) \leq \exp\left(-\frac{1}{\delta^2}W\left(\frac{2}{C_\alpha(A)}\right)\int_0^\delta \xi d\xi\right) = \exp\left(-\frac{1}{2}W\left(\frac{2}{C_\alpha(A)}\right)\right),$$

which concludes the proof. □

3.3. An Infinite Series Representation

In (3.2), we defined an Hpd Gamma process $\mathbf{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$ on $\mathcal{X}$ based on the infinitely divisible Hpd Gamma distribution. The mean measure ν from (3.1) of the underlying Poisson process Π was chosen as a mixture of Lévy measures of Hpd Gamma distributions, where the distributional parameters α and β are allowed to change over $\mathcal{X}$. For a practical usage (as e.g. in Bayesian nonparametric inference), an almost surely convergent series representation of $\mathbf{\Phi}$ (and Π respectively) involving iid random variables is of great usefulness. As an example of such a representation, the Dirichlet process obeys the well-known stick-breaking (1.5), which expands the process as an infinite sum involving iid random variables drawn from a Beta distribution and the base measure G_0. In this section, we derive an almost surely convergent series representation for $\mathbf{\Phi}$. It is not based on stick-breaking, but on LePage's method (LePage, 1981), which can be conceived as a generalization of the Inverse CDF Transform Sampling for non-uniform random variables (Devroye, 1986). The almost sure convergence of such series has been discussed in Rosiński (2001). The following Lemma summarizes the results that are needed for our construction.

Lemma 3.12 (LePage's Method)**.** *Let $\mathcal{Y}$ be a Borel space. Consider a Poisson process Π on $\mathcal{Y}\times[0,\infty)$ with σ-finite mean measure ν being of the form*

$$\nu(dy, dr) = \rho(dr|y)\alpha^*(dy), \tag{3.15}$$

where α^* is a probability measure on $\mathcal{Y}$ and $\{\rho(\cdot|y)\colon y\in\mathcal{Y}\}$ is a family of absolutely continuous measures on $[0,\infty)$, such that for all measurable $B\subset[0,\infty)$, the mapping $[\mathcal{Y}\ni y\mapsto\rho(B|y)]$ is measurable. Denote by

$$\rho^-(w|y) = \inf\{r>0\colon \rho([r,\infty)|y) < w\} \tag{3.16}$$

the inverse of the tail of $\rho(\cdot|y)$. Then the following almost sure representation for Π holds:

$$\Pi \overset{a.s.}{=} \{(y_j, \rho^-(w_j|y_j))\}_{j\geq 1}, \quad y_j \overset{iid}{\sim} \alpha^*, \quad w_j = \sum_{i=1}^{j} v_i, \quad v_i \overset{iid}{\sim} \mathrm{Exp}(1), \tag{3.17}$$

where $\{y_j\}_{j\geq1}$ and $\{v_j\}_{j\geq1}$ are independent.

Proof. The proof is a minor adaption of Section 3(B) in Rosiński (2001). We start with a Poisson Process Π_0 on $(0,\infty)$ with mean measure $\nu_0(dr) = dr$ being equal to the Lebesgue measure. From the Interval Theorem (see Theorem B.20 in the appendix), we can conclude the representation in distribution

$$\Pi_0 \overset{d}{=} \{w_j\}_{j\geq1}, \quad w_j = \sum_{i=1}^{j} v_i, \quad v_i \overset{\text{iid}}{\sim} \mathrm{Exp}(1). \tag{3.18}$$

With the Marking Theorem (see Theorem B.21 in the appendix), we can use Π_0 to construct a Poisson Process Π_1 on $\mathcal{Y}\times(0,\infty)$ having mean measure $\nu_1(dy,dr) = \alpha^*(dy)dr$ as follows:

$$\Pi_1 \overset{d}{=} \{(y_j, w_j)\}_{j\geq1}, \quad y_j \overset{\text{iid}}{\sim} \alpha^*,$$

where the sequence $\{w_j\}_{j\geq 1}$ is as in (3.18) and $\{y_j\}_{j\geq 1}$ being independent thereof. Now consider the mapping $h\colon \mathcal{Y}\times(0,\infty)\to\mathcal{Y}\times[0,\infty)$, defined as $h(y,r)=(y,\rho^-(r|y))$ with ρ^- from (3.16). For $s>0$, let $h^{-1}(y,s):=(y,\rho([s,\infty)|y))$. Then for $A\subset\mathcal{Y}$ measurable and $0<a<b$ it holds

$$(\nu_1\circ h^{-1})(A\times[a,b)) = \int_A \int_{[\rho([b,\infty)|y),\rho([a,\infty)|y))} dr\alpha^*(dy) = \int_A \rho([a,b)|y)\alpha^*(dy),$$

which is equal to $\nu(A\times[a,b))$ with ν from (3.15). Thus $\nu=\nu_1\circ h^{-1}$ by Carathéodory's extension theorem (recall that the measure extension is unique, because ν is assumed to be σ-finite). Applying the almost sure mapping theorem (see Theorem B.22 in the appendix) to Π_1 and h yields the representation (3.17) for Π, with possibly different sequences $\{v_j\},\{y_j\}$ which are, however, equal in distribution. □

The result from Lemma 3.12 can now be applied to derive a series representation for the Gamma process $\mathbf{\Phi}$ from (3.2). To do so, we will need the following strengthened version of Assumption GP2:

Assumption GP2'. *The integral $C_\alpha := \int_{\mathcal{X}} \alpha(x,\bar{\mathbb{S}}_d^+)dx$ is finite and there exists a constant $\beta_0>0$ such that $\beta(x,\boldsymbol{U})\geq\beta_0$ holds for all $\boldsymbol{U}\in\bar{\mathbb{S}}_d^+$ and all $x\in\mathcal{X}\setminus N$, where $N\subset\mathcal{X}$ is a null set.*

Lemma 3.13. *Let $\mathcal{X}$ fulfill assumption $\mathcal{X}1$. Let $\mathbf{\Phi}\sim \mathrm{CRM}_{d\times d}(\nu)$ with ν as in (3.1) fulfilling Assumptions GP1 and GP2'. Then the following series representation holds for $\mathbf{\Phi}$:*

$$\mathbf{\Phi} \overset{a.s.}{=} \sum_{j\geq 1} \delta_{x_j} r_j \boldsymbol{U}_j, \quad (x_j,\boldsymbol{U}_j)\overset{iid}{\sim}\alpha^*,$$

$$r_j = \rho^-(w_j|C_\alpha,\beta(x_j,\boldsymbol{U}_j)), \quad w_j=\sum_{i=1}^{j} v_i, \quad v_i\overset{iid}{\sim}\mathrm{Exp}(1),$$

where α^* is the probability measure on $\mathcal{X}\times\bar{\mathbb{S}}_d^+$ induced by α, i.e. $\alpha^*(dx,d\boldsymbol{U})=\frac{\alpha(x,d\boldsymbol{U})dx}{C_\alpha}$ with C_α as in Assumption GP2' and ρ^- is the inverse of the tail of the radial Gamma measure ρ:

$$\rho^-(w|a,b)=\inf\{r>0\colon \rho([r,\infty]|a,b)<w\}, \quad \rho(dr|a,b)=a\frac{\exp(-br)}{r}dr.$$

Proof. We first observe that Assumption GP2' implies Assumption GP2: Indeed, we find

$$\begin{aligned}
&\int_{\mathcal{X}}\int_{\bar{\mathbb{S}}_d^+}\int_0^\infty \min(1,r)\nu(dx,d\boldsymbol{U},dr)\\
&=\int_{\mathcal{X}}\int_{\bar{\mathbb{S}}_d^+}\int_0^1 r\nu(dx,d\boldsymbol{U},dr)+\int_{\mathcal{X}}\int_{\bar{\mathbb{S}}_d^+}\int_1^\infty \nu(dx,d\boldsymbol{U},dr)\\
&=:\int_{\mathcal{X}}\int_{\bar{\mathbb{S}}_d^+} R_1(x,\boldsymbol{U})\alpha(x,d\boldsymbol{U})dx+\int_{\mathcal{X}}\int_{\bar{\mathbb{S}}_d^+} R_2(x,\boldsymbol{U})\alpha(x,d\boldsymbol{U})dx
\end{aligned}$$

with

$$R_1(x,\boldsymbol{U})=\int_0^1 \exp(-\beta(x,\boldsymbol{U})r)dr\leq\int_0^1 dr=1$$

and

$$R_2(x,\boldsymbol{U})=\int_1^\infty \frac{\exp(-\beta(x,\boldsymbol{U})r)}{r}dr\leq\int_1^\infty \exp(-\beta_0 r)dr=\frac{\exp(-\beta_0)}{\beta_0}.$$

This yields

$$\int_{\mathcal{X}} \int_{\bar{\mathbb{S}}_d^+} \int_0^\infty \min(1, r)\nu(dx, d\boldsymbol{U}, dr) \leq C_\alpha \left(1 + \frac{\exp(-\beta_0)}{\beta_0}\right) < \infty,$$

with C_α from Assumption GP2'. Hence Assumption GP2 is satisfied and $\boldsymbol{\Phi}$ is well-defined by Theorem 3.2. Denote by $\Pi \sim \mathrm{PP}(\nu)$ the underlying Poisson Process. Let $\mathcal{Y} := \mathcal{X} \times \bar{\mathbb{S}}_d^+$ be endowed with the product σ-algebra and write $y = (x, \boldsymbol{U})$ as well as $dy = (dx, d\boldsymbol{U})$. Denote by $\tilde{\alpha}$ the measure on $\mathcal{Y}$ given as $\tilde{\alpha}(dy) = \alpha(x, d\boldsymbol{U})dx$. Then ν can be written as

$$\nu(dy, dr) = \frac{\exp(-\beta(y)r)}{r} dr\tilde{\alpha}(dy) = \rho(dr|C_\alpha, \beta(y))\tilde{\alpha}^*(dy),$$

where $\tilde{\alpha}^*(dy) = \frac{\tilde{\alpha}(dy)}{C_\alpha}$ denotes the probability measure on $\mathcal{Y}$ induced by $\tilde{\alpha}$. Since $\rho(dr|a, b)$ is a σ-finite measure for all $a, b > 0$, it follows that ν is σ-finite and an application of Lemma 3.12 to Π yields

$$\Pi \overset{\text{a.s.}}{=} \{(y_j, \rho^-(w_j|C_\alpha, \beta(y_j)))\}_{j\geq 1}, \quad y_j = (x_j, \boldsymbol{U}_j) \overset{\text{iid}}{\sim} \tilde{\alpha}^*$$

and $\{w_j\}_{j\geq 1}$ as in the claim and independent of $\{y_j\}$. Recalling the connection between $\boldsymbol{\Phi}$ and Π from (3.2) concludes the proof. □

Remark 3.14. *The result from Lemma 3.13 requires the computation of the total mass C_α of the measure α on $\mathcal{X} \times \bar{\mathbb{S}}_d^+$. In many cases, this can be derived analytically. For example, when employing the $A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ distribution from Section 2.4 with $\eta \equiv \eta_0 > d - 1$ and $\omega = \omega(x) > 0$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(x) \in \mathcal{S}_d^+$ for $x \in \mathcal{X}$, it holds $\int_{\bar{\mathbb{S}}_d^+} \alpha(x, d\boldsymbol{U}) = \omega(x) \int_{\bar{\mathbb{S}}_d^+} \alpha_{\eta, \boldsymbol{\Sigma}(x)}(d\boldsymbol{U}) = \omega(x)$, and hence $C_\alpha = \int_{\mathcal{X}} \omega(x)dx$.*

There is no analytically closed form available for the inverse tail of the Gamma measure $\rho^-(\cdot|a, b)$. However, we have $\rho^-(w|a, b) = b^{-1}E_1^-(a^{-1}w)$, where E_1^- is defined as the inverse exponential integral function:

$$E_1^-(w) := \inf\{r > 0 \colon E_1(r) < w\}, \quad E_1(r) = \int_r^\infty \frac{\exp(-z)}{z} dz.$$

As noted in Wolpert and Ickstadt (1998), the values of E_1 can be approximated numerically by tail probabilities of rescaled χ^2 distributions and the values of E_1^- by the corresponding quantiles:

$$E_1(r) = \lim_{\delta \to 0} \frac{2}{\delta} \mathbb{P}(\chi_\delta^2 > 2r) \approx \frac{2}{\delta_0} \mathbb{P}(\chi_{\delta_0}^2 > 2r), \quad E_1^-(z) \approx \frac{1}{2} q_{\chi_{\delta_0}^2}(1 - z\delta_0/2) \tag{3.19}$$

for some *small* value δ_0 (e.g. $\delta_0 = 10^{-12}$), with $q_{\chi_k^2}(p)$ denoting the p-quantile of the Chi Squared distribution with k degrees of freedom.

3.4. Numerical Simulation of Prior Samples

In this Section, we discuss the non-trivial task of drawing random samples from the $\mathrm{Ga}_{d\times d}(\alpha, \beta)$ process. We restrict our attention to the $A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ distribution from Section 2.4, for the

following two reasons. On the one hand, it offers a great amount of prior modeling flexibility while retaining a simple parametrization in terms of only the three parameters $\eta, \omega, \mathbf{\Sigma}$. On the other hand, the density of its Lévy measure is available in analytical form (see (2.17)) and this property is in fact necessary for the methods we will present in Section 3.4.2. As discussed at the end of this Section, the techniques can be extended to many other $\mathrm{Ga}_{d\times d}(\alpha, \beta)$ processes. Some illustrations will be shown in Section 3.4.3 and a straightforward extension to simulation of random samples from the $A\Gamma$ distribution (rather than the process) is presented in Section 3.4.4. The methods rely on Markov Chain Monte Carlo (MCMC) algorithms such as the Gibbs sampler and the Metropolis-Hastings algorithm. A description of both algorithms (along with a comprehensive introduction to MCMC) can be found in Section 6.3 in Christensen et al. (2011). We start our consideration in Section 3.4.1 with a convenient and practical parametrization for elements of the sphere $\mathbb{S}_d^+$.

3.4.1. Parametrization of the $\boldsymbol{U}_l$'s

In this section, we derive a practical parametrization of Hpd matrices $\boldsymbol{U} \in \mathbb{S}_d^+$ with unit trace (see (2.3)). We use a minor adaptation of the *hyperspherical coordinates* approach derived in Mittelbach et al. (2012). To elaborate, let $\boldsymbol{U} = \boldsymbol{U}_L \boldsymbol{U}_L^*$ denote the Cholesky decomposition and denote the elements of $\boldsymbol{U}_L$ as follows:

$$\boldsymbol{U}_L = \begin{pmatrix} y_1 & & & & & & \\ y_2 - iy_3 & y_4 & & & & & \\ y_5 - iy_6 & y_7 - iy_8 & y_9 & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ y_{(d-1)^2+1} - iy_{(d-1)^2+2} & \dots & \dots & \dots & y_{d^2-2} - iy_{d^2-1} & y_{d^2} \end{pmatrix}.$$

The components $y_1, \dots, y_{d^2}$ are represented in hyperspherical coordinates $\varphi_1, \dots, \varphi_{d^2-1}$, i.e.

$$y_j = \begin{cases} \cos(\varphi_j) \prod_{l=1}^{j-1} \sin(\varphi_l), & j = 1, \dots, d^2 - 1 \\ \prod_{l=1}^{j-1} \sin(\varphi_l), & j = d^2, \end{cases} \tag{3.20}$$

where the range of φ_j is given by

$$\mathcal{I}_j = \begin{cases} (0, \pi/2), & j = l^2 \text{ for } l = 1, \dots, (d-1) \\ (0, \pi), & \text{else}, \end{cases} \tag{3.21}$$

for $j = 1, \dots, d^2 - 1$, see (60)-(63) in Mittelbach et al. (2012). This yields the parametrization $\underline{\varphi} := (\varphi_1, \dots, \varphi_{d^2-1})$ for $\boldsymbol{U}$. Denote by $T \colon \mathbb{S}_d^+ \to \mathcal{I} := \bigotimes_{j=1}^{d^2-1} \mathcal{I}_j$ the transformation $T(\boldsymbol{U}) = \underline{\varphi}$. Then the absolute value of the determinant of the Jacobian $\boldsymbol{J}_{T^{-1}}$ of the inverse mapping T^{-1} is given by (see Section IV.A and Section II.B in Mittelbach et al. (2012))

$$|\boldsymbol{J}_{T^{-1}}(\underline{\varphi})| = \prod_{j=1}^{d^2-1} \cos^{p_j}(\varphi_j) \sin^{q_j}(\varphi_j) \tag{3.22}$$

with the exponents

$$p_j = \begin{cases} (d+1) - l, & j = l^2 \text{ for } l = 1, \dots, d-1 \\ 0, & \text{else} \end{cases}$$

and

$$q_j = d^2 - (l-1)d - l - 1 - m, \quad \text{for } j = l^2 + m,\ j = 1, \dots, d-1,\ m = 1, \dots, l,$$

see (27)-(31) in Mittelbach et al. (2012).

3.4.2. The Algorithm

First, let us briefly revisit the $A\Gamma$ distribution. Let $\eta_0 > d-1$, $\omega_0 > 0$ and $\mathbf{\Sigma}_0 \in \mathcal{S}_d^+$. Recall the following distributional properties of the $A\Gamma(\eta_0, \omega_0, \mathbf{\Sigma}_0)$ distribution (see Lemma 2.7 and the proof thereof, and Lemma 2.9):

- The $A\Gamma(\eta_0, \omega_0, \mathbf{\Sigma}_0)$ distribution is the $\mathrm{Ga}_{d\times d}(\omega_0 \alpha_{\eta_0, \mathbf{\Sigma}_0}, \beta_{\mathbf{\Sigma}_0})$ distribution with the *probability measure* $\alpha_{\eta_0, \mathbf{\Sigma}_0}$ on $\mathbb{S}_d^+$ and the mapping $\beta_{\mathbf{\Sigma}_0} \colon \mathbb{S}_d^+ \to (0, \infty)$ from Lemma 2.7.
- For $\boldsymbol{X} \sim A\Gamma(\eta_0, \omega_0, \mathbf{\Sigma}_0)$ it holds $\mathrm{E}\boldsymbol{X} = \frac{\omega_0}{d}\mathbf{\Sigma}_0$ and $\mathrm{Cov}\boldsymbol{X} \propto \omega_0 \mathbf{\Sigma}_0 \otimes \mathbf{\Sigma}_0$, where the proportionality is understood component-wise.

Let $\mathcal{X}$ fulfill Assumption $\mathcal{X}1$. We define an $A\Gamma(\eta, \omega, \mathbf{\Sigma})$ process on $\mathcal{X}$ under the following assumptions on the process parameters $\eta, \omega, \mathbf{\Sigma}$:

Assumption $A\Gamma\mathbf{1}$. *Let $\eta \equiv \eta_0 > d-1$ and $\omega \colon \mathcal{X} \to (0, \infty)$ be measurable such that $\int_{\mathcal{X}} \omega(x)dx < \infty$. Let $\mathbf{\Sigma} \colon \mathcal{X} \to \mathcal{S}_d^+$ be measurable such that $\sup_{x\in\mathcal{X}} \lambda_{\max}(\mathbf{\Sigma}(x)) \leq \tau < \infty$.*

Consider the Poisson mean measure ν on $\mathcal{X} \times \bar{\mathbb{S}}_d^+ \times (0, \infty)$ given as

$$\nu(dx, d\boldsymbol{U}, dr) = \frac{\exp(-\beta_{\mathbf{\Sigma}(x)}(\boldsymbol{U})r)}{r} dr\omega(x)\alpha_{\eta, \mathbf{\Sigma}(x)}(d\boldsymbol{U})dx. \tag{3.23}$$

and let $\mathbf{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$. First observe that by Theorem 3.2 and Lemma 2.9 it holds

$$\mathrm{E}[\mathbf{\Phi}(A)] = \frac{1}{d}\int_A \omega(x)\mathbf{\Sigma}(x)dx, \quad A \subset \mathcal{X} \text{ measurable.} \tag{3.24}$$

Denote by α the measure on $\mathcal{X} \times \bar{\mathbb{S}}_d^+$ defined as $\alpha(dx, d\boldsymbol{U}) = \omega(x)\alpha_{\eta, \mathbf{\Sigma}(x)}(d\boldsymbol{U})dx$. Let $C_\alpha := \alpha(\mathcal{X}, \mathbb{S}_d^+) = \int_{\mathcal{X}} \omega(x)dx < \infty$ by Assumption $A\Gamma1$ (where the last equality is due to Remark 3.14) denote its total mass and $\alpha^* := \frac{\alpha}{C_\alpha}$ its normalization. Under Assumption $A\Gamma1$, the assumptions of Lemma 3.13 are fulfilled since $C_\alpha < \infty$ and, using Lemma B.4 in the Appendix, it holds

$$\beta_{\mathbf{\Sigma}(x)}(\boldsymbol{U}) = \mathrm{tr}\left(\mathbf{\Sigma}(x)^{-1}\boldsymbol{U}\right) \geq \lambda_{\min}(\mathbf{\Sigma}(x)^{-1})\,\mathrm{tr}(\boldsymbol{U}) = \frac{1}{\lambda_{\max}(\mathbf{\Sigma}(x))} \geq \frac{1}{\tau}$$

for all $x \in \mathcal{X}$ and all $\boldsymbol{U} \in \mathbb{S}_d^+$. From Lemma 3.13, we obtain

$$\begin{gathered}\mathbf{\Phi} \overset{\text{a.s.}}{=} \sum_{j=1}^{\infty} \delta_{x_j} r_j \boldsymbol{U}_j, \quad (x_j, \boldsymbol{U}_j) \overset{\text{iid}}{\sim} \alpha^*, \quad r_j = \rho^-(w_j | C_\alpha, \beta_{\mathbf{\Sigma}(x_j)}(\boldsymbol{U}_j)), \\ w_j = \sum_{i=1}^{j} v_i, \quad v_i \overset{\text{iid}}{\sim} \mathrm{Exp}(1),\end{gathered} \tag{3.25}$$

where the v_i's are independent of the $(x_j, \boldsymbol{U}_j)$'s and the inverse tail ρ^- of the Gamma measure as in Lemma 3.13. In practice, the representation (3.25) is truncated at some large positive

integer L, yielding the approximate form $\mathbf{\Phi} \approx \sum_{j=1}^{L} \delta_{x_j} r_j \boldsymbol{U}_j$ (the part after (5.5) in Section 5.2 for a more detailed discussion of the truncation issue).

In this approximation, $\mathbf{\Phi}$ is represented by the $3L$ parameters $(x_1, \ldots, x_L, r_1, \ldots, r_L, \boldsymbol{U}_1, \ldots, \boldsymbol{U}_L)$. To generate random samples of $\mathbf{\Phi}$, one needs to generate random samples from the probability distribution α^* on $\mathcal{X} \times \mathbb{S}_d^+$. Let $\omega^*(x) = \frac{\omega(x)}{C_\alpha}$. Denote by

$$g^*(x, \boldsymbol{U}) = \frac{\Gamma(\eta d)\omega^*(x)|\boldsymbol{U}|^{\eta-d}}{\tilde{\Gamma}_d(\eta)|\mathbf{\Sigma}(x)|^\eta \operatorname{tr}(\mathbf{\Sigma}(x)^{-1}\boldsymbol{U})^{d\eta}}, \quad x \in \mathcal{X}, \boldsymbol{U} \in \mathbb{S}_d^+$$

the probability density of α^*. We observe

$$g^*(x, \boldsymbol{U}) = \omega^*(x) g^*(\boldsymbol{U}|x), \quad x \in \mathcal{X}, \boldsymbol{U} \in \mathbb{S}_d^+ \tag{3.26}$$

with

$$g^*(\boldsymbol{U}|x) = \frac{\Gamma(\eta d)|\boldsymbol{U}|^{\eta-d}}{\tilde{\Gamma}_d(\eta)|\mathbf{\Sigma}(x)|^\eta \operatorname{tr}(\mathbf{\Sigma}(x)^{-1}\boldsymbol{U})^{d\eta}}, \quad \boldsymbol{U} \in \mathbb{S}_d^+$$

being the Lebesgue density of the $\alpha_{\eta,\mathbf{\Sigma}(x)}$ measure on $\mathbb{S}_d^+$ for fixed $x \in \mathcal{X}$. Assume that the function ω is such that we can draw random samples from (the probability distribution with density) ω^*. Then it follows from (3.26) that a random sample $(x, \boldsymbol{U})$ from α^* can be generated by first drawing x from ω^*, and then – given x – drawing $\boldsymbol{U}$ from $\alpha_{\eta,\mathbf{\Sigma}(x)}$. This idea is summarized in Algorithm 1. In view of Algorithm 1, a reasonable idea is to set ω^* to the (known) probability

Algorithm 1: Generate a draw from the $A\Gamma(\eta, \omega, \mathbf{\Sigma})$ process

input : $L \in \mathbb{N}$, $\eta > d-1$, functions $\omega\colon \mathcal{X} \to (0, \infty)$ and $\mathbf{\Sigma}\colon \mathcal{X} \to \mathcal{S}_d^+$

Compute $C_\alpha \leftarrow \int_{\mathcal{X}} \omega(x)dx$;
Draw $v_1, \ldots, v_L \overset{\text{iid}}{\sim} \text{Exp}(1)$ and compute $w_1, \ldots, w_L$ from (3.25);
Draw $x_1, \ldots, x_L$ independently with pdf $\omega^* = \frac{\omega}{C_\alpha}$;
for $j \leftarrow 1$ **to** L **do**
    Draw $\boldsymbol{U}_j$ from $\alpha_{\eta,\mathbf{\Sigma}(x_j)}$, e.g. with Algorithm 2;
    Compute $r_j \leftarrow \rho^-(w_j | C_\alpha, \beta_{\mathbf{\Sigma}(x_j)}(\boldsymbol{U}_j))$;
end

output: $\mathbf{\Phi} \leftarrow \sum_{j=1}^{L} \delta_{x_j} r_j \boldsymbol{U}_j$

density function of a continuous random variable on $\mathcal{X}$, such that samples from ω^* can readily be obtained. A common example for the case $\mathcal{X} = [0, 1]$ are the Beta densities and finite mixtures thereof. If however ω^* does not belong to a known family, a numerical sampling method such as Rejection Sampling or MCMC techniques (see Chapter 6 in Christensen et al. (2011)) can be employed to draw from ω^*.

Algorithm 1 relies on random draws from the probability distribution $\alpha_{\eta,\mathbf{\Sigma}_0}$ on $\mathbb{S}_d^+$ for some $\mathbf{\Sigma}_0 \in \mathcal{S}_d^+$. An efficient method for this purpose can be derived with the following key observation: The measure $\alpha_{\eta_0,\mathbf{\Sigma}_0}$ is generated by the measure $\alpha_{\eta_0} = \alpha_{\eta_0,\boldsymbol{I}_d}$ from (2.14) under the transformation $\mathbb{S}_d^+ \ni \boldsymbol{V} \mapsto \frac{\mathbf{\Sigma}_0^{1/2}\boldsymbol{V}\mathbf{\Sigma}_0^{1/2}}{\operatorname{tr}(\mathbf{\Sigma}_0\boldsymbol{V})} \in \mathbb{S}_d^+$, see (2.15) for the precise argument. This idea is summarized in Algorithm 2, and it remains to derive a method to draw random samples from α_η.

Algorithm 2: Generate a draw from $\alpha_{\eta_0,\boldsymbol{\Sigma}_0}$

input : $\eta_0 > d-1$, $\boldsymbol{\Sigma}_0 \in \mathcal{S}_d^+$

Compute Hpd matrix square root $\boldsymbol{\Sigma}_0^{1/2}$;

Draw random matrix $\boldsymbol{V}$ from α_{η_0}, e.g. with Algorithm 3;

output: $\boldsymbol{U} \leftarrow \frac{\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{V}\boldsymbol{\Sigma}_0^{1/2}}{\mathrm{tr}(\boldsymbol{\Sigma}_0\boldsymbol{V})}$

Since α_η has the Lebesgue density

$$g_0^*(\boldsymbol{U}) = \frac{\Gamma(\eta d)}{\tilde{\Gamma}_d(\eta)}|\boldsymbol{U}|^{\eta-d}, \quad \boldsymbol{U} \in \mathbb{S}_d^+, \tag{3.27}$$

random samples from α_η can be obtained efficiently with the Metropolis Hastings algorithm (see Section 6.3.3 in Christensen et al. (2011)), with the parametrization for elements of the sphere $\mathbb{S}_d^+$ from Section 3.4.1. This is illustrated in Algorithm 3. Note that the Jacobian of the re-parametrization has to be taken into account for the target density (see Line 3). Also note that uniform proposals are used under the re-parametrization (see Line 4). If $\eta_0 = d$, the α_{η_0}

Algorithm 3: Generate a draw from α_{η_0} (Metropolis-Hastings)

input : $\eta_0 > d-1$, $N_{\text{burn}} \in \mathbb{N}$

1 Initialize $\underline{\varphi}^{(1)} \leftarrow \mathrm{Unif}(\mathcal{I})$ with $\mathcal{I} = \bigotimes_{j=1}^{d^2-1} \mathcal{I}_j$ from (3.21);
2 Set $\boldsymbol{U}^{(1)} \leftarrow T^{-1}(\underline{\varphi}^{(1)})$ with T from (3.22);
3 Let $p_{\text{target}}(\underline{\varphi}) = g_0^*(T^{-1}(\underline{\varphi}))|\boldsymbol{J}_{T^{-1}}(\underline{\varphi})|$;
4 Let $p_{\text{prop}}(\underline{\varphi}) \propto \mathbb{1}_{\mathcal{I}}(\underline{\varphi})$;
5 **for** $i \leftarrow 1, \ldots, N_{\text{burn}}$ **do**
6 Retrieve $\underline{\varphi}^{(i+1)}$ from an MH step with target density p_{target}, proposal density p_{prop} and previous value $\underline{\varphi}^{(i)}$;
7 Set $\boldsymbol{U}^{(i+1)} \leftarrow T^{-1}(\underline{\varphi}^{(i+1)})$;
8 **end**

output: $\boldsymbol{U}^{(N_{\text{burn}}+1)}$

measure corresponds to the uniform probability measure on $\mathbb{S}_d^+$. In this case, even more efficient methods for generating random samples exist (see Mittelbach et al. (2012)).

3.4.3. Illustration

As an example, we will consider three $A\Gamma$-processes $\boldsymbol{\Phi}_i$ on $\mathcal{X} = [0,1]$ in dimension $d = 2$, given by the following parameters $\omega_i, \boldsymbol{\Sigma}_i$ for $i = 1,2,3$ and $b(x|j,k)$ denoting the probability density of the Beta(j,k) distribution:

$$\omega_1(x) \equiv 2, \quad \boldsymbol{\Sigma}_1(x) \equiv \boldsymbol{I}_2, \tag{3.28}$$

$$\omega_2(x) = 2(1 + 9b(x|1,10)), \quad \boldsymbol{\Sigma}_2(x) = \frac{2}{\omega_2(x)}\boldsymbol{I}_2, \tag{3.29}$$

$$\omega_3(x) \equiv 2, \quad \boldsymbol{\Sigma}_3(x) = \boldsymbol{f}_{\text{VAR}}(\pi x|\boldsymbol{B},\boldsymbol{S}), \tag{3.30}$$

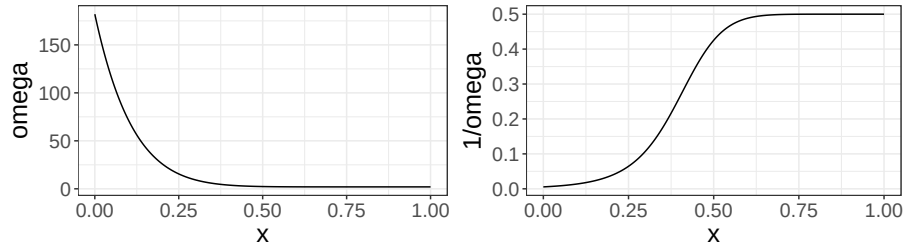


Figure 3.1.: Visualization of $\omega_2(x)$ and $\frac{1}{\omega_2(x)}$ from (3.29).

for $0 \leq x \leq 1$ and $\boldsymbol{f}_{\mathrm{VAR}}(\cdot|\boldsymbol{B}, \boldsymbol{S})$ denoting the spectral density of a VAR(1) model

$$\underline{Z}_t = \boldsymbol{B}\underline{Z}_{t-1} + \underline{e}_t, \quad \underline{e}_t \sim N_d(\underline{0}, \boldsymbol{S}),$$

(the formula of $\boldsymbol{f}_{\mathrm{VAR}}$ can be found in (6.7)), with VAR parameter $\boldsymbol{B}$ and covariance matrix $\boldsymbol{S}$ given by

$$\boldsymbol{B} = \begin{pmatrix} 0.5 & 0 \\ 0 & -0.7 \end{pmatrix}, \quad \boldsymbol{S} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}.$$

A visualization of $\omega_2(x)$ is shown in Figure 3.1 and a visualization of $\boldsymbol{\Sigma}_3(x)$ is included in Figure 5.1(c) and Figure 5.1(d).

The increments $d\boldsymbol{\Phi}_i(x)$ of $\boldsymbol{\Phi}_i$ among $0 \leq x \leq 1$ are independent and $A\Gamma(\eta, \omega_i(x), \boldsymbol{\Sigma}_i(x))$ distributed, see (3.5). In particular, it follows from (3.24) that $\mathrm{E}[d\boldsymbol{\Phi}_i(x)] = \frac{\omega_i(x)}{2}\boldsymbol{\Sigma}_i(x)dx$ and hence

$$\mathrm{E}[d\boldsymbol{\Phi}_1(x)] = \mathrm{E}[d\boldsymbol{\Phi}_2(x)] = \boldsymbol{I}_2 dx, \quad \mathrm{E}[d\boldsymbol{\Phi}_3(x)] = \boldsymbol{f}_{\mathrm{VAR}}(\pi x|\boldsymbol{B}, \boldsymbol{S})dx.$$

Furthermore, since by Lemma 2.9 it holds $\mathrm{Cov}[d\boldsymbol{\Phi}_i(x)] \propto \omega_i(x)\boldsymbol{\Sigma}_i(x) \otimes \boldsymbol{\Sigma}_i(x)$ (the proportionality being understood component-wise) it follows that the deviation of the increments of $\boldsymbol{\Phi}_1$ are homogeneous among x, whereas the deviation of the increments of $\boldsymbol{\Phi}_2$ is proportional to $\omega_2(x)^{-1}$, i.e. strictly increasing as x approaches 1, see Figure 3.1. The (cumulated) mean curve of the process $\boldsymbol{\Phi}_i$ is given as

$$\mathrm{E}[\boldsymbol{\Phi}_i([0, x])] = \frac{1}{2}\int_0^x \omega_i(t)\boldsymbol{\Sigma}_i(t)dt, \quad 0 \leq x \leq 1. \tag{3.31}$$

A visualization of the pointwise distribution of the $\boldsymbol{\Phi}_i$'s is shown in Figure 3.2. It can be seen that, while the mean curve of $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ coincide – the two processes clearly differ by the second order structure, since the increase of uncertainty is stronger at the right boundary of $\boldsymbol{\Phi}_2$, which is in line with the choice of $\omega_2(x)$. A higher truncation threshold of $L = 300$ is needed for the approximation of $\boldsymbol{\Phi}_2$, to render the strong inhomogeneity properly, whereas a value of $L = 60$ was sufficient for $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_3$.

A visualization of two $A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ processes with $\eta \neq 2$ is shown in Figure 3.3. Recall from Remark 2.10 that η is a regularity parameter of the underlying $A\Gamma$ distribution. With small values of η, more probability mass is put towards the boundary of $\mathbb{S}_d^+$ (favoring nearly-singular matrices). This behavior is mirrored in the corresponding $A\Gamma$-process. It can indeed be seen

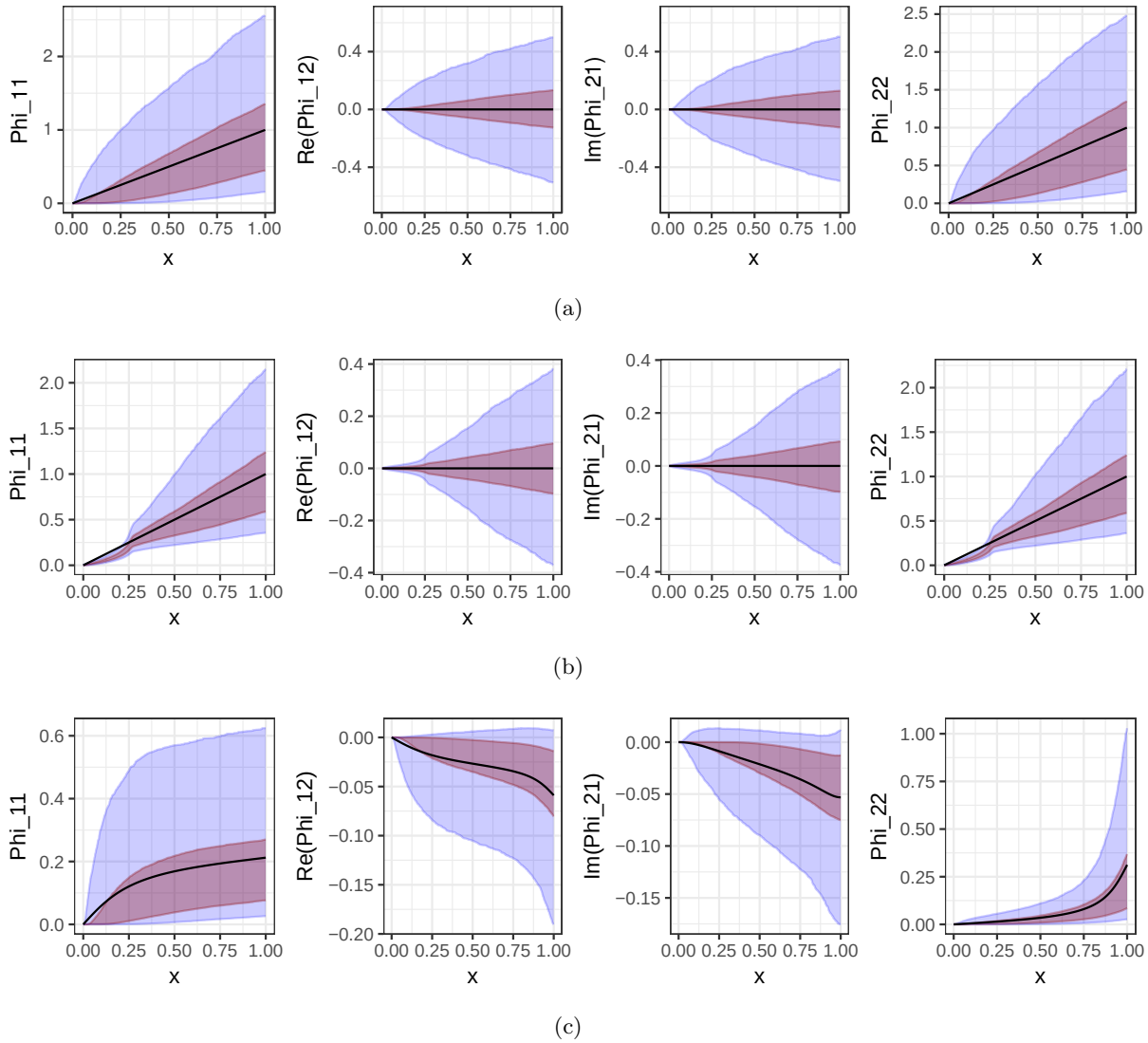
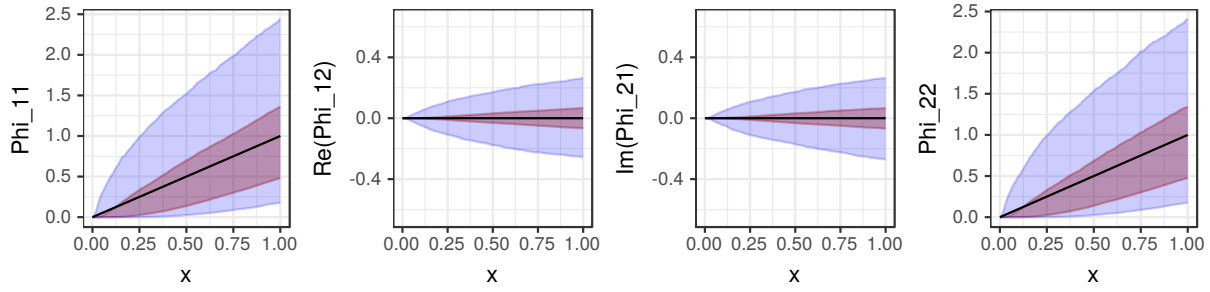
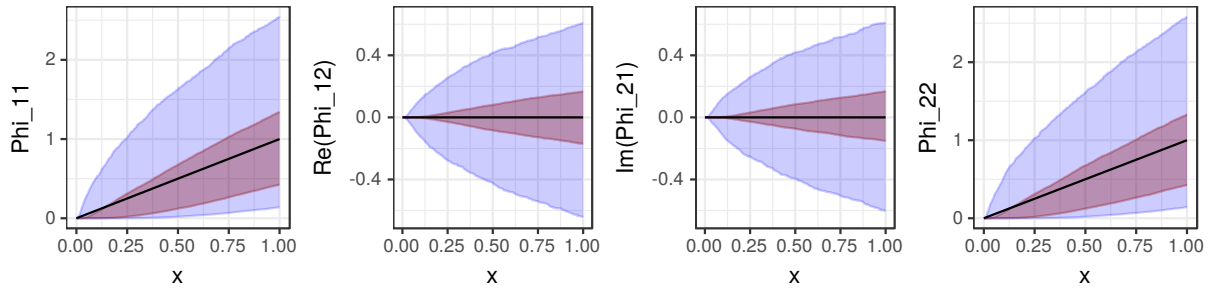


Figure 3.2.: Visualization of the pointwise (among $0 \leq x \leq 1$) distribution of the cumulated process $\mathbf{\Phi}_i([0, x])$ for (a) $\mathbf{\Phi}_1$ from (3.28), (b) $\mathbf{\Phi}_2$ from (3.29) and (c) $\mathbf{\Phi}_3$ from (3.30). The pointwise mean is drawn in solid black, the area between pointwise upper and lower quartile in shaded red and the area between pointwise 0.95 and 0.05 quantile in shaded blue. The visualizations are based on 10,000 samples respectively generated with Algorithm 1 with $L = 60$ ($L = 300$ for (b)).



(a)



(b)

Figure 3.3.: Visualization of the pointwise (among $0 \leq x \leq 1$) distribution of the cumulated process $\boldsymbol{\Phi}_1([0,x])$ with $\boldsymbol{\Phi}_1$ from (3.28) for (a) $\eta = 8$ and (b) $\eta = 1.1$

in the larger scale of the off-diagonal entries in Figure 3.3(b) in comparison to Figure 3.2(a). For $\eta = 8$, the $A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ distribution is more regular, yielding a smaller off-diagonal scale in Figure 3.3(a).

3.4.4. Numerical Simulation of Hpd Gamma Matrices

Algorithm 1 can readily be adapted to draw random $A\Gamma$ matrices. Indeed, for the homogeneous process case $\omega \equiv \omega_0 > 0$ and $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}_0 \in \mathcal{S}_d^+$, we recall from Remark 3.3 that $\boldsymbol{\Phi}([0,1]) \sim \mathrm{Ga}_{d\times d}(\alpha_{\eta_0,\boldsymbol{\Sigma}_0}, \beta_{\boldsymbol{\Sigma}_0})$. On the other hand, we get $\boldsymbol{\Phi}([0,1]) \approx \sum_{j=1}^L r_j \boldsymbol{U}_j$ from (3.25). Accordingly, Algorithm 4 describes how random samples from the $A\Gamma$ distribution can be drawn.

Algorithm 4: Generate a draw from the $A\Gamma(\eta_0, \omega_0, \boldsymbol{\Sigma}_0)$ distribution

input : $L \in \mathbb{N}$, $\eta_0 > d-1, \omega_0 > 0$, $\boldsymbol{\Sigma}_0 \in \mathcal{S}_d^+$

1 Draw $v_1, \ldots, v_L \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$ and compute $w_1, \ldots, w_L$ from (3.25);

2 Draw $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_L \overset{\text{iid}}{\sim} \alpha_{\eta_0,\boldsymbol{\Sigma}_0}$, e.g. with Algorithm 2;

3 **for** $j \leftarrow 1$ **to** L **do** Compute $r_j \leftarrow \rho^-(w_j | C_\alpha, \beta_{\boldsymbol{\Sigma}_0}(\boldsymbol{U}_j))$;

output: $\boldsymbol{Z} \leftarrow \sum_{j=1}^L r_j \boldsymbol{U}_j$

Algorithm 1 and Algorithm 4 are not restricted to the $A\Gamma$ distribution, but can be extended to other $\mathrm{Ga}_{d\times d}$ distributions, as long as the finite measure α on $\mathcal{X} \times \bar{\mathbb{S}}_d^+$ has a Lebesgue density g,

the local spherical mass function $\omega(x) := \int_{\mathbb{S}_d^+} g_\alpha(x, \boldsymbol{U}) d\boldsymbol{U}$ is available and there exists a method to draw random samples from (the probability distribution with Lebesgue density)

$$g(\boldsymbol{U}|x) := \frac{g(x, \boldsymbol{U})}{\int_{\mathbb{S}_d^+} g(x, \boldsymbol{V}) d\boldsymbol{V}}, \quad \boldsymbol{U} \in \mathbb{S}_d^+$$

for all $x \in \mathcal{X}$.

Part II.

Spectral Density Inference

4.

Whittle's Likelihood

4.1. Introduction

Consider a Gaussian stationary time series $\{\underline{Z}_t\}$ in $\mathbb{R}^d$ with $\mathrm{E}\underline{Z}_t = \underline{0}$. Denote by $\boldsymbol{\Gamma}(h) := \mathrm{E}[\underline{Z}_t \underline{Z}_{t+h}^T] \in \mathbb{R}^{d\times d}$ for $h \in \mathbb{Z}$ the autocovariance function of $\{\underline{Z}_t\}$. For a sample $\underline{z}_1, \ldots, \underline{z}_n$ of size n, denote by $\underline{z} := (\underline{z}_1^T, \ldots, \underline{z}_n^T)^T \in \mathbb{R}^{nd}$ the vector of stacked observations. The Gaussian likelihood P^n is given in terms of the Lebesgue density

$$p^n(\underline{z}) = \frac{1}{\sqrt{(2\pi)^{nd}|\boldsymbol{\Gamma}_{nd}|}} \exp\left(-\frac{1}{2}\underline{z}^T \boldsymbol{\Gamma}_{nd}^{-1} \underline{z}\right), \quad \underline{z} \in \mathbb{R}^d, \tag{4.1}$$

with covariance matrix $\boldsymbol{\Gamma}_{nd} = \mathrm{E}[\underline{Z}\underline{Z}^T] \in \mathbb{R}^{nd\times nd}$ given by

$$\boldsymbol{\Gamma}_{nd} = \begin{pmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}(1) & \ldots & \boldsymbol{\Gamma}(n-1) \\ \boldsymbol{\Gamma}(-1) & \boldsymbol{\Gamma}(0) & \ldots & \\ \vdots & & \ddots & \\ \boldsymbol{\Gamma}(-n+1) & \ldots & & \boldsymbol{\Gamma}(0) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}(1) & \ldots & \boldsymbol{\Gamma}(n-1) \\ \boldsymbol{\Gamma}(1)^T & \boldsymbol{\Gamma}(0) & \ldots & \\ \vdots & & \ddots & \\ \boldsymbol{\Gamma}(n-1)^T & \ldots & & \boldsymbol{\Gamma}(0) \end{pmatrix} \tag{4.2}$$

with the last equality following from $\boldsymbol{\Gamma}(h) = \boldsymbol{\Gamma}(-h)^T$ for $h \in \mathbb{Z}$. Since $\boldsymbol{\Gamma}_{nd}$ is consisting of n^2 blocks of size $d \times d$, we will also follow the notational convention of writing

$$\boldsymbol{\Gamma}_{nd} = \begin{pmatrix} \boldsymbol{\Gamma}_{nd}(1,1) & \ldots & \boldsymbol{\Gamma}_{nd}(1,n) \\ \vdots & & \vdots \\ \boldsymbol{\Gamma}_{nd}(n,1) & \ldots & \boldsymbol{\Gamma}_{nd}(n,n) \end{pmatrix} \tag{4.3}$$

with $\boldsymbol{\Gamma}_{nd}(j,k) = \boldsymbol{\Gamma}(-j+k)$. This notation will become particularly useful when comparing $\boldsymbol{\Gamma}_{nd}$ with other block matrices. There are several issues with the likelihood (4.1) that make it unappealing in computational applications. First of all, it relies on the inversion of the $nd \times nd$ covariance matrix $\boldsymbol{\Gamma}_{nd}$ – a task of considerable computational cost (with $O(d^3n^2)$ floating point operations, see Akaike (1973)), which is also prone to numerical instabilities even for moderate values of nd. Aside from this, and relevant in the context of spectral inference, the likelihood P^n depends on the spectral density $\boldsymbol{f}$ (given its existence) only indirectly through $\boldsymbol{\Gamma}_{nd}$ in terms of the Herglotz Lemma representation

$$\boldsymbol{\Gamma}(h) = \int_0^{2\pi} \boldsymbol{f}(\omega) \exp(-ih\omega) d\omega, \quad h \in \mathbb{Z}. \tag{4.4}$$

This indirect dependence on $\boldsymbol{f}$ comes at an additional computational cost of computing $\boldsymbol{\Gamma}(-n+1), \ldots, \boldsymbol{\Gamma}(n-1)$ from $\boldsymbol{f}$ through (4.4) to get $\boldsymbol{\Gamma}_{nd}$.

A different idea constitutes the usage of a frequency domain representation of the data itself. To elaborate, consider the *Fourier Coefficients*

$$\tilde{\underline{Z}}_j := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \underline{Z}_t \exp(-it\omega_j), \quad \omega_j := \frac{2\pi j}{n}, \quad j = 0, \ldots, \left\lfloor \frac{n}{2} \right\rfloor. \tag{4.5}$$

The values $\omega_0, \ldots, \omega_{\lfloor n/2 \rfloor}$ are called the *Fourier Frequencies*. We observe that $\tilde{\underline{Z}}_0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \underline{Z}_t$ corresponds to the (rescaled) sample mean and, for n even, $\tilde{\underline{Z}}_{n/2} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (-1)^t \underline{Z}_t$ to the (rescaled) alternating sample mean. There is a one-to-one relation between the Fourier Coefficients (which, in what follows, we will also equivalently call *frequency domain observations*) and the sample $\underline{Z}_1, \ldots, \underline{Z}_n$ (which we will also call *time domain observations*). In fact, it is well known that the Fourier inversion formula

$$\underline{Z}_t = \frac{1}{2\pi\sqrt{n}} \sum_{j=0}^{n-1} \tilde{\underline{Z}}_j \exp(it\omega_j), \quad t = 1, \ldots, n$$

holds, where for $j > \lfloor n/2 \rfloor$, the values of $\tilde{\underline{Z}}_j$ are defined as $\tilde{\underline{Z}}_j := \tilde{\underline{Z}}^*_{n-j}$ for n even and $\tilde{\underline{Z}}_j := \tilde{\underline{Z}}^*_{n-j+1}$ for n odd. Thus the Fourier coefficients contain the same amount of statistical information as the time domain observations. Asymptotically, the Fourier coefficients are in a certain sense independent and normally distributed. To elaborate, for $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$, denote by $CN_d(0, \boldsymbol{\Sigma})$ the complex multivariate normal distribution, as given by the Lebesgue density

$$p(\underline{z}) = \frac{1}{\pi^d |\boldsymbol{\Sigma}|} \exp\left(-\underline{z}^* \boldsymbol{\Sigma}^{-1} \underline{z}\right), \quad \underline{z} \in \mathbb{C}^d. \tag{4.6}$$

Then, as n tends to infinity, for all $m \in \mathbb{N}$ and all sequences $\omega_{l_{1,n}}, \ldots, \omega_{l_{m,n}}$ of Fourier frequencies with $\omega_{l_{j,n}} \to \omega_j \in [0, \pi]$ for $j = 1, \ldots, m$, the joint distribution of $(\tilde{\underline{Z}}_{l_{1,n}}, \ldots, \tilde{\underline{Z}}_{l_{1,n}})$ converges to the distribution of an m-vector of independent normals, with the following limiting distributions:

$$\begin{cases} N_d(0, 2\pi \boldsymbol{f}(0)), & \omega_j = 0, \\ CN_d(0, 2\pi \boldsymbol{f}(\omega_{j_i})), & \omega_j \in (0, \pi) \\ N_d(0, 2\pi \boldsymbol{f}(\pi)), & \omega_j = \pi. \end{cases}$$

This statement holds true for a wide class of time series models, even beyond Gaussianity, see e.g. Theorem 13 in Chapter 4 of Hannan (1970). For Gaussian time series, we will derive an even stronger result, involving measure contiguity rather than asymptotic distributional equality, in the upcoming Section 4.3. Closely related to the Fourier coefficients is the *periodogram* matrix $\tilde{\boldsymbol{J}}_j := \frac{1}{2\pi} \tilde{\underline{Z}}_j \tilde{\underline{Z}}_j^*$. The above results yield that (any finite selection of) the periodogram ordinates are asymptotically independent with asymptotic mean being equal to the spectral density matrix:

$$\mathrm{E}\tilde{\boldsymbol{J}}_j \approx \boldsymbol{f}(\omega_j), \quad j = 0, \ldots, \lfloor n/2 \rfloor. \tag{4.7}$$

Motivated by these considerations, Whittle's Likelihood P_W^n is defined to resemble the asymptotic distribution of the Fourier coefficients, i.e. in terms of the Lebesgue density

$$p_W^n(\tilde{\underline{z}}_0, \ldots, \tilde{\underline{z}}_{\lfloor n/2 \rfloor} | \boldsymbol{f}) = \prod_{j=0}^{\lfloor n/2 \rfloor} p_W^{j,n}(\tilde{\underline{z}}_j | \boldsymbol{f}) \tag{4.8}$$

with

$$p_W^{0,n}(\tilde{\underline{z}}_0) = \frac{1}{\sqrt{(2\pi)^d |2\pi \boldsymbol{f}(0)|}} \exp\left(-\frac{1}{4\pi} \tilde{\underline{z}}_0^T \boldsymbol{f}(0)^{-1} \tilde{\underline{z}}_0\right),$$
$$p_W^{j,n}(\tilde{\underline{z}}_j) = \frac{1}{\pi^d |2\pi \boldsymbol{f}(\omega_j)|} \exp\left(-\frac{1}{2\pi} \tilde{\underline{z}}_j^* \boldsymbol{f}(\omega_j)^{-1} \tilde{\underline{z}}_j\right), \quad j = 1, \ldots, \left\lceil \frac{n}{2} \right\rceil - 1,$$
$$p_W^{n/2,n}(\tilde{\underline{z}}_{n/2}) = \frac{1}{\sqrt{(2\pi)^d |2\pi \boldsymbol{f}(\pi)|}} \exp\left(-\frac{1}{4\pi} \tilde{\underline{z}}_{n/2}^T \boldsymbol{f}(\pi)^{-1} \tilde{\underline{z}}_{n/2}\right), \quad \text{for } n \text{ even.}$$

Whittle's Likelihood has several conceptual advantages over the full Gaussian likelihood P^n. Being a product likelihood, it comes with much less computational costs to evaluate (i.e. $O(nd^3)$), since only n $d \times d$ matrices have to be inverted rather than one $nd \times nd$ matrix. This also makes the evaluation much more numerically stable. Furthermore, since it depends directly on the spectral density $\boldsymbol{f}$, it lends itself naturally to spectral inference. Since it is only asymptotically justified, it may be noted that there are situations when the Whittle likelihood approximation tends to produce biased results, in particular in situations where the effective sample size is small and this may even be the case for moderate sample sizes if the auto-dependence in the time series is strong, see Contreras-Cristán et al. (2006). There exist methods to mitigate this effect to some extent (see Sykulski et al. (2016)) which are however beyond the scope of this work.

4.2. A Real Valued Formulation

We will sometimes work with a real-valued formulation of Whittle's likelihood, which will be described in the following. Let $N := \left\lceil \frac{n}{2} \right\rceil - 1$. Let

$$\tilde{\mathbb{D}}_n := \begin{cases} \mathbb{R}^d \times \mathbb{C}^{dN} \times \mathbb{R}^d, & n \text{ even}, \\ \mathbb{R}^d \times \mathbb{C}^{dN}, & n \text{ odd}, \end{cases}$$

and consider the multivariate Discrete Fourier Transform (mDFT) operator $\mathcal{F}_{nd} \colon \mathbb{R}^{n \times d} \to \tilde{\mathbb{D}}_n$, $\mathcal{F}_{nd}((\underline{Z}_1, \ldots, \underline{Z}_n)^T) = (\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2 \rfloor})$. We will first derive a real-valued version of $\mathcal{F}_{nd}$, which will be more convenient in many subsequent considerations. To elaborate, consider the isomorphism $\mathrm{vec}_{\mathbb{R}} \colon \tilde{\mathbb{D}}_n \to \mathbb{R}^{nd}$ defined by stacking real and imaginary parts of $(\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2 \rfloor}) \in \tilde{\mathbb{D}}_n$ suitably below each other:

$$\begin{aligned} \tilde{\underline{Z}} &:= \mathrm{vec}_{\mathbb{R}}(\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2 \rfloor}) \\ &= \begin{cases} \left(\tilde{\underline{Z}}_0^T, \sqrt{2} \begin{pmatrix} \Re \tilde{\underline{Z}}_1 \\ \Im \tilde{\underline{Z}}_1 \end{pmatrix}^T, \ldots, \sqrt{2} \begin{pmatrix} \Re \tilde{\underline{Z}}_N \\ \Im \tilde{\underline{Z}}_N \end{pmatrix}^T, \tilde{\underline{Z}}_{N+1}^T \right)^T, & n \text{ even}, \\ \left(\tilde{\underline{Z}}_0^T, \sqrt{2} \begin{pmatrix} \Re \tilde{\underline{Z}}_1 \\ \Im \tilde{\underline{Z}}_1 \end{pmatrix}^T, \ldots, \sqrt{2} \begin{pmatrix} \Re \tilde{\underline{Z}}_N \\ \Im \tilde{\underline{Z}}_N \end{pmatrix}^T \right)^T, & n \text{ odd}, \end{cases} \end{aligned} \tag{4.9}$$

recalling that $\Re \underline{Z}$ and $\Im \underline{Z}$ denote the real- and imaginary parts of a complex vector $\underline{Z}$. Denote by $\underline{Z} := (\underline{Z}_1^T, \ldots, \underline{Z}_n^T)^T \in \mathbb{R}^{nd}$ the column vector of stacked time domain observations. The

following result shows that the real-valued frequency domain vector $\underline{\tilde{Z}}$ from (4.9) can be obtained from $\underline{Z}$ by applying a linear, orthogonal transformation.

Lemma 4.1. *There exists an orthogonal matrix $\boldsymbol{F}_{nd} \in \mathbb{R}^{nd\times nd}$ such for all $\underline{Z}_1, \dots, \underline{Z}_n \in \mathbb{R}^d$ with $\underline{Z} := (\underline{Z}_1^T, \dots, \underline{Z}_n^T)^T \in \mathbb{R}^{nd}$ and $\underline{\tilde{Z}} = \mathrm{vec}_{\mathbb{R}}(\underline{\tilde{Z}}_0, \dots, \underline{\tilde{Z}}_{\lfloor n/2 \rfloor}) \in \mathbb{R}^{nd}$ it holds*

$$\underline{\tilde{Z}} = \boldsymbol{F}_{nd}\underline{Z}. \tag{4.10}$$

Proof. We start with the case $d = 1$. It is well known (see e.g. Section 10.1 in Brockwell and Davis (1991) or Section 2.1 in Kirch et al. (2017)) that in this case the transformation matrix $\boldsymbol{F}_n := \boldsymbol{F}_{n1}$ can be obtained as follows: With

$$\underline{e}_j := \frac{1}{\sqrt{n}}\left(e^{-2\pi ij/n}, e^{-4\pi ij/n}, \dots, e^{-2n\pi ij/n}\right)^T, \quad j = 0, \dots, \left\lfloor \frac{n}{2} \right\rfloor$$

and $\underline{c}_j := \sqrt{2}\Re\underline{e}_j$, $\underline{s}_j := \sqrt{2}\Im\underline{e}_j$, we define

$$\boldsymbol{F}_n := \begin{cases} (\underline{e}_0, \underline{c}_1, \underline{s}_1, \dots, \underline{c}_N, \underline{s}_N, \underline{e}_{n/2})^T, & n \text{ even}, \\ (\underline{e}_0, \underline{c}_1, \underline{s}_1, \dots, \underline{c}_N, \underline{s}_N)^T, & n \text{ odd}. \end{cases}$$

It is also well known that $\boldsymbol{F}_n$ is orthogonal. From $\boldsymbol{F}_n$ we can construct the transformation matrix $\boldsymbol{F}_{nd}$ for the case $d > 1$ as follows. Denote by $\boldsymbol{R}_{nd}$ the matrix fulfilling for all $X_1, \dots, X_{nd} \in \mathbb{R}$ the following equality:

$$\boldsymbol{R}_{nd}(X_1, X_2, \dots, X_{nd}) = (X_1, X_{d+1}, \dots, X_{(n-1)d+1}, X_2, X_{d+2}, \dots, X_{nd}), \tag{4.11}$$

i.e.

$$(\boldsymbol{R}_{nd}(X_1, \dots, X_{nd}))_{(r-1)n+i} = X_{(i-1)d+r} \tag{4.12}$$

for all $r = 1, \dots, d$ and $i = 1, \dots, n$. Being a permutation matrix, it follows that $\boldsymbol{R}_{nd}$ is orthogonal. Define the block diagonal matrix $\tilde{\boldsymbol{F}}_{nd} = \boldsymbol{I}_d \otimes \boldsymbol{F}_n$. Since $\boldsymbol{F}_n$ is the one-dimensional transformation matrix, the frequency domain vector of the j'th component is given by $(\tilde{Z}_{1,j}, \dots, \tilde{Z}_{n,j})^T = \boldsymbol{F}_n(Z_{1,j}, \dots, Z_{n,j})^T$ for $j = 1, \dots, d$, hence

$$\tilde{\boldsymbol{F}}_{nd}(Z_{1,1}, \dots, Z_{n,1}, Z_{1,2}, \dots, Z_{n,n})^T = (\tilde{Z}_{1,1}, \dots, \tilde{Z}_{n,1}, \tilde{Z}_{1,2}, \dots, \tilde{Z}_{n,d})^T. \tag{4.13}$$

Note also that $\tilde{\boldsymbol{F}}_{nd}$ is orthogonal, since, by Lemma B.1 in the Appendix,

$$(\boldsymbol{I}_d \otimes \boldsymbol{F}_n)(\boldsymbol{I}_d \otimes \boldsymbol{F}_n)^T = (\boldsymbol{I}_d \otimes \boldsymbol{F}_n)(\boldsymbol{I}_d \otimes \boldsymbol{F}_n^T) = \boldsymbol{I}_d \otimes (\boldsymbol{F}_n\boldsymbol{F}_n^T) = \boldsymbol{I}_d \otimes \boldsymbol{I}_n = \boldsymbol{I}_{nd}$$

(and $(\boldsymbol{I}_d \otimes \boldsymbol{F}_n)^T(\boldsymbol{I}_d \otimes \boldsymbol{F}_n) = \boldsymbol{I}_{nd}$ is shown analogously). On the other hand, the right hand side of (4.13) can also be obtained as $\boldsymbol{R}_{nd}\underline{\tilde{Z}}$, which is equivalent to $\underline{\tilde{Z}} = \boldsymbol{R}_{nd}^T\tilde{\boldsymbol{F}}_{nd}\boldsymbol{R}_{nd}\underline{Z}$. Thus the matrix

$$\boldsymbol{F}_{nd} := \boldsymbol{R}_{nd}^T\tilde{\boldsymbol{F}}_{nd}\boldsymbol{R}_{nd} = \boldsymbol{R}_{nd}^T(\boldsymbol{I}_d \otimes \boldsymbol{F}_n)\boldsymbol{R}_{nd} \tag{4.14}$$

fulfills (4.10) and noting that $\boldsymbol{F}_{nd}$ is orthogonal concludes the proof. □

The following algebra isomorphism describes the connection between the covariance matrix of a complex normal random vector and the covariance matrix of its real-valued counterpart.

Lemma 4.2. *Consider the mapping*

$$\mathcal{B}\colon \mathbb{C}^{d\times d} \to \mathbb{R}^{2d\times 2d}, \quad \mathcal{B}\boldsymbol{A} = \begin{pmatrix} \Re\boldsymbol{A} & -\Im\boldsymbol{A} \\ \Im\boldsymbol{A} & \Re\boldsymbol{A} \end{pmatrix}. \tag{4.15}$$

Then it holds

(a) $\mathcal{B}$ *is an algebra isomorphism.*

(b) $\boldsymbol{A} \in \mathbb{C}^{d\times d}$ *is Hermitian if and only if* $\mathcal{B}\boldsymbol{A}$ *is symmetric. In this case the absolute value of the determinant of* $\mathcal{B}\boldsymbol{A}$ *fulfills* $|\mathcal{B}\boldsymbol{A}| = |\boldsymbol{A}|^2$.

Proof. The mapping $\mathcal{B}$ has been considered in the proof of Theorem 13 in Chapter IV (*The Law of Large Numbers*) in Hannan (1970). The results are also derived there. □

The algebra isomorphism $\mathcal{B}$ establishes a direct link between the complex multivariate normal distribution and the multivariate normal distribution of double dimensionality, as the following result shows.

Lemma 4.3. *Let* $\underline{Z}$ *be a random element in* $\mathbb{C}^d$ *with mean zero and* $\mathrm{E}\left[\|\underline{Z}\|^2\right]$ *being finite. Denote by* $\boldsymbol{\Sigma} := \mathrm{E}\left[(\underline{Z} - \mathrm{E}\underline{Z})(\underline{Z} - \mathrm{E}\underline{Z})^*\right]$ *the covariance matrix of* $\underline{Z}$*. Assume that* $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$*. Consider the real-valued version* $\underline{X} := (\sqrt{2}\Re\underline{Z}^T, \sqrt{2}\Im\underline{Z}^T)^T \in \mathbb{R}^{2d}$ *of* $\underline{Z}$ *obtained by stacking (and rescaling) the imaginary part below the real part. Then* $\underline{Z} \sim CN_d(0, \boldsymbol{\Sigma})$ *if and only if* $\underline{X} \sim N_{2d}(0, \mathcal{B}\boldsymbol{\Sigma})$.

Proof. It is shown in the proof of Theorem 13 in Chapter IV in Hannan (1970) that for $\underline{z} \in \mathbb{C}^d$ and $\underline{x} := (\sqrt{2}\Re\underline{z}^T, \sqrt{2}\Im\underline{z}^T)^T \in \mathbb{R}^{2d}$, it holds

$$\underline{z}^*\boldsymbol{\Sigma}\underline{z} = \frac{1}{2}\underline{x}^T(\mathcal{B}\boldsymbol{\Sigma})\underline{x}.$$

Thus, denoting by $p_{\underline{Z}}$ and $p_{\underline{X}}$ the probability densities of $\underline{Z}$ and $\underline{X}$ respectively, this yields

$$p_{\underline{Z}}(\underline{z}) \propto \exp\left(-\underline{z}^*\boldsymbol{\Sigma}^{-1}\underline{z}\right) \iff p_{\underline{X}}(\underline{x}) \propto \exp\left(-\frac{1}{2}\underline{x}^T(\mathcal{B}\boldsymbol{\Sigma})^{-1}\underline{x}\right),$$

concluding the proof. □

As an immediate consequence of Lemma 4.3, the formula for Whittle's likelihood in terms of the real-valued frequency domain observations $\underline{\tilde{Z}}$ is given by the following Lebesgue density:

$$p_W^n(\underline{\tilde{z}}|\boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^{nd}|\boldsymbol{D}_{nd}|}} \exp\left(-\frac{1}{2}\underline{\tilde{z}}^T\boldsymbol{D}_{nd}^{-1}\underline{\tilde{z}}\right), \quad \underline{\tilde{z}} \in \mathbb{R}^{nd} \tag{4.16}$$

with the following $nd \times nd$ block diagonal matrix (note that the block sizes are visualized by the size of the surrounding squares, where the block size for $\boldsymbol{f}(0)$ and – if applicable – $\boldsymbol{f}(\pi)$ is

given by d and by $2d$ for all other blocks):

$$\boldsymbol{D}_{nd} = \boldsymbol{D}_{nd}[\boldsymbol{f}] := 2\pi \cdot \begin{bmatrix} \boldsymbol{f}(0) & & & & \\ & \mathcal{B}\boldsymbol{f}(\omega_1) & & & \\ & & \ddots & & \\ & & & \mathcal{B}\boldsymbol{f}(\omega_N) & \\ & & & & \boldsymbol{f}(\pi) \end{bmatrix}$$

for n even and

$$\boldsymbol{D}_{nd} = \boldsymbol{D}_{nd}[\boldsymbol{f}] := 2\pi \cdot \begin{bmatrix} \boldsymbol{f}(0) & & & \\ & \mathcal{B}\boldsymbol{f}(\omega_1) & & \\ & & \ddots & \\ & & & \mathcal{B}\boldsymbol{f}(\omega_N) \end{bmatrix}$$

for n odd.

4.3. Mutual Contiguity of Whittle's Likelihood and Full Gaussian Likelihood

Two sequences (P_n) and (Q_n) of measures on measurable spaces $\mathcal{X}_n$ are called *mutually contiguous*, if for every sequence (A_n) of measurable sets it holds: $P_n(A_n) \to 0$ if and only if $Q_n(A_n) \to 0$. In this case, convergence in probability under P_n is equivalent to convergence in probability under Q_n. In this section we will show that for stationary multivariate Gaussian time series, the full Gaussian likelihood and Whittle's likelihood are mutually contiguous. We will make the following assumption on the underlying true spectral density matrix $\boldsymbol{f}$:

Assumption $\boldsymbol{f}$1. *The eigenvalues of $\boldsymbol{f}(\omega)$ are uniformly bounded and uniformly bounded away from 0. That is, there exist positive constants b_0, b_1 such that*

$$\lambda_{\min}(\boldsymbol{f}(\omega)) \geq b_0, \quad \lambda_{\max}(\boldsymbol{f}(\omega)) \leq b_1, \quad 0 \leq \omega \leq \pi.$$

Assumption $\boldsymbol{f}$2. *The autocovariance function* $\boldsymbol{\Gamma}(h) = \int_0^{2\pi} \boldsymbol{f}(\omega)\exp(ih\omega)d\omega$ *of* $\boldsymbol{f}$ *fulfills*

$$\sum_{h\in\mathbb{Z}} \|\boldsymbol{\Gamma}(h)\||h|^a < \infty,$$

for some $a > 1$.

In Choudhuri et al. (2004b), mutual contiguity of Whittle's Likelihood P_W^n and the full Gaussian Likelihood P^n has been established in the case $d = 1$ under Assumptions $\boldsymbol{f}$1-$\boldsymbol{f}$2. A related result (also in the univariate case) has been derived in Cai et al. (2013), where the authors showed an asymptotic equivalence result (in the Le Cam sense) for estimation of the log spectral density.

Let us briefly discuss Assumptions $\boldsymbol{f}$1–$\boldsymbol{f}$2. An application of Lemma B.15 in the Appendix yields that Assumption $\boldsymbol{f}$2 implies that $\boldsymbol{f}$ is continuously differentiable with derivative being Hölder of order $a - 1 > 0$. In fact, this already implies (since $\boldsymbol{f}$ is a continuous function defined on a compact interval) that the part $\lambda_{\max}(\boldsymbol{f}(\omega)) \leq b_1$ of Assumption $\boldsymbol{f}$1 holds for some positive constant b_1. The additional assumption $\lambda_{\min}(\boldsymbol{f}(\omega)) \geq b_0$ is in this case equivalent to $|\boldsymbol{f}(\omega)| \neq 0$. The boundedness of the eigenvalues away from 0 has a statistical interpretation in terms of decay of so-called linear dependence coefficients, see Theorem 1.7 in Bradley (2002), whereas the Assumption $\boldsymbol{f}$2 concerns the decay of the autocovariance function.

It is of course possible to conduct inference under more general assumptions. As an example, in Rousseau et al. (2012), a Bayesian nonparametric approach to spectral inference for *long-memory* processes is presented, allowing for spectral densities having a pole (of certain order) as $\omega \to 0$. However, it is known that for such models the Whittle approximation P_W^n is no longer justified (see Section 3 in Robinson (1995)) and these models have to be treated differently. The following Theorem is the main result of this section.

Theorem 4.4 (Contiguity of Gaussian and Whittle measure in multivariate case)**.** *Let* $\{\underline{Z}_t\}$ *be an* $\mathbb{R}^d$*-valued time series with mean zero, spectral density matrix* $\boldsymbol{f}$ *and autocovariance function* $\boldsymbol{\Gamma}(h)$, $h \in \mathbb{Z}$. *Consider the multivariate Discrete Fourier Transform (mDFT) coefficients* $\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2\rfloor}$ *from* (4.5). *Denote by* P^n *the joint distribution of* $(\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2\rfloor})$. *Denote by* $P_W^n = P_W^n(\cdot|\boldsymbol{f})$ *the joint distribution of* $(\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2\rfloor})$ *under the Whittle likelihood, that is,* $\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2\rfloor}$ *are independent and*

$$\begin{aligned}
\tilde{\underline{Z}}_0 &\overset{P_W^n}{\sim} N_d(\underline{0}, 2\pi\boldsymbol{f}(0)),\\
\tilde{\underline{Z}}_j &\overset{P_W^n}{\sim} CN_d(\underline{0}, 2\pi\boldsymbol{f}(\omega_j)), \quad \text{for } j = 1, \ldots, \left\lceil \frac{n}{2} \right\rceil - 1,\\
\tilde{\underline{Z}}_{n/2} &\overset{P_W^n}{\sim} N_d(\underline{0}, 2\pi\boldsymbol{f}(\pi)), \quad \text{if } n \text{ is even.}
\end{aligned}$$

Let $\boldsymbol{f}$ *and* $\boldsymbol{\Gamma}$ *fulfill assumptions* $\boldsymbol{f}$*1-*$\boldsymbol{f}$*2. Then* P^n *and* P_W^n *are mutually contiguous.*

We have the following immediate corollary from Theorem 4.4.

Corollary 4.5. *Let the assumptions of Theorem 4.4 be fulfilled. Denote by* $\tilde{P}^n$ *the joint distribution of* $(\tilde{\underline{Z}}_1, \ldots, \tilde{\underline{Z}}_{\lceil n/2\rceil - 1})$. *Denote by* $\tilde{P}_W^n$ *the joint distribution of* $\lceil n/2\rceil - 1$ *independent complex* d*-variate normal random vectors with mean zero and covariance matrix* $2\pi\boldsymbol{f}(\omega_j)$, $j = 1, \ldots, \lceil n/2\rceil - 1$. *Then* $\tilde{P}^n$ *and* $\tilde{P}_W^n$ *are mutually contiguous.*

Despite being a very powerful result, the applicability of Theorem 4.4 is restricted to Gaussian time series and the contiguity result does not hold for non-Gaussian time series. As a very simple example, consider $Z_1, \dots, Z_n \overset{\text{iid}}{\sim} P$ with a univariate distribution P having mean zero and variance σ^2. In this case, the spectral density is $f(\omega) \equiv \frac{\sigma^2}{2\pi}$ for $0 \leq \omega \leq \pi$. Consider the sample excess kurtosis

$$\hat{\kappa}_n := \frac{1}{n\hat{\sigma}_n^4} \sum_{j=1}^{n} (Z_j - \hat{\mu}_n)^4 - 3, \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^{n} (Z_j - \hat{\mu}_n)^2, \quad \hat{\mu}_n = \frac{1}{n} \sum_{j=1}^{n} Z_j.$$

It is known that (under regularity assumptions on the existence of moments), $\hat{\kappa}_n$ is a consistent estimator of the excess kurtosis $\kappa = \frac{\mathrm{E}Z_1^4}{\sigma^4} - 3$, a result that in fact holds true under much more general conditions (see Theorem 3 in Bai and Ng (2005)). If P is leptokurtic (e.g. Student t, Laplace) then $\kappa > 0$ and if P is platykurtic (e.g. uniform) then $\kappa < 0$ and in these cases it holds $\hat{\kappa}_n \overset{P^n}{\to} \kappa \neq 0$. Under $P_W^n = P_W^n(\cdot|f)$ however, $Z_1, \dots, Z_n$ are iid Gaussian (see Section 2.1 in Kirch et al. (2017)) with excess kurtosis zero, such that $\hat{\kappa}_n \overset{P_W^n}{\to} 0$. This shows that P^n and P_W^n cannot be mutually contiguous.

The *boundary frequencies* $\omega_0 = 0$ and $\omega_{n/2} = \pi$ (the latter occurring for n even) have to be taken into account or left away consistently to get mutual contiguity. As an example, under the assumptions of Theorem 4.4, denote by P^n the full Gaussian likelihood in the frequency domain, i.e. the joint distribution of $(\underline{\tilde{Z}}_0, \dots, \underline{\tilde{Z}}_{\lfloor n/2 \rfloor})$, and by $\tilde{P}^n$ the sample mean centered (and sample alternating mean centered) version thereof, i.e. the joint distribution of $(\underline{\tilde{Z}}_1, \dots, \underline{\tilde{Z}}_{\lceil n/2 \rceil - 1})$ from Corollary 4.5. Then P^n and $\tilde{P}^n$ are not mutually contiguous, even though $\{\underline{Z}_t\}$ has mean zero. To see why this is the case, consider the distribution of $\sqrt{n}\bar{\underline{Z}}_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \underline{Z}_j$. Under P^n, $\sqrt{n}\bar{\underline{Z}}_n$ is asymptotically normally distributed, with mean zero and asymptotic covariance equal to $2\pi \boldsymbol{f}(0)$, see Section 11.2 in Brockwell and Davis (1991). Under $\tilde{P}^n$ however, it holds $\sqrt{n}\bar{\underline{Z}}_n = 0 \to 0$ as $n \to \infty$. Thus P^n and $\tilde{P}^n$ cannot be mutually contiguous and similarly, P^n and $\tilde{P}_W^n$ cannot be mutually contiguous.

The proof of Theorem 4.4 will be presented in the upcoming Section 4.3.3. For the proof, a version of Szegö's strong limit theorem is needed. This will be presented and proven in the upcoming Section 4.3.1. Furthermore, a result on the convergence speed of frequency domain covariance matrices is needed, which will be discussed in Section 4.3.2.

4.3.1. Szegö's Strong Limit Theorem

Being a classical result in matrix theory, Szegö's strong limit theorem relates the asymptotic determinant of the time-domain covariance matrix with the integrated log spectral density. In the multivariate setting, the covariance matrix is of block Toeplitz structure (see (4.2)) and the spectral density function is matrix-valued. The result has been generalized to this setting in Widom (1974) and Widom (1975) (see also Widom (1976)). Although the result can be stated in more generality, the following formulation is sufficient for our purposes.

Theorem 4.6 (Szegö's strong limit theorem for Block Toeplitz matrices)**.** *Let* $\boldsymbol{f}\colon [0, 2\pi] \to \mathcal{S}_d^+$ *be measurable and fulfill Assumptions* ***f*** *1-****f*** *2. For* $n > 0$*, define the* $nd \times nd$ *Block Toeplitz*

matrix generated by $\boldsymbol{f}$ *as*

$$\boldsymbol{T}_n := (\boldsymbol{\Gamma}(-i+j))_{i,j=0}^{n-1} = \begin{pmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}(1) & \cdots & \boldsymbol{\Gamma}(n-1) \\ \boldsymbol{\Gamma}(-1) & \boldsymbol{\Gamma}(0) & \cdots & \\ \vdots & & \ddots & \\ \boldsymbol{\Gamma}(-n+1) & \cdots & & \boldsymbol{\Gamma}(0) \end{pmatrix}. \tag{4.17}$$

with $\boldsymbol{\Gamma}(h)$ *for* $h \in \mathbb{Z}$ *as in Assumption* **f2**.

(a) Let $G := \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \log|2\pi \boldsymbol{f}(\omega)|d\omega\right) > 0$. *Then there exists a positive constant* E, *such that*

$$\lim_{n\to\infty} \frac{|\boldsymbol{T}_n|}{G^n} = E. \tag{4.18}$$

(b) Denote by $\boldsymbol{T}_n^-$ *the* $nd \times nd$ *block Toeplitz matrix generated by the function* $\omega \mapsto \boldsymbol{f}(\omega)^{-1}$, *i.e.* $\boldsymbol{T}_n^- := (\boldsymbol{\Gamma}^-(-i+j))_{i,j=0}^{n-1}$ *with* $\boldsymbol{\Gamma}^-(h) = \int_0^{2\pi} \boldsymbol{f}(\omega)^{-1}\exp(ih\omega)d\omega$ *for* $h \in \mathbb{Z}$. *Then*

$$\|\boldsymbol{I}_{nd} - \boldsymbol{T}_n\boldsymbol{T}_n^-\| = O(1), \quad \text{as } n \to \infty, \tag{4.19}$$

where $\boldsymbol{I}_{nd}$ *denotes the* $nd \times nd$*-dimensional identity matrix.*

The remainder of this Section is devoted to the proof of Theorem 4.6. First, we need some background from the theory of Toeplitz operators on Banach spaces. We start with the definition of an infinite-dimensional version of the block Toeplitz matrices $\boldsymbol{T}_n$. Denote by $\mathbb{L}_d^2$ the complex Hilbert space of $\mathbb{C}^d$-valued functions on the interval $[0, 2\pi]$ with square integrable components, endowed with the inner product $\mathbb{L}_d^2 \ni \phi, \psi \mapsto \langle \underline{\phi}, \underline{\psi}\rangle := \int_0^{2\pi} \underline{\phi}(x)^* \underline{\psi}(x)dx$. Recall that any $\phi \in \mathbb{L}_d^2$ can be written in terms of its Fourier series

$$\underline{\phi} = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \underline{\phi}_h \exp(-ih\cdot), \quad \underline{\phi}_h = \int_0^{2\pi} \underline{\phi}(x)\exp(ihx)dx, \quad h \in \mathbb{Z},$$

the series being convergent in $\mathbb{L}_d^2$. Denote by $\mathbb{H}_d^2$ the *Hardy Space* on $[0, 2\pi]$, i.e. the space of all $\underline{\phi} \in \mathbb{L}_d^2$ such that $\underline{\phi}_h = 0$ holds for all $h < 0$. Note that $\mathbb{H}_d^2$ is a closed subspace of $\mathbb{L}_d^2$. Denote by Π_d the projection of $\mathbb{L}_d^2$ onto $\mathbb{H}_d^2$. The matrix-vector multiplication with $\boldsymbol{f}$ (from left) can be conceived as a bounded linear operator $\mathbb{L}_d^2 \ni \underline{\phi} \mapsto \boldsymbol{f}\underline{\phi} \in \mathbb{L}_d^2$. The *Toeplitz Operator* on $\mathbb{H}_d^2$ generated by $\boldsymbol{f}$ is defined as

$$T\colon \mathbb{H}_d^2 \to \mathbb{H}_d^2, \quad T\underline{\phi} := \Pi_d(\boldsymbol{f}\underline{\phi}). \tag{4.20}$$

The operator T can be thought of as an infinite-dimensional version of the matrix $\boldsymbol{T}_n$. Indeed, the latter can as well be conceived as a bounded linear operator acting on the Hardy Space $\mathbb{H}_d^2$ through matrix multiplication with the vector of the first n Fourier coefficients $\boldsymbol{\phi}_n := (\underline{\phi}_0^T, \ldots, \underline{\phi}_{n-1}^T)^T \in \mathbb{C}^{nd}$:

$$T_n\colon \mathbb{H}_d^2 \to \mathbb{H}_d^2, \quad (\underline{\phi}_h)_{h\geq 0} \mapsto ((T_n\underline{\phi})_h)_{h\geq 0} = (\boldsymbol{T}_n\boldsymbol{\phi}_n, 0, 0, \ldots). \tag{4.21}$$

In Section 7.5 in Böttcher (2006), it is shown that

$$\|T_n - T\| \to 0, \quad n \to \infty \tag{4.22}$$

holds, where $\|\cdot\|$ denotes the operator norm.

Denote by $\mathbb{L}^\infty_{d\times d}$ the Banach algebra of $\mathbb{C}^{d\times d}$-valued functions on $[0,2\pi]$ with bounded components, endowed with pointwise matrix-matrix multiplication and the maximum Frobenius norm $\mathbb{L}^\infty_{d\times d} \ni \boldsymbol{g} \mapsto \|\boldsymbol{g}\|_{F,\infty} := \sup_{0\le\omega\le\pi} \|\boldsymbol{g}(\omega)\|$. Define the *Besov Space* $\mathbb{B}_{d\times d}$ as the space of $\mathbb{C}^{d\times d}$-valued functions on $[0,2\pi]$ with components belonging to

$$\left\{ g \in \mathbb{L}^2 \colon \sum_{h\in\mathbb{Z}} |h||g_h|^2 < \infty, \text{ with } g_h = \int_0^{2\pi} g(\omega)\exp(ih\omega)d\omega\rangle \right\}. \tag{4.23}$$

Finally, we define the *Krein Algebra* $\mathbb{K}_{d\times d}$ as

$$\mathbb{K}_{d\times d} := \left\{ \boldsymbol{f} \in \mathbb{L}^\infty_{d\times d} \colon (I_{d\times d} - \Pi_{d\times d})\boldsymbol{f} \in \mathbb{B}_{d\times d} \text{ and } \Pi_{d\times d}\boldsymbol{f} \in \mathbb{B}_{d\times d} \right\},$$

where $I_{d\times d}$ denotes the identity operator on $\mathbb{L}^\infty_{d\times d}$ and $\Pi_{d\times d}$ denotes the projection onto the closed subspace $\mathbb{H}^2_{d\times d}$ of $\mathbb{L}^2_{d\times d}$ (with $\mathbb{H}^2_{d\times d}$ being defined analogously to $\mathbb{H}^2_d$). We will use the following result to show Theorem 4.6.

Theorem 4.7 (Szegö's strong limit theorem, abstract formulation)**.** *Let $\boldsymbol{f} \in \mathbb{K}_{d\times d}$ and the Toeplitz operator T be defined as in (4.20). Let the Toeplitz matrices $\boldsymbol{T}_n$ be as in (4.17) and $G := \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \log|2\pi \boldsymbol{f}(\omega)|d\omega\right) > 0$. Assume that T is invertible. Then there exists a constant $E \neq 0$ such that*

$$\lim_{n\to\infty} \frac{|\boldsymbol{T}_n|}{G^n} = E.$$

Proof. See Theorem 10.30 in Böttcher (2006). □

For part (b) of Theorem 4.6, we will also need the following result.

Lemma 4.8. *Let $\boldsymbol{f}\colon [0,2\pi] \to \mathcal{S}_d^+$ such that $\boldsymbol{f} \in \mathbb{B}_{d\times d}$ and $\boldsymbol{f}^{-1} \in \mathbb{B}_{d\times d}$. Let the $nd \times nd$ block Toeplitz matrices $\boldsymbol{T}_n$ and $\boldsymbol{T}_n^-$ be defined as in Theorem 4.6. Then it holds*

$$\|\boldsymbol{I}_{nd} - \boldsymbol{T}_n\boldsymbol{T}_n^-\| = O(1), \quad \text{as } n \to \infty.$$

Proof. The proof for the univariate case $d = 1$ can be found in Lemma A1.4 in Dzhaparidze and Kotz (2012) and the proof for the multivariate case $d > 1$ follows along the same lines. □

Now we have all the tools at hand to show Theorem 4.6.

Proof (of Theorem 4.6). To derive part (a), we will verify that the assumptions of Theorem 4.7 are fulfilled. To this end, it suffices to verify that $\boldsymbol{f}$ belongs to $\mathbb{K}_{d\times d}$ and that the Toeplitz operator T from (4.20) is invertible. We first observe that each component of $\boldsymbol{\Gamma}$ fulfills $|\boldsymbol{\Gamma}(h)_{rs}| \le C|h|^{-a}$ for some universal constant $C > 0$, uniformly in the components $r, s = 1,\dots,d$ by Assumption ***f***2 and since $|\boldsymbol{\Gamma}(h)_{rs}| \le \|\boldsymbol{\Gamma}(h)\|$. In particular, since $a > 1$,

$$\sum_{h\ge 0} h\,|\boldsymbol{\Gamma}(h)_{rs}|^2 < \infty \quad \text{and} \quad \sum_{h<0} (-h)\,|\boldsymbol{\Gamma}(h)_{rs}|^2 < \infty, \tag{4.24}$$

which shows that $\Pi_{d\times d}\boldsymbol{f}$ and $(I_{d\times d}-\Pi_{d\times d})\boldsymbol{f}$ are both belonging to $\mathbb{B}_{d\times d}$. This shows $\boldsymbol{f}\in\mathbb{K}_{d\times d}$. For $\underline{0}\neq\underline{z}=(\underline{z}_0^T,\dots,\underline{z}_{n-1}^T)^T\in\mathbb{C}^{nd}$, it holds

$$\underline{z}^*\boldsymbol{T}_n\underline{z}=\sum_{j,k=0}^{n-1}\underline{z}_j^*\boldsymbol{\Gamma}(k-j)\underline{z}_k=\frac{1}{2\pi}\int_0^{2\pi}\tilde{\underline{z}}(\omega)^*\boldsymbol{f}(\omega)\tilde{\underline{z}}(\omega)d\omega,$$

with $\tilde{\underline{z}}(\omega)=\sum_{j=0}^{n-1}\underline{z}_j\exp(ij\omega)$. Since $\boldsymbol{f}(\omega)$ is positive definite for $0\leq\omega\leq2\pi$, this implies that $\boldsymbol{T}_n$ is positive definite (in particular invertible) for all n. By Assumption $\boldsymbol{f}1$, the eigenvalues of $\boldsymbol{f}(\omega)$ stay uniformly bounded away from 0 and so do the eigenvalues of $\boldsymbol{T}_n$ as $n\to\infty$. From (4.22), this concludes that T is invertible. Now an application of Theorem 4.7 yields (4.18) for a constant $E\neq0$. Recalling that $\boldsymbol{T}_n$ is positive definite for all n yields $|\boldsymbol{T}_n|>0$ for all n and since $G>0$, it holds $E>0$, concluding (a).

To show (b), we first observe that $\boldsymbol{f}\in\mathbb{B}_{d\times d}$ holds by (4.24). Using Lemma B.16 in the Appendix, we find that the function $\boldsymbol{f}^{-1}$ also fulfills Assumptions $\boldsymbol{f}1$-$\boldsymbol{f}2$ and hence $\boldsymbol{f}^{-1}\in\mathbb{B}_{d\times d}$ can be concluded by the same argument as for (4.24). The result now follows from Lemma 4.8. □

Remark. *The constant E in (4.18) is given by $E=\prod_{j\geq0}(1+\lambda_j(R))$, with $(\lambda_j(R))_j$ denoting the sequence of eigenvalues of the operator $R:=I-TT^-$ for the Toeplitz operator T from (4.20) (and T^- being defined analogously with $\boldsymbol{f}^{-1}$ instead of $\boldsymbol{f}$). Although there is no closed-form expression known for E for the case $d\geq2$, it is sufficient for our purpose to know $E>0$. See Section 1.6 and Remark 10.29 in Böttcher (2006) for further information.*

4.3.2. Convergence Rate of Frequency Domain Covariance Matrix

The following result quantifies the speed of convergence of the covariance matrix in the frequency domain. It is very similar in spirit to Theorem 1.1 in Hannan and Wahlberg (1989), where the result is shown in a complex valued formulation. See also Section 3 in Davies (1973), where a related result is shown with a slower rate of convergence but under more general assumptions.

Theorem 4.9. *Let $\{\underline{Z}_t\}$ be a stationary $\mathbb{R}^d$-valued time series with mean zero and spectral density matrix $\boldsymbol{f}$ fulfilling Assumptions $\boldsymbol{f}1$-$\boldsymbol{f}2$. Denote by $\boldsymbol{\Gamma}_{nd}\in\mathbb{R}^{nd\times nd}$ the time domain covariance matrix of $\underline{Z}_1,\dots,\underline{Z}_n$ from (4.2). Let $\boldsymbol{H}_{nd}:=\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}\boldsymbol{F}_{nd}^T-\boldsymbol{D}_{nd}$ with the block diagonal matrix $\boldsymbol{D}_{nd}$ from (4.16) and the Fourier transformation matrix $\boldsymbol{F}_{nd}$ from Lemma 4.1. Denote the $d\times d$-blocks of $\boldsymbol{H}_{nd}$ by $\boldsymbol{H}_{nd}(i,j)\in\mathbb{R}^{d\times d}$ for $i,j=1,\dots,n$ (as in (4.3)). Then it holds*

$$\sup_{i,j=1,\dots,n}\|\boldsymbol{H}_{nd}(i,j)\|\lesssim n^{-1}$$

for all $n\in\mathbb{N}$, with proportionality constants not depending on n.

The proof of Theorem 4.9 relies on an approximation of the covariance matrix $\boldsymbol{\Gamma}_{nd}$ by a so-called block circulant matrix. To elaborate, a matrix $\boldsymbol{M}\in\mathbb{R}^{nd\times nd}$ is called *block circulant* if there

exist $\boldsymbol{m}(0), \dots, \boldsymbol{m}(n-1) \in \mathbb{R}^{d\times d}$ such that

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{m}(0) & \boldsymbol{m}(1) & \dots & \boldsymbol{m}(n-1) \\ \boldsymbol{m}(n-1) & \boldsymbol{m}(0) & \dots & \boldsymbol{m}(n-2) \\ \boldsymbol{m}(n-2) & \boldsymbol{m}(n-1) & \dots & \boldsymbol{m}(n-3) \\ \vdots & \vdots & & \vdots \\ \boldsymbol{m}(1) & \boldsymbol{m}(2) & \dots & \boldsymbol{m}(0) \end{pmatrix}.$$

In the case $d = 1$, $\boldsymbol{M}$ is called *circulant*. There is a close connection between block circulant matrices and the multivariate Discrete Fourier transform. Indeed, the eigenvalues of a block circulant matrix $\boldsymbol{M}$ are related to the Fourier coefficients of $\boldsymbol{m}(0), \dots, \boldsymbol{m}(n-1)$ and $\boldsymbol{M}$ is block diagonalized by the Fourier basis (see part (b) of the upcoming proof of Theorem 4.9). In view of this insight, it is not too surprising that circulants play an important role for the asymptotic behavior of the frequency domain covariance matrix of a stationary time series. A comprehensive introduction into circulant matrices can be found in Gray (2006). The following proof follows the ideas from the univariate case, which is presented in Proposition 5.2 in Brockwell and Davis (1991).

Proof of Theorem 4.9. We will follow the notational convention from (4.3) of denoting the $d \times d$ blocks of any matrix $\boldsymbol{A} \in \mathbb{R}^{nd\times nd}$ by $\boldsymbol{A}(i,j) \in \mathbb{R}^{d\times d}$ for $i, j = 1, \dots, n$. Consider the following symmetric block circulant matrix in $\mathbb{R}^{nd\times nd}$:

$$\boldsymbol{C}_{nd} := \begin{pmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}(1) & \dots & \boldsymbol{\Gamma}\left(\lfloor n/2 \rfloor\right) & \boldsymbol{\Gamma}\left(\lceil n/2 \rceil - 1\right)^T & \dots & \boldsymbol{\Gamma}(2)^T & \boldsymbol{\Gamma}(1)^T \\ \boldsymbol{\Gamma}(1)^T & \boldsymbol{\Gamma}(0) & \dots & \boldsymbol{\Gamma}\left(\lfloor n/2 \rfloor - 1\right) & \boldsymbol{\Gamma}\left(\lfloor n/2 \rfloor\right) & \dots & \boldsymbol{\Gamma}(3)^T & \boldsymbol{\Gamma}(2)^T \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{\Gamma}(1) & \boldsymbol{\Gamma}(2) & \dots & \boldsymbol{\Gamma}\left(\lceil n/2 \rceil - 1\right)^T & \boldsymbol{\Gamma}\left(\lceil n/2 \rceil - 2\right)^T & \dots & \boldsymbol{\Gamma}(1)^T & \boldsymbol{\Gamma}(0)^T \end{pmatrix}.$$

We will show the claim by verifying the following three assertions:

(a) With $\boldsymbol{G}_{nd} := \boldsymbol{F}_{nd}(\boldsymbol{\Gamma}_{nd} - \boldsymbol{C}_{nd})\boldsymbol{F}_{nd}^T$ it holds

$$\sup_{i,j=1,\dots,n} \|\boldsymbol{G}_{nd}(i,j)\| \lesssim n^{-1}$$

for all $n \in \mathbb{N}$, with proportionality constants not depending on n.

(b) It holds $\boldsymbol{F}_{nd}\boldsymbol{C}_{nd}\boldsymbol{F}_{nd}^T = \boldsymbol{\Lambda}_{nd}$ with $\boldsymbol{\Lambda}_{nd} = \boldsymbol{D}_{nd}[\boldsymbol{f}_n]$, where

$$\boldsymbol{f}_n(\omega) := \frac{1}{2\pi} \sum_{|h| \le \lfloor n/2 \rfloor} \boldsymbol{\Gamma}(h) \exp(-ih\omega), \quad 0 \le \omega \le 2\pi,$$

denotes a truncated series representation of $\boldsymbol{f}$ and $\boldsymbol{D}_{nd}[\boldsymbol{f}_n]$ is the block diagonal matrix from (4.16) corresponding to $\boldsymbol{f}_n$.

(c) With $\boldsymbol{\Lambda}_{nd}$ from (b), it holds

$$\sup_{i,j=1,\dots,n} \|\boldsymbol{\Lambda}_{nd}(i,j) - \boldsymbol{D}_{nd}(i,j)\| \lesssim n^{-1}$$

for all n, with proportionality constants not depending on n.

Assertion (a) states that the full covariance matrix $\boldsymbol{\Gamma}_{nd}$ can be approximated (in the frequency domain) by the block circulant matrix $\boldsymbol{C}_{nd}$. Assertion (b) states that the frequency domain representation thereof is of block diagonal structure $\boldsymbol{\Lambda}_{nd}$, which in fact closely resembles the asymptotic frequency domain covariance structure (under Whittle's Likelihood) and this resembling is made mathematically rigorous in Assertion (c). Once we have shown (a)-(c), the assertion of the theorem readily follows from

$$\|\boldsymbol{H}_{nd}(i,j)\| = \|\boldsymbol{G}_{nd}(i,j) + \boldsymbol{\Lambda}_{nd}(i,j) - \boldsymbol{D}_{nd}(i,j)\| \leq \|\boldsymbol{G}_{nd}(i,j)\| + \|\boldsymbol{\Lambda}_{nd}(i,j) - \boldsymbol{D}_{nd}(i,j)\|.$$

To show (a), first note that for $1 \leq i,j \leq n$ it holds

$$\boldsymbol{G}_{nd}(i,j) = \sum_{k,l=1}^{n} \boldsymbol{F}_{nd}(i,k)(\boldsymbol{\Gamma}_{nd}(k,l) - \boldsymbol{C}_{nd}(k,l))\boldsymbol{F}_{nd}(j,l)^T.$$

Since the entries of $\boldsymbol{F}_{nd}$ are bounded in absolute value by $(2/n)^{1/2}$ (see the definition of $\boldsymbol{F}_{nd}$ in the proof of Lemma 4.1), it follows $\|\boldsymbol{F}_{nd}(i,j)\| \lesssim n^{-1/2}$ and hence, using $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$ (see Lemma B.4 in the Appendix)

$$\begin{aligned}
\|\boldsymbol{G}_{nd}(i,j)\| &\lesssim \frac{1}{n}\sum_{k,l=1}^{n} \|\boldsymbol{\Gamma}_{nd}(k,l) - \boldsymbol{C}_{nd}(k,l)\| \\
&= \frac{1}{n}\left(\sum_{m=1}^{\lceil n/2\rceil-1} m\left\|\boldsymbol{\Gamma}(n-m) - \boldsymbol{\Gamma}(m)^T\right\| + \sum_{m=1}^{\lceil n/2\rceil-1} m\left\|\boldsymbol{\Gamma}(n-m)^T - \boldsymbol{\Gamma}(m)\right\|\right) \\
&\lesssim \frac{1}{n}\left(\sum_{m=1}^{\lceil n/2\rceil-1} m\|\boldsymbol{\Gamma}(m)\| + N\sum_{l=n-\lceil n/2\rceil+1}^{n-1} \|\boldsymbol{\Gamma}(l)\|\right) \\
&\lesssim \frac{1}{n}\sum_{m=1}^{n} m\|\boldsymbol{\Gamma}(m)\| \lesssim \frac{1}{n}\sum_{m=1}^{\infty} m\|\boldsymbol{\Gamma}(m)\| \lesssim \frac{1}{n},
\end{aligned}$$

uniformly in i,j, where the last inequality follows from Assumption ***f*2**. This concludes (a).

Let us now prove (b) and investigate the frequency domain transformed version of $\boldsymbol{C}_{nd}$. Recalling the representation $\boldsymbol{F}_{nd} = \boldsymbol{R}_{nd}^T(\boldsymbol{I}_d \otimes \boldsymbol{F}_n)\boldsymbol{R}_{nd}$ of the Fourier transformation matrix from (4.14) with the permutation matrix $\boldsymbol{R}_{nd}$ from (4.11), part (b) is equivalent to

$$(\boldsymbol{I}_d \otimes \boldsymbol{F}_n)\tilde{\boldsymbol{C}}_{nd} = \tilde{\boldsymbol{\Lambda}}_{nd}(\boldsymbol{I}_d \otimes \boldsymbol{F}_n), \tag{4.25}$$

with $\tilde{\boldsymbol{C}}_{nd} = \boldsymbol{R}_{nd}\boldsymbol{C}_{nd}\boldsymbol{R}_{nd}^T$ and $\tilde{\boldsymbol{\Lambda}}_{nd} = \boldsymbol{R}_{nd}\boldsymbol{\Lambda}_{nd}\boldsymbol{R}_{nd}^T$. For a matrix $\boldsymbol{A} \in \mathbb{R}^{nd\times nd}$ with block components $\boldsymbol{A}(i,j) \in \mathbb{R}^{d\times d}$ for $i,j = 1,\dots,n$, denote by $\boldsymbol{A}^{(r,s)} := (\boldsymbol{A}(i,j)_{r,s})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ the matrix consisting of the (r,s)-th entry of all block components, for $r,s = 1,\dots,d$.

We will first investigate the structure of the transformation $\boldsymbol{A} \mapsto \boldsymbol{R}_{nd}\boldsymbol{A}\boldsymbol{R}_{nd}^T$. For this purpose, denote the elements of $\boldsymbol{A}$ by $a_{k,l}$ for $k,l = 1,\dots,nd$. Write $k = (r-1)n + i$ and $l = (s-1)n + j$ with $r,s \in \{1,\dots,d\}$ and $i,j \in \{1,\dots,n\}$. Since $\boldsymbol{R}_{nd}$ is a permutation matrix, the operation $\boldsymbol{A} \mapsto \boldsymbol{R}_{nd}\boldsymbol{A}$ only reorders the elements within the columns of $\boldsymbol{A}$ and similarly, the operation $\boldsymbol{A} \mapsto \boldsymbol{A}\boldsymbol{R}_{nd}^T$ only reorders within the rows. By (4.12), the (k,l)-th entry of $\boldsymbol{R}_{nd}\boldsymbol{A}\boldsymbol{R}_{nd}^T$ is

$$(\boldsymbol{R}_{nd}\boldsymbol{A}\boldsymbol{R}_{nd}^T)_{k,l} = \left((\boldsymbol{R}_{nd}\boldsymbol{A})_{k*}\boldsymbol{R}_{nd}^T\right)_l$$

with $(\boldsymbol{R}_{nd}\boldsymbol{A})_{k*} \in \mathbb{R}^{1\times nd}$ denoting the k-th row of $\boldsymbol{R}_{nd}\boldsymbol{A}$. Using (4.12), we find

$$(\boldsymbol{R}_{nd}\boldsymbol{A})_{k*} = \big(a_{(i-1)d+r,1}, \dots, a_{(i-1)d+r,nd}\big).$$

The l-th entry of $(\boldsymbol{R}_{nd}\boldsymbol{A})_{k*}\boldsymbol{R}_{nd}^T = [\boldsymbol{R}_{nd}(\boldsymbol{R}_{nd}\boldsymbol{A})_{k*}^T]^T$ is $a_{(i-1)d+r,(j-1)d+s}$ by (4.12). On the other hand, the (i,j)-th entry of $\boldsymbol{A}^{(r,s)}$ is given by $\boldsymbol{A}(i,j)_{r,s} = a_{(i-1)d+r,(j-1)d+s}$ by definition. We have just established the following equality:

$$\boldsymbol{R}_{nd}\boldsymbol{A}\boldsymbol{R}_{nd}^T = \boldsymbol{R}_{nd}\begin{pmatrix}\boldsymbol{A}(1,1) & \dots & \boldsymbol{A}(1,n)\\ \vdots & & \vdots \\ \boldsymbol{A}(n,1) & \dots & \boldsymbol{A}(n,n)\end{pmatrix}\boldsymbol{R}_{nd}^T = \begin{pmatrix}\boldsymbol{A}^{(1,1)} & \dots & \boldsymbol{A}^{(1,d)}\\ \vdots & & \vdots \\ \boldsymbol{A}^{(d,1)} & \dots & \boldsymbol{A}^{(d,d)}\end{pmatrix}.$$

This shows that by multiplication from left and right, the permutation matrix $\boldsymbol{R}_{nd}$ reorders the elements of $\boldsymbol{A}$ from a *d-nesting* to an *n-nesting*. Thus (4.25) is equivalent to

$$\begin{pmatrix}\boldsymbol{F}_n\boldsymbol{C}_{nd}^{(1,1)} & \dots & \boldsymbol{F}_n\boldsymbol{C}_{nd}^{(1,d)}\\ \vdots & & \vdots \\ \boldsymbol{F}_n\boldsymbol{C}_{nd}^{(d,1)} & \dots & \boldsymbol{F}_n\boldsymbol{C}_{nd}^{(d,d)}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Lambda}_{nd}^{(1,1)}\boldsymbol{F}_n & \dots & \boldsymbol{\Lambda}_{nd}^{(1,d)}\boldsymbol{F}_n\\ \vdots & & \vdots \\ \boldsymbol{\Lambda}_{nd}^{(d,1)}\boldsymbol{F}_n & \dots & \boldsymbol{\Lambda}_{nd}^{(d,d)}\boldsymbol{F}_n,\end{pmatrix}. \tag{4.26}$$

Note that the matrices $\boldsymbol{C}_{nd}^{(r,s)} \in \mathbb{R}^{n\times n}$ are circulant for $1 \le r,s \le d$ and symmetric and circulant for $r = s$. Denoting the entries of $\boldsymbol{f}_n(\omega) \in \mathcal{S}_d^+$ by $f_n^{(r,s)}(\omega)$ for $r,s = 1,\dots,d$, it follows from the structure of $\boldsymbol{D}_{nd}$ (as defined after (4.16)) that

$$\boldsymbol{\Lambda}_{nd}^{(r,r)} = 2\pi \operatorname{diag}\Big(f_n^{(r,r)}(0), f_n^{(r,r)}(\omega_1), f_n^{(r,r)}(\omega_1), \dots, f_n^{(r,r)}(\omega_{\lfloor n/2\rfloor})\Big)$$

as well as, for $r \neq s$ and n even,

$$\boldsymbol{\Lambda}_{nd}^{(r,s)} = 2\pi\begin{pmatrix} f_n^{(r,s)}(0) & & & & \\ & \Re f_n^{(r,s)}(\omega_1) & -\Im f_n^{(r,s)}(\omega_1) & & \\ & \Im f_n^{(r,s)}(\omega_1) & \Re f_n^{(r,s)}(\omega_1) & & \\ & & & \ddots & \\ & & & & f_n^{(r,s)}(\pi)\end{pmatrix}$$

and for n odd,

$$\boldsymbol{\Lambda}_{nd}^{(r,s)} = 2\pi\begin{pmatrix} f_n^{(r,s)}(0) & & & & & \\ & \Re f_n^{(r,s)}(\omega_1) & -\Im f_n^{(r,s)}(\omega_1) & & & \\ & \Im f_n^{(r,s)}(\omega_1) & \Re f_n^{(r,s)}(\omega_1) & & & \\ & & & \ddots & & \\ & & & & \Re f_n^{(r,s)}(\omega_{\lfloor n/2\rfloor}) & -\Im f_n^{(r,s)}(\omega_{\lfloor n/2\rfloor}) \\ & & & & \Im f_n^{(r,s)}(\omega_{\lfloor n/2\rfloor}) & \Re f_n^{(r,s)}(\omega_{\lfloor n/2\rfloor})\end{pmatrix}.$$

It is well known (see e.g. Proposition 4.5.1 in Brockwell and Davis (1991) or Theorem 3.1 in Gray (2006)) that $\boldsymbol{F}_n\boldsymbol{C}_{nd}^{(r,s)} = \boldsymbol{\Lambda}_{nd}^{(r,s)}\boldsymbol{F}_n$ holds for $r,s = 1,\dots,d$, concluding representation (4.26) and in particular (b).

To derive (c), we use $\boldsymbol{f}(\omega) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}} \boldsymbol{\Gamma}(h)\exp(-ih\omega)$ to obtain

$$n\|\boldsymbol{f}_n(\omega) - \boldsymbol{f}(\omega)\| \leq \frac{n}{2\pi} \sum_{h>\lfloor n/2\rfloor} \|\boldsymbol{\Gamma}(h)\| \lesssim \sum_{h>\lfloor n/2\rfloor} |h|\|\boldsymbol{\Gamma}(h)\|,$$

which is bounded from above by a positive constant by Assumption ($\boldsymbol{f}2$). This shows $\|\boldsymbol{f}_n(\omega) - \boldsymbol{f}(\omega)\| \lesssim n^{-1}$ uniformly for $0 \leq \omega \leq 2\pi$, yielding $\|\boldsymbol{\Lambda}_{nd}(i,j) - \boldsymbol{D}_{nd}(i,j)\| \lesssim n^{-1}$ uniformly for $i,j = 1,\ldots,n$, which concludes (c). □

4.3.3. Proof of Contiguity

The main idea to establish contiguity is summarized in the following Lemma, which gives sufficient conditions for mutual contiguity and is well known in the literature. Since there was no proof readily available, it is presented for the sake of completeness.

Lemma 4.10. *Let $\{P_n\},\{Q_n\}$ be sequences of probability measures, with P_n and Q_n being defined on a measurable space $\mathcal{X}_n$, such that P_n and Q_n are mutually absolutely continuous for all n. Let $\Lambda_n := \log\frac{dP_n}{dQ_n}$ be the log Radon-Nikodym derivative. Assume that $\{E\Lambda_n\}$ and $\{\mathrm{Var}\Lambda_n\}$ are bounded sequences under both P_n and Q_n. Then P_n and Q_n are mutually contiguous.*

Proof. We first show that under the assumption of mutually absolutely continuity, it suffices that Λ_n is tight under both P_n and Q_n in order to obtain mutual contiguity. To see this, we employ the following characterization of contiguity from Le Cam's first lemma (see Lemma 6.4 in van der Vaart (2000)):

$$Q_n \triangleleft P_n \iff \text{If for a subsequence } m = m(n) : \frac{dP_m}{dQ_m} \overset{Q_m}{\to} U, \text{ then } P(U>0) = 1, \tag{4.27}$$

where $Q_n \triangleleft P_n$ is the notation for Q_n being contiguous with respect to P_n, and $A_n \overset{Q_n}{\to} B$ means that the sequence of random variables A_n converges weakly under Q_n to B.

We now show that tightness of Λ_n implies contiguity. To do so, assume that Λ_n is tight under Q_n. Take any subsequence $m = m(n)$ such that $\frac{dP_m}{dQ_m} \overset{Q_m}{\to} U$ holds for some random variable U. Using the tightness of Λ_n, there exists a subsequence $\tilde{m}(n)$ of $m(n)$ such that $\Lambda_{\tilde{m}} \overset{Q_{\tilde{m}}}{\to} V$ holds for some random variable V. An application of the continuous mapping theorem reveals $U = \exp(V)$ and hence $P(U>0) = P(V > -\infty) = 1$. By characterization (4.27), this yields $Q_n \triangleleft P_n$. With the same argument for $\frac{dQ_m}{dP_m} = (\frac{dP_m}{dQ_m})^{-1}$ and using the tightness of $-\Lambda_m$ (which clearly follows from the tightness of Λ_m), we also obtain $P_n \triangleleft Q_n$, i.e. mutual contiguity of P_n and Q_n.

It thus remains to show that boundedness of $\mathrm{E}\Lambda_n$ and $\mathrm{Var}\Lambda_n$ implies tightness of Λ_n under both P_n and Q_n. To do so, denote $\mu_{\max} := \sup_n |\mathrm{E}_{P_n}\Lambda_n| < \infty$ and $\sigma^2_{\max} := \sup_n \mathrm{Var}_{P_n}\Lambda_n < \infty$. Let $\varepsilon > 0$ and choose the compact interval

$$K_\varepsilon := \left[-\mu_{\max} - \frac{\sigma_{\max}}{\sqrt{\varepsilon}}, \mu_{\max} + \frac{\sigma_{\max}}{\sqrt{\varepsilon}}\right].$$

Then, with

$$K_{P_n} := \left[\mathrm{E}_{P_n}\Lambda_n - \frac{\sqrt{\mathrm{Var}_{P_n}\Lambda_n}}{\sqrt{\varepsilon}}, \mathrm{E}_{P_n}\Lambda_n + \frac{\sqrt{\mathrm{Var}_{P_n}\Lambda_n}}{\sqrt{\varepsilon}}\right]$$

it holds $K_{P_n} \subset K_\varepsilon$ for all n and an application of Chebyshev's inequality yields $1 - P_n(\Lambda_n \in K_\varepsilon) \leq 1 - P_n(\Lambda_n \in K_{P_n}) \leq \varepsilon$ for all n, i.e. Λ_n is tight under P_n. By the same argument, we also obtain that Λ_n is tight under Q_n if $|\mathrm{E}_{Q_n}\Lambda_n|$ and $\mathrm{Var}_{Q_n}|\Lambda_n|$ are bounded. □

Now we can prove the main result of this section.

Proof of Theorem 4.4. The proof follows the ideas of the proof for the univariate case from Choudhuri et al. (2004b). Some of the arguments from there will be repeated here for the sake of completeness and readability. Similar to the real-valued frequency domain formulation of Whittle's Likelihood P_W^n from (4.16), the true Gaussian likelihood can also be expressed under the real-valued frequency domain transformation, in terms of the Lebesgue density

$$p^n(\tilde{\underline{z}}) = \frac{1}{\sqrt{(2\pi)^{nd}|\boldsymbol{\Gamma}_{nd}|}} \exp\left\{-\frac{1}{2}\tilde{\underline{z}}^T(\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}\boldsymbol{F}_{nd}^T)^{-1}\tilde{\underline{z}}\right\}, \quad \tilde{\underline{z}} \in \mathbb{R}^{nd} \tag{4.28}$$

with $\boldsymbol{\Gamma}_{nd}$ being the time-domain covariance matrix from (4.2). Since $\mathrm{vec}_{\mathbb{R}}$ is an isomorphism, it is sufficient to show mutual contiguity in the real valued formulation. It is evident from (4.28) that the covariance matrix of $\tilde{\underline{Z}} = \boldsymbol{F}_{nd}\underline{Z}$ under P^n is given by $\tilde{\boldsymbol{\Gamma}}_{nd} := \boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}\boldsymbol{F}_{nd}^T \in \mathbb{R}^{nd\times nd}$. Let us have a more detailed look at the structure of $\tilde{\boldsymbol{\Gamma}}_{nd}$. Denote by $\tilde{\boldsymbol{\Gamma}}_n[j,k]$ the covariance matrix between the jth and kth Fourier coefficient in the real valued formulation. Then $\tilde{\boldsymbol{\Gamma}}_{nd}$ is of the following block structure – with block sizes visualized by the size of the surrounding rectangles, where the small squares are $d \times d$ and the large squares are $2d \times 2d$:

$$\tilde{\boldsymbol{\Gamma}}_{nd} = \left[\begin{array}{cc|c|cc} \tilde{\boldsymbol{\Gamma}}_n[0,0] & \tilde{\boldsymbol{\Gamma}}_n[0,1] & \dots & \tilde{\boldsymbol{\Gamma}}_n[0,\frac{n}{2}-1] & \tilde{\boldsymbol{\Gamma}}_n[0,\frac{n}{2}] \\ \tilde{\boldsymbol{\Gamma}}_n[1,0] & \tilde{\boldsymbol{\Gamma}}_n[1,1] & \dots & \tilde{\boldsymbol{\Gamma}}_n[1,\frac{n}{2}-1] & \tilde{\boldsymbol{\Gamma}}_n[1,\frac{n}{2}] \\ \vdots & \ddots & & \vdots & \vdots \\ \hline \tilde{\boldsymbol{\Gamma}}_n[\frac{n}{2},0] & \tilde{\boldsymbol{\Gamma}}_n[\frac{n}{2},1] & \dots & \tilde{\boldsymbol{\Gamma}}_n[\frac{n}{2},\frac{n}{2}-1] & \tilde{\boldsymbol{\Gamma}}_n[\frac{n}{2},\frac{n}{2}] \end{array}\right]$$

for n even, and a corresponding structure for n odd. To show that P^n and P_W^n are mutually contiguous, we first note that P^n and P_W^n are mutually absolutely continuous for all n. Hence,

by Lemma 4.10, it suffices to show that the log likelihood ratio sequence

$$\begin{aligned}\Lambda_n := \log \frac{p_W^n(\tilde{\underline{Z}})}{p^n(\tilde{\underline{Z}})} &= \frac{1}{2}\left(\log|\boldsymbol{\Gamma}_{nd}| - \log|\boldsymbol{D}_{nd}|\right) + \frac{1}{2}\tilde{\underline{Z}}^T\left(\tilde{\boldsymbol{\Gamma}}_{nd}^{-1} - \boldsymbol{D}_{nd}^{-1}\right)\tilde{\underline{Z}} \\ &=: \frac{1}{2}\Lambda_n^{(1)} + \frac{1}{2}\Lambda_n^{(2)}\end{aligned} \tag{4.29}$$

has bounded mean and variance under both P^n and P_W^n.

We start by showing the boundedness of $\Lambda_n^{(1)}$ from (4.29). Using $|\mathcal{B}\boldsymbol{A}| = |\boldsymbol{A}|^2$ from Lemma 4.2 (b) and $|\boldsymbol{f}(\pi + t)| = |\boldsymbol{f}(\pi - t)^T| = |\boldsymbol{f}(\pi - t)|$, we find $\log|\boldsymbol{D}_{nd}| = \sum_{j=0}^{n-1}\log|2\pi\boldsymbol{f}(\omega_j)|$. By Assumption $\boldsymbol{f}2$, $\boldsymbol{f}$ is (component-wise) continuously differentiable (see Lemma B.15 in the Appendix). Since the eigenvalues of $\boldsymbol{f}$ are uniformly bounded away from 0 by Assumption $\boldsymbol{f}1$, the function $\omega \mapsto |\boldsymbol{f}(\omega)|$ is continuously differentiable and bounded away from 0 (see Theorem 1 in Section 3 of Chapter 8 in Magnus and Neudecker (2007)). Hence, the function $\psi\colon [0, 2\pi] \to \mathbb{R}, \psi(\omega) := \log|2\pi\boldsymbol{f}(\omega)|$ is continuously differentiable. An application of Lemma B.23 in the Appendix reveals that the Riemann sum converges at rate n^{-1}:

$$\left|\frac{1}{2\pi}\int_0^{2\pi}\log|2\pi\boldsymbol{f}(\omega)|d\omega - \frac{1}{n}\sum_{j=0}^{n-1}\log|2\pi\boldsymbol{f}(\omega_j)|\right| \lesssim \frac{\max_{0\le\omega\le 2\pi}|\psi'(\omega)|}{n}$$

yielding

$$\log|\boldsymbol{D}_{nd}| = \sum_{j=0}^{n-1}\log|2\pi\boldsymbol{f}(\lambda_j)| = n\log G + O(1) \tag{4.30}$$

as $n \to \infty$, with

$$G := \exp\left\{\frac{1}{2\pi}\int_0^{2\pi}\log|2\pi\boldsymbol{f}(\lambda)|d\lambda\right\} > 0.$$

On the other hand, an application of Lemma 4.6 (a) shows that there exists a positive constant E such that

$$\log|\boldsymbol{\Gamma}_{nd}| = n\log G + \log E + o(1) = n\log G + O(1) \tag{4.31}$$

as $n \to \infty$. Putting together (4.30) and (4.31) shows $\Lambda_n^{(1)} = \log|\boldsymbol{\Gamma}_{nd}| - \log|\boldsymbol{D}_{nd}| = O(1)$.

For the second summand $\Lambda_n^{(2)}$ from (4.29) we compute the mean and variance under P^n and P_W^n using Lemma B.24 (a) as

$$\mathrm{E}_{P^n}\Lambda_n^{(2)} = \operatorname{tr}\left(\boldsymbol{I}_{nd} - \tilde{\boldsymbol{\Gamma}}_{nd}\boldsymbol{D}_{nd}^{-1}\right), \tag{4.32}$$

$$\mathrm{Var}_{P^n}\Lambda_n^{(2)} = 2\operatorname{tr}\left(\left(\boldsymbol{I}_{nd} - \tilde{\boldsymbol{\Gamma}}_{nd}\boldsymbol{D}_{nd}^{-1}\right)^2\right), \tag{4.33}$$

$$\mathrm{E}_{P_W^n}\Lambda_n^{(2)} = \operatorname{tr}\left(\boldsymbol{D}_{nd}\tilde{\boldsymbol{\Gamma}}_{nd}^{-1} - \boldsymbol{I}_{nd}\right), \tag{4.34}$$

$$\mathrm{Var}_{P_W^n}\Lambda_n^{(2)} = 2\operatorname{tr}\left(\left(\boldsymbol{D}_{nd}\tilde{\boldsymbol{\Gamma}}_{nd}^{-1} - \boldsymbol{I}_{nd}\right)^2\right). \tag{4.35}$$

Now consider the matrix $\boldsymbol{H}_{nd} := \tilde{\boldsymbol{\Gamma}}_{nd} - \boldsymbol{D}_{nd} \in \mathbb{R}^{nd\times nd}$ as in Theorem 4.9. We partition the matrix $\boldsymbol{H}_{nd}$ into blocks of the same size as those of $\tilde{\boldsymbol{\Gamma}}_{nd}$, and denote these blocks by $\boldsymbol{H}_n[j,k]$

for $j,k = 0,\dots,\lfloor n/2\rfloor$. From Theorem 4.9, we conclude that $\|\boldsymbol{H}_{nd}[j,k]\| \lesssim n^{-1}$ uniformly in j,k. To avoid the notational case distinction between $\omega \in \{0,\pi\}$ and $\omega \in (0,\pi)$, we introduce the symbolic notation

$$\tilde{\boldsymbol{f}}(\omega) := \begin{cases} \boldsymbol{f}(\omega), & \omega \in \{0,\pi\} \\ \mathcal{B}\boldsymbol{f}(\omega), & \text{else.} \end{cases}$$

Using the result $|\operatorname{tr}(\boldsymbol{AB})| \le \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|$ for $\boldsymbol{A},\boldsymbol{B} \in \mathbb{R}^{d\times d}$ from Lemma B.4 in the Appendix, we obtain from (4.32)

$$\begin{aligned} \left|\mathrm{E}_{P^n}\Lambda_n^{(2)}\right| = \left|\operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{D}_{nd}^{-1}\right)\right| &= 2\pi\left|\sum_{j=0}^{\lfloor n/2\rfloor}\operatorname{tr}\left(\boldsymbol{H}_n[j,j]\tilde{\boldsymbol{f}}^{-1}(\omega_j)\right)\right| \\ &\lesssim 2\pi\sum_{j=0}^{\lfloor n/2\rfloor}\|\boldsymbol{H}_n[j,j]\|\left\|\tilde{\boldsymbol{f}}^{-1}(\omega_j)\right\|_2 \\ &\lesssim (\lfloor n/2\rfloor+1)n^{-1}b_0^{-1} = O(1), \end{aligned} \tag{4.36}$$

with b_0 from Assumption ***f***1, where $\|\mathcal{B}\boldsymbol{A}\| \lesssim \|\boldsymbol{A}\|$ was used. Similarly, we compute for (4.33)

$$\begin{aligned} 0 \le \operatorname{tr}\left(\left(\boldsymbol{H}_{nd}\boldsymbol{D}_{nd}^{-1}\right)^2\right) &= 4\pi^2\sum_{j,k=0}^{\lfloor n/2\rfloor}\operatorname{tr}\left(\boldsymbol{H}_n[j,k]\tilde{\boldsymbol{f}}^{-1}(\omega_k)\boldsymbol{H}_n[k,j]\tilde{\boldsymbol{f}}^{-1}(\omega_j)\right) \\ &\lesssim \sum_{j,k=0}^{\lfloor n/2\rfloor}\|\boldsymbol{H}_n[j,k]\|\|\boldsymbol{H}_n[k,j]\|\|\tilde{\boldsymbol{f}}^{-1}(\omega_j)\|_2\|\tilde{\boldsymbol{f}}^{-1}(\omega_k)\|_2 \\ &\lesssim (\lfloor n/2\rfloor+1)^2 n^{-2}b_0^{-2} = O(1), \end{aligned}$$

showing $\operatorname{Var}_{P^n}\Lambda_n^{(2)} = O(1)$.

Now consider a time series with spectral density matrix $\boldsymbol{f}(\omega)^{-1}$, $0 \le \omega \le \pi$. Denote by $\boldsymbol{\Gamma}_{nd}^-$ the associated block Toeplitz covariance matrix in the time domain:

$$\boldsymbol{\Gamma}_{nd}^- = \begin{pmatrix} \boldsymbol{\Gamma}^-(0) & \boldsymbol{\Gamma}^-(1) & \dots & \boldsymbol{\Gamma}^-(n-1) \\ \boldsymbol{\Gamma}^-(-1) & \boldsymbol{\Gamma}^-(0) & \dots & \\ \vdots & & \ddots & \\ \boldsymbol{\Gamma}^-(-n+1) & \dots & & \boldsymbol{\Gamma}^-(0) \end{pmatrix}, \quad \boldsymbol{\Gamma}^-(h) = \int_0^{2\pi}\boldsymbol{f}(\omega)^{-1}\exp(ih\omega)d\omega.$$

Clearly, $\boldsymbol{f}^{-1}$ fulfills Assumption ***f***1. Furthermore, from Lemma B.16 in the Appendix, we obtain that $\boldsymbol{f}^{-1}$ and $\boldsymbol{\Gamma}_{nd}^-$ also fulfill Assumption ***f***2. Let $\boldsymbol{H}_{nd}^- := \boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^-\boldsymbol{F}_{nd}^T - \boldsymbol{D}_{nd}^{-1}$ and denote (similar to $\boldsymbol{H}_{nd}$) the blocks of $\boldsymbol{H}_{nd}^-$ by $\boldsymbol{H}_n^-[j,k]$ for $0 \le j,k \le \lfloor n/2\rfloor$. Applying Theorem 4.9 to $\boldsymbol{\Gamma}_{nd}^-$ and $\boldsymbol{f}(\omega)^{-1}$ yields $\|\boldsymbol{H}_{nd}^-[j,k]\| \lesssim n^{-1}$ uniformly in j,k. Thus we compute from (4.34)

$$\begin{aligned} \left|\mathrm{E}_{P_W^n}\Lambda_n^{(2)}\right| &= \left|\operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^{-1}\boldsymbol{F}_{nd}^T\right)\right| \\ &\le \left|\operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^-\boldsymbol{F}_{nd}^T\right)\right| + \left|\operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\left(\boldsymbol{\Gamma}_{nd}^{-1}-\boldsymbol{\Gamma}_{nd}^-\right)\boldsymbol{F}_{nd}^T\right)\right| \end{aligned} \tag{4.37}$$

Since the Frobenius norm values of the blocks $\boldsymbol{H}_n[j,k]$ of $\boldsymbol{H}_{nd}$ and of the blocks $\boldsymbol{H}_n^-[j,k]$ of $\boldsymbol{H}_{nd}^-$

are uniformly in $O(n^{-1})$, we get

$$\begin{aligned}\|\boldsymbol{H}_{nd}\|^2 = \operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{H}_{nd}^T\right) &= \sum_{j,k=0}^{\lfloor n/2\rfloor} \operatorname{tr}\left(\boldsymbol{H}_n[j,k]\boldsymbol{H}_n[j,k]^T\right) \\ &\leq \sum_{j,k=0}^{\lfloor n/2\rfloor} \|\boldsymbol{H}_n[j,k]\|^2 = O(1)\end{aligned} \tag{4.38}$$

and similarly $\|\boldsymbol{H}_{nd}^-\| = O(1)$. For the first summand in (4.37), we obtain

$$\operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^-\boldsymbol{F}_{nd}^T\right) = \operatorname{tr}(\boldsymbol{H}_{nd}\boldsymbol{D}_{nd}^{-1}) + \operatorname{tr}(\boldsymbol{H}_{nd}\boldsymbol{H}_{nd}^-) = O(1),$$

since $\operatorname{tr}(\boldsymbol{H}_{nd}\boldsymbol{D}_{nd}^{-1}) = O(1)$ (see (4.36)) and $|\operatorname{tr}(\boldsymbol{H}_{nd}\boldsymbol{H}_{nd}^-)| \leq \|\boldsymbol{H}_{nd}\|\|\boldsymbol{H}_{nd}^-\| = O(1)$. For the second summand in (4.37), we obtain, using that $\boldsymbol{F}_{nd}$ is orthogonal and hence $\|\boldsymbol{F}_{nd}^T\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\| = \|\boldsymbol{H}_{nd}\|$ (by Lemma B.2 in the Appendix) and $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|$ (by Lemma B.4 in the Appendix)

$$\left|\operatorname{tr}\left(\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\left(\boldsymbol{\Gamma}_{nd}^{-1} - \boldsymbol{\Gamma}_{nd}^-\right)\boldsymbol{F}_{nd}^T\right)\right| \leq \|\boldsymbol{H}_{nd}\|\left\|\boldsymbol{\Gamma}_{nd}^{-1} - \boldsymbol{\Gamma}_{nd}^-\right\| \leq \|\boldsymbol{H}_{nd}\|\left\|\boldsymbol{\Gamma}_{nd}^{-1}\right\|_2\left\|\boldsymbol{I}_{nd} - \boldsymbol{\Gamma}_{nd}\boldsymbol{\Gamma}_{nd}^-\right\|.$$

By (4.38), it holds $\|\boldsymbol{H}_{nd}\| = O(1)$. Furthermore, it is known (see Lemma 2.1 in Hannan and Wahlberg (1989)) that

$$\left\|\boldsymbol{\Gamma}_{nd}^{-1}\right\|_2 \leq \max_{0\leq\omega\leq\pi}\|\boldsymbol{f}(\omega)^{-1}\|_2 \leq b_0^{-1} = O(1),$$

with b_0 from Assumption ***f*** 1, It thus remains to show the boundedness of $\left\|\boldsymbol{I}_{nd} - \boldsymbol{\Gamma}_{nd}\boldsymbol{\Gamma}_{nd}^-\right\|$, which follows from part (b) of Theorem 4.6.

Finally, the variance part (4.35) under Whittle's likelihood can be computed as

$$\frac{1}{2}\operatorname{Var}P_W^n\Lambda_n^{(2)} = \operatorname{tr}\left(\left(\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^{-1}\boldsymbol{F}_{nd}^T\right)^2\right) = \left\|\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^{-1}\boldsymbol{F}_{nd}^T\right\|^2$$

and

$$\left\|\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^{-1}\boldsymbol{F}_{nd}^T\right\| \leq \left\|\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^-\boldsymbol{F}_{nd}^T\right\| + \left\|\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\left(\boldsymbol{\Gamma}_{nd}^{-1} - \boldsymbol{\Gamma}_{nd}^-\right)\boldsymbol{F}_{nd}^T\right\|. \tag{4.39}$$

The first summand in (4.39) is

$$\left\|\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\boldsymbol{\Gamma}_{nd}^-\boldsymbol{F}_{nd}^T\right\| = \left\|\boldsymbol{H}_{nd}\left(\boldsymbol{D}_{nd}^{-1} + \boldsymbol{H}_{nd}^-\right)\right\| \leq \left\|\boldsymbol{H}_{nd}\boldsymbol{D}_{nd}^{-1}\right\| + \left\|\boldsymbol{H}_{nd}\boldsymbol{H}_{nd}^-\right\| = O(1),$$

since $\left\|\boldsymbol{H}_{nd}\boldsymbol{D}_{nd}^{-1}\right\| = O(1)$ (see (4.36)) and $\left\|\boldsymbol{H}_{nd}\boldsymbol{H}_{nd}^-\right\| \leq \|\boldsymbol{H}_{nd}\|\left\|\boldsymbol{H}_{nd}^-\right\| = O(1)$. The second summand in (4.39) is

$$\left\|\boldsymbol{H}_{nd}\boldsymbol{F}_{nd}\left(\boldsymbol{\Gamma}_{nd}^{-1} - \boldsymbol{\Gamma}_{nd}^-\right)\boldsymbol{F}_{nd}^T\right\| \leq \|\boldsymbol{H}_{nd}\|\|\boldsymbol{\Gamma}_{nd}^{-1} - \boldsymbol{\Gamma}_{nd}^-\| = O(1).$$

□

5.

Bayesian Nonparametric Method for Spectral Density Inference

5.1. Definition of Prior and Method

In this section, we use the Hpd matrix Gamma process from Chapter 3 in conjunction with Whittle's likelihood from Chapter 4 to construct a novel method for Bayesian nonparametric spectral inference for multivariate time series. To elaborate, let $\mathcal{X} = [0, \pi]$ and let $\mathbf{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$, with the Hpd Gamma measure ν from (3.1) fulfilling Assumptions GP1-GP2. For $k > 0$, let the equidistant interval partition of $[0, \pi]$ of size k be denoted by

$$I_{j,k} = \left(\frac{(j-1)\pi}{k}, \frac{j\pi}{k} \right], \quad j = 1, \ldots, k. \tag{5.1}$$

We define the *Bernstein-Hpd-Gamma prior* for the spectral density matrix $\boldsymbol{f}$ as

$$\begin{aligned} \boldsymbol{f}(\pi x) &:= \sum_{j=1}^{k} \mathbf{\Phi}\left(I_{j,k}\right) b(x|j, k-j+1), \quad 0 \le x \le 1, \\ k &\sim p(k), \end{aligned} \tag{5.2}$$

with the Bernstein polynomial basis functions $b(\cdot|j, k-j+1)$ as defined in (B.20) in the Appendix. It is assumed that the prior distribution of the polynomial degree $k \in \mathbb{N}$ (as given by the probability mass function $p(k)$) and the Hpd-Gamma process $\mathbf{\Phi}$ are independent. The definition (5.2) can be conceived as an extension of the Bernstein-Dirichlet prior (1.1) from the univariate case to the multivariate case.

Remark. *Recall the following approximation property of Bernstein polynomials: Let $\boldsymbol{f}_0 \colon \mathcal{X} \to \mathcal{S}_d^+$ be continuous. Let $\boldsymbol{F}_0$ be the spectral measure corresponding to $\boldsymbol{f}_0$, i.e. $\boldsymbol{F}_0(A) = \int_A \boldsymbol{f}_0(x)dx$ for $A \subset \mathcal{X}$ measurable. Then the convergence*

$$\sum_{j=1}^{k} \boldsymbol{F}_0\left(I_{j,k}\right) b(x|j, k-j+1) \to \boldsymbol{f}_0(\pi x), \quad 0 \le x \le 1,$$

holds uniformly and component-wise as $k \to \infty$ (see Lemma B.12 in the Appendix). On the

other hand, the prior mean of $\boldsymbol{f}$ *as in* (5.2) *conditioned on* k *is given by*

$$\mathrm{E}\left[\boldsymbol{f}(\pi x)|k\right] = \sum_{j=1}^{k} \mathrm{E}\boldsymbol{\Phi}\left(I_{j,k}\right) b(x|j, k-j+1).$$

In this sense, the Gamma process $\boldsymbol{\Phi}$ *is directly connected to prior modeling of* $\boldsymbol{F}_0$*. If an a priori guess of* $\boldsymbol{f}_0$ *is available, it is reasonable to choose the prior parameters of* $\boldsymbol{\Phi}$ *in such a way that* $\mathrm{E}\boldsymbol{\Phi}(A) = \boldsymbol{F}_0(A)$ *holds for any measurable* $A \subset \mathcal{X}$*. As an example, when using the* $A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ *process from Section 3.4.2, this can be achieved by choosing the process parameters* ω *and* $\boldsymbol{\Sigma}$ *in such a way that* $\frac{1}{d}\omega(x)\boldsymbol{\Sigma}(x) = \boldsymbol{f}_0(x)$ *holds, see* (3.24)*.*

To complete the Bayesian model specification, we employ the version $\tilde{P}_W^n$ of Whittle's Likelihood from Corollary 4.5, which is defined in terms of the following Lebesgue density for the Fourier coefficients $\underline{\tilde{Z}}_1, \ldots, \underline{\tilde{Z}}_N$ with $N = \lceil n/2 \rceil - 1$:

$$\tilde{p}_W^n(\underline{\tilde{z}}_1, \ldots, \underline{\tilde{z}}_N|\boldsymbol{f}) = \prod_{j=1}^{N} \frac{1}{\pi^d |2\pi \boldsymbol{f}(\omega_j)|} \exp\left(-\frac{1}{2\pi} \underline{\tilde{z}}_j^* \boldsymbol{f}(\omega_j)^{-1} \underline{\tilde{z}}_j\right), \tag{5.3}$$

for $\underline{\tilde{z}}_1, \ldots, \underline{\tilde{z}}_N \in \mathbb{C}^d$. Observe that in contrast to (4.8), the boundary frequencies $\omega = \{0, \pi\}$ are not considered in (5.3). The reason is that we assume the time series to be mean centered. Since the boundary frequencies represent the sample mean and the sample alternating mean respectively, we exclude them from inference. This can be done since we are only interested in spectral density inference and ignoring any structure that is related to the mean. However, if one is interested in additionally estimating the mean or more generally in inference for semiparametric models (see the upcoming Chapter 8), the boundary frequencies will have to be included in the likelihood as well.

A common choice for the probability mass function of k is $p(k) = C \exp(-ck \log k)$ for $k \in \mathbb{N}$ for some positive constants c, C. This choice is indeed motivated from asymptotic considerations, since it yields the "just right" rate of decay (of $p(k)$ as $k \to \infty$), which is needed to obtain consistency and contraction rate results (see Theorem 7.3 and Theorem 7.20, as well as Remark 7.2 and Remark 7.19). Note that the constant C is only needed for normalization (making all values $p(k)$ sum up to 1), such that the actual distributional parameter is c. Given c, there is no analytically closed form available to determine the normalizing constant $C^{-1} = \sum_{l \in \mathbb{N}} p(l)$, the prior expected value $\mathrm{E}[k] = \sum_{l \in \mathbb{N}} lp(l)$ or the prior variance $\mathrm{Var}[k] = \sum_{l \in \mathbb{N}} (l - \mathrm{E}[k])^2 p(l)$. However, these values can be approximated numerically. As an example, Table 5.1 contains numerical approximations for some values of c. These have been obtained by drawing 100,000 random samples from the discrete set $k \in \{1, \ldots, k_{\max}\}$ with probability mass function $p(k)$, where $k_{\max} = 10,000$.

To draw random samples from the prior (5.2), we can use an extension of Algorithm 1 from Section 3.4. Indeed, since k and $\boldsymbol{\Phi}$ are a priori *independent*, we only have to extend Algorithm 1 by random draws of $p(k)$, which can readily be obtained if we restrict k to a finite set $\{1, \ldots, k_{\max}\}$ for some large integer $k_{\max}$, since k is discrete. A visualization of draws from the prior (5.2) is shown in Figure 5.1, where the process priors $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2$ and $\boldsymbol{\Phi}_3$ from (3.28)-(3.30) have been used (see also Figure 3.2 for a visualization of the processes $\boldsymbol{\Phi}_i$ themselves for $i = 1, 2, 3$). Again, it

c	1	0.5	0.1	0.05	0.01	0.005	0.001
C^{-1}	1.29	1.78	4.87	7.99	27.28	47.53	180.66
$\mathrm{E}[k]$	1.26	1.66	4.14	6.66	22.76	40.00	155.96
$\mathrm{Var}[k]$	0.27	0.85	10.04	29.82	410.05	1314.881	21024.78

Table 5.1.: Numerical approximations for the normalizing constant C^{-1}, the expected value and the variance of the distribution $p(k) = C\exp(-ck\log k)$, $k \geq 1$.

can be seen that the pointwise uncertainty of the mixture for $\mathbf{\Phi}_2$ is monotonically increasing, whereas the mixture mean coincides with the mixture mean for $\mathbf{\Phi}_1$. For the VARMA process $\mathbf{\Phi}_3$ mixture mean curve, a choice of $c = 0.01$ of the prior parameter for k is not enough to ensure that the mean curve is rendered properly a priori (see the discrepancy at the left and right boundary of the dotted blue VARMA mean curve and the solid black mixture mean curve in Figure 5.1(c)). This effect is mitigated by the choice of $c = 0.001$ (see Figure 5.1(d)), which allows for higher values of k (c.f. Table 5.1) and thus a more detailed a priori resolution.

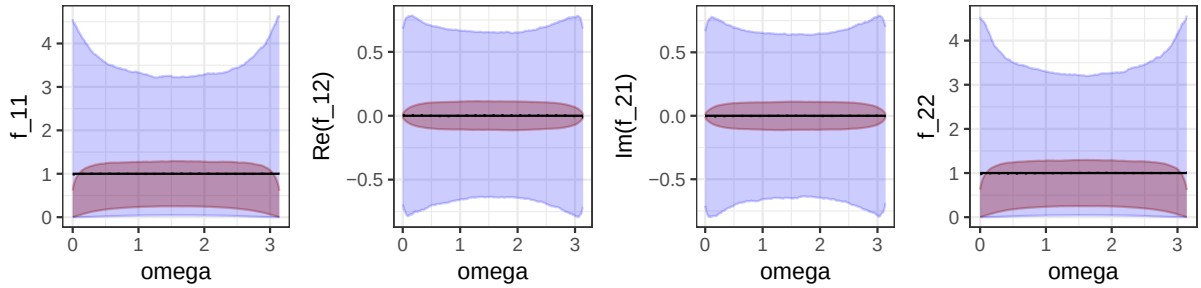
It can be seen that the pointwise quartiles of the Bernstein mixtures in Figure 5.1 are biased towards zero at the boundary of $\mathcal{X}$. This artifact is systematically introduced by the usage of Bernstein polynomials. It has to be kept in mind when using the mixture prior for spectral inference. This issue becomes particular severe in a semiparametric context, when the posterior distribution of the parameter of interest depends on $\boldsymbol{f}(0)$ or $\boldsymbol{f}(\pi)$ (see the examples in the upcoming Section 8.1). One possible remedy to this artifact constitutes the usage of *truncated* Bernstein polynomials, which yield more robust mixture properties at the boundary. To elaborate, the beta densities $b(\cdot|j, k-j+1)$ in (5.2) are replaced by their truncated and dilated counterparts

$$b_{\tau_l}^{\tau_r}(x|j, k-j+1) := b(\tau_l + x(\tau_r - \tau_l)|j, k-j+1), \quad 0 \leq x \leq 1, \tag{5.4}$$

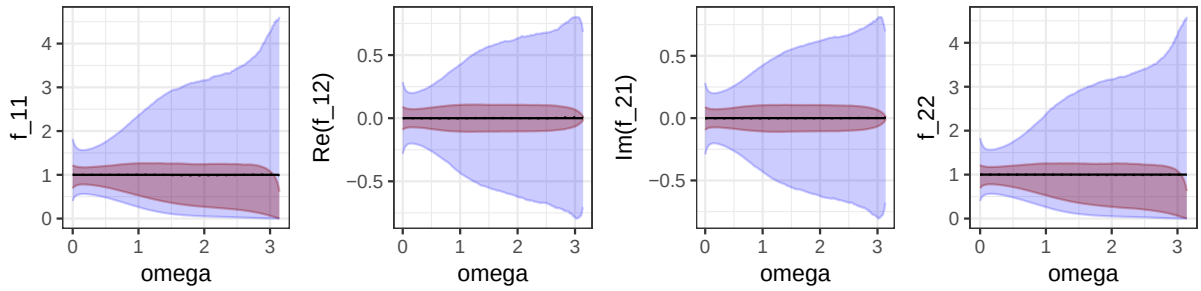
for some $0 < \tau_l < \tau_r < 1$. As an example, Figure 5.2 depicts mixtures of the same processes as in Figure 5.1, with Bernstein polynomials truncated at $\tau_l = 0.1$ and $\tau_r = 0.9$. It can be seen that while the artifacts at the boundary are removed, the truncation enforces more effort to render a priori information at the boundary, in particular in Figure 5.2(c). This effect also comes from a different approximation behavior of truncated Bernstein polynomials. Indeed, to establish an asymptotic approximation result as in Lemma B.12 in the Appendix, the truncation bounds have to be chosen to fulfill $\tau_l \to 0$ and $\tau_r \to 1$ as $k \to \infty$. In the numerical illustrations in this work, we will employ truncated Bernstein polynomials due to their greater robustness at the boundary, with fixed values $\tau_l = 0.1$ and $\tau_r = 0.9$. Asymptotic consideration with truncated Bernstein polynomials are beyond the scope of this work. It may be noted that the rendering effect can to some extent be taken into account by choosing a smaller value of c, as in Figure 5.2(d).

5.2. Numerical Simulation of Posterior Samples

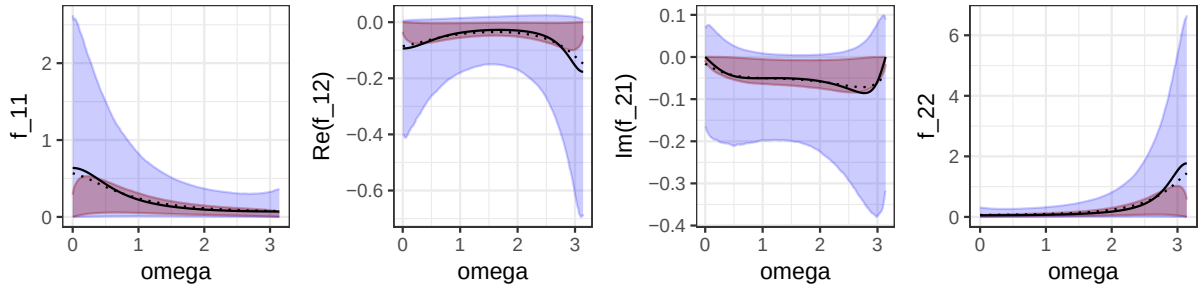
We start by describing the computational methods to generate posterior samples under the prior (5.2) on the spectral density matrix $\boldsymbol{f}$ in conjunction with Whittle's likelihood (5.3).



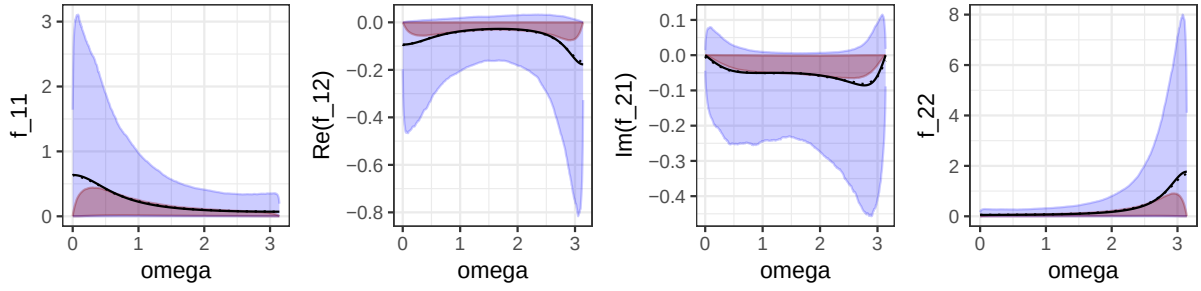
(a)



(b)

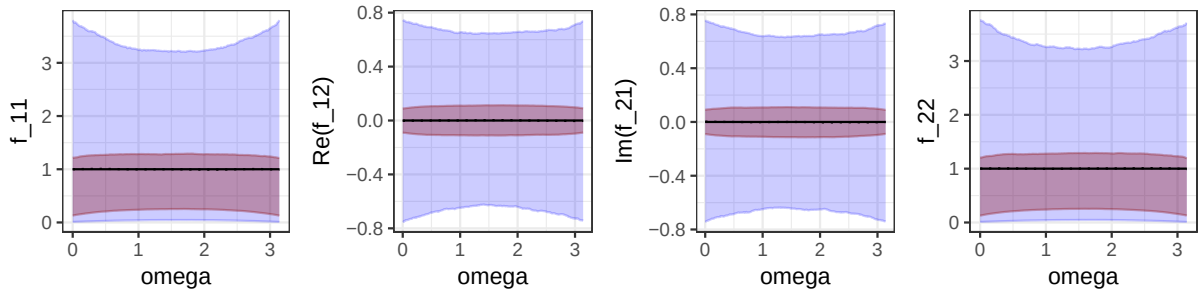


(c)

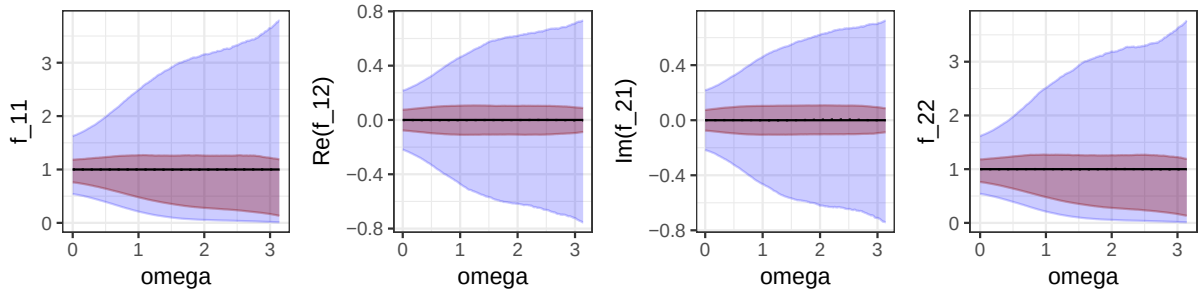


(d)

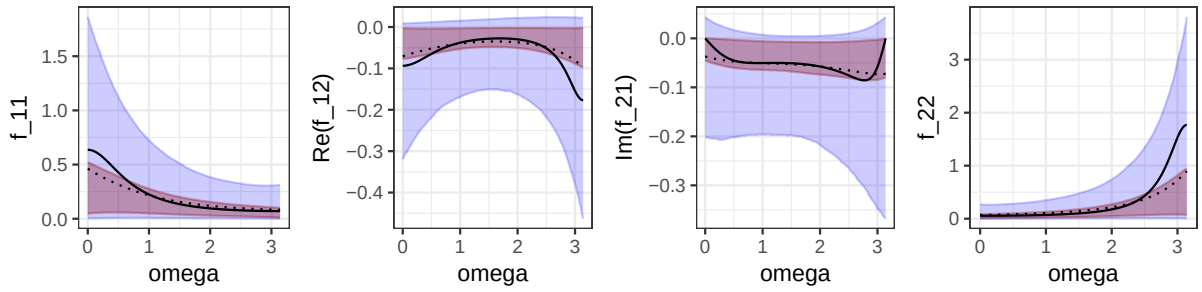
Figure 5.1.: Visualization of the pointwise (among $0 \leq \omega \leq \pi$) distribution of the Bernstein-Hpd-Gamma prior with (a) $\mathbf{\Phi}_1$ from (3.28) and $c = 0.01$, (b) $\mathbf{\Phi}_2$ from (3.29) and $c = 0.01$, (c) $\mathbf{\Phi}_3$ from (3.30) and $c = 0.01$ and (d) $\mathbf{\Phi}_3$ from (3.30) and $c = 0.001$. The pointwise mean is drawn in dashed black, the area between pointwise upper and lower quartile in shaded red and the area between pointwise 0.95 and 0.05 quantile in shaded blue. The curve $\frac{1}{2}\omega_i(x)\mathbf{\Sigma}_i(x)$ is shown in solid black. The visualizations are based on 50,000 samples respectively from Algorithm 1 with $L = 30$ ($L = 300$ for (b)) and k sampled directly from $\{1, \ldots, k_{\max} = 500\}$.



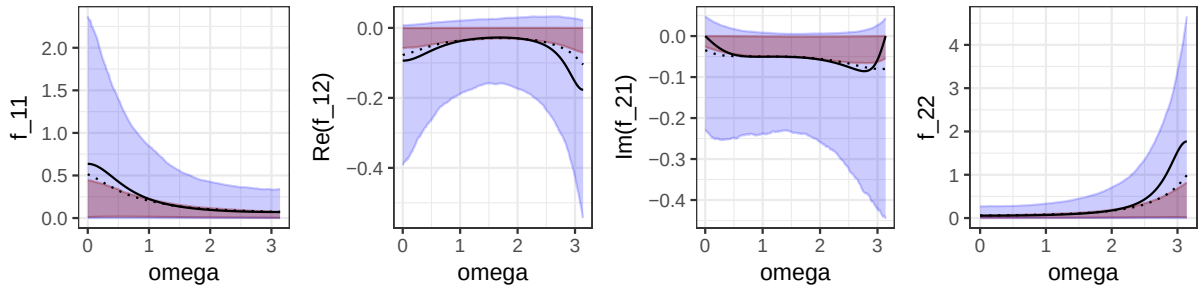
(a)



(b)



(c)



(d)

Figure 5.2.: Visualization of the pointwise (among $0 \leq \omega \leq \pi$) distribution of the Bernstein-Hpd-Gamma prior as in Figure 5.1, with Bernstein polynomials truncated at $\tau_l = 0.1$ and $\tau_r = 0.9$.

Let $\mathcal{X} = [0, \pi]$ and let $\mathbf{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$, such that Assumptions GP1 and GP2' and GP4 are fulfilled. Recall the representation for $\mathbf{\Phi}$ from Lemma 3.13. Since the series representation for $\mathbf{\Phi}$ is almost surely convergent, it can be well approximated by a truncation

$$\mathbf{\Phi} \approx \sum_{l=1}^{L} \delta_{x_l} r_l \boldsymbol{U}_l \tag{5.5}$$

at some large integer L. The value of L should depend on the data and sample size and can e.g. be determined numerically according to the desired tolerance for error in computation. From our experience in simulation studies, a conservative choice can be formulated as $L = d^2 \max\{20, n^{1/3}\}$, where in fact in most considered scenarios a choice of $L = \max\{20, dn^{1/3}\}$ was already sufficient. See Muliere and Tardella (1998) and Choudhuri et al. (2004a) for discussions on the similar problem of truncating the stick-breaking representation of the Dirichlet process.

In the following, we assume that the probability measure $\alpha^* = \frac{\alpha}{C_\alpha}$ with $C_\alpha = \int_0^\pi \alpha(x, \bar{\mathbb{S}}_d^+)dx < \infty$ has a Lebesgue density on $[0, \pi] \times \mathcal{S}_d^+$, which we denote by $g^* = g^*(x, \boldsymbol{U})$. The truncated series approximation (5.5) is parametrized by the $3L$ parameters

$$\underline{\Theta}_{\mathbf{\Phi}} := (v_1, \ldots, v_L, x_1, \ldots, x_L, \boldsymbol{U}_1, \ldots, \boldsymbol{U}_L) \tag{5.6}$$

with $v_1, \ldots, v_L \overset{\text{iid}}{\sim} \mathrm{Exp}(1)$ and $(x_1, \boldsymbol{U}_1), \ldots, (x_L, \boldsymbol{U}_L) \overset{\text{iid}}{\sim} \alpha^*$ as in Lemma 3.13. Accordingly, the spectral density matrix in the prior model (5.2) is parametrized by $(\underline{\Theta}_{\mathbf{\Phi}}, k)$. From Lemma 3.13, we obtain that the prior probabilities of $\underline{\Theta}_{\mathbf{\Phi}}$ are given by the Lebesgue density

$$p(\underline{\Theta}_{\mathbf{\Phi}}) = \exp\left(-\sum_{l=1}^{L} [v_l - \log g^*(x_l, \boldsymbol{U}_l)]\right).$$

Along with Whittle's likelihood $\tilde{p}_W^n$ from (5.3) for the frequency domain observation $\underline{\tilde{Z}}_1, \ldots, \underline{\tilde{Z}}_N$, this specifies a full Bayesian model with posterior distribution

$$p(\underline{\Theta}_{\mathbf{\Phi}}, k | \underline{\tilde{Z}}_1, \ldots, \underline{\tilde{Z}}_N) \propto \tilde{p}_W^n(\underline{\tilde{Z}}_1, \ldots, \underline{\tilde{Z}}_N | \underline{\Theta}_{\mathbf{\Phi}}, k) p(\underline{\Theta}_{\mathbf{\Phi}}) p(k). \tag{5.7}$$

Since this posterior distribution is in general not tractable analytically, we employ computational methods to generate random samples from it. To do so, we first derive a practically suitable parametrization of $\underline{\Theta}_{\mathbf{\Phi}}$ in Section 5.2.1, before presenting a Metropolis-within-Gibbs algorithm to sample from (5.7) in Section 5.2.2. The implementation in `R` is briefly discussed in Section 5.2.3.

5.2.1. Parametrization of the radial parts

Instead of the parametrization (5.6) with $v_1, \ldots, v_L$, we use the parameters $r_1, \ldots, r_L$ from Lemma 3.13, due to both numerical stability and a clearer interpretation – recalling that the r_l's are the radial parts of the mass atoms of the Hpd Gamma process $\mathbf{\Phi}$. Furthermore, we will employ the $\underline{\varphi}_l$ parametrization from Section 3.4.1 for the $\boldsymbol{U}_l$'s. This yields the representation

$$\underline{\tilde{\Theta}}_{\mathbf{\Phi}} := (r_1, \ldots, r_L, x_1, \ldots, x_L, \underline{\varphi}_1, \ldots, \underline{\varphi}_L). \tag{5.8}$$

To work with this parametrization, we derive the Jacobian of the mapping $\tilde{T}\colon [0,\infty)^L \to [0,\infty)^L, (r_1,\dots,r_L) \mapsto (v_1,\dots,v_L)$. Let $w_1,\dots,w_L$ be as in Lemma 3.13. First observe that $v_1 = w_1$ and

$$v_l = w_l - w_{l-1}, \quad w_l = \rho\left([r_l,\infty)|C_\alpha, \beta(x_l, \boldsymbol{U}_l)\right), \quad l = 2,\dots,L.$$

Since w_l only depends on r_l, it follows that v_l only depends on $r_1,\dots,r_l$ (and not on $r_{l+1},\dots,r_L$), such that the Jacobian $\boldsymbol{J}_{\tilde{T}}$ of $\tilde{T}$ is a lower triangular matrix, and the determinant is given by the product of its diagonal entries:

$$|\boldsymbol{J}_{\tilde{T}}(r_1,\dots,r_L)| = \prod_{l=1}^{L}\left|\frac{\partial v_l}{\partial r_l}\right| = \prod_{l=1}^{L}\left|\frac{\partial w_l}{\partial r_l}\right| = \prod_{l=1}^{L}\left|\frac{d}{dr}\left[\rho\left([r,\infty)|C_\alpha,\beta(x_l,\boldsymbol{U}_l)\right)\right]_{r=r_l}\right|$$

Since

$$\left|\frac{d}{dr}\rho\left([r,\infty)|C_\alpha,\beta(x,\boldsymbol{U})\right)\right| = C_\alpha\frac{\exp(-\beta(x,\boldsymbol{U})r)}{r},$$

we arrive at

$$|\boldsymbol{J}_{\tilde{T}}(r_1,\dots,r_L)| = C_\alpha^L\prod_{l=1}^{L}\frac{\exp(-\beta(x_l,\boldsymbol{U}_l)r_l)}{r_l}. \tag{5.9}$$

In practice, the transformation $\tilde{T}$ is evaluated using the numerical approximation for $\rho\left([r,\infty)|a,b\right)$ from (3.19).

5.2.2. A Metropolis-within-Gibbs sampler

We can now present a Markov Chain Monte Carlo method to obtain random samples from the joint posterior distribution of $\boldsymbol{\Phi}$ and k. It is based on the Metropolis-within-Gibbs algorithm. Consider the parametrization $\underline{\tilde{\Theta}}_{\boldsymbol{\Phi}}$ from (5.8) for $\boldsymbol{\Phi}$. The posterior distribution in this parametrization is given by

$$p(\underline{\tilde{\Theta}}_{\boldsymbol{\Phi}}, k|\underline{Z}_1,\dots,\underline{Z}_n) = \left(\prod_{l=1}^{L}|\boldsymbol{J}_{T^{-1}}(\underline{\varphi}_l)|\right)|\boldsymbol{J}_{\tilde{T}}(r_1,\dots,r_L)|p(\Theta_{\boldsymbol{\Phi}}, k|\underline{Z}_1,\dots,\underline{Z}_n) \tag{5.10}$$

with $p(\Theta_{\boldsymbol{\Phi}}, k|\underline{Z}_1,\dots,\underline{Z}_n)$ as in (5.7) and the Jacobian determinants of the employed transformations T and $\tilde{T}$ as in (3.22) and (5.9). To obtain random samples $\underline{\Theta}^{(1)},\dots,\underline{\Theta}^{(N)}$ from (5.10), we employ the Gibbs sampler (see Section 6.3.2 in Christensen et al. (2011)). Since the full conditionals do not belong in general to a known class of distributions, each parameter is updated with a Metropolis-Hastings (MH) step (Section 6.3.3 in Christensen et al. (2011)).

The starting value $\underline{\Theta}^{(1)} = (\underline{\tilde{\Theta}}_{\boldsymbol{\Phi}}^{(1)}, k^{(1)})$ for the Markov Chain is as follows: Some large integer for $k^{(1)}$ and $r_1^{(1)},\dots,r_L^{(1)} \overset{\text{iid}}{\sim} \text{Exp}(1)$ and $x_1^{(1)},\dots,x_L^{(1)} \overset{\text{iid}}{\sim} \text{Unif}([0,\pi])$ and $\underline{\varphi}_1^{(1)},\dots,\underline{\varphi}_L^{(1)} \overset{\text{iid}}{\sim} \text{Unif}(\mathcal{I})$ with $\mathcal{I} = \otimes_{j=1}^{d^2-1} I_j$, see (3.21). In the $i+1$-th iteration of the Gibbs sampler, the proposal value $k^{*,(i+1)}$ for the Metropolis-Hastings step of k is

$$k^{*,(i+1)} := \min\left\{k_{\max},\ \max\{1,\ k^{(i)} + \lfloor\epsilon_k^{(i+1)}\rfloor\}\right\}, \quad \epsilon_k^{(i+1)} \sim \text{Cauchy}(0,1)$$

and some large integer $k_{\max} < \infty$ which is introduced for computational speed, enabling pre-computation and storage of the beta densities of degree $1, ..., k_{\max}$. A sufficiently large value of $k_{\max}$ depends on the specific posterior distribution and can be determined numerically by a preliminary pilot run of the algorithm. For many situations (including all the numerical illustrations in this work), a value of $k_{\max} = 300$ was found to be sufficient. Although sampling from the (approximate) full conditional of k is possible, a Metropolis-Hastings step is used to avoid the computationally expensive task of evaluating $p(\tilde{\underline{\Theta}}_{\boldsymbol{\Phi}}^{(i)}, k | \underline{Z}_1, \ldots, \underline{Z}_n)$ for $k = 1, \ldots, k_{\max}$ in every iteration of the Gibbs sampler.

The radial parts $r_1, \ldots, r_L$ are updated one at a time by individual MH steps. The Metropolis proposal values for the r_l's are drawn from a Log-Normal distribution, centered around the previous value:

$$r_l^{*,(i+1)} \sim LN\left(\log(r_l^{(i)}),\ \sigma_{r_l}^2\right), \quad l = 1, \ldots, L.$$

The proposal scaling parameters $\sigma_{r_l}^2$ are determined *adaptively* during the burn-in period, to ensure an efficient mixing of the Markov Chain. To elaborate, every $B \in \mathbb{N}$ iterations during burn-in, the proposal log scaling parameters $s_{r_l}^{(i)} := \log(\sigma_{r_l}^{(i)})$ are updated, depending on the acceptance rate $a_{r_l,B}^{(i)}$ of the B previous samples $r_l^{(i-B+1)}, \ldots, r_l^{(i)}$. If the acceptance rate is too high, the log scaling parameter is increased, and vice versa:

$$s_{r_l}^{(i+1)} := \begin{cases} s_{r_l}^{(i)} + \delta_{\text{adapt}}^{(i)}, & a_{r_l,B}^{(i)} > a^* \\ s_{r_l}^{(i)} - \delta_{\text{adapt}}^{(i)}, & a_{r_l,B}^{(i)} < a^*, \end{cases}$$

where $a^* \in (0,1)$ is a *target acceptance rate* and $\delta_{\text{adapt}}^{(i)} > 0$ is an *adaption step width*. Note that the proposal scaling parameters depend on the iteration i only during the burn-in. After a burn-in period of length N_{burn}, they are fixed to $\sigma_{r_l}^2 = \exp(2s_{r_l}^{(N_{\text{burn}})})$. For all adaptive proposal scalings in this work, we follow the general recommendations from Roberts and Rosenthal (2009) and use a target acceptance rate of $a^* = 0.44$, an adaption step width of $\delta_{\text{adapt}}^{(i)} = \min(0.05, 1/\sqrt{i})$ and an *adaption batch size* of $B = 50$. See Section 3 in Roberts and Rosenthal (2009) for further details on Adaptive Metropolis-Within-Gibbs.

The x_l's are also with individual MH steps, with proposal values $x_l^{*,(i+1)}$ on $[0, \pi]$ given by

$$\tilde{x}_l^{*,(i+1)} := x_l^{(i)} + \epsilon_{x_l}, \quad \epsilon_{x_l} \sim \text{Unif}([-\delta_{x_l}, \delta_{x_l}]),$$

$$x_l^{*,(i+1)} := \begin{cases} \tilde{x}_l^{*,(i+1)} + \pi, & \tilde{x}_l^{*,(i+1)} < 0, \\ \tilde{x}_l^{*,(i+1)} - \pi, & \tilde{x}_l^{*,(i+1)} > \pi, \\ \tilde{x}_l^{*,(i+1)}, & \text{else}, \end{cases}$$

and proposal scaling parameters $\delta_{x_l} := \pi l/(l + 2\sqrt{n})$. This proposal suggestion is due to Choudhuri et al. (2004a), where it has been used for the mass location parameters of a Dirichlet process. Whereas these proposals yielded feasible mixing properties in all considered examples, it may be of interest for future research whether the mixing could be improved even further by sampling the x_l's in a blocked MH-step, with proposal covariance matrix being continuously estimated based on the samples obtained so far (see Section 2 in Roberts and Rosenthal (2009) for further details on this idea).

The components $\varphi_{l,1}, \ldots, \varphi_{l,d^2-1}$ of $\vec{\varphi}_l$ are sampled *blocked*, i.e. the Metropolis-Hastings step is conducted for the vectors $\vec{\varphi}_l$ and not for each of their components. Besides a computational advantage (in terms of fewer likelihood evaluations), the blocked sampling is used to take the posterior correlation between the $\varphi_{l,j}$'s into account. The proposal $\underline{\varphi}_l^{*,(i+1)}$ for $\underline{\varphi}_l$ on $\mathcal{I}$ is given as

$$\tilde{\underline{\varphi}}_l^{*,(i+1)} := \underline{\varphi}_l^{(i)} + \underline{\epsilon}_{\underline{\varphi}_l}, \quad \epsilon_{\underline{\varphi}_l,j} \overset{\text{ind.}}{\sim} \text{Unif}\left([-\delta_{\underline{\varphi}_l} \max(\mathcal{I}_j), \delta_{\underline{\varphi}_l} \max(\mathcal{I}_j)]\right)$$

$$\varphi_{l,j}^{*,(i+1)} := \begin{cases} \tilde{\varphi}_{l,j}^{*,(i+1)} + \max(\mathcal{I}_j), & \tilde{\varphi}_{l,j}^{*,(i+1)} < 0, \\ \tilde{\varphi}_{l,j}^{*,(i+1)} - \max(\mathcal{I}_j), & \tilde{\varphi}_{l,j}^{*,(i+1)} > \max(\mathcal{I}_j), \\ \tilde{\varphi}_{l,j}^{*,(i+1)}, & \text{else} \end{cases}$$

for $l = 1, \ldots, L$ and $j = 1, \ldots, d^2 - 1$. The proposal scaling parameters $\delta_{\vec{\varphi}_l} > 0$ are determined adaptively during burn-in. For the sake of simplicity and to save computational costs, the proposal scaling parameters do not depend on j.

5.2.3. Implementation

The MCMC algorithm of Section 5.2.2 is implemented in the `R` programming language (R Core Team, 2018). The computationally most demanding part is the evaluation of Whittle's Likelihood that has to be performed in every evaluation of the posterior density in the Metropolis-Hastings steps within the Gibbs sampler. To conduct the evaluation, it is also needed to reconstruct the spectral density matrix $\boldsymbol{f}$ from the parametrization $\underline{\Theta} = (\tilde{\underline{\Theta}}_{\boldsymbol{\Phi}}, k)$. To make the runtime of the algorithm feasible, all performance critical parts have been written in C++ and incorporated in `R` with the `Rcpp` and `RcppArmadillo` (Eddelbuettel and Francois, 2011) packages. Extensive usage has been made of the efficient linear algebra routines and the data structures provided by the `RcppArmadillo` package. As a prominent example, the data structure `arma::cx_cube` has been used to store the values of the spectral density matrix at the Fourier frequencies $(\boldsymbol{f}(\omega_1), \ldots, \boldsymbol{f}(\omega_{\lceil n/2 \rceil - 1}))$ as a complex array of dimension $(\lceil n/2 \rceil - 1) \times d \times d$. Similarly, the data structure for storing the current value of $(\boldsymbol{U}_1, \ldots, \boldsymbol{U}_L)$ is a complex array of dimension $L \times d \times d$. The current values of $(r_1, \ldots, r_L)$ are stored in a regular vector of length L.

By (5.5), the mixture weight matrices for $\boldsymbol{f}$ from (5.2) are given by $\boldsymbol{\Phi}(\mathcal{I}_{j,k}) = \sum_{l=1}^{L} \mathbb{1}_{\mathcal{I}_{j,k}}(x_l) r_l \boldsymbol{U}_l$, which can readily be computed for $j = 1, \ldots, k$ in $O(n+k)$ steps. To speed up the computation of the mixture, we pre-compute and store the values of the beta densities $b(\xi_j, i, k-i+1)$ (or their truncated versions $b_{\tau_1}^{\tau_2}(\xi_j, i, k-i+1)$ from (5.4), respectively) with $\xi_j = \omega_j/\pi$ for $j = 1, \ldots, \lceil n/2 \rceil - 1$ and $i = 1, \ldots, k$ and $k = 1, \ldots, k_{\max}$.

The Fourier coefficients $\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2 \rfloor}$ are computed once at the beginning of the algorithm in $O(n \log n)$ complexity with the Fast Fourier Transform algorithm, of which an implementation is provided in `R` in the `fft` function of the `stats` package. For the evaluation of Whittle's Likelihood, the matrices $\boldsymbol{f}(\omega_j)$ have to be inverted. To keep the algorithm numerically stable, we keep track of the condition numbers $\kappa(\boldsymbol{f}(\omega_j)) = \frac{\lambda_{\max}(\boldsymbol{f}(\omega_j))}{\lambda_{\min}(\boldsymbol{f}(\omega_j))}$. If a parameter $\underline{\Theta}^*$ is proposed in a Metropolis-Hastings step such that for the corresponding spectral density matrix $\boldsymbol{f}^*$ the condition number at a Fourier frequency is too high, i.e. $\kappa(\boldsymbol{f}^*(\omega_j)) > C_{\text{cond}}$ for some j, we reject $\underline{\Theta}^*$ in the Metropolis-Hastings for the sake of numerical stability. By default, we use $C_{\text{cond}} = 10^{12}$.

The implementation of the MCMC algorithm of Section 5.2.2, along with the VAR procedure (see the upcoming Section 6.2) and the semiparametric extension from the upcoming Section 8.2 comprises around 6,500 lines of `R` code and around `650` lines of `C++` code.

5.3. Forecasting

The Bayesian framework lends itself naturally to forecasting. Recently, Bayesian time series forecasting has gained new momentum with several leading tech companies publishing software packages for this purpose. These include, among others, the `prophet` package (Taylor and Letham, 2017) by Facebook (implementing a Bayesian additive seasonal time series model), the `CausalImpact` package (Brodersen et al., 2015) by Google (implementing a Bayesian approach to causal impact estimation in time series) and the `anomalyDetection` package (Boehmke et al., 2018) by Twitter for network anomaly detection. However, all of those approaches rely on structural time series models such as ARIMA or VARMA models and are intrinsically parametric. We will show in the following that forecasting can readily be incorporated in the nonparametric approach from this work as well.

To produce a one-step forecast in the Bayesian framework, one is interested in the predictive density $p(\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n)$ which can be written as a mixture of the conditional predictive density (given $\boldsymbol{f}$) with respect to the posterior distribution of $\boldsymbol{f}$:

$$p(\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n) = \int p(\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f})P(d\boldsymbol{f}|\underline{Z}_1,\dots,\underline{Z}_n). \tag{5.11}$$

Assume that we have a sample $\boldsymbol{f}^{(1)},\dots,\boldsymbol{f}^{(M)}$ from the posterior distribution $P(d\boldsymbol{f}|\underline{Z}_1,\dots,\underline{Z}_n)$ at hand, as e.g. obtained by the MCMC method of Section 5.2. Assume further that for each $\boldsymbol{f}^{(i)}$, we can draw a sample $\underline{Z}_{n+1}^{(i)}$ from the conditional predictive distribution with density $p(\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f})$. Then it follows from (5.11) that $\underline{Z}_{n+1}^{(1)},\dots,\underline{Z}_{n+1}^{(M)}$ is a sample from the predictive distribution $P(d\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n)$. In the following, we will show that the conditional predictive distribution is in fact multivariate normal, such that random samples thereof can readily be obtained.

Denote by $\underline{X}_{n+1} := \mathrm{vec}(\underline{Z}_1,\dots,\underline{Z}_{n+1}) \in \mathbb{R}^{(n+1)d}$. Recall the real-valued formulation (4.16) of Whittle's likelihood. If $n+1$ observations are involved in the likelihood, the frequency domain covariance matrix is $\boldsymbol{D}_{(n+1)d} = \boldsymbol{D}_{(n+1)d}[\boldsymbol{f}]$, which is defined as in (4.16) using $\boldsymbol{f}$ evaluated at the Fourier frequencies $\omega_{j,n+1} = \frac{2\pi j}{n+1}$ for $j = 0,\dots,\lfloor\frac{n+1}{2}\rfloor$. A back-transformation of Whittle's likelihood in the time domain yields

$$p(\underline{X}_{n+1}|\boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^{(n+1)d}|\boldsymbol{D}_{(n+1)d}|}} \exp\left(-\frac{1}{2}\underline{X}_{n+1}^T\boldsymbol{G}_{(n+1)d}\underline{X}_{n+1}\right) \tag{5.12}$$

with the inverse $(n+1)d \times (n+1)d$ time domain covariance matrix

$$\boldsymbol{G}_{(n+1)d} = \boldsymbol{F}_{(n+1)d}\boldsymbol{D}_{(n+1)d}^{-1}\boldsymbol{F}_{(n+1)d}^T.$$

We denote by $\boldsymbol{G}(i,j) \in \mathbb{R}^{d\times d}$ the block components of $\boldsymbol{G}_{(n+1)d}$ for $i,j = 1,\dots,n+1$ (dropping the subindex $(n+1)d$ of the block components for the sake of notational convenience). From

Lemma 4.15 (b) it follows that $\boldsymbol{D}_{(n+1)d}$ is symmetric, which readily implies that $\boldsymbol{G}_{(n+1)d}$ is symmetric and hence $\boldsymbol{G}(i,j) = \boldsymbol{G}(j,i)^T$ holds for all $i,j = 1,\dots,n+1$. Furthermore, $\boldsymbol{G}(i,i)$ is symmetric positive definite for $i = 1,\dots,n+1$. Conditioning (5.12) on $\underline{Z}_1,\dots,\underline{Z}_n$, we obtain

$$\begin{aligned} p(\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f}) &\propto \exp\Bigg(-\frac{1}{2}\Big[\underline{Z}_{n+1}^T\boldsymbol{G}(n+1,n+1)\underline{Z}_{n+1} \\ &\qquad + \sum_{i=1}^n \underline{Z}_i^T\boldsymbol{G}(i,n+1)\underline{Z}_{n+1} + \sum_{j=1}^n \underline{Z}_{n+1}^T\boldsymbol{G}(n+1,j)\underline{Z}_j\Big]\Bigg) \\ &= \exp\Bigg(-\frac{1}{2}\Big[\underline{Z}_{n+1}^T\boldsymbol{G}(n+1,n+1)\underline{Z}_{n+1} \\ &\qquad + 2\sum_{i=1}^n \underline{Z}_i^T\boldsymbol{G}(i,n+1)\underline{Z}_{n+1}\Big]\Bigg). \end{aligned}$$

Now consider the transformation $Y_{n+1} := \boldsymbol{G}(n+1,n+1)^{1/2}\underline{Z}_{n+1}$. With this transformation and with

$$\tilde{\underline{b}}_{n+1} := \sum_{i=1}^n \boldsymbol{G}(n+1,n+1)^{-1/2}\boldsymbol{G}(n+1,i)\underline{Z}_i \in \mathbb{R}^d$$

it follows

$$\begin{aligned} p(Y_{n+1}|Z_1,\dots,Z_n,\boldsymbol{f}) &\propto \exp\left(-\frac{1}{2}\left[\underline{Y}_{n+1}^T\underline{Y}_{n+1} + 2\tilde{\underline{b}}_{n+1}^T\underline{Y}_{n+1}\right]\right) \\ &\propto \exp\left(-\frac{1}{2}\left[\underline{Y}_{n+1}^T\underline{Y}_{n+1} + 2\tilde{\underline{b}}_{n+1}^T\underline{Y}_{n+1} + \tilde{\underline{b}}_{n+1}^T\tilde{\underline{b}}_{n+1}\right]\right) \\ &= \exp\left(-\frac{1}{2}\left(\underline{Y}_{n+1} + \tilde{\underline{b}}_{n+1}\right)^T\left(\underline{Y}_{n+1} + \tilde{\underline{b}}_{n+1}\right)\right) \end{aligned}$$

and conducting a back-substitution $\underline{Z}_{n+1} = \boldsymbol{G}(n+1,n+1)^{-1/2}\underline{Y}_{n+1}$ reveals

$$\begin{aligned} &p(\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f}) \\ &\propto \exp\left(-\frac{1}{2}\left(\boldsymbol{G}(n+1,n+1)^{1/2}\underline{Z}_{n+1} + \tilde{\underline{b}}_{n+1}\right)^T\left(\boldsymbol{G}(n+1,n+1)^{1/2}Z_{n+1} + \tilde{\underline{b}}_{n+1}\right)\right) \\ &= \exp\left(-\frac{1}{2}\left(\underline{Z}_{n+1} + \underline{b}_{n+1}\right)^T\boldsymbol{G}(n+1,n+1)\left(\underline{Z}_{n+1} + \underline{b}_{n+1}\right)\right) \end{aligned}$$

with

$$\underline{b}_{n+1} = \boldsymbol{G}(n+1,n+1)^{-1/2}\tilde{\underline{b}}_{n+1} = \sum_{i=1}^n \boldsymbol{G}(n+1,n+1)^{-1}\boldsymbol{G}(n+1,i)\underline{Z}_i.$$

This shows that the conditional distribution of $\underline{Z}_{n+1}$ given $\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f}$ is multivariate normal with mean $-\underline{b}_{n+1}$ and covariance matrix $\boldsymbol{G}(n+1,n+1)^{-1}$:

$$\underline{Z}_{n+1}|\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f} \sim N_d\left(-\underline{b}_{n+1},\boldsymbol{G}(n+1,n+1)^{-1}\right). \tag{5.13}$$

The above approach can readily be extended to k-step ahead forecasts for any $k \geq 1$. Indeed, similar to (5.11), we first observe

$$p(\underline{Z}_{n+1},\dots,\underline{Z}_{n+k}|\underline{Z}_1,\dots,\underline{Z}_n) = \int p(\underline{Z}_{n+1},\dots,\underline{Z}_{n+k}|\underline{Z}_1,\dots,\underline{Z}_n,\boldsymbol{f})dP(\boldsymbol{f}|\underline{Z}_1,\dots,\underline{Z}_n)$$

and thus it suffices again to draw from $p(\underline{Z}_{n+1}, \ldots, \underline{Z}_{n+k} | \underline{Z}_1, \ldots, \underline{Z}_n, \boldsymbol{f})$ for any $\boldsymbol{f}$. On the other hand, we also have the equality

$$p(\underline{Z}_{n+1}, \ldots, \underline{Z}_{n+k} | \underline{Z}_1, \ldots, \underline{Z}_n, \boldsymbol{f}) = p(\underline{Z}_{n+1} | \underline{Z}_1, \ldots, \underline{Z}_n, \boldsymbol{f}) p(\underline{Z}_{n+2} | \underline{Z}_1, \ldots, \underline{Z}_{n+1}, \boldsymbol{f}) \\ \times \ldots \times p(\underline{Z}_{n+k} | \underline{Z}_1, \ldots, \underline{Z}_{n+k-1}, \boldsymbol{f}),$$

giving rise to an *iterative* procedure to draw from $p(\underline{Z}_{n+1}, \ldots, \underline{Z}_{n+k} | \underline{Z}_1, \ldots, \underline{Z}_n, \boldsymbol{f})$: Start with drawing $\underline{Z}_{n+1}$ from $p(\underline{Z}_{n+1} | \underline{Z}_1, \ldots, \underline{Z}_n, \boldsymbol{f})$ using (5.13). Then draw $\underline{Z}_{n+2}$ from

$$\underline{Z}_{n+2} | \underline{Z}_1, \ldots, \underline{Z}_{n+1}, \boldsymbol{f} \sim N_d \left(-\underline{b}_{n+2}, \tilde{\boldsymbol{G}}(n+2, n+2)^{-1} \right)$$

with

$$\underline{b}_{n+2} = \sum_{i=1}^{n+1} \tilde{\boldsymbol{G}}(n+2, n+2)^{-1} \tilde{\boldsymbol{G}}(n+2, i) \underline{Z}_i$$

and the inverse $(n+2)d \times (n+2)d$ time domain covariance matrix

$$\boldsymbol{G}_{(n+2)d} = \boldsymbol{F}_{(n+2)d} \boldsymbol{D}_{(n+2)d}^{-1} \boldsymbol{F}_{(n+2)d}^T$$

and its block components $\tilde{\boldsymbol{G}}(i, j) \in \mathbb{R}^{d \times d}$ for $i, j = 1, \ldots, n+2$. This scheme can be continued k times to finally obtain a sample of $p(\underline{Z}_{n+1}, \ldots, \underline{Z}_{n+k} | \underline{Z}_1, \ldots, \underline{Z}_n, \boldsymbol{f})$.

5.4. Missing Values

It is conceptually straightforward to accommodate missing values within the Bayesian framework. Assume that we have data $\underline{Z}_1, \ldots, \underline{Z}_n$ and that there are missing values at positions $\mathcal{I} \subset \{1, \ldots, n\}$. Denote the missing values by $\underline{Z}_{\mathcal{I}} := \{\underline{Z}_i \colon i \in \mathcal{I}\}$ and the actually observed data by $\underline{Z}_{-\mathcal{I}} = \{\underline{Z}_1 \ldots, \underline{Z}_n\} \setminus \underline{Z}_{\mathcal{I}}$. Following McCulloch and Tsay (1994), the missing values can be treated as random and included in the inference as latent variables. The object of interest is then the joint posterior distribution $P(d\boldsymbol{f}, d\underline{Z}_{\mathcal{I}} | \underline{Z}_{-\mathcal{I}})$. To sample from it, the MCMC algorithm from Section 5.2 can readily be extended by an additional Gibbs step to draw from the full conditional $P(d\underline{Z}_i | \underline{Z}_{-i}, \boldsymbol{f})$ for each missing value $\underline{Z}_i \in \underline{Z}_{\mathcal{I}}$, where $\underline{Z}_{-i} = \{\underline{Z}_1, \ldots, \underline{Z}_{i-1}, \underline{Z}_{i+1}, \ldots, \underline{Z}_n\}$. With the same calculations as in Section 5.3, one can derive the following closed-form expression for the full conditionals of the missing values:

$$\underline{Z}_i | \underline{Z}_{-i}, \boldsymbol{f} \sim N_d \left(-\underline{b}_i, \boldsymbol{G}_{nd}(i, i)^{-1} \right), \quad \underline{b}_i := \sum_{j \neq i} \boldsymbol{G}_{nd}(i, i)^{-1} \boldsymbol{G}_{nd}(i, j) \underline{Z}_j,$$

with inverse $nd \times nd$ time domain covariance matrix $\boldsymbol{G}_{nd} = \boldsymbol{F}_{nd} \boldsymbol{D}_{nd}^{-1} \boldsymbol{F}_{nd}^T$ and its block components $\boldsymbol{G}_{nd}(i, j) \in \mathbb{R}^{d \times d}$.

6.

Numerical Illustration

6.1. Methodology

We start by explaining how an estimate of the true spectral density $\boldsymbol{f}_0$ can be generated from a posterior sample. The idea is to take the pointwise posterior median of all components (or of their real and imaginary parts in case of complex values). To elaborate, consider the vectorization operator $\mathcal{G}\colon \mathcal{S}_d \to \mathbb{R}^{d^2}$ that maps each hermitian matrix $\boldsymbol{A} = (a_{rs})_{r,s=1}^d$ to a vector $\mathcal{G}\boldsymbol{A}$ consisting of the diagonal elements $a_{11}, \dots, a_{dd}$ and the real and imaginary parts of the entries $\{a_{rs}\colon r < s\}$ above the diagonal. Denote the corresponding inverse transformation by $\mathcal{G}^{-1}\colon \mathcal{G}(\mathcal{S}_d) \to \mathcal{S}_d$, where $\mathcal{G}(\mathcal{S}_d) \subset \mathbb{R}^{d^2}$ denotes the image of $\mathcal{G}$.

Bayes Estimators

Assume that we have a sample $\boldsymbol{f}^{(1)}, \dots, \boldsymbol{f}^{(N)}\colon [0,\pi] \to \mathcal{S}_d^+$ of spectral density functions at hand. Consider the vectorized versions $\underline{g}^{(j)} := \mathcal{G}\boldsymbol{f}^{(j)}\colon [0,\pi] \to \mathbb{R}^{d^2}$ for $j = 1, \dots, N$. Denote the components of $\underline{g}^{(j)}$ by $(g_1^{(j)}, \dots, g_{d^2}^{(j)})$. Denote the pointwise sample median function by $\hat{\underline{g}} = (\hat{g}_1, \dots, \hat{g}_{d^2})$, i.e. $\hat{g}_r(\omega)$ is defined as the median of $\{g_r^{(1)}(\omega), \dots, g_r^{(N)}(\omega)\}$ for $0 \le \omega \le \pi$ and $r = 1, \dots, d^2$. Then an estimate of $\boldsymbol{f}_0$ can be obtained by

$$\hat{\boldsymbol{f}}_0(\omega) := \mathcal{G}^{-1}\hat{\underline{g}}(\omega), \quad 0 \le \omega \le \pi. \tag{6.1}$$

We call $\hat{\boldsymbol{f}}_0$ the *pointwise median spectral density* of $\boldsymbol{f}^{(1)}, \dots, \boldsymbol{f}^{(N)}$. In subsequent Sections, we will often use $\hat{\boldsymbol{f}}_0$ as a Bayes estimator of $\boldsymbol{f}_0$ (as already notationally indicated by the circumflex). If $\boldsymbol{f}_0$ is known, we can gauge the goodness of $\hat{\boldsymbol{f}}_0$ in terms of the $\mathbb{L}^1$-error

$$\|\hat{\boldsymbol{f}}_0 - \boldsymbol{f}_0\|_1 := \int_0^\pi \|\hat{\boldsymbol{f}}_0(\omega) - \boldsymbol{f}_0(\omega)\| d\omega \approx \frac{1}{\lceil n/2 \rceil - 1} \sum_{j=1}^{\lceil n/2 \rceil - 1} \|\hat{\boldsymbol{f}}_0(\omega_j) - \boldsymbol{f}_0(\omega_j)\| \tag{6.2}$$

and the $\mathbb{L}^2$-error

$$\begin{aligned} \|\hat{\boldsymbol{f}}_0 - \boldsymbol{f}_0\|_2 &:= \left(\int_0^\pi \|\hat{\boldsymbol{f}}_0(\omega) - \boldsymbol{f}_0(\omega)\|^2 d\omega \right)^{1/2} \\ &\approx \left(\frac{1}{\lceil n/2 \rceil - 1} \sum_{j=1}^{\lceil n/2 \rceil - 1} \|\hat{\boldsymbol{f}}_0(\omega_j) - \boldsymbol{f}_0(\omega_j)\|^2 \right)^{1/2}. \end{aligned} \tag{6.3}$$

Pointwise Credible Regions

Besides an estimator of $\boldsymbol{f}_0$, we are also interested in investigating how much variability the posterior distribution contains. One might ask e.g. whether most of the posterior probability mass is allocated within a small neighborhood of $\hat{\boldsymbol{f}}_0$, or whether the probability mass is spread out widely in the function space. One possible way of investigating this constitutes the computation of the *pointwise a-quantile spectral density* $\hat{\boldsymbol{f}}_0^{[a]}$ functions for $0 < a < 1$. This is done by the very same construction as $\hat{\boldsymbol{f}}_0$ (employing the a-quantile of $\{g_r^{(1)}(\omega), \dots, g_r^{(N)}(\omega)\}$ rater than the median). As an example, we will often consider $\hat{\boldsymbol{f}}_0^{[0.05]}$ and $\hat{\boldsymbol{f}}_0^{[0.95]}$. The region

$$\mathcal{C}_{\mathrm{pw}}(\omega|0.9) := \left\{ t\hat{\boldsymbol{f}}_0^{[0.05]}(\omega) + (1-t)\hat{\boldsymbol{f}}_0^{[0.95]}(\omega) \colon 0 \leq t \leq 1 \right\}, \quad 0 \leq \omega \leq \pi, \tag{6.4}$$

between the function graphs of $\hat{\boldsymbol{f}}_0^{[0.05]}$ and $\hat{\boldsymbol{f}}_0^{[0.95]}$ has the interpretation of being a *pointwise 90% credibility region* – that is, for every $0 \leq \omega \leq \pi$, it holds true that the posterior probability of $\boldsymbol{f}(\omega)$ to lie in $\mathcal{C}_{\mathrm{pw}}(\omega|0.9)$ is at least 90%. It is important to note that this statement does not hold true uniformly over $0 \leq \omega \leq \pi$. In particular, for the graph area $\mathcal{C}_{\mathrm{pw}}(0.9) := \{\mathcal{C}_{\mathrm{pw}}(\omega|0.9) \colon 0 \leq \omega \leq \pi\}$ it does *not* hold that $\boldsymbol{f} \in \mathcal{C}_{\mathrm{pw}}(0.9)$ with posterior probability at least 90%. The reason is that pointwise credible regions do not take the multiple testing problem into account that arises when considering multiple frequencies ω simultaneously.

Uniform Credible Regions

One possible remedy for the aforementioned drawback of pointwise credible regions constitutes the usage of *uniform credible regions*. These have been used in Häfner and Kirch (2017) (see also Neumann and Polzehl (1998) and Kirch et al. (2017)) for univariate functions and we will present a generalization to the matrix-valued case. To elaborate, denote by $\mathcal{H} \colon \mathcal{S}_d^+ \to \mathbb{R}^{d^2}$ the transformation that maps each Hermitian positive definite matrix $\boldsymbol{A} = (a_{rs})_{r,s=1}^d$ to a vector $\mathcal{H}\boldsymbol{A}$ consisting of the logarithmized diagonal elements $\log a_{11}, \dots, \log a_{dd}$ and the (non-logarithmized) real and imaginary parts of the entries $\{a_{rs} \colon r < s\}$ above the diagonal. For the transformed versions $\underline{h}^{(j)} = (h_1^{(j)}, \dots, h_{d^2}^{(j)}) := \mathcal{H}\boldsymbol{f}^{(j)}$ for $j = 1, \dots, N$, denote the pointwise sample median function by $\hat{\underline{h}} := (\hat{h}_1, \dots, \hat{h}_{d^2})$. Let $\hat{\underline{\sigma}} := (\hat{\sigma}_1, \dots, \hat{\sigma}_{d^2})$ with $\hat{\sigma}_r(\omega)$ being the median absolute deviation of $\{h_r^{(1)}(\omega), \dots, h_r^{(N)}(\omega)\}$ for $0 \leq \omega \leq \pi$ and $r = 1, \dots, d^2$. For $0 < a < 1/2$, determine ξ_a as the smallest positive number such that

$$\frac{1}{N}\sum_{j=1}^{N} \mathbb{1}\left\{ \max_{\substack{0\leq\omega\leq\pi \\ r=1,\dots,d^2}} \frac{\left|h_r^{(j)}(\omega) - \hat{h}_r(\omega)\right|}{\hat{\sigma}_r(\omega)} \leq \xi_a \right\} \geq 1 - a.$$

Let $\hat{\underline{h}}^{[a/2]} := \hat{\underline{h}} - \xi_a\hat{\underline{\sigma}}$ and $\hat{\underline{h}}^{[1-a/2]} := \hat{\underline{h}} + \xi_a\hat{\underline{\sigma}}$ and $\tilde{\boldsymbol{f}}_0^{[a/2]} := \mathcal{H}^{-1}\hat{\underline{h}}^{[a/2]}$ as well as $\tilde{\boldsymbol{f}}_0^{[1-a/2]} := \mathcal{H}^{-1}\hat{\underline{h}}^{[1-a/2]}$. Then the region between the function graphs of $\tilde{\boldsymbol{f}}_0^{[a/2]}$ and $\tilde{\boldsymbol{f}}_0^{[1-a/2]}$ is called *uniform* $(1-a)$*-credibility region* and will be denoted by $\mathcal{C}_{\mathrm{uni}}(\omega|1-a)$. As an example, we will often consider $a = 0.1$, in which case

$$\mathcal{C}_{\mathrm{uni}}(\omega|0.9) := \left\{ t\tilde{\boldsymbol{f}}_0^{[0.05]}(\omega) + (1-t)\tilde{\boldsymbol{f}}_0^{[0.95]}(\omega) \colon 0 \leq t \leq 1 \right\}, \quad 0 \leq \omega \leq \pi. \tag{6.5}$$

By construction, it holds that $\boldsymbol{f} \in \mathcal{C}_{\text{uni}}(\omega|0.9)$ with (empirical) posterior probability of at least 0.9. It is of interest to investigate the frequentist coverage properties of the uniform credible sets, in particular whether the posterior credibility matches (at least approximately) the empirical coverage.

6.2. Simulation Study

In this section, we evaluate the performance of our procedure with simulated data. We employ the Bernstein-Hpd-Gamma prior from (5.2) for $\boldsymbol{f}$, where the prior probability mass function for the polynomial degree k is chosen as $p(k) \propto \exp(-0.01k \log k)$ (c.f. Table 5.1). As a prior for $\boldsymbol{\Phi}$, we use an $A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$ process (as considered in Section 3.4.2). We choose process parameters $\eta \equiv d = 2$ and $\omega \equiv d = 2$ and $\boldsymbol{\Sigma} \equiv 10^4 \boldsymbol{I}_2$. Recall from (3.24) that the prior mean of $\boldsymbol{\Phi}$ corresponds to the spectral measure of White Noise, with constant spectral density. Furthermore, the underlying Poisson Process mean measure ν from (3.23) of $\boldsymbol{\Phi}$ simplifies in this case to

$$\nu(dx, d\boldsymbol{U}, dr) = \frac{2\Gamma(4)}{\tilde{\Gamma}_2(2)} \frac{\exp(-10^{-4}r)}{r} dr = \frac{12}{\pi} \frac{\exp(-10^{-4}r)}{r} dr,$$

where the right hand side does not depend on $\boldsymbol{U}$. In this sense, $\boldsymbol{\Phi}$ is *isotropic*, i.e. all "directions" $\boldsymbol{U} \in \mathbb{S}_d^+$ are equally likely to contribute to $\boldsymbol{\Phi}$ under the prior. The Bernstein polynomial basis is truncated as in (5.4), with $\tau_l = 0.1$ and $\tau_r = 0.9$, to improve the mixture properties at the boundary. Along with this prior, we employ Whittle's Likelihood (5.3). Inference is conducted with the Markov Chain Monte Carlo (MCMC) algorithm from Section 5.2.

The Markov Chain is run for a total number of 80,000 iterations, where the first 30,000 iterations are discarded as *burn-in* period to ensure that the chain "converged" (i.e. reached states of actually sampling from the posterior distribution, see Section 6.3 in Christensen et al. (2011)). These numbers have been determined with convergence diagnostics that have been conducted in preliminary pilot runs for a few individual realizations. These diagnostics include a visual inspection of the log posterior trace plot and a validation of the results from chains with different starting values. It may be noted that a visual inspection of the parameter trace plots is not well suited for convergence diagnostics, because the marginal posterior of the parameters is often multimodal.

To decrease the dependence within the sample and to reduce memory consumption, the remaining 50,000 are (equidistantly) *thinned* by a factor of 5, such that a sample of size 10,000 remains (see Section 6.3 in Christensen et al. (2011)). By preliminary pilot runs, we found that values of $k_{\max} = 300$ for the maximum polynomial degree and $L = 20$ for the series truncation parameter of $\boldsymbol{\Phi}$ were sufficient for all the considered examples. This procedure for spectral density inference will be called *the NP procedure* (where NP stands for nonparametric method) in the following.

Comparison Method: Bayesian Vector Autoregression

We compare the NP procedure to a Bayesian Vector Autoregression (VAR). To elaborate, a (Gaussian) VAR model of order $p \geq 0$ is formulated as follows:

$$\underline{Z}_t = \sum_{j=1}^{p} \boldsymbol{B}_j \underline{Z}_{t-1} + \underline{e}_t, \quad \underline{e}_t \overset{\text{iid}}{\sim} N_d(\underline{0}, \boldsymbol{\Sigma}), \quad t = p+1, \ldots, n \tag{6.6}$$

with innovation covariance matrix $\boldsymbol{\Sigma} \in \mathcal{S}_d^+(\mathbb{R})$ and coefficient matrices $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_p \in \mathbb{R}^{d \times d}$, where $\boldsymbol{B}_p \neq \boldsymbol{0}$. For the model (6.6) to be stationary, the constraint $|\boldsymbol{I}_d - \sum_{j=1}^{p} \boldsymbol{B}_j z^j| \neq 0$ is assumed for all $z \in \mathbb{C}$ with $|z| \leq 1$ (see Section 11.3 in Brockwell and Davis (1991)). In this case, the spectral density is given by

$$\boldsymbol{f}_{\text{VAR}}(\omega | \boldsymbol{B}_1, \ldots, \boldsymbol{B}_p, \boldsymbol{\Sigma}) = \frac{1}{2\pi} \left(\boldsymbol{I}_d - \sum_{j=1}^{p} \boldsymbol{B}_j e^{ij\omega} \right)^{-1} \boldsymbol{\Sigma} \left(\boldsymbol{I}_d - \sum_{j=1}^{p} \boldsymbol{B}_j^T e^{-ij\omega} \right)^{-1} \tag{6.7}$$

for $0 \leq \omega \leq \pi$. The first p observations $\underline{Z}_1, \ldots, \underline{Z}_p$ in model (6.6) are considered to be fixed. The conditional likelihood for the remaining observations $\underline{Z}_{p+1}, \ldots, \underline{Z}_n$ is

$$\begin{aligned} &p_{\text{VAR}}(\underline{Z}_{p+1} \ldots, \underline{Z}_n | \underline{Z}_1, \ldots, \underline{Z}_p, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_p, \boldsymbol{\Sigma}) \\ &= \prod_{j=p+1}^{n} p_{\text{VAR}}(\underline{Z}_j | \underline{Z}_{j-1}, \ldots, \underline{Z}_{j-p}, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_p, \boldsymbol{\Sigma}) \end{aligned}$$

with

$$\begin{aligned} &p_{\text{VAR}}(\underline{Z}_j | \underline{Z}_{j-1}, \ldots, \underline{Z}_{j-p-1}, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_p, \boldsymbol{\Sigma}) \\ &= \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp \left\{ -\frac{1}{2} \left(\underline{Z}_j - \sum_{l=1}^{p} \boldsymbol{B}_l \underline{Z}_{j-l} \right)^T \boldsymbol{\Sigma}^{-1} \left(\underline{Z}_j - \sum_{l=1}^{p} \boldsymbol{B}_l \underline{Z}_{j-l} \right) \right\} \end{aligned}$$

for $j = p+1, \ldots, n$. The VAR equation (6.6) can also be written in an equivalent vectorized notation (see Section 2.2.3 in Koop and Korobilis (2010)) as

$$\underline{Z}_t = \boldsymbol{Y}_t \underline{b} + \underline{e}_t, \quad t = p+1, \ldots, n, \tag{6.8}$$

with regressor matrix

$$\boldsymbol{Y}_t = \begin{pmatrix} \underline{Y}_t^T & 0 & \ldots & 0 \\ 0 & \underline{Y}_t^T & & \\ \vdots & & \ddots & \\ 0 & \ldots & & \underline{Y}_t^T \end{pmatrix} \in \mathbb{R}^{d \times pd^2}.$$

and regressors $\underline{Y}_t = (\underline{Z}_{t-1}^T, \ldots, \underline{Z}_{t-p}^T)^T \in \mathbb{R}^{pd}$ and vectorized coefficients

$$\underline{b} = (B_{1,1,1}, \ldots, B_{1,1,d}, B_{2,1,1}, \ldots, B_{p,d,d}) \in \mathbb{R}^{pd^2}.$$

The order p is not included into the Bayesian inference, but determined in a preliminary model selection step with Akaike's Information Criterion (AIC, see Akaike (1974)), based on a Yule-Walker estimate and the full likelihood of $\underline{Z}_1, \ldots, \underline{Z}_n$ for each VAR order in consideration. For

the parameters of the VAR(p) model, we employ the *Independent Normal-Inverse-Wishart* prior in parametrization (6.8), under which $\underline{b}$ and $\boldsymbol{\Sigma}$ are independent and distributed as

$$\underline{b} \sim N_{pd^2}(\underline{0}, \boldsymbol{V}), \quad \boldsymbol{\Sigma} \sim \text{Wish}^{-1}_{d\times d}(\nu, \boldsymbol{S}),$$

where the inverse Wishart distribution $\text{Wish}^{-1}_{d\times d}$ is as defined in (B.2) in the Appendix. Note that this prior specification does not enforce causality or invertibility of the parametric working model. The prior parameters are chosen as $\boldsymbol{V} = 10^4 \boldsymbol{I}_d$ and $\nu = 10^{-4}$ and $\boldsymbol{S} = 10^{-4}\boldsymbol{I}_d$. This choice can be interpreted as "vague", since the prior covariance $\boldsymbol{V}$ of $\underline{b}$ is "large" and the prior of $\boldsymbol{\Sigma}$ is improper. Posterior inference is conducted with a Gibbs sampler, where the full conditionals are available due to conjugacy properties of the prior and the Gaussian likelihood (see Section 2.2.3 in Koop and Korobilis (2010)). The Markov Chain is run for a total number of 80,000 iterations, with a burn-in period of 30,000 iterations and a thinned factor of 5, yielding a sample of size 10,000. We will refer to this inference method as *the VAR procedure* in the following.

Simulated Data

We consider simulated data drawn from the following bivariate VAR(2) model:

$$\underline{Z}_t = \begin{pmatrix} 0.5 & 0 \\ 0 & -0.3 \end{pmatrix} \underline{Z}_{t-1} + \begin{pmatrix} 0 & 0 \\ 0 & -0.5 \end{pmatrix} \underline{Z}_{t-2} + \underline{e}_t, \quad \underline{e}_t \overset{\text{iid}}{\sim} N_2 \left(\underline{0}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} \right). \tag{6.9}$$

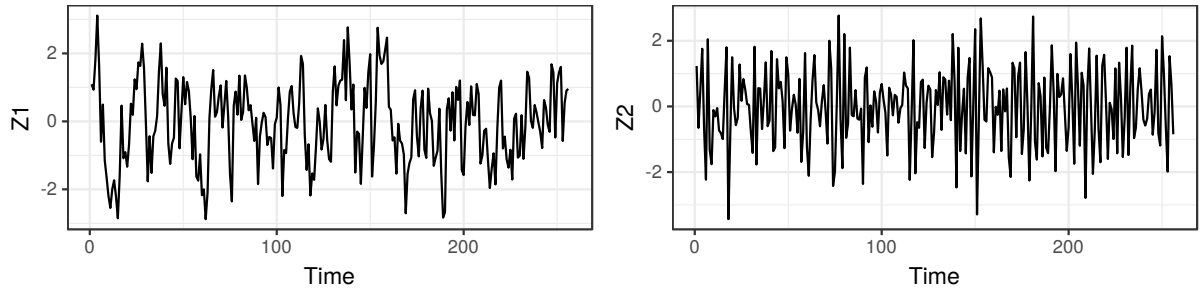
This model is chosen because it has already been analyzed in the literature (see Rosen and Stoffer (2007)). Furthermore, we will also consider the following bivariate VMA(1) model:

$$\underline{Z}_t = \underline{e}_t + \begin{pmatrix} -0.75 & 0.5 \\ 0.5 & 0.75 \end{pmatrix} \underline{e}_{t-1}, \quad \underline{e}_t \overset{\text{iid}}{\sim} N_2 \left(\underline{0}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right), \tag{6.10}$$

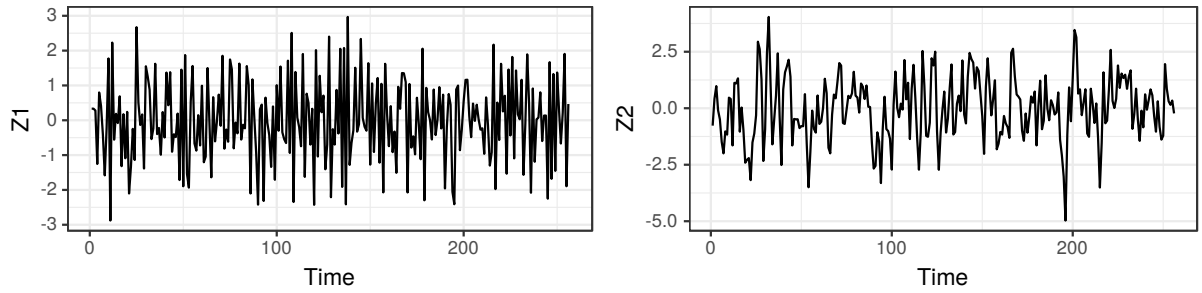
which is chosen because it is a simple example of a linear multivariate time series that does not belong to the family of VAR models. Exemplary realizations from models (6.9)–(6.10) are shown in Figure 6.1. We consider the sample sizes $n = 256$, $n = 512$ and $n = 1024$. For each sample size, we generate $N = 500$ independent realizations of model (6.9) and model (6.10) and compare the inference results of NP and VAR procedure.

Results

First, let us compare the Bayes estimators. We will use the pointwise median spectral density (6.1) for this purpose. Exemplary visualizations of the estimates from the NP and VAR procedure for $n = 256$ are shown in Figure 6.2. First note that the individual spectra are visualized on a logarithmic scale, whereas real and imaginary parts of the cross spectra are shown on a regular scale. It can be seen in Figure 6.2 (a) that both procedures yield reasonable spectral estimates for VAR(2) data. However, the VAR procedure fits better than the NP estimate to the true spectral density. This is not too surprising, since the fitted VAR model is well-specified. For the VMA(1) data in Figure 6.2 (b), it can be seen that the VAR procedure struggles to captures the rough shape of the ground truth spectral density (introducing erroneous wavy bumps).



(a)



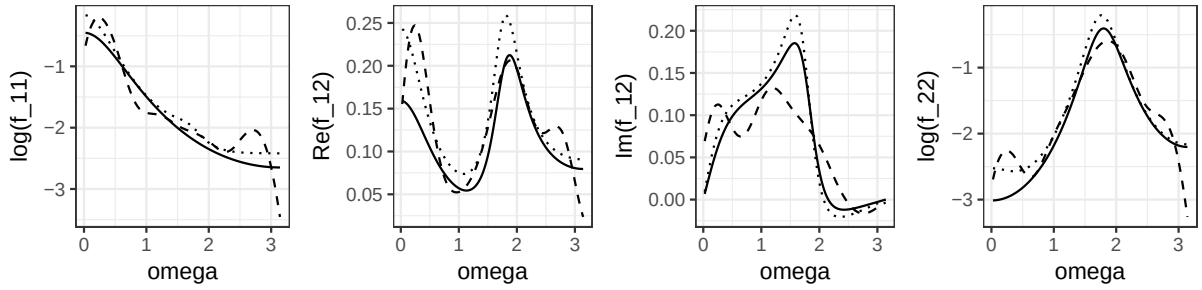
(b)

Figure 6.1.: Realization of (a) VAR(2) model (6.9) and (b) VMA(1) model (6.10) of respective length $n = 256$. The left panel depicts the first component and the right panel the second component of the time series.

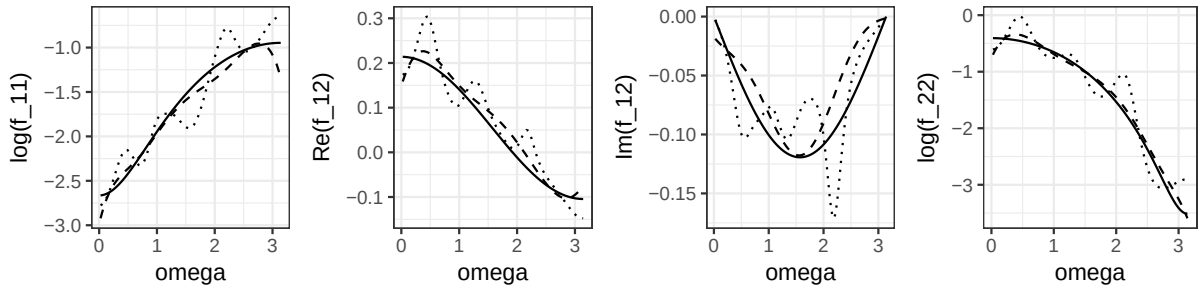
This comes from the fact that the VAR order p is determined in a preliminary model selection step based on AIC and in the misspecified case, high orders are preferred to account for the model bias. On the other hand, the fit of the NP procedure looks much closer with appropriate smoothness. This exemplifies the power of using the NP procedure, which is (in comparison to the parametric VAR procedure) to a much less extent susceptible to misspecification.

The corresponding 90%-credibility regions (pointwise regions from (6.4) and uniform regions from (6.5)) are shown in Figure 6.3 for the VAR(2) example and in Figure 6.4 for the VMA(1) example. It can be seen that both procedures yield reasonable regions for the VAR(2) example, where the NP procedure struggles slightly more to catch the main features (yielding wavy regions) than the well-specified VAR procedure. For the VMA(1) example however, it can be seen that the VAR procedure infers a much larger degree of uncertainty (in terms of size of credible regions), which – similar to the corresponding spectral estimate from Figure 6.2 (b) – look very wavy and bumpy. This does not hold true for the NP procedure, where the regions adhere a degree of smoothness that seems appropriate for the ground truth of this example.

Exemplary runtimes (as measured for one respective realization) are shown in Table 6.1. It comes as no surprise that the NP procedure – due to its larger number of parameters – is computationally more demanding than the parametric VAR procedure. In fact, the runtimes are roughly one order of magnitude larger. Considering the increased robustness of the NP procedure subject to model misspecification, this can be considered as an acceptable tradeoff. It can also be observed that the runtimes of the NP procedure are larger for VAR(2) data than



(a)

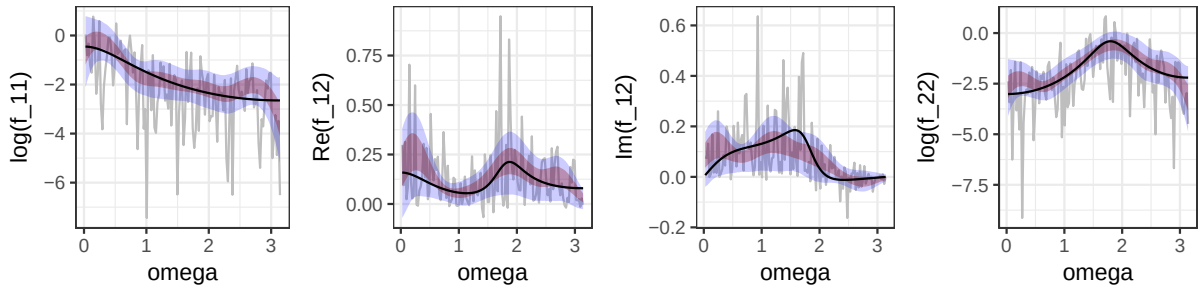


(b)

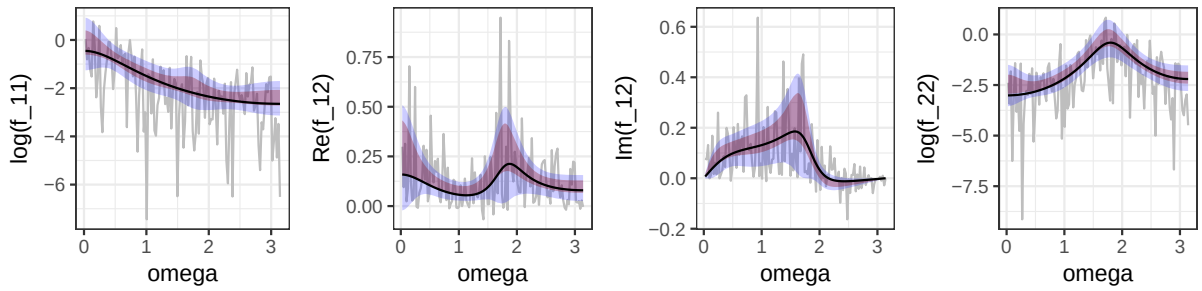
Figure 6.2.: Spectral estimates for a realization of length $n = 256$ of (a) the VAR(2) model (6.9) and (b) the VMA(1) model (6.10) for the NP procedure (dashed) and the VAR procedure (dotted), where the true spectral density is shown as solid black line.

	Time (s)					
	$n = 256$		$n = 512$		$n = 1024$	
	NP	VAR	NP	VAR	NP	VAR
VAR(2) data	82.71	14.48	137.17	27.33	360.91	50.01
VMA(1) data	70.92	20.79	108.63	37.36	303.93	89.70

Table 6.1.: Average runtime of 1000 MCMC iterations for NP procedure and VAR procedure for one respective realization of model (6.9) and (6.10).

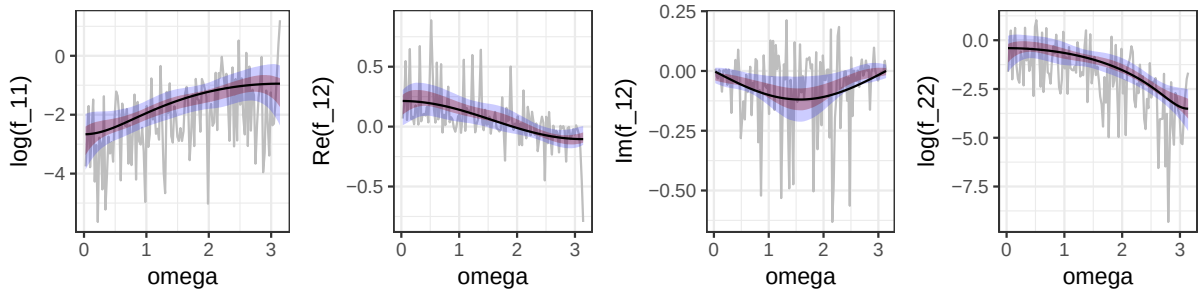


(a)

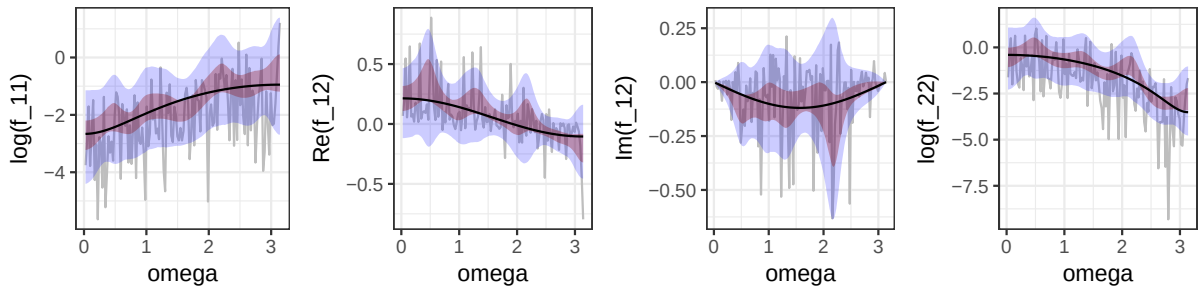


(b)

Figure 6.3.: Posterior credibility regions for a realization of length $n = 256$ of the VAR(2) model (6.9) for (a) NP and (b) VAR. Pointwise 90% region is visualized in shaded red and uniform 90% region in shaded blue and the true spectral density is shown as solid black line, whereas the periodogram is shown in gray.



(a)



(b)

Figure 6.4.: Posterior credibility regions for a realization of length $n = 256$ of the VMA(1) model (6.10) for (a) NP and (b) VAR. Pointwise 90% region is visualized in shaded red and uniform 90% region in shaded blue and the true spectral density is shown as solid black line, whereas the periodogram is shown in gray.

for VMA(1) data. This is due the fact that the VMA(1) spectral density is *smoother* than the VAR(2) spectral density in consideration, hence the NP procedure needs to employ less effort to capture the spectral features of the data. Indeed, for the length $n = 256$, the average posterior mean of the polynomial degree k is 45.86 for VAR(2) data and 14.32 for VMA(1) data, with average posterior variance of 85.22 and 14.94 respectively. As for the VAR procedure, the behavior is the other way round. This is due the fact that for VAR(2) data, the model is well-specified, whereas for VMA(1) data, the model complexity has to be increased considerably to capture the misspecified spectral features. Indeed, for the length $n = 256$, the average VAR order p (as chosen by the preliminary AIC-based model selection step) is 2.24 for the VAR(2) data and 7.50 for the VMA(1) data.

A more systematic insight in the results can be gained from Table 6.2, which contains the average $\mathbb{L}^1$-error and average $\mathbb{L}^2$-error (as defined in (6.2) and (6.3)) of the Bayes estimates among the 500 independent realizations from each model, for different sample sizes. The table also contains the empirical coverage of the uniform 90% region and the median (among replications) of the median (among frequencies) width of the components of the uniform 90% region. It can be seen that the VAR model yields smaller errors for the well-specified VAR(2) data and performs worse for the misspecified model, in which case the NP procedure yields superior results (in terms of smaller errors). This is in line with the above discussion from Figure 6.2. In the well-specified case, the VAR procedure yields *honest* credibility regions, i.e. the empirical coverage matches (approximately) the corresponding posterior probability mass of the regions, 90% in this case (we borrowed the notion of *honesty* from Szabó et al. (2015)). In the misspecified case, coverage of the VAR procedure is much larger, indicating a higher degree of posterior volatility, which matches the observations from Figure 6.2.

The NP procedure yields less coverage in all examples. The investigation of frequentist coverage properties within the Bayesian nonparametric framework still remains a delicate and difficult matter (see e.g. the discussion in Section 3.5 in Rousseau (2016)), with only few results known to this day. To achieve honest credible sets, a balance has to be achieved between enough prior flexibility on the one hand (yielding coverage) and enough penalization of prior complexity on the other hand (yielding consistency). This is closely related to the commonly known bias/variance-tradeoff. In other words, following the insights from Szabó et al. (2015), the prior should be chosen to slightly undersmooth the truth (for coverage), but not too much (for consistency). It is conjectured that the Bernstein polynomial prior used in the NP procedure is not suitable for this idea because Bernstein polynomials tend to oversmooth the truth, as suggested by their suboptimal approximation rates, see Lemma B.13. It will be of great interest for future research to consider different basis functions and to investigate whether it is possible to establish conditions under which the posterior distribution yields (asymptotically) honest confidence sets.

6.3. Analysis of the Southern Oscillation Index

In this Section we analyze the Southern Oscillation Index and Recruitment series from Shumway and Stoffer (2010), which have also been analyzed in Rosen and Stoffer (2007). Both series consist of monthly data for 452 months ranging over the years 1950-1987 and are available as datasets

	VAR(2) data					
	$n = 256$		$n = 512$		$n = 1024$	
	NP	VAR	NP	VAR	NP	VAR
$\mathbb{L}^1$-error	0.105	0.071	0.081	0.050	0.064	0.034
$\mathbb{L}^2$-error	0.133	0.094	0.107	0.066	0.085	0.045
Coverage	0.52	0.90	0.37	0.90	0.26	0.91
Width $\boldsymbol{f}_{11}$	0.287	0.200	0.166	0.116	0.107	0.074
Width $\Re\boldsymbol{f}_{12}$	0.177	0.111	0.124	0.069	0.088	0.046
Width $\Im\boldsymbol{f}_{12}$	0.225	0.129	0.154	0.078	0.108	0.051
Width $\boldsymbol{f}_{22}$	0.419	0.183	0.246	0.106	0.160	0.068
	VMA(1) data					
	$n = 256$		$n = 512$		$n = 1024$	
	NP	VAR	NP	VAR	NP	VAR
$\mathbb{L}^1$-error	0.095	0.155	0.070	0.121	0.053	0.091
$\mathbb{L}^2$-error	0.113	0.187	0.084	0.144	0.064	0.108
Coverage	0.61	0.99	0.44	0.98	0.27	0.96
Width $\boldsymbol{f}_{11}$	0.266	1.233	0.184	0.623	0.131	0.395
Width $\Re\boldsymbol{f}_{12}$	0.122	0.485	0.090	0.327	0.077	0.225
Width $\Im\boldsymbol{f}_{12}$	0.200	0.584	0.139	0.390	0.100	0.273
Width $\boldsymbol{f}_{22}$	0.408	1.889	0.281	0.952	0.206	0.598

Table 6.2.: Average $\mathbb{L}^1$-error, average $\mathbb{L}^2$-error, empirical coverage and median width of uniform 90% credibility regions of the NP procedure and the VAR procedure for 500 realizations of model (6.9) and (6.10).

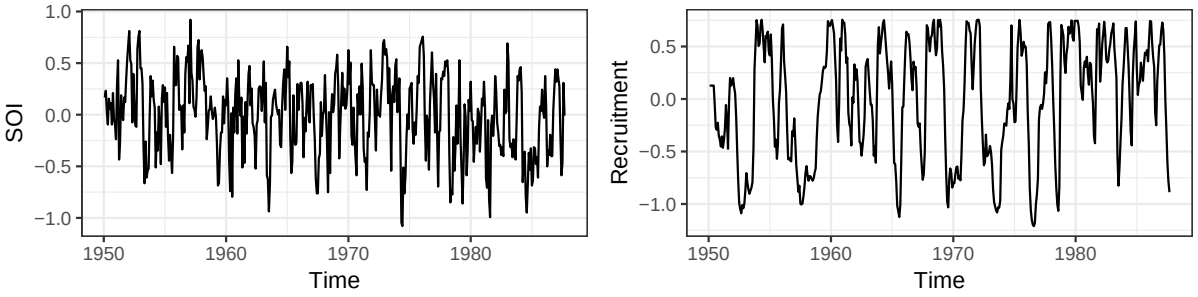


Figure 6.5.: The SOI and Recruitment series after mean-centering and re-scaling.

`soi` and `rec` in the `R` package `astsa` Stoffer (2017).

The Southern Oscillation Index (SOI) is defined as the normalized difference in air pressure between Tahiti (French Polynesia) and Darwin (Northern Territory, Australia). It constitutes a key indicator for warming or cooling effects of the central and eastern Pacific ocean known as *El Niño* and *La Niña*, see Bureau of Meteorology of the Australian Government (2018a). It is known that the SOI obeys a cyclical behavior, with cycles being classified in different meteorological phases (see Bureau of Meteorology of the Australian Government (2018b)). Due to this cyclical behavior, it can be expected that the observations in the SOI time series are not independent, but obey a time-dependence structure. One possible way to handle this would be an explicit modeling of the cyclical component (along with possible seasonal effects), e.g. in terms of cyclostationary time series or seasonal ARMA models. However, similar to the analysis in Rosen and Stoffer (2007), we make the simplifying assumption that all these effects can (at least approximately) be described in terms of the autocovariance structure.

The Recruitment Series consists of the number of new spawned fish in a population in the Pacific Ocean. Since the fish are known to spawn better in colder waters (Rosen and Stoffer, 2007), it can be expected that there also exists a cross-correlation between the SOI and the Recruitment series. The number of new fish is intrinsically integer-valued and this could be taken explicitly into account with integer-valued time series models. The data as provided in the dataset `rec` in the `astsa` package is re-scaled to the interval $[0, 100]$ and has been treated as real-valued in Rosen and Stoffer (2007) and we will follow the same approximation. We analyze the sample mean centered version of the data, where we also divide the Recruitment series data `rec` by 50 to ensure that the values of both series are in the same order of magnitude. The last re-scaling step is not necessary, but introduced for the sake of more numerical stability within the MCMC algorithms. In fact, it ensures that the spectra of both time series are (roughly) of the same order of magnitude, which stabilizes the numerical evaluation of Whittle's Likelihood – recall that the spectral density matrices at the Fourier frequencies have to be inverted for this purpose. Furthermore, since we are considering linear dependence structure, the inference results could easily be scaled back. See Figure 6.5 for a visualization of the transformed data with which we are working in the following.

We employ the NP procedure, with the same prior and MCMC parameters as discussed in Section 6.2. We parse the data in terms of a bivariate time series, with SOI constituting the first and Recruitment the second components. Note however that (unlike procedures that work with

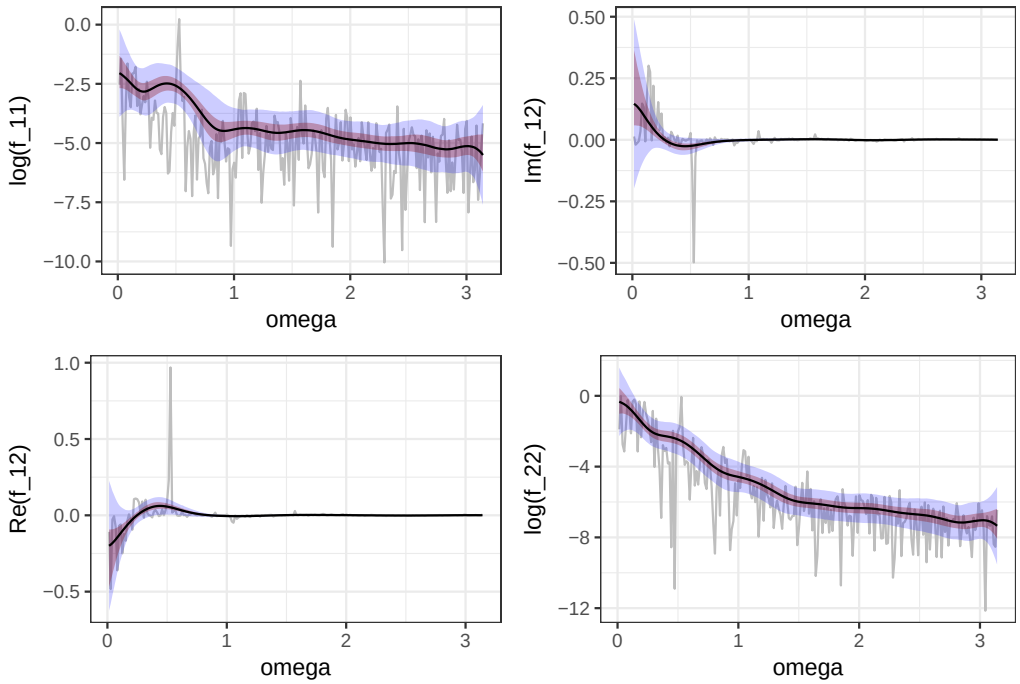


Figure 6.6.: Estimated spectra for the SOI and Recruitment series from the NP procedure. The posterior median spectral density is shown as solid black line, the pointwise 90% region is visualized in shaded red and the uniform 90% region in shaded blue, whereas the periodogram is shown in gray.

the Cholesky decomposition of $\boldsymbol{f}$) the results are the same if the components were swapped, since the employed prior is isotropic.

The results are shown in Figure 6.6. It can be seen that a spectral peak at $\omega_{\text{yearly}} := 2\pi/12 \approx 0.52$ is estimated in the first individual spectrum (which belongs to the SOI time series). Since we have monthly data, the frequency ω_{yearly} corresponds to a temporal distance of 12 months. The peak thus has the interpretation of the SOI series having a prominent annual (i.e. at lag 12) covariance. It can also be seen that cross-periodogram ordinates peak at ω_{yearly}. However, this seems to affect the spectral estimate only to a minor degree and the NP procedure *smooths out* this feature. This is because this cross-periodogram peak only consists of one single periodogram ordinate. To get a deeper insight in the inferred cross dependence structure, we investigate the *squared coherency* function $|\kappa|^2$ between the components, where the coherency function κ is defined as

$$\kappa(\omega|\boldsymbol{f}) := \frac{f_{12}(\omega)}{(f_{11}(\omega)f_{22}(\omega))^{1/2}}, \quad 0 \leq \omega \leq \pi.$$

The coherency can be thought of as a frequency-domain version of the cross-correlation and it holds $|\kappa(\omega|\boldsymbol{f})|^2 \leq 1$ for $0 \leq \omega \leq \pi$, see Section 11.6 in Brockwell and Davis (1991). We compute the squared coherency functions $|\kappa(\cdot, \boldsymbol{f}^{(j)})|^2$ for all of the posterior samples $\boldsymbol{f}^{(1)}, \ldots, \boldsymbol{f}^{(10{,}000)}$. Figure 6.7 shows the results. It can be seen that the posterior squared coherency also has a peak at ω_{yearly}, concluding that a prominent annual cross-correlation between the time series has been inferred. Furthermore, the squared coherency shows several minor peaks, which are

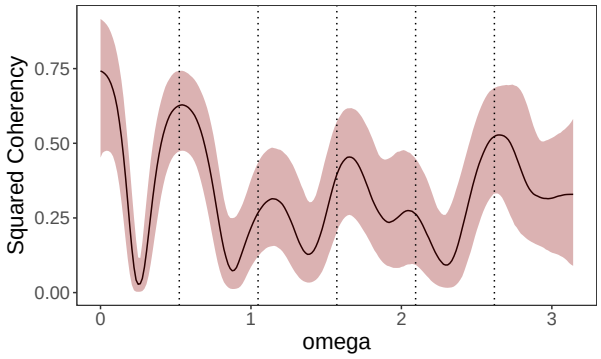


Figure 6.7.: Estimated squared coherency (pointwise posterior median) between SOI and Recruitment series. The pointwise 90% credible area is shown as shaded red area. The dotted vertical lines indicate the annual frequency $\omega_{\text{yearly}} = 2\pi/12 \approx 0.52$ and its harmonics $l\omega_{\text{yearly}}$, $l = 2, \ldots, 5$.

approximately located at the harmonic (i.e. integer multiple) frequencies of ω_{yearly}. These observations are in line with the findings that have been discussed in Rosen and Stoffer (2007). In fact, the squared coherency of the NP procedure shares the same qualitative features as the in the referenced literature (Rosen and Stoffer, 2007), except for having a higher degree of smoothness. One further observation which is also in line with previously known findings, is that there is another peak indicated in the low-frequency part (i.e. ω approaching 0) of both f_{11} and $|\kappa|^2$. This corresponds to a time-dependence in the order of magnitude of several years and can possibly be explained by the aforementioned meteorological cycles.

6.4. Discussion

We investigated the performance of the proposed NP procedure with both simulated and real data. In Section 6.2, we have seen that the procedure performs well for simulated data, taking full advantage over parametric procedures by not being susceptible to misspecification. The examples have also shown that there is a need for further theoretical investigation of frequentist properties to make "honest" uncertainty quantification possible (i.e. credible sets that are also confidence sets, at least asymptotically).

As for real data, we have seen in the Southern Oscillation Index example from Section 6.3 that the NP procedure yields results that are in line with previous findings from the literature, with curve estimates that may be *smoother* than for some other approaches. This property is due to the usage of Bernstein polynomials, which are known to have suboptimal approximation rates (see Remark B.14 in the Appendix) and have the tendency to produce over-smoothed results (see the simulation section in Edwards et al. (2017)). If this property is not desirable, there are several options to remedy. One obvious option constitutes the use of a different polynomial basis. As an example, in Edwards et al. (2017) a B-Spline basis with flexible choice of knot locations is employed to improve the estimation of sharp peaks of the spectral density of a univariate time series. It is conceptually straightforward to extend the NP procedure from this

work in the same direction, although great care will have to be taken to provide an efficient software implementation, since the precluding evaluation of the basis functions further increases the computational and numerical complexity of the MCMC algorithm. Another option is to generalize Whittle's Likelihood to incorporate a parametric working model that can help to estimate rough spectral features. This has been done in the univariate case in Kirch et al. (2017).

7.

Asymptotic Properties

For existing approaches to Bayesian nonparametric analysis of multivariate stationary time series, it is not known if posterior consistency holds. In the following Section 7.1, we will settle this question to the affirmative for the proposed nonparametric method from Chapter 5. We will also derive contraction rates of the posterior distribution in Section 7.2.

7.1. Posterior Consistency

We start our considerations with a brief overview on posterior consistency results in Section 7.1.1. While the famous Bernstein-von-Mises theorem (see van der Vaart (2000), Section 10.2) ensures asymptotic normality – and in particular consistency in case of well-specification – of the posterior distribution in the parametric setting under regularity and model identifiability assumptions, this result does not translate to the infinite-dimensional setting of Bayesian nonparametrics in general. In fact, counterexamples exist where the validity of a Bernstein-von-Mises theorem or even posterior consistency fails in infinite dimensions (see Diaconis and Freedman (1986) and Freedman (1999)). The consistency theorem and the assumptions are formulated in Section 7.1.2, whereas the proof will be developed in Sections 7.1.3-7.1.5. Section 7.1.6 contains a discussion of our findings.

7.1.1. Introduction

Let $(Z_1, \dots, Z_n) \sim P_\theta^n$ with $\theta \in \Theta$ and Θ being some measurable space endowed with a metric d. Assume that some prior distribution on Θ is specified. Then the posterior distribution of θ is called *consistent at* $\theta_0 \in \Theta$ with respect to d, if for every $\varepsilon > 0$ it holds

$$P(B_\varepsilon^c(\theta_0)|Z_1, \dots, Z_n) \to 0, \quad \text{in } P_{\theta_0}^n \text{ probability as } n \to \infty$$

with $B_\varepsilon(\theta_0) = \{\theta \in \Theta \colon d(\theta, \theta_0) < \varepsilon\}$ and $B_\varepsilon^c(\theta_0)$ denoting its complement set in Θ. In other words, posterior consistency formalizes the convergence of the posterior distribution towards a degenerate measure at the *true* value θ_0. A famous and very general result due to Doob (see van der Vaart (2000), Section 10.4, or Ghosal and van der Vaart (2017), Section 6.2) ensures posterior consistency for a possibly infinite-dimensional model, requiring appropriate

model identifiability and measurability. However, this result is restricted to a subset of the parameter space with prior mass one. Null-sets in infinite dimensions may be *large*: In fact, a null-set may even contain a dense subspace – think e.g. about the space of polynomials which is densely embedded in $\mathbb{L}^2$, but may attain prior mass 0 if e.g. a Fourier basis prior is employed. Thus the result by Doob – despite its generality – is too pessimistic for many situations. See the discussion in Section 6.2 in Ghosal and van der Vaart (2017) for another example.

A more practical result for the iid setting is the famous theorem from Schwartz (1965) and in particular the extensions by Barron et al. (1999) and Ghosal et al. (1999). Besides full prior support in terms of Kullback-Leibler neighborhoods, a stronger form of model identifiability, i.e. exponentially powerful testability, is required to get consistency of the posterior distribution. The extended formulation from Barron et al. (1999) relies on sieves of the parameter space and is of great practical usefulness. In fact, it is known (see e.g. Section 5 in Barron (1989)) that an exponentially powerful test sequence (φ_n) for testing $H_0 \colon p \in U$ against $H_1 \colon p \notin U$ does not exist for neighborhoods in many relevant topologies (e.g. $\mathbb{L}^1$ or Hellinger) in general. This renders the fact that a sieve is needed to get consistency in these topologies.

The following result is another extension of Schwartz's theorem to the case of independent, non-identically distributed random variables. It will be the key tool for proving posterior consistency of the spectral density under Whittle's likelihood.

Theorem 7.1 (Extended version of Schwarz' Theorem for non-iid observations, Theorem A.1 in Choudhuri et al. (2004a))**.** *Let $Z_{1,n}, \dots, Z_{n,n}$ be independently distributed having density $p_{i,n}(\cdot|\theta)$ with respect to a σ-finite measure on a Borel space for $i = 1, \dots, n$, where $\theta \in \Theta$ and Θ is some measurable space. Let $\theta_0 \in \Theta$. Denote the joint distribution of $(Z_{1,n}, \dots, Z_{n,n})$ under θ by P_θ^n. Denote the Kullback-Leibler divergence from θ_0 to θ at $Z_{i,n}$ and the associated variance by*

$$K_{i,n}(\theta_0, \theta) := \mathrm{E}_{\theta_0} \log \frac{p_{i,n}(Z_{i,n}|\theta_0)}{p_{i,n}(Z_{i,n}|\theta)}, \quad V_{i,n}(\theta_0, \theta) := \mathrm{Var}_{\theta_0} \log \frac{p_{i,n}(Z_{i,n}|\theta_0)}{p_{i,n}(Z_{i,n}|\theta)}. \tag{7.1}$$

Let $K_n(\theta_0, \theta) := \frac{1}{n} \sum_{i=1}^n K_{i,n}(\theta_0, \theta)$. Consider a sieve sequence (Θ_n) with $\Theta_n \subset \Theta$ and let $U \subset \Theta$ denote a neighborhood of Θ_0. Let a prior P be given on Θ. Let the following support assumptions on the prior and testability assumptions on the model be fulfilled:

- *Prior positivity of neighborhoods: There exists $B \subset \Theta$ with*
  - *(a)* $P(B) > 0$,
  - *(b)* $\liminf_{n\to\infty} P\left(\{\theta \in B \colon K_n(\theta_0, \theta) < \varepsilon\}\right) > 0$ *for all* $\varepsilon > 0$,
  - *(c)* $\frac{1}{n^2} \sum_{i=1}^n V_{i,n}(\theta_0, \theta) \to 0$ *for all* $\theta \in B$ *as* $n \to \infty$.
- *Existence of uniformly exponentially powerful tests: There exists a sequence (φ_n) of tests and constants c_1, c_2, c_3 not depending on n, such that*
  - *(d)* $\mathrm{E}_{\theta_0} \varphi_n \to 0$ *as* $n \to \infty$,
  - *(e)* $\sup_{\theta \in U^c \cap \Theta_n} E_\theta(1 - \varphi_n) \leq \exp(-c_1 n)$ *and*
  - *(f)* $P(\Theta_n^c) \leq c_2 \exp(-c_3 n)$.

Then

$$P(U^c|Z_{1,n},\dots,Z_{n,n}) \to 0, \quad \text{in } P_{\theta_0}^n \text{ probability as } n \to \infty$$

with $U_n^c = \Theta \setminus U_n$.

Let $g\colon \mathcal{X} \to \mathbb{R}$ be measurable and bounded. Then for a *weak* neighborhood of the form

$$U = U(g) = \left\{p\colon \left|\int_{\mathcal{X}} g(x)p(x)\mu(dx) - \int_{\mathcal{X}} g(x)p_0(x)\mu(dx)\right| < \varepsilon\right\}, \tag{7.2}$$

a test sequence (φ_n) fulfilling the assumptions (d)-(f) from Theorem 7.1 can always be constructed, at least for the special case of iid observations (see Example 6.20 in Ghosal and van der Vaart (2017)). However, since only one fixed function g (or at most finitely many g's) are considered, the topology is too weak, i.e that the neighborhoods $U(g)$ are too large, as e.g. argued in Section 1 in Barron et al. (1999). For consistency in stronger topologies, it is typically necessary to explicitly derive an appropriate test sequence (φ_n). Under the prior positivity of Kullback-Leibler neighborhoods assumption, the exponentially powerful testability condition is sufficient and necessary for $P(U_n^c|Z_1,\dots,Z_n)$ to decay to 0 exponentially (see Theorem 3.11 in Choi and Ramamoorthi (2008)).

Figure 7.1 depict an illustrative example of data from a VAR(2) process and the corresponding pointwise and uniform spectral posterior credibility regions inferred by the NP procedure from Section 6.2. It can be seen that, with growing sample size, the size of the credibility regions is shrinking and that their location is approaching the ground truth spectral density $\boldsymbol{f}_0$, indicating that all posterior mass is asymptotically allocated near $\boldsymbol{f}_0$.

7.1.2. Posterior Consistency for the Spectral Density Matrix

In this section, we will establish posterior consistency for the spectral density for stationary Gaussian time series under the Bernstein-Hpd-Gamma prior from Section 5.1 and Whittle's likelihood. The proof extends the ideas from Choudhuri et al. (2004a) (see also Section 7.3.3 in Ghosal and van der Vaart (2017)), where a similar result has been shown for the normalized spectral density of a univariate stationary Gaussian time series with a Bernstein-Dirichlet prior and Whittle's likelihood. See also Kirch et al. (2017), where posterior consistency has been established for a semiparametric procedure with a Bernstein-Dirichlet prior on a nonparametric correction of the spectral density of a univariate stationary Gaussian time series under a parametric likelihood. These approaches rely on the assumption that the normalizing constant (i.e. the integral of the true spectral density) is a priori known. Our method of proof does not rely on this assumption, but instead requires a pointwise upper bound for the (eigenvalues of the) spectral density to be known a priori. Indeed, to achieve exponentially powerful model testability, we will conduct a *truncation* of the Bernstein-Hpd-Gamma prior such that the eigenvalues are bounded by a universal constant with prior probability one. Note that this is in line with Assumption $\boldsymbol{f}1$ for the true spectral density $\boldsymbol{f}_0$. The truncation will be done by a suitable restriction of the prior parameters $(k, \boldsymbol{\Phi})$. There should be no confusion between the truncated Bernstein-Hpd-Gamma prior (which bounds the eigenvalues of $\boldsymbol{f}$ by a universal

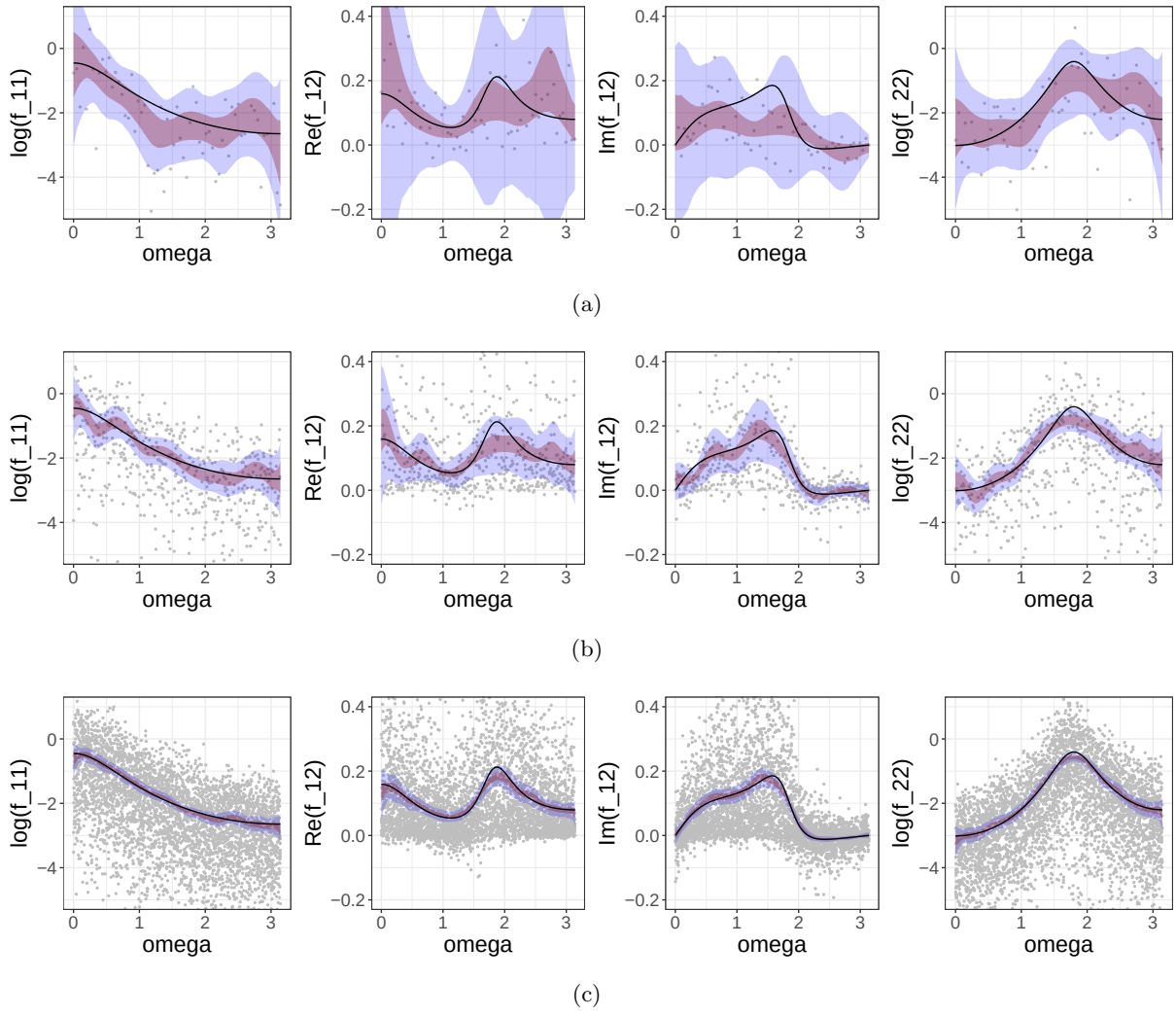


Figure 7.1.: Visualization of posterior consistency: Credibility regions of the NP procedure for data drawn from the VAR(2) process (6.9) of length (a) $n = 128$, (b) $n = 1024$ and (c) $n = 8192$, where the periodogram is visualized in gray.

constant and which will be derived in the following) and the truncated Bernstein polynomials from (5.4) (which are merely a restriction of the polynomial basis functions to get better mixture properties at the boundary of $[0, \pi]$).

To elaborate, let the positive integers $\mathbb{N}$ be endowed with the power σ-algebra $\mathcal{F}_{\mathbb{N}}$. Denote by $\mathcal{M}_{d\times d}$ the set of all Hpsd measures $\mathbf{\Phi}$ on $[0, \pi]$ fulfilling $\|\mathbf{\Phi}([0, \pi])\| < \infty$, endowed with the smallest σ-algebra $\mathcal{F}_{\mathcal{M}_{d\times d}}$ such that the mapping

$$\mathcal{M}_{d\times d} \ni \mathbf{\Phi} \mapsto \mathbf{\Phi}(A) \in \bar{\mathcal{S}}_d^+$$

is measurable for all measurable $A \subset [0, \pi]$. Let the product space $\mathbb{N} \times \mathcal{M}_{d\times d}$ be endowed with the product σ-algebra $\mathcal{F}_{\mathbb{N}\times\mathcal{M}_{d\times d}}$ of $\mathcal{F}_{\mathbb{N}}$ and $\mathcal{F}_{\mathcal{M}_{d\times d}}$. Consider the space $\mathcal{D}_{d\times d}$ of continuous $\bar{\mathcal{S}}_d^+$-valued functions on $[0, \pi]$, endowed with the σ-algebra $\mathcal{F}_{\mathcal{D}_{d\times d}}$ induced by the maximum Frobenius norm

$$\|\boldsymbol{f}\|_{F,\infty} := \max_{0\le\omega\le\pi} \|\boldsymbol{f}(\omega)\|, \quad \boldsymbol{f} \in \mathcal{D}_{d\times d}. \tag{7.3}$$

The Bernstein-Hpd-Gamma prior (5.2) is actually a prior on the parametrization $(k, \mathbf{\Phi}) \in \mathbb{N} \times \mathcal{M}_{d\times d}$. It can also be conceived as a prior P on $\boldsymbol{f} \in \mathcal{D}_{d\times d}$ via the mapping

$$\mathfrak{B}\colon \mathbb{N} \times \mathcal{M}_{d\times d} \to \mathcal{D}_{d\times d}, \quad \mathfrak{B}(k, \mathbf{\Phi}) := \sum_{j=1}^{k} \mathbf{\Phi}(I_{j,k}) b(\cdot/\pi|j, k), \tag{7.4}$$

with $I_{j,k}$ the interval partition from (5.1) and $b(\cdot, j, k)$ the Bernstein polynomial basis functions of degree k for $j = 1, \ldots, k$ from (B.20). It is argued in Section 2 in Petrone (1999) that $\mathfrak{B}$ is $(\mathcal{F}_{\mathbb{N}\times\mathcal{M}_{d\times d}}, \mathcal{F}_{\mathcal{D}_{d\times d}})$-measurable. Let $\tau > 0$ and consider the set

$$\mathcal{C}_\tau := \{\boldsymbol{f} \in \mathcal{D}_{d\times d}\colon \lambda_{\max}(\boldsymbol{f}(\omega)) \le \tau \text{ for all } 0 \le \omega \le \pi\}. \tag{7.5}$$

Note that $\mathcal{C}_\tau$ is measurable, i.e. $\mathcal{C}_\tau \in \mathcal{F}_{\mathcal{D}_{d\times d}}$. Assume that the prior P on $\boldsymbol{f}$ fulfills $P(\mathcal{C}_\tau) > 0$ (this is e.g. the case if the prior on $\mathbf{\Phi}$ fulfills Assumption GP3 and if the prior on k fulfills Assumption $k1$ below, see the upcoming Lemma 7.4). Then we define the *truncated Bernstein-Hpd-Gamma* prior P_τ as the restriction of P to $\mathcal{C}_\tau$:

$$P_\tau(F) := \frac{P(F \cap \mathcal{C}_\tau)}{P(\mathcal{C}_\tau)}, \quad \text{for } F \subset \mathcal{D}_{d\times d} \text{ measurable.} \tag{7.6}$$

Note that k and $\mathbf{\Phi}$ are no longer independent under P_τ. In what follows, we will conceive P_τ as a prior on $\boldsymbol{f}$ or (equivalently) as a prior on $(k, \mathbf{\Phi})$, whatever is more convenient. For our main theorem of this Section, we will make the following assumption:

Assumption $k1$. *There exist positive constants c, C such the prior probability mass function of k fulfills*

$$0 < p(k) \le C \exp(-ck \log k), \quad k \in \mathbb{N}.$$

Let us briefly discuss this prior assumption on k.

Remark 7.2. *Clearly, any prior probability mass function of the form* $p(k) = C\exp(-ck\log k)$ *for* $k \in \mathbb{N}$ *and positive constants* c, C *fulfills Assumption k1. Many other prior choices are of course possible and Assumption k1 states full support and a decay condition on the tails. As another example, if* $k \sim \text{Poi}(C)$ *a priori, i.e. with probability mass function* $p(k) = \frac{C^k}{k!}\exp(-C)$ *for* $k \in \mathbb{N}$ *and a positive constant* C*, then Assumption k1 is fulfilled. Indeed, it clearly holds* $p(k) > 0$ *and an application of Stirling's Formula (see Abramowitz and Stegun (1964), p. 257) reveals* $\log(k!) = k\log k - k + O(\log k) \gtrsim k\log k$ *and hence*

$$p(k) \lesssim \frac{C^k}{k!} = \exp\left(k\log C - \log(k!)\right) \leq \exp\left(-k\left(\log\frac{1}{C} + c\log k\right)\right) \leq \exp(-\tilde{c}k\log k)$$

for positive constants $c, \tilde{c}$*.*

Denote by $P^n_{W,\tau}(\boldsymbol{f}|\underline{Z}_1,\ldots,\underline{Z}_n)$ the pseudo-posterior distribution of the spectral density matrix $\boldsymbol{f}$ when employing the truncated Bernstein-Hpd-Gamma prior P_τ from (7.6) in conjunction with Whittle's likelihood $\tilde{P}^n_W$ from (5.3). Recall that the Fourier coefficients occurring in $\tilde{P}^n_W$ are

$$\underline{\tilde{Z}}_j = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\underline{Z}_t\exp(-it\omega_j), \quad \omega_j = \frac{2\pi j}{n}, \quad n = 1,\ldots,N, \quad N := \left\lceil\frac{n}{2}\right\rceil - 1. \tag{7.7}$$

The following result establishes consistency for the spectral density matrix in the $\mathbb{L}^1$ topology.

Theorem 7.3. *Let* $\{\underline{Z}_t\}$ *be a Gaussian stationary time series in* $\mathbb{R}^d$ *with mean zero, true spectral density matrix* $\boldsymbol{f}_0$ *and corresponding true autocovariance function* $\boldsymbol{\Gamma}_0$ *fulfilling Assumptions f1-f2. Denote by* $P^n_{\boldsymbol{f}_0}$ *the joint distribution of the Fourier coefficients* $\underline{\tilde{Z}}_1,\ldots,\underline{\tilde{Z}}_N$ *from (7.7) under* $\boldsymbol{f}_0$*. Let* $\tau \in (b_1,\infty)$*, with the upper bound* b_1 *on the eigenvalues of* $\boldsymbol{f}_0$ *from Assumption f1. Let the prior on the spectral density matrix* $\boldsymbol{f}$ *be given by the truncated Bernstein-Hpd-Gamma prior* P_τ *from (7.6), with the prior on* $\boldsymbol{\Phi}$ *fulfilling Assumptions GP1, GP2 and GP3 and the prior on* k *fulfilling Assumption k1. Then, for all* $\varepsilon > 0$ *and* $U_\varepsilon := \{\int_0^\pi \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\|d\omega < \varepsilon\}$ *it holds*

$$P^n_{W,\tau}(U^c_\varepsilon|\underline{Z}_1,\ldots,\underline{Z}_n) \to 0, \quad \text{in } P^n_{\boldsymbol{f}_0} \text{ probability as } n\to\infty.$$

We will prove Theorem 7.3 by showing that the assumptions of Theorem 7.1 are fulfilled. As a first important observation, it is sufficient to prove Theorem 7.3 under Whittle's likelihood. This follows from the mutual contiguity result of Corollary 4.5, which implies that convergence of $P^n_{W,\tau}(U^c_\varepsilon|\underline{Z}_1,\ldots,\underline{Z}_n)$ in $P^n_{\boldsymbol{f}_0}$ probability is equivalent to convergence in $\tilde{P}^n_W(\cdot|\boldsymbol{f}_0)$ probability. We may thus assume that the Fourier coefficients $\underline{\tilde{Z}}_1,\ldots,\underline{\tilde{Z}}_N$ are distributed as under $\tilde{P}^n_W(\cdot|\boldsymbol{f}_0)$, that is, independent and complex multivariate normal with mean zero and covariance matrix $2\pi\boldsymbol{f}(\omega_j)$.

7.1.3. Prior Positivity of Neighborhoods

We start the proof with the prior positivity of neighborhoods. The idea is to show prior positivity of a $\|\cdot\|_{F,\infty}$-neighborhood and then show that the former is contained in a Kullback-Leibler neighborhood. The upcoming results from Lemma 7.4 and Lemma 7.6 are generalizations of Section B1 in Choudhuri et al. (2004a) and some steps in the proofs are similar.

Lemma 7.4. *Let the assumptions of Theorem 7.3 be fulfilled. Let $\varepsilon > 0$,*

$$B_\varepsilon := \{\boldsymbol{f} \in \mathcal{D}_{d\times d} \colon \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} < \varepsilon\} \tag{7.8}$$

and denote the non-truncated Bernstein-Hpd-Gamma prior from (5.2) by P.

(a) Then it holds $P(B_\varepsilon) > 0$.

(b) Furthermore, for $0 < \varepsilon < \tau - b_1$, it holds $B_\varepsilon \subset \mathcal{C}_\tau$ for the prior truncation set $\mathcal{C}_\tau$ from (7.5). In particular, it holds $P(\mathcal{C}_\tau) > 0$.

(c) It holds $P_\tau(B_\varepsilon) > 0$.

Proof. Let $\boldsymbol{F}_0(A) := \int_A \boldsymbol{f}_0(\omega)d\omega$ for $A \subset [0,\pi]$ denote the spectral measure matrix corresponding to the spectral density matrix $\boldsymbol{f}_0$ (where the integral is understood component-wise). Choose k_0 such that $\|\boldsymbol{f}_0 - \mathfrak{B}(k, \boldsymbol{F}_0)\|_{F,\infty} < \frac{\varepsilon}{2}$ holds for all $k \geq k_0$, with the Bernstein polynomial expansion operator $\mathfrak{B}$ from (7.4). This can be done according to the uniform approximation property of Bernstein polynomials, see Theorem B.12 in the Appendix. Then for $k \geq k_0$ and $\boldsymbol{f} = \mathfrak{B}(k, \boldsymbol{\Phi})$, we use $\|b(\cdot, j, k-j+1)\|_\infty \leq k$ (see (B.21) in the Appendix) and get

$$\|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} \leq \|\boldsymbol{f} - \mathfrak{B}(k, \boldsymbol{F}_0)\|_{F,\infty} + \frac{\varepsilon}{2} \leq k \sum_{j=1}^{k} \|\boldsymbol{\Phi}(I_{j,k}) - \boldsymbol{F}_0(I_{j,k})\| + \frac{\varepsilon}{2},$$

where the interval partition $\{I_{j,k} \colon j = 1, \ldots, k\}$ is given by (5.1). Since k and $\boldsymbol{\Phi}$ are independent under P, this yields

$$P(B_\varepsilon) \geq \sum_{k \geq k_0} p(k) P\left(\max_{j=1,\ldots,k} \|\boldsymbol{\Phi}(I_{j,k}) - \boldsymbol{F}_0(I_{j,k})\| < \frac{\varepsilon}{2k^2}\right).$$

Since the prior on k has full support on $\mathbb{N}$ (i.e. $p(k) > 0$ for all $k \in \mathbb{N}$) by Assumption *k*1, it suffices to show that there exists $\tilde{k} \geq k_0$ such that

$$P\left(\max_{j=1,\ldots,\tilde{k}} \left\|\boldsymbol{\Phi}(I_{j,\tilde{k}}) - \boldsymbol{F}_0(I_{j,\tilde{k}})\right\| < \frac{\varepsilon}{2\tilde{k}^2}\right) > 0 \tag{7.9}$$

holds. For $\boldsymbol{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$, the probability on the left hand side of (7.9) simplifies to

$$\prod_{j=1}^{\tilde{k}} P\left(\left\|\boldsymbol{\Phi}(I_{j,\tilde{k}}) - \boldsymbol{F}_0(I_{j,\tilde{k}})\right\| < \frac{\varepsilon}{2\tilde{k}^2}\right)$$

by the independence property of $\boldsymbol{\Phi}$ (see Remark 3.1). Since $\int_{I_{j,\tilde{k}}} dx = \pi/\tilde{k} > 0$, an application of Theorem 3.4 yields $\mathrm{supp}(\boldsymbol{\Phi}(I_{j,\tilde{k}})) = \bar{\mathcal{S}}_d^+$ for all $j = 1, \ldots, \tilde{k}$. This leads to

$$P\left(\left\|\boldsymbol{\Phi}(I_{j,\tilde{k}}) - \boldsymbol{F}_0(I_{j,\tilde{k}})\right\| < \frac{\varepsilon}{2\tilde{k}^2}\right) > 0, \quad j = 1, \ldots, \tilde{k},$$

and in particular $P(B_\varepsilon) > 0$, concluding (a). By assumption it holds $\tau > b_1$. Recalling $\|\boldsymbol{f}_0(\omega)\|_2 = \lambda_{\max}(\boldsymbol{f}_0(\omega))$ from (B.9) in the Appendix, it follows from Lemma B.4 (a) in the Appendix that

$$\lambda_{\max}(\boldsymbol{f}(\omega)) \leq \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\|_2 + \lambda_{\max}(\boldsymbol{f}_0(\omega)) \leq \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\| + \lambda_{\max}(\boldsymbol{f}_0(\omega)),$$

yielding $B_\varepsilon \subset \mathcal{C}_\tau$ for $0 < \varepsilon < \tau - b_1$. In particular, we have $P(\mathcal{C}_\tau) \geq P(B_\varepsilon) > 0$ by part (a), concluding (b). Recalling $P_\tau(B_\varepsilon) = \frac{P(B_\varepsilon \cap \mathcal{C}_\tau)}{P(\mathcal{C}_\tau)}$, it also readily follows $P_\tau(B_\varepsilon) = \frac{P(B_\varepsilon)}{P(\mathcal{C}_\tau)} > 0$, i.e. (c). □

The following result links the Kullback-Leibler terms of mean zero complex multivariate normals to the distance of their covariance matrices.

Lemma 7.5. *Let $\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1 \in \mathcal{S}_d^+$ and denote by $p_i(\underline{z})$ the Lebesgue density of the $CN_d(\underline{0}, \boldsymbol{\Sigma}_i)$ distribution for $i = 0, 1$. Consider the Kullback-Leibler divergence and variance term*

$$K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) := \int_{\mathbb{C}^d} \log \frac{p_0(\underline{z})}{p_1(\underline{z})} p_0(\underline{z}) d\underline{z},$$
$$V(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) := \int_{\mathbb{C}^d} \left(\log \frac{p_0(\underline{z})}{p_1(\underline{z})} - K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}) \right)^2 p_0(\underline{z}) d\underline{z}.$$

(a) Then it holds

$$V(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \lambda_{\max}(\boldsymbol{\Sigma}_0)^2 \|\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}\|^2$$

and

$$V(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \frac{1}{\lambda_{\min}(\boldsymbol{\Sigma}_1)^2} \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|^2,$$

with $\lambda_{\min}(\boldsymbol{\Sigma})$ and $\lambda_{\max}(\boldsymbol{\Sigma})$ being the smallest and largest eigenvalue of $\boldsymbol{\Sigma} \in \mathcal{S}_d$.

(b) Let $\boldsymbol{Q} := \boldsymbol{\Sigma}_1^{-1/2} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1/2}$. Then $\lambda_{\min}(\boldsymbol{Q}) \geq \frac{1}{2}$ implies $K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \|\boldsymbol{Q} - \boldsymbol{I}_d\|^2$. In particular, in this case it holds

$$K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \lambda_{\max}(\boldsymbol{\Sigma}_0)^2 \|\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}\|^2$$

and

$$K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \frac{1}{\lambda_{\min}(\boldsymbol{\Sigma}_1)^2} \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|^2.$$

Proof. First we observe that

$$\log \frac{p_0(\underline{z})}{p_1(\underline{z})} = \log \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} + \underline{z}^* \left(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1} \right) \underline{z}, \quad \underline{z} \in \mathbb{C}^d.$$

For $\underline{Z} \sim CN_d(\underline{0}, \boldsymbol{\Sigma}_0)$ and $\boldsymbol{A} \in \mathcal{S}_d$, an application of Lemma B.24 in the Appendix yields $\mathrm{E}(\underline{Z}^* \boldsymbol{A} \underline{Z}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_0)$ and $\mathrm{Var}(\underline{Z}^* \boldsymbol{A} \underline{Z}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_0 \boldsymbol{A}\boldsymbol{\Sigma}_0)$. This yields, using Lemma B.5 and Lemma B.2 in the Appendix,

$$K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) = \log \frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} + \mathrm{tr}\left(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0 - \boldsymbol{I}_d \right) = \mathrm{tr}(\boldsymbol{Q} - \boldsymbol{I}_d) - \log |\boldsymbol{Q}| \tag{7.10}$$

and

$$V(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) = \mathrm{tr}\left(\left((\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1}) \boldsymbol{\Sigma}_0 \right)^2 \right). \tag{7.11}$$

From (7.11), we derive the bound, using Lemma B.4 (c) in the Appendix

$$V(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \|(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\Sigma}_0\|^2 \leq \|\boldsymbol{\Sigma}_0\|_2^2 \left\|\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1}\right\|^2 = \lambda_{\max}(\boldsymbol{\Sigma}_0)^2 \left\|\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1}\right\|^2$$

and similarly, using that $\lambda_{\max}(\boldsymbol{A}^{-1}) = \lambda_{\min}(\boldsymbol{A})^{-1}$ for $\boldsymbol{A} \in \mathcal{S}_d^+$,

$$V(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) = \operatorname{tr}\left(\left(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1)\right)^2\right) \leq \lambda_{\min}(\boldsymbol{\Sigma}_1)^{-2} \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|^2,$$

concluding part (a). Denote by $\lambda_1, \ldots, \lambda_d$ the eigenvalues of $\boldsymbol{Q}$ and assume that $\lambda_{\min}(\boldsymbol{Q}) \geq \frac{1}{2}$, i.e. $\lambda_i \geq \frac{1}{2}$ for $i = 1, \ldots, d$. From (7.10) we obtain

$$K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) = \sum_{i=1}^{d} (\lambda_i - 1 - \log \lambda_i) = \sum_{i=1}^{d} \left(\tilde{\lambda}_i - \log(1 + \tilde{\lambda}_i)\right)$$

with $\tilde{\lambda}_i = \lambda_i - 1$. Note that $\tilde{\lambda}_i \geq -\frac{1}{2}$ by assumption. Using the inequality $x - \log(1 + x) \leq x^2$ for $x \geq -\frac{1}{2}$, we get

$$K(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \leq \sum_{i=1}^{d} \tilde{\lambda}_i^2 = \sum_{i=1}^{d} (\lambda_i - 1)^2 = \|\boldsymbol{Q} - \boldsymbol{I}_d\|^2.$$

Finally, applying Lemma B.6 to $\boldsymbol{Q} \in \mathcal{S}_d^+$, we also get

$$\|\boldsymbol{Q} - \boldsymbol{I}_d\| = \|\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{I}_d\| = \|\boldsymbol{\Sigma}_0(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1})\| \leq \lambda_{\max}(\boldsymbol{\Sigma}_0)\|\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1}\|$$

and similarly $\|\boldsymbol{Q} - \boldsymbol{I}_d\| \leq \lambda_{\min}(\boldsymbol{\Sigma}_1)^{-1} \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|$, concluding (b). □

The following Lemma shows prior positivity of Kullback-Leibler neighborhoods under the truncated Bernstein-Hpd-Gamma.

Lemma 7.6. *Let the assumptions of Theorem 7.3 be fulfilled. Let $\underline{\tilde{Z}}_1, \ldots, \underline{\tilde{Z}}_N$ be the multivariate Fourier coefficients as in (7.7) with $N = \lceil n/2 \rceil - 1$. Let*

$$p_{j,N}(\underline{\tilde{z}}_j|\boldsymbol{f}) = \frac{1}{(2\pi^2)^d |\boldsymbol{f}(\omega_j)|} \exp\left(-\frac{1}{2\pi} \underline{\tilde{z}}_j^* \boldsymbol{f}(\omega_j)^{-1} \underline{\tilde{z}}_j\right), \quad \underline{\tilde{z}}_j \in \mathbb{C}^d, \quad j = 1, \ldots, N \tag{7.12}$$

denote the probability density of $\underline{\tilde{Z}}_j$ under Whittle's Likelihood $\tilde{P}_W^n(\cdot|\boldsymbol{f})$ and

$$K_N(\boldsymbol{f}_0, \boldsymbol{f}) = \frac{1}{N} \sum_{j=1}^{N} K_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}), \quad K_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) = \mathrm{E}_{\boldsymbol{f}_0} \log \frac{p_{j,N}(\underline{\tilde{Z}}_j|\boldsymbol{f}_0)}{p_{j,N}(\underline{\tilde{Z}}_j|\boldsymbol{f})}, \tag{7.13}$$

$$V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) = \mathrm{Var}_{\boldsymbol{f}_0} \log \frac{p_{j,N}(\underline{\tilde{Z}}_j|\boldsymbol{f}_0)}{p_{j,N}(\underline{\tilde{Z}}_j|\boldsymbol{f})}, \tag{7.14}$$

for $j = 1, \ldots, N$, where $\mathrm{E}_{\boldsymbol{f}_0}$ and $\mathrm{Var}_{\boldsymbol{f}_0}$ denotes the mean and variance under $\tilde{P}_W^n(\cdot|\boldsymbol{f}_0)$. Then for ε small enough and B_ε from (7.8) it holds

$$\liminf_{N \to \infty} P_\tau \left(\left\{\boldsymbol{f} \in B_\varepsilon \colon K_N(\boldsymbol{f}_0, \boldsymbol{f}) < \frac{4\varepsilon^2}{b_0^2}\right\}\right) > 0, \tag{7.15}$$

$$\frac{1}{N^2} \sum_{j=1}^{N} V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) \to 0, \quad \textit{for all } \boldsymbol{f} \in B_\varepsilon \textit{ as } N \to \infty. \tag{7.16}$$

Proof. By Lemma 7.4, we have $P_\tau(B_\varepsilon) > 0$. Let $\boldsymbol{f} \in B_\varepsilon$ for $\varepsilon > 0$ small enough. We start by showing the following two auxiliary results:

$$\min_{0\le\omega\le\pi} \lambda_{\min}(\boldsymbol{f}(\omega)) \ge \frac{b_0}{2}, \tag{7.17}$$

$$\min_{0\le\omega\le\pi} \lambda_{\min}\left(\boldsymbol{f}(\omega)^{-1/2}\boldsymbol{f}_0(\omega)\boldsymbol{f}(\omega)^{-1/2}\right) \ge \frac{1}{2}, \tag{7.18}$$

with b_0 from Assumption $\boldsymbol{f}1$. From the Min-Max principle of Courant-Fisher (see Lemma B.3 in the Appendix), it follows $\lambda_{\min}(\boldsymbol{A}+\boldsymbol{B}) \ge \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{B})$ for $\boldsymbol{A},\boldsymbol{B} \in \mathcal{S}_d$, and by Lemma B.4 (b) in the Appendix this yields $\lambda_{\min}(\boldsymbol{A}+\boldsymbol{B}) \ge \lambda_{\min}(\boldsymbol{A}) - \max\{\lambda_{\max}(\boldsymbol{B}), -\lambda_{\min}(\boldsymbol{B})\} \ge \lambda_{\min}(\boldsymbol{A}) - \|\boldsymbol{B}\|$ and hence

$$\lambda_{\min}(\boldsymbol{f}(\omega)) = \lambda_{\min}(\boldsymbol{f}_0(\omega) + \boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)) \ge \lambda_{\min}(\boldsymbol{f}_0(\omega)) - \|\boldsymbol{f}-\boldsymbol{f}_0\| \ge b_0 - \varepsilon$$

for $0 \le \omega \le \pi$, yielding (7.17) for $\varepsilon \le b_0/2$. Define $\tilde{\boldsymbol{Q}}(\omega) := \boldsymbol{f}(\omega)^{-1/2}\boldsymbol{f}_0(\omega)\boldsymbol{f}(\omega)^{-1/2}$ (the notation $\tilde{\boldsymbol{Q}}$ is chosen to avoid confusion with the similarly defined function $\boldsymbol{Q}(\omega)$ as in the upcoming Corollary 7.10). Note that $\tilde{\boldsymbol{Q}}(\omega)$ is Hermitian positive definite. Since (7.18) is equivalent to $\lambda_{\min}(\tilde{\boldsymbol{Q}}(\omega))-1 \ge -1/2$, it suffices to the show the stronger statement $(\lambda_{\min}(\tilde{\boldsymbol{Q}}(\omega))-1)^2 \le 1/4$. In fact, to show (7.18), we will show the even stronger statement

$$\max_{0\le\omega\le\pi} \sum_{i=1}^{d} \left(\lambda_i(\tilde{\boldsymbol{Q}}(\omega)) - 1\right)^2 \le \frac{1}{4}, \tag{7.19}$$

where $\lambda_1(\boldsymbol{A}),\dots,\lambda_d(\boldsymbol{A}) \in \mathbb{R}$ denote the eigenvalues of $\boldsymbol{A} \in \mathcal{S}_d$, counted with their respective multiplicity. To show (7.19), first recall from (B.8) in the Appendix that the representation

$$\begin{aligned}\sum_{i=1}^{d} \left(\lambda_i(\tilde{\boldsymbol{Q}}(\omega)) - 1\right)^2 &= \sum_{i=1}^{d} \left(\lambda_i\left(\tilde{\boldsymbol{Q}}(\omega) - \boldsymbol{I}_d\right)\right)^2 = \|\tilde{\boldsymbol{Q}}(\omega) - \boldsymbol{I}_d\|^2 \\ &= \left\|\boldsymbol{f}(\omega)^{-1/2}\left(\boldsymbol{f}_0(\omega) - \boldsymbol{f}(\omega)\right)\boldsymbol{f}(\omega)^{-1/2}\right\|^2\end{aligned}$$

holds. We use (7.17) and $\|\boldsymbol{AB}\| \le \lambda_{\max}(\boldsymbol{A})\|\boldsymbol{B}\| = \lambda_{\min}(\boldsymbol{A})^{-1}\|\boldsymbol{B}\|$ for $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$ and $\boldsymbol{B} \in \mathcal{S}_d$ from Lemma B.4 (d) in the Appendix to get

$$\left\|\boldsymbol{f}(\omega)^{-1/2}\left(\boldsymbol{f}_0(\omega) - \boldsymbol{f}(\omega)\right)\boldsymbol{f}(\omega)^{-1/2}\right\|^2 \le \lambda_{\min}(\boldsymbol{f}(\omega))^{-2}\|\boldsymbol{f}_0(\omega) - \boldsymbol{f}(\omega)\|^2 \le \frac{4}{b_0^2}\varepsilon^2 \le \frac{1}{4}$$

for $0 \le \omega \le \pi$ for ε small enough ($\varepsilon < \frac{b_0}{4}$). This shows (7.19) and in particular (7.18). An application of part (b) of Lemma 7.5 yields, since (7.18) holds,

$$K_{j,N}(\boldsymbol{f}_0,\boldsymbol{f}) \le \frac{1}{\lambda_{\min}(\boldsymbol{f}(\omega_j))^2}\|\boldsymbol{f}_0(\omega_j) - \boldsymbol{f}(\omega_j)\|^2 \le \frac{4\varepsilon^2}{b_0^2},$$

for $j = 1,\dots,N$, where (7.17) was used in the last step. This yields $K_N(\boldsymbol{f}_0,\boldsymbol{f}) < \frac{4\varepsilon^2}{b_0^2}$ for every $\boldsymbol{f} \in B_\varepsilon$, i.e.

$$P_\tau\left(\left\{\boldsymbol{f} \in B_\varepsilon \colon K_N(\boldsymbol{f}_0,\boldsymbol{f}) < \frac{4\varepsilon^2}{b_0^2}\right\}\right) \ge P_\tau(B_\varepsilon) > 0$$

by Lemma 7.4, concluding the proof of (7.15), since $P_\tau(B_\varepsilon)$ does not depend on N. Similarly, an application of part (a) of Lemma 7.5 yields $V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) \leq \frac{4\varepsilon^2}{b_0^2}$ for $j = 1, \dots, N$, which implies

$$\frac{1}{N^2} \sum_{j=1}^{N} V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) \leq \frac{1}{N} \left(\frac{4\varepsilon}{b_0} \right)^2 \to 0$$

as $N \to \infty$, concluding (7.16). □

7.1.4. Exponential Testability

Now we show the existence of a sieve (Θ_n) of the parameter space $\Theta = \mathcal{D}_{d\times d}$ and a sequence of tests (φ_n) fulfilling the assumptions (d)-(f) of Theorem 7.1. For $k > 0$, let

$$\mathcal{S}_d^{+k} := \left\{ \underline{\boldsymbol{W}} \colon \underline{\boldsymbol{W}} = (\boldsymbol{W}_1, \dots, \boldsymbol{W}_k) \text{ with } \boldsymbol{W}_j \in \mathcal{S}_d^+ \text{ for } j = 1, \dots, k \right\}.$$

For $\underline{\boldsymbol{W}} = (\boldsymbol{W}_1, \dots, \boldsymbol{W}_k) \in \mathcal{S}_d^{+k}$, let $\mathfrak{B}(k, \underline{\boldsymbol{W}}) := \sum_{j=1}^{k} \boldsymbol{W}_j b(\cdot | j, k-j+1)$. Recalling the truncation set $\mathcal{C}_\tau$ from (7.5), we now define the sieve Θ_n as

$$\Theta_n := \bigcup_{k=1}^{k_n} \left\{ \mathfrak{B}(k, \underline{\boldsymbol{W}}) \colon \underline{\boldsymbol{W}} \in \mathcal{S}_d^{+k} \right\} \cap \mathcal{C}_\tau, \quad k_n := \left\lfloor \frac{\delta n}{\log n} \right\rfloor, \tag{7.20}$$

with $\delta > 0$ to be specified later. We need the following auxiliary result.

Lemma 7.7. *Let $\underline{\boldsymbol{W}} = (\boldsymbol{W}_1, \dots, \boldsymbol{W}_k) \in \mathcal{S}_d^{+k}$. Then it holds*

$$\|\boldsymbol{W}_j\| \leq \sqrt{d} \|\mathfrak{B}(k, \underline{\boldsymbol{W}})\|_{F,\infty}, \quad j = 1, \dots, k,$$

where $\|\cdot\|_{F,\infty}$ denotes the maximum Frobenius norm from (7.3).

Proof. Let $\boldsymbol{W}_0 := \sum_{j=1}^{k} \boldsymbol{W}_j \in \mathcal{S}_d^+$ and let $j \in \{1, \dots, k\}$. It is clear that $\boldsymbol{W}_j \leq \boldsymbol{W}_0$ (in the Hpsd sense, i.e. $\boldsymbol{W}_0 - \boldsymbol{W}_j \in \bar{\mathcal{S}}_d^+$). Clearly, this implies $\lambda_{\max}(\boldsymbol{W}_0) \geq \lambda_{\max}(\boldsymbol{W}_j)$ (because otherwise it would hold $\underline{z}^*(\boldsymbol{W}_0 - \boldsymbol{W}_j)\underline{z} < 0$ for $\underline{0} \neq \underline{z}$ being an eigenvector of $\boldsymbol{W}_j$ corresponding to $\lambda_{\max}(\boldsymbol{W}_j)$, which would contradict $\boldsymbol{W}_j \leq \boldsymbol{W}_0$). Using this along with Lemma B.4 in the Appendix yields

$$\|\boldsymbol{W}_j\| \leq \sqrt{d} \lambda_{\max}(\boldsymbol{W}_j) \leq \sqrt{d} \lambda_{\max}(\boldsymbol{W}_0) \leq \sqrt{d} \|\boldsymbol{W}_0\|. \tag{7.21}$$

Furthermore, using $\int_0^\pi b(\omega/\pi | j, k-j+1) d\omega = \pi \int_0^1 b(x | j, k-j+1) dx = \pi$ it holds

$$\boldsymbol{W}_0 = \frac{1}{\pi} \sum_{j=1}^{k} \boldsymbol{W}_j \int_0^\pi b(\omega | j, k-j+1) d\omega = \frac{1}{\pi} \int_0^\pi \mathfrak{B}(k, \underline{\boldsymbol{W}})[\omega] d\omega$$

and in particular

$$\|\boldsymbol{W}_0\| \leq \frac{1}{\pi} \int_0^\pi \|\mathfrak{B}(k, \underline{\boldsymbol{W}})[\omega]\| d\omega \leq \|\mathfrak{B}(k, \underline{\boldsymbol{W}})\|_{F,\infty}.$$

Combining this with (7.21) concludes the proof. □

The following Lemma quantifies the metric entropy of Θ_n with respect to the maximum Frobenius norm $\|\cdot\|_{F,\infty}$ in terms of the ε-covering number. Recall from (B.28) in the Appendix that the ε-covering number is defined as the smallest number of ε-balls needed to cover Θ_n. This result will be needed for the construction of tests, where we will use the technique of covering the alternative set with small balls – and the bound for the entropy is needed to derive consistency of the test under the null hypothesis.

Lemma 7.8. *The ε-covering number $N(\varepsilon, \Theta_n, \|\cdot\|_{F,\infty})$ of Θ_n from (7.20) in the maximum Frobenius norm $\|\cdot\|_{F,\infty}$ is bounded by*

$$\log N(\varepsilon, \Theta_n, \|\cdot\|_{F,\infty}) \leq k_n \left(4d^2 \log k_n + 2d^2 \log \frac{6d^{3/2}\tau}{\varepsilon} + 1 \right)$$

with k_n as in (7.20).

Proof. The proof is an adaption of the proof of Lemma B4 in Choudhuri et al. (2004a). Let $\tilde{\Theta}_{k,\tau} := \{\mathfrak{B}(k, \underline{\boldsymbol{W}}) \colon \underline{\boldsymbol{W}} \in \mathcal{S}_d^{+k}\} \cap \mathcal{C}_\tau$. Since $\Theta_n = \cup_{k=1}^{k_n} \tilde{\Theta}_{k,\tau}$ it holds

$$N(\varepsilon, \Theta_n, \|\cdot\|_{F,\infty}) \leq \sum_{k=1}^{k_n} N(\varepsilon, \tilde{\Theta}_{k,\tau}, \|\cdot\|_{F,\infty}). \tag{7.22}$$

Let $k \leq k_n$ and let $\underline{\boldsymbol{W}}_1, \underline{\boldsymbol{W}}_2 \in \mathcal{S}_d^{+k}$, such that $\boldsymbol{f}_i := \mathfrak{B}(k, \underline{\boldsymbol{W}}_i) \in \tilde{\Theta}_{k,\tau}$ for $i = 1, 2$. We consider the norm $\|\underline{\boldsymbol{W}}_i\|_1 := \sum_{j=1}^{k} \|\boldsymbol{W}_{ij}\|_1$ on $\mathcal{S}_d^{+k}$, with the 1-norm $\|\boldsymbol{W}_{ij}\|_1$ as defined in (B.4). By Lemma B.4 in the Appendix, it holds $\|\boldsymbol{A}\|_1 \leq d\|\boldsymbol{A}\|$ and together with Lemma 7.7, this yields

$$\|\underline{\boldsymbol{W}}_i\|_1 = \sum_{j=1}^{k} \|\boldsymbol{W}_{ij}\|_1 \leq d \sum_{j=1}^{k} \|\boldsymbol{W}_{ij}\| \leq d^{3/2} k \|\boldsymbol{f}_i\|_{F,\infty} \leq d^{3/2} k\tau. \tag{7.23}$$

Recalling $\|b(\cdot, j, k-j+1)\|_\infty \leq k$ from (B.21) in the Appendix and $\|\boldsymbol{A}\| \leq \|\boldsymbol{A}\|_1$ from Lemma B.4 in the Appendix, we obtain

$$\|\boldsymbol{f}_1 - \boldsymbol{f}_2\|_{F,\infty} \leq k \sum_{j=1}^{k} \|\boldsymbol{W}_{1j} - \boldsymbol{W}_{2j}\| \leq k \sum_{j=1}^{k} \|\boldsymbol{W}_{1j} - \boldsymbol{W}_{2j}\|_1 = k \|\underline{\boldsymbol{W}}_1 - \underline{\boldsymbol{W}}_2\|_1. \tag{7.24}$$

Combining (7.23) and (7.24) with Lemma B.32 (b), we get

$$\begin{aligned} N(\varepsilon, \tilde{\Theta}_{k,\tau}, \|\cdot\|_{F,\infty}) &\leq N\left(\frac{\varepsilon}{k}, \left\{\underline{\boldsymbol{W}} \in \mathcal{S}_d^{+k} \colon \|\underline{\boldsymbol{W}}\|_1 \leq d^{3/2} k\tau\right\}, \|\cdot\|_1\right) \\ &\leq N\left(\frac{\varepsilon}{k}, \left\{\underline{z} \in \mathbb{C}^{kd^2} \colon \|\underline{z}\|_1 \leq d^{3/2} k\tau\right\}, \|\cdot\|_1\right), \end{aligned}$$

where we conceived $\mathcal{S}_d^{+k}$ as a subset of $\mathbb{C}^{kd^2}$ in the last step. Noting for $z = x + iy \in \mathbb{C}$ (with $x, y \in \mathbb{R}$) the inequality $|z| \leq |x| + |y|$, it follows that the right hand side is bounded from above by

$$N\left(\frac{\varepsilon}{2k}, \{\underline{x} \in \mathbb{R}^{2kd^2} \colon \|\underline{x}\|_1 \leq d^{3/2} k\tau\}, \|\cdot\|_1\right) \leq \left(\frac{6d^{3/2}k^2\tau}{\varepsilon}\right)^{2kd^2} \leq \left(\frac{6d^{3/2}k_n^2\tau}{\varepsilon}\right)^{2k_n d^2},$$

where the first inequality was obtained with an application of Lemma B.31 in the Appendix. From (7.22), this yields

$$\begin{aligned}\log N(\varepsilon, \Theta_n, \|\cdot\|_{F,\infty}) &\leq \log k_n + 4d^2 k_n \log k_n + 2d^2 k_n \log \frac{6d^{3/2}\tau}{\varepsilon} \\ &\leq k_n \left(4d^2 \log k_n + 2d^2 \log \frac{6d^{3/2}\tau}{\varepsilon} + 1\right).\end{aligned}$$

□

To construct uniformly exponentially powerful tests, we need the following results. They translate the integral condition $\int_0^\pi \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\| d\omega > \varepsilon$ (occurring in the test alternatives) to suitable *pointwise* conditions that hold at sufficiently many Fourier frequencies – making testing possible at these frequencies. The results are similar in spirit to Lemma B2 in Choudhuri et al. (2004a), where the univariate case under the assumption $\int_0^\pi f(\omega) - f_0(\omega) d\omega = 0$ is treated. Some arguments of our proof of the upcoming Lemma 7.9 are similar to the proof of Lemma B2 in Choudhuri et al. (2004a) and are included for the sake of completeness and readability.

Lemma 7.9. *Let $-\infty < a < b < \infty$ and $\boldsymbol{f}_0 \colon [a,b] \to \mathcal{S}_d$ be continuous. Let $\varepsilon > 0$. Let $\Omega_n(a,b) := \{a, a + (b-a)/n, \ldots, b - (b-a)/n, b\}$. Then there exists $k_0 \in \mathbb{N}$ such that for every $k \geq k_0$ and every $\boldsymbol{f} \colon [a,b] \to \mathcal{S}_d$ with components f_{rs}, $r,s = 1,\ldots,d$ being polynomials of degree k, it holds true that $\int_a^b \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\| dx > \varepsilon$ implies*

$$\# \left\{ x \in \Omega_n(a,b) \colon \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\| > \frac{\varepsilon}{4(b-a)} \right\} \geq \frac{n\varepsilon}{8\rho(b-a)} - 4k,$$

with $\#E$ denoting the cardinality of a set E and $\rho := \max\{\|\boldsymbol{f}\|_{F,\infty}, \|\boldsymbol{f}_0\|_{F,\infty}\}$.

Proof. Put $A := \left\{ x \in [a,b] \colon \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\| > \frac{\varepsilon}{4(b-a)} \right\}$. By the uniform approximation property of Bernstein polynomials (see Lemma B.12 in the Appendix) there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ the Bernstein polynomial approximation of degree k

$$\tilde{\boldsymbol{f}}_0(x) := \sum_{j=1}^{k} \left(\int_{a+\frac{j-1}{k}(b-a)}^{a+\frac{j}{k}(b-a)} \boldsymbol{f}_0(t) dt \right) b\left(\frac{x-a}{b-a} \,\middle|\, j, k-j+1 \right), \quad a \leq x \leq b,$$

fulfills

$$\max_{a \leq x \leq b} \|\boldsymbol{f}_0(x) - \tilde{\boldsymbol{f}}_0(x)\| < \frac{\varepsilon}{4(b-a)}. \tag{7.25}$$

Put $\tilde{A} := \left\{ x \in [a,b] \colon \|\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x)\| > \frac{\varepsilon}{2(b-a)} \right\}$. Because of

$$\|\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x)| \leq \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\| + \frac{\varepsilon}{4(b-a)},$$

it holds $\tilde{A} \subset A$ and we continue by bounding the cardinality of $\tilde{A} \cap \Omega_n(a,b)$ from below. The integrated Frobenius norm of $\boldsymbol{f} - \tilde{\boldsymbol{f}}_0$ decomposes as

$$\int_a^b \|\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x)\| dx = \int_{\tilde{A}} \|\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x)\| dx + \int_{\tilde{A}^c} \|\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x)\| dx \leq 2\rho \mathcal{L}(\tilde{A}) + \frac{\varepsilon}{2},$$

where $\mathcal{L}$ denotes the Lebesgue measure on $\mathbb{R}$. Combining this with the assumption $\int_a^b \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\|dx > \varepsilon$ and (7.25) yields

$$\varepsilon < \int_a^b \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\|dx \leq \int_a^b \|\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x)\|dx + \frac{\varepsilon}{4} \leq 2\rho\mathcal{L}(\tilde{A}) + \frac{3\varepsilon}{4}$$

and thus $\mathcal{L}(\tilde{A}) \geq \frac{\varepsilon}{8\rho}$. Since $\tilde{A} = \left\{x \in [a,b] \colon t(x) > \frac{\varepsilon^2}{4(b-a)^2}\right\}$ with $t(x) := \operatorname{tr}\left((\boldsymbol{f}(x) - \tilde{\boldsymbol{f}}_0(x))^2\right)$ being a polynomial of degree $2k$, it follows that $\tilde{A}$ is the union of at most $2k$ open intervals. Denote these intervals by $\tilde{A}_1, \ldots, \tilde{A}_m$ with $m \leq 2k$ and let h_j be the length of interval $\tilde{A}_j$ for $j = 1, \ldots, m$. Since any subinterval of $[a,b]$ having length h contains at least $\frac{nh}{b-a} - 2$ many x's from $\Omega_n(a,b)$, it follows that the cardinality of $\tilde{A} \cap \Omega_n(a,b)$ is bounded from below by

$$\sum_{j=1}^{m}\left(\frac{nh_j}{b-a} - 2\right) = \frac{n\mathcal{L}(\tilde{A})}{b-a} - 2m \geq \frac{n\varepsilon}{8\rho(b-a)} - 4k.$$

□

For the construction of tests, a slightly different characterization is needed. This follows readily from Lemma 7.9 under the additional assumption that all eigenvalues are strictly positive, and is formulated in the following corollary.

Corollary 7.10. *Let the assumptions of Lemma 7.9 be fulfilled. Assume additionally that* $\boldsymbol{f}_0(x) \in \mathcal{S}_d^+$ *for* $a \leq x \leq b$. *Let* k_0 *and* $\boldsymbol{f}$ *be as in Lemma 7.9. Let* $\boldsymbol{Q}(x) := \boldsymbol{f}_0(x)^{-1/2}\boldsymbol{f}(x)\boldsymbol{f}_0(x)^{-1/2}$. *Then* $\int_a^b \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\|dx > \varepsilon$ *implies*

$$\#\left\{x \in \Omega_n(a,b) \colon \lambda_{\max}(\boldsymbol{Q}(x)) > 1 + \tilde{\varepsilon} \text{ or } \lambda_{\min}(\boldsymbol{Q}(x)) < 1 - \tilde{\varepsilon}\right\} \geq \frac{n\varepsilon}{8\rho(b-a)} - 4k$$

with $\tilde{\varepsilon} = \frac{\varepsilon}{4(b-a)\rho\sqrt{d}}$ *and* $\rho := \max\{\|\boldsymbol{f}\|_{F,\infty}, \|\boldsymbol{f}_0\|_{F,\infty}\}$.

Proof. Denote the sorted eigenvalues of $\boldsymbol{Q}(x)$ by $\lambda_1(x) \leq \ldots \leq \lambda_d(x)$ for $a \leq x \leq b$. First observe that for every $\delta > 0$, $\sum_{i=1}^d (\lambda_i(x) - 1)^2 > \delta^2$ implies either $\lambda_d(x) > 1 + \frac{\delta}{\sqrt{d}}$ or $\lambda_1(x) < 1 - \frac{\delta}{\sqrt{d}}$, as otherwise $|\lambda_i(x) - 1| \leq \frac{\delta}{\sqrt{d}}$ for $i = 1, \ldots, d$. Therefore, we will bound the cardinality of the set

$$A := \left\{x \in \Omega_n(a,b) \colon \sum_{i=1}^{d}(\lambda_i(x) - 1)^2 > d\tilde{\varepsilon}^2\right\}.$$

To do so, we compute for $a \leq x \leq b$

$$\begin{aligned}\sum_{i=1}^{d}(\lambda_i(x) - 1)^2 &= \operatorname{tr}\left((\boldsymbol{Q}(x) - \boldsymbol{I}_d)(\boldsymbol{Q}(x) - \boldsymbol{I}_d)\right) = \|\boldsymbol{Q}(x) - \boldsymbol{I}_d\|^2 \\ &= \|\boldsymbol{f}_0^{-1/2}(\boldsymbol{f}(x) - \boldsymbol{f}_0(x))\boldsymbol{f}_0^{-1/2}\|^2.\end{aligned}$$

From Lemma B.4 in the Appendix, we get

$$\|\boldsymbol{f}_0^{-1/2}(\boldsymbol{f}(x) - \boldsymbol{f}_0(x))\boldsymbol{f}_0^{-1/2}\|^2 \geq \lambda_{\max}(\boldsymbol{f}_0(x))^{-2}\|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\|^2 \geq \frac{1}{\rho^2}\|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\|^2.$$

This shows $\tilde{A} \subset A$ with $\tilde{A} := \left\{x \in \Omega_n(a,b) \colon \|\boldsymbol{f}(x) - \boldsymbol{f}_0(x)\| > \sqrt{d}\rho\tilde{\varepsilon}\right\}$. An application of Lemma 7.9 shows that the cardinality of $\tilde{A}$ is at least $\frac{n\varepsilon}{8\rho(b-a)} - 4k$, concluding the proof. □

Now we can construct a uniformly exponentially powerful test as follows: First, we will derive a test that is exponentially powerful for a fixed alternative. This will be done in Lemma 7.14. The proof makes use of the previous pointwise distinguishably results of alternatives at sufficiently many Fourier frequencies. The exponential rate is due to a large deviation result for complex multivariate normal random variables, which is presented in the upcoming Theorem 7.11. The proof of Theorem 7.11 is involved and relies on some auxiliary results, which are presented in Lemma 7.12 and Lemma 7.13. Afterwards, we will construct a uniformly exponentially powerful test by covering the whole alternative by small balls in Lemma 7.15, where the covering number is controlled by the entropy result from Lemma 7.8.

Theorem 7.11. *Let $\underline{Y}_1, \dots, \underline{Y}_m$ be independent with $\underline{Y}_j \sim CN_d(\underline{0}, \mathbf{\Sigma}_j)$ and $\mathbf{\Sigma}_j \in \mathcal{S}_d^+$ for $j = 1, \dots, m$. With $\mathbf{\Sigma}_{0j}, \mathbf{\Sigma}_{1j} \in \mathcal{S}_d^+$, consider testing*

$$H_0 \colon \mathbf{\Sigma}_j = \mathbf{\Sigma}_{0j} \text{ for } j = 1, \dots, m \quad \text{against} \quad H_1 \colon \mathbf{\Sigma}_j = \mathbf{\Sigma}_{1j} \text{ for } j = 1, \dots, m$$

with either

$$\lambda_{\max}\left(\mathbf{\Sigma}_{0j}^{-1/2}\mathbf{\Sigma}_{1j}\mathbf{\Sigma}_{0j}^{-1/2}\right) > 1 + \varepsilon, \quad j = 1, \dots, m, \tag{7.26}$$

or

$$\lambda_{\min}\left(\mathbf{\Sigma}_{0j}^{-1/2}\mathbf{\Sigma}_{1j}\mathbf{\Sigma}_{0j}^{-1/2}\right) < 1 - \varepsilon, \quad j = 1, \dots, m, \tag{7.27}$$

where $\boldsymbol{A}^{1/2}$ denotes the Hermitian positive definite matrix square root of $\boldsymbol{A} \in \mathcal{S}_d^+$ and $\varepsilon > 0$ does not depend on m. Then there exists a test φ_m fulfilling

$$\mathrm{E}_{H_0}\varphi_m \leq \exp(-c_0 m) \tag{7.28}$$

$$\mathrm{E}_{H_1}(1 - \varphi_m) \leq \exp(-c_1 m) \tag{7.29}$$

for all m, with positive constants c_0, c_1 depending only on ε.

To prove the result of Theorem 7.11, we need the following two technical Lemmas.

Lemma 7.12. *Let $\mathbf{\Sigma}_0, \mathbf{\Sigma}_1 \in \mathcal{S}_d^+$. For $\underline{0} \neq \underline{a} \in \mathbb{C}^d$, define $\psi(\underline{a}) := \frac{\underline{a}^* \mathbf{\Sigma}_1 \underline{a}}{\underline{a}^* \mathbf{\Sigma}_0 \underline{a}}$. Denote by $\underline{a}_{\max}$ and $\underline{a}_{\min}$ an eigenvector corresponding to the largest (and smallest) eigenvalue $\lambda_{\max}$ (and $\lambda_{\min}$) of the matrix $\mathbf{\Sigma}_0^{-1/2}\mathbf{\Sigma}_1\mathbf{\Sigma}_0^{-1/2}$. Then*

$$\sup_{\underline{a} \neq \underline{0}} \psi(\underline{a}) = \psi(\mathbf{\Sigma}_0^{-1/2}\underline{a}_{\max}) = \lambda_{\max}, \quad \inf_{\underline{a} \neq \underline{0}} \psi(\underline{a}) = \psi(\mathbf{\Sigma}_0^{-1/2}\underline{a}_{\min}) = \lambda_{\min}.$$

Proof. It holds

$$\psi(\mathbf{\Sigma}_0^{-1/2}\underline{a}_{\max}) = \frac{\underline{a}_{\max}^* \mathbf{\Sigma}_0^{-1/2}\mathbf{\Sigma}_1\mathbf{\Sigma}_0^{-1/2}\underline{a}_{\max}}{\|\underline{a}_{\max}\|^2} = \lambda_{\max} = \sup_{\underline{a} \neq \underline{0}} \psi(\underline{a}),$$

where the characterization of $\lambda_{\max}$ from the Minmax-theorem of Courant-Fisher (see Lemma B.3 in the Appendix) was used in the last step. The representation for $\lambda_{\min}$ follows analogously. □

Lemma 7.13. *Let $\boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}} \in \mathcal{S}_d^+$ and $Z \sim CN_d(\underline{0}, \boldsymbol{\Sigma})$. Let $\underline{0} \neq \underline{a} \in \mathbb{C}^d$. Consider the random quadratic form $\Psi_{\underline{a},\tilde{\boldsymbol{\Sigma}}}(\underline{Z}) := \frac{\underline{a}^* \underline{Z}\underline{Z}^* \underline{a}}{\underline{a}^* \tilde{\boldsymbol{\Sigma}} \underline{a}}$. Then*

$$\Psi_{\underline{a},\tilde{\boldsymbol{\Sigma}}}(\underline{Z}) \sim \frac{\underline{a}^* \boldsymbol{\Sigma} \underline{a}}{2\underline{a}^* \tilde{\boldsymbol{\Sigma}} \underline{a}} \chi_2^2,$$

where χ_k^2 denotes the Chi-squared distribution with k degrees of freedom.

Proof. We first observe that, by the cyclic property of the trace (see Lemma B.2 in the Appendix)

$$\underline{a}^* \underline{Z}\underline{Z}^* \underline{a} = \operatorname{tr}(\underline{a}^* \underline{Z}\underline{Z}^* \underline{a}) = \operatorname{tr}(\underline{Z}^* \boldsymbol{A} \underline{Z}) = \underline{Z}^* \boldsymbol{A} \underline{Z},$$

with the Hermitian matrix $\boldsymbol{A} := \underline{a}\underline{a}^*$. By Lemma B.24, we obtain $\underline{Z}^* \boldsymbol{A} \underline{Z} \overset{d}{=} \sum_{i=1}^d \lambda_i |X_i|^2$, where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\boldsymbol{\Sigma}\boldsymbol{A}$ and $X_1, \ldots, X_d \overset{\text{iid}}{\sim} CN(0,1)$. Since $\boldsymbol{\Sigma}\boldsymbol{A}$ is of rank 1, $d-1$ eigenvalues are equal to 0. Without loss of generality, assume $\lambda_1 \neq 0$. This implies the representation $\underline{Z}^* \boldsymbol{A} \underline{Z} \sim \lambda_1 |X_1|^2$. On the other hand, we have

$$\underline{a}^* \boldsymbol{\Sigma} \underline{a} = \operatorname{tr}(\underline{a}^* \boldsymbol{\Sigma} \underline{a}) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) = \sum_{i=1}^d \lambda_i = \lambda_1,$$

which shows $\Psi_{\underline{a},\boldsymbol{\Sigma}}(\underline{Z}) \sim |X_1|^2$. From Lemma 4.3, we know that $X_1 \sim CN(0,1)$ is equivalent to $\Re(X_1), \Im(X_1) \overset{\text{iid}}{\sim} N(0, 1/2)$, yielding $|X_1|^2 \sim \frac{1}{2}\chi_2^2$. Since $\Psi_{\underline{a},\tilde{\boldsymbol{\Sigma}}}(\underline{Z}) = \frac{\underline{a}^* \boldsymbol{\Sigma} \underline{a}}{\underline{a}^* \tilde{\boldsymbol{\Sigma}} \underline{a}} \Psi_{\underline{a},\boldsymbol{\Sigma}}(\underline{Z})$, the result follows. □

Proof (of Theorem 7.11). The proof is an adaption of Lemma B3 in Choudhuri et al. (2004a) and Lemma A2 in Kirch et al. (2017). We first consider the case (7.26). Let $\boldsymbol{Q}_j := \boldsymbol{\Sigma}_{0j}^{-1/2} \boldsymbol{\Sigma}_{1j} \boldsymbol{\Sigma}_{0j}^{-1/2}$ for $j = 1, \ldots, m$. Denote by $\underline{0} \neq \underline{b}_j \in \mathbb{C}^d$ an eigenvector of $\boldsymbol{Q}_j$ corresponding the largest eigenvalue $\lambda_{\max}(\boldsymbol{Q}_j)$ and let $\underline{a}_j := \boldsymbol{\Sigma}_{0j}^{-1/2} \underline{b}_j$. Consider the test statistic

$$T_m := \sum_{j=1}^m \Psi_{\underline{a}_j,\boldsymbol{\Sigma}_{0j}}(\underline{Y}_j), \quad \Psi_{\underline{a}_j,\boldsymbol{\Sigma}_{0j}}(\underline{Y}_j) := \frac{\underline{a}_j^* \underline{Y}_j \underline{Y}_j^* \underline{a}_j}{\underline{a}_j^* \boldsymbol{\Sigma}_{0j} \underline{a}_j}$$

and define the corresponding test φ_m as follows:

$$\varphi_m = 1 \text{ if } T_m > m\left(1 + \frac{\varepsilon}{2}\right), \quad \varphi_m = 0 \text{ else.}$$

Then, with $1 < x := 1 + \frac{\varepsilon}{2} < \frac{3}{2}$ (for $\varepsilon < 1$), we obtain for arbitrary $z > 0$ with Markov's inequality

$$\mathrm{E}_{H_0} \varphi_m = P_{H_0}(T_m > mx) = P_{H_0}(e^{T_m z} > e^{mxz}) \leq e^{-mxz} \mathrm{E}_{H_0} e^{T_m z}.$$

Under H_0, it holds $\Psi_{\underline{a}_j,\boldsymbol{\Sigma}_{0j}}(\underline{Y}_j) \overset{\text{iid}}{\sim} \frac{\chi_2^2}{2}$, $j = 1, \ldots, m$, by Lemma 7.13, i.e. $T_m \overset{H_0}{\sim} \frac{\chi_{2m}^2}{2}$. Thus the moment generating function ψ_m of T_m exists and is given as $\psi_m(z) = \mathrm{E}_{H_0} e^{T_m z} = (1-z)^{-m}$ for $z < 1$, yielding

$$\mathrm{E}_{H_0} \varphi_m \leq \exp\left(-m(\log(1-z) + xz)\right), \quad \text{for all } 0 < z < 1.$$

The function $g(z) = \log(1-z) + xz$ attains its maximum at $z^* := 1 - \frac{1}{x} \in (0,1)$ and $g(z^*) = \log\frac{1}{x} - 1 + x = x - (1 + \log x) > 0$ since $x > 1$. This yields $\mathrm{E}_{H_0}\varphi_m \leq \exp(-c_0 m)$ with $c_0 = g(z^*)$, concluding (7.28).

Under H_1, it holds

$$T_m \sim \sum_{j=1}^{m} \frac{\underline{a}_j^* \boldsymbol{\Sigma}_{1j} \underline{a}_j}{\underline{a}_j^* \boldsymbol{\Sigma}_{0j} \underline{a}_j} X_j = \sum_{j=1}^{m} \lambda_{\max}(\boldsymbol{Q}_j) X_j$$

with $X_1, \ldots, X_m \overset{\text{iid}}{\sim} \frac{\chi_2^2}{2}$ by Lemma 7.13 and Lemma 7.12. Using assumption (7.26), this yields, with $S_m \sim \frac{\chi_{2m}^2}{2}$,

$$\mathrm{E}_{H_1}(1-\varphi_m) = P_{H_1}\left(T_m \leq m\left(1+\frac{\varepsilon}{2}\right)\right) \leq P\left((1+\varepsilon)S_m \leq m\left(1+\frac{\varepsilon}{2}\right)\right) = P(S_m \leq m\tilde{x})$$

with $\tilde{x} := \frac{1+\varepsilon/2}{1+\varepsilon} \in \left(\frac{1}{2}, 1\right)$. With the same arguments as under H_0, we get for $0 < \tilde{z} < 1$,

$$P(S_m \leq m\tilde{x}) = P(e^{-S_m\tilde{z}} \geq e^{-m\tilde{x}\tilde{z}}) \leq \exp(-m(\log(1+\tilde{z}) - \tilde{x}\tilde{z})).$$

The function $\tilde{g}(\tilde{z}) = \log(1+\tilde{z}) - \tilde{x}\tilde{z}$ attains its maximum at $\tilde{z}^* := \frac{1}{\tilde{x}} - 1 \in (0,1)$ and $\tilde{g}(\tilde{z}^*) = \log\frac{1}{\tilde{x}} - 1 + \tilde{x} > 0$ since $0 < \tilde{x} < 1$. This yields (7.29) with $c_1 = \tilde{g}(\tilde{z}^*)$, concluding the proof for the case (7.26).

The proof for the case (7.27) is similar, using the test statistic

$$\tilde{T}_m := \sum_{j=1}^{m} \Psi_{\tilde{\underline{a}}_j, \boldsymbol{\Sigma}_{0j}}(\underline{Y}_j), \quad \Psi_{\tilde{\underline{a}}_j, \boldsymbol{\Sigma}_{0j}}(Y_j) := \frac{\tilde{\underline{a}}_j^* \underline{Y}_j \underline{Y}_j^* \tilde{\underline{a}}_j}{\tilde{\underline{a}}_j^* \boldsymbol{\Sigma}_{0j} \tilde{\underline{a}}_j},$$

where $\tilde{\underline{a}}_j := \boldsymbol{\Sigma}_{0j}^{-1/2} \tilde{\underline{b}}_j$ and $\underline{0} \neq \tilde{\underline{b}}_j \in \mathbb{C}^d$ being an eigenvector of $\boldsymbol{Q}_j$ corresponding to the smallest eigenvalue $\lambda_{\min}(\boldsymbol{Q}_j)$. The test φ_m is then defined as

$$\varphi_m = 1 \text{ if } T_m < m\left(1 - \frac{\varepsilon}{2}\right), \quad \varphi_m = 0 \text{ else}$$

and the proof of (7.28) and (7.29) is similar as for the case (7.26). □

Together with Corollary 7.10, we can now use Theorem 7.11 to construct an exponentially consistent and powerful test against a fixed alternative.

Lemma 7.14. *Let $\boldsymbol{f}_0 \colon [0,\pi] \to \mathcal{S}_d^+$ be continuous and fulfill Assumption **f** 1. Let $\tau > b_1$ with b_1 from Assumption **f** 1. Let $\boldsymbol{f} \in \Theta_n \cap U_\varepsilon^c$, with $U_\varepsilon = \left\{\boldsymbol{f} \colon \int_0^\pi \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\| d\omega < \varepsilon\right\}$ and Θ_n from (7.20). Let $\underline{\tilde{Z}}_1, \ldots, \underline{\tilde{Z}}_N$ be distributed as the multivariate Fourier coefficients under the Whittle likelihood $\tilde{P}_W^n(\cdot|\tilde{\boldsymbol{f}})$, i.e. $\underline{\tilde{Z}}_j \sim CN_d(\underline{0}, 2\pi\tilde{\boldsymbol{f}}(\omega_j))$ for $j = 1, \ldots, N = \left\lceil \frac{n}{2} \right\rceil - 1$. Then there exists a test $\varphi_{n,\boldsymbol{f},\varepsilon}$, depending only on $\boldsymbol{f}$ and $\boldsymbol{f}_0$ to test*

$$H_0 \colon \tilde{\boldsymbol{f}} = \boldsymbol{f}_0 \quad \textit{against} \quad H_1 \colon \tilde{\boldsymbol{f}} = \boldsymbol{f}$$

such that

$$\mathrm{E}_{H_0}\varphi_{n,\boldsymbol{f},\varepsilon} \leq 2\exp(-c_0 n), \quad \mathrm{E}_{H_1}(1 - \varphi_{n,\boldsymbol{f},\varepsilon}) \leq \exp(-c_1 n)$$

for all n, with positive constants c_0, c_1 depending only on ε, τ and d.

Proof. For $j = 1, \dots, N$, let $\boldsymbol{Q}_j := \boldsymbol{f}_0(\omega_j)^{-1/2} \boldsymbol{f}(\omega_j) \boldsymbol{f}_0(\omega_j)^{-1/2}$ and $\underline{a}_j := \boldsymbol{f}_0(\omega_j)^{-1/2} \tilde{\underline{a}}_j$ with $\underline{0} \neq \tilde{\underline{a}}_j \in \mathbb{C}^d$ being an eigenvector of $\boldsymbol{Q}_j$ corresponding to the largest eigenvalue $\lambda_{\max}(\boldsymbol{Q}_j)$. Similarly, let $\underline{b}_j := \boldsymbol{f}_0(\omega_j)^{-1/2} \tilde{\underline{b}}_j$ with $0 \neq \tilde{\underline{b}}_j \in \mathbb{C}^d$ being an eigenvector of $\boldsymbol{Q}_j$ corresponding to the smallest eigenvalue $\lambda_{\min}(\boldsymbol{Q}_j)$. Let $\tilde{\varepsilon} := \frac{\varepsilon}{4\pi\tau\sqrt{d}}$. Consider the two tests $\varphi^+_{n,\boldsymbol{f},\varepsilon}$ and $\varphi^-_{n,\boldsymbol{f},\varepsilon}$ defined as

$$\varphi^+_{n,\boldsymbol{f},\varepsilon} = 1 \text{ if } \sum_{j \in I^+_{n,\boldsymbol{f}}} \frac{\underline{a}_j^* \tilde{\underline{Z}}_j \tilde{\underline{Z}}_j^* \underline{a}_j}{\underline{a}_j^* \boldsymbol{f}_0(\omega_j) \underline{a}_j} > m^+ \left(1 + \frac{\tilde{\varepsilon}}{2}\right), \quad \varphi^+_{n,\boldsymbol{f},\varepsilon} = 0 \text{ else},$$

$$\varphi^-_{n,\boldsymbol{f},\varepsilon} = 1 \text{ if } \sum_{j \in I^-_{n,\boldsymbol{f}}} \frac{\underline{b}_j^* \tilde{\underline{Z}}_j \tilde{\underline{Z}}_j^* \underline{b}_j}{\underline{b}_j^* \boldsymbol{f}_0(\omega_j) \underline{b}_j} < m^- \left(1 - \frac{\tilde{\varepsilon}}{2}\right), \quad \varphi^-_{n,\boldsymbol{f},\varepsilon} = 0 \text{ else},$$

with

$$I^+_{n,\boldsymbol{f}} := \{j \colon \lambda_{\max}(\boldsymbol{Q}_j) > 1 + \tilde{\varepsilon}\}, \quad m^+ := \# I^+_{n,\boldsymbol{f}},$$
$$I^-_{n,\boldsymbol{f}} := \{j \colon \lambda_{\min}(\boldsymbol{Q}_j) < 1 - \tilde{\varepsilon}\}, \quad m^- := \# I^-_{n,\boldsymbol{f}}.$$

By Corollary 7.10, it holds $m^+ + m^- \geq \frac{N\varepsilon}{8\tau(b-a)} - 4k_n \geq n\left(\frac{\varepsilon}{16\tau(b-a)} - \delta\right) > 0$ by the choice of k_n in (7.20), if δ is chosen small enough. Thus, with $\tilde{\delta} := \frac{1}{2}\left(\frac{\varepsilon}{16\tau(b-a)} - \delta\right) > 0$, it holds either $m^+ \geq \tilde{\delta} n$ or $m^- \geq \tilde{\delta} n$. We define the test as

$$\varphi_{n,\boldsymbol{f},\varepsilon} := \max\left\{\mathbb{1}_{\{m^+ \geq \tilde{\delta} n\}} \varphi^+_{n,\boldsymbol{f},\varepsilon}, \mathbb{1}_{\{m^- \geq \tilde{\delta} n\}} \varphi^-_{n,\boldsymbol{f},\varepsilon}\right\}.$$

Applying Theorem 7.11 to $\{\tilde{\underline{Z}}_j \colon j \in I^+_{n,\boldsymbol{f}}\}$ and $\{\tilde{\underline{Z}}_j \colon j \in I^-_{n,\boldsymbol{f}}\}$ yields

$$\mathrm{E}_{H_0} \varphi^\dagger_{n,\boldsymbol{f},\varepsilon} \leq 2 \exp(-c_0^\dagger m^\dagger), \quad \mathrm{E}_{H_1}(1 - \varphi^\dagger_{n,\boldsymbol{f},\varepsilon}) \leq \exp(-c_1^\dagger m^\dagger),$$

for $\dagger \in \{+, -\}$ and constants $c_i^\dagger$ only depending on $\tilde{\varepsilon}$, i.e. depending on ε and τ and d. This yields the desired properties of $\varphi_{n,\boldsymbol{f},\varepsilon}$ with $c_0 = \tilde{\delta} \min\{c_0^+, c_0^-\}$ and $c_1 = \tilde{\delta} \max\{c_1^+, c_1^-\}$. □

Now we can construct a uniformly exponentially powerful test, by combining the tests against fixed alternatives from Lemma 7.14 with the bound for the covering number from Lemma 7.8.

Lemma 7.15. *Let $\boldsymbol{f}_0 \colon [0, \pi] \to \mathcal{S}_d^+$ be continuous and fulfill Assumption **f** 1. Let $\tau > b_1$ with b_1 from Assumption **f** 1. Let $U_\varepsilon = \{\boldsymbol{f} \colon \int_0^\pi \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\| < \varepsilon\}$ for $\varepsilon > 0$ small enough and let Θ_n and k_n be as in (7.20) with $\delta > 0$ small enough. Let $\tilde{\underline{Z}}_1, \dots, \tilde{\underline{Z}}_N$ be as in Lemma 7.14. Then there exists a test φ_n to test*

$$H_0 \colon \tilde{\boldsymbol{f}} = \boldsymbol{f}_0 \quad \textit{against} \quad H_1 \colon \tilde{\boldsymbol{f}} \in U_\varepsilon^c \cap \Theta_n$$

such that

$$\mathrm{E}_{\boldsymbol{f}_0} \varphi_n \to 0 \quad \textit{and} \quad \sup_{\boldsymbol{f} \in U_\varepsilon^c \cap \Theta_n} \mathrm{E}_{\boldsymbol{f}}(1 - \varphi_n) \leq \exp(-c_1 n)$$

as $n \to \infty$ for some constant $c_1 > 0$, where $\mathrm{E}_{\tilde{\boldsymbol{f}}}$ denotes the expected value under the joint distribution of $\tilde{Z}_1, \dots, \tilde{Z}_N$.

Proof. The proceeding is as described e.g. in Section 6.4 in Ghosal and van der Vaart (2017). We start by covering $U_\varepsilon^c \cap \Theta_n$ with balls of radius $\frac{\varepsilon}{2}$ in $\|\cdot\|_{F,\infty}$, with centers $\boldsymbol{f}_1, \dots, \boldsymbol{f}_{N_\varepsilon}$, where $N_\varepsilon = N\left(\frac{\varepsilon}{2}, \Theta_n, \|\cdot\|_{F,\infty}\right)$ is the $\varepsilon/2$-covering number. Then we use the tests $\varphi_{n,j} := \varphi_{n,\boldsymbol{f}_j,\varepsilon/2}$, $j = 1, \dots, N_\varepsilon$ from Lemma 7.14 against the respective simple alternative $\boldsymbol{f}_j$ to define a test

$$\varphi_n := \max\{\varphi_{n,1}, \dots, \varphi_{n,N_\varepsilon}\}.$$

For $\boldsymbol{f} \in U_\varepsilon^c \cap \Theta_n$, let $j(\boldsymbol{f})$ be such that $\|\boldsymbol{f}_{j(\boldsymbol{f})} - \boldsymbol{f}\|_{F,\infty} < \frac{\varepsilon}{2}$. Then by Lemma 7.14, we readily obtain that φ_n is uniformly exponentially powerful:

$$\sup_{\boldsymbol{f} \in U_\varepsilon^c \cap \Theta_n} \mathrm{E}_{\boldsymbol{f}}(1-\varphi_n) \le \sup_{\boldsymbol{f} \in U_\varepsilon^c \cap \Theta_n} \mathrm{E}_{\boldsymbol{f}}\left(1-\varphi_{n,j(\boldsymbol{f})}\right) \le \sup_{\boldsymbol{f} \in U_\varepsilon^c \cap \Theta_n} \exp(-c_1 n) = \exp(-c_1 n),$$

because c_1 does not depend on $\boldsymbol{f} \in U_\varepsilon^c \cap \Theta_n$. The size of φ_n can be bounded with an upper bound for the covering number N_ε, which has been given in Lemma 7.8. Together with the size obtained for the $\varphi_{n,j}$'s in Lemma 7.14, this yields

$$\mathrm{E}_{\boldsymbol{f}_0}\varphi_n \le \sum_{j=1}^{N_\varepsilon} \mathrm{E}_{\boldsymbol{f}_0}\varphi_{n,\boldsymbol{f}_j} \le 2N_\varepsilon \exp(-c_0 n) \le 2\exp\left(k_n\left(4d^2 \log k_n + \tilde{C}\right) - c_0 n\right)$$

where $\tilde{C}$ is a constant that depends only on τ, ε and d. Now recalling from (7.20) that $k_n = \lfloor \delta n / \log n \rfloor$ with $\delta > 0$ sufficiently small, we obtain for sufficiently large values of n

$$k_n(4d^2 \log k_n + \tilde{C}) - c_0 n \le 8d^2 k_n \log k_n - c_0 n \le 8d^2 \delta n \left(\frac{\log \delta}{\log n} + 1 - \frac{\log\log n}{\log n} - \frac{c_0}{8d^2\delta}\right)$$

which is bounded from above by $-\tilde{c}n$ for a positive constant $\tilde{c}$ if $\delta < \min\{1, \frac{c_0}{8d^2}\}$, because in this case

$$\frac{\log \delta}{\log n} + 1 - \frac{\log\log n}{\log n} - \frac{c_0}{8d^2\delta} < 1 - \frac{c_0}{8d^2\delta} < 0.$$

This shows $\mathrm{E}_{\boldsymbol{f}_0}\varphi_n \lesssim \exp(-\tilde{c}n) \to 0$ as $n \to \infty$. □

7.1.5. Proof of Consistency Theorem

The following result bounds the prior mass of the sieve complement.

Lemma 7.16. *Under the assumptions of Theorem 7.3, there exist positive constants $C, \tilde{c}$ such that the complement $\Theta_n^c = \mathcal{D}_{d\times d} \setminus \Theta_n$ of Θ_n from (7.20) fulfills*

$$P_\tau(\Theta_n^c) \le C\exp(-\tilde{c}n)$$

for all n.

Proof. Since

$$P_\tau(\Theta_n^c) = \frac{P(\Theta_n^c \cap \mathcal{C}_\tau)}{P(\mathcal{C}_\tau)} \le \frac{1}{P(\mathcal{C}_\tau)} P(\Theta_n^c) = \frac{1}{P(\mathcal{C}_\tau)} \sum_{k>k_n} p(k)$$

it suffices to show the existence of a positive constant $\tilde{c}$ such that $\sum_{k>k_n} p(k) \lesssim \exp(-\tilde{c}n)$, with the prior $p(k)$ fulfilling Assumption *k*1. We start with

$$\sum_{k>k_n} p(k) \leq C \sum_{k>k_n} \exp(-ck \log k) \leq C \int_{k_n}^{\infty} \exp(-cx \log x) dx,$$

where in the last step it was used that the function $(0, \infty) \ni x \mapsto \exp(-cx \log x)$ is monotonically decreasing. We continue with

$$\int_{k_n}^{\infty} \exp(-cx \log x) dx \leq \int_{k_n}^{\infty} \exp(-cx \log k_n) dx = \frac{1}{c \log k_n} \exp(-ck_n \log k_n).$$

Clearly it holds $\frac{1}{c \log k_n} \leq 1$ for n large enough. Furthermore, using $k_n = \left\lfloor \frac{\delta n}{\log n} \right\rfloor \geq \frac{\tilde{\delta} n}{\log n}$ for any $0 < \tilde{\delta} < \delta$ (for n large enough), it follows

$$\begin{aligned}
\exp(-ck_n \log k_n) &\leq \exp\left(-\frac{c\tilde{\delta} n}{\log n} \left(\log \tilde{\delta} + \log n - \log \log n \right) \right) \\
&= \exp\left(-c\tilde{\delta} n + \frac{c\tilde{\delta} n}{\log n} \left(\log \log n - \log \tilde{\delta} \right) \right) \\
&\leq \exp\left(-c\tilde{\delta} n + \frac{2c\tilde{\delta} n}{\log n} \log \log n \right) \leq \exp(-\tilde{c} n)
\end{aligned}$$

for n large enough for a positive constant $\tilde{c}$. □

Now we can collect all the previous results and present the proof of the consistency theorem.

Proof (of Theorem 7.3). We show that all assumptions of Theorem 7.1 are satisfied. By Lemma 7.4 and Lemma 7.6, it follows that $B := \{\boldsymbol{f} \in \mathcal{D}_{d\times d} \colon \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} < \varepsilon\}$ fulfills the assumptions (a)-(c) of Theorem 7.1. Let the sieve Θ_n be as in (7.20), with $\delta > 0$ small enough (as specified in the proof of Lemma 7.15). Then it follows from Lemma 7.15 that the testability assumptions (d) and (e) of Theorem 7.1 are satisfied. Finally, assumption (f) of Theorem 7.1 is fulfilled by Lemma 7.16, concluding the proof. □

7.1.6. Discussion

We have shown posterior consistency of the spectral density of a multivariate time series. Our proof technique relies on the assumption that the prior on $\boldsymbol{f}$ is supported on a set $\{\boldsymbol{f} \in \mathcal{D}_{d\times d} \colon \|\boldsymbol{f}\|_2 < \tau\}$ for some $\tau > 0$ sufficiently large to cover the ground truth $\boldsymbol{f}_0$. This assumption is needed for our proof of exponentially powerful model testability, which would not work for arbitrarily high-peaked alternatives. Prior restrictions of this type are common practice in the theoretical analysis of Bayesian nonparametrics (see e.g. Section 5 in Shen and Wasserman (2001), Section 4 in Wu and Ghosal (2010), Section 2.2 in van der Vaart and van Zanten (2009) or Section 6 in Ghosal and Van Der Vaart (2007), among others). However, one may ask if it is possible to relax the prior restriction, e.g. to allow for the truncation bound to grow with the sample size, i.e. $\tau_n \to \infty$ as $n \to \infty$. Such assumptions (involving τ_n prior restrictions

on both $\boldsymbol{f}$ and its derivative) have been considered in Ghosal and Roy (2006) and Choi and Schervish (2007) in the context of nonparametric regression and it will be of interest for future research if they can also be employed in the scope of this work.

It may be noted that, although the formulation (and the proof) of Theorem 7.3 involves the Frobenius matrix norm, the result can equivalently be stated for any other matrix norm, say $\|\cdot\|_*$. This is due to the equivalence of matrix norms. The statement of the Theorem then translates to neighborhoods of the form $U_{\varepsilon,*} := \{\boldsymbol{f} \in \mathcal{D}_{d\times d}\colon \int_0^\pi \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\|_* d\omega < \varepsilon\}$. In fact, the Frobenius norm has been chosen mainly for mathematical convenience to conduct the proofs – due to its close connection to eigenvalues and the trace.

One further interesting aspect for future research is to widen the class of stationary time series for which consistency holds beyond Gaussianity. Our proof technique relies heavily on the mutual contiguity of the exact likelihood and Whittle's likelihood approximation for Gaussian time series. For different classes of distributions, this result is not true in general (see the discussion after Corollary 4.5) and hence a different proof technique will be needed to establish consistency in the non-Gaussian case. Another interesting extension would be to consider the high-dimensional asymptotic setting, in which the dimension d of the observations is allowed to depend on n and to grow to infinity as n tends to infinity.

7.2. Posterior Contraction Rates

A posterior consistency result states that the posterior distribution converges asymptotically to the degenerate measure at the true parameter θ_0. In other words, for every arbitrarily small (but fixed) neighborhood U of θ_0, all posterior probability mass will eventually aggregate within U as the sample size grows to infinity. It is natural to ask *how fast* this convergence takes place. The idea of how to measure this constitutes the usage of *shrinking* (instead of fixed) neighborhoods U_n of θ_0. The question is then, how fast the radii of these neighborhoods can shrink to 0, such that they still capture all posterior mass asymptotically. This idea is made mathematically rigorous by the notion of contraction rates. A brief introduction into posterior contraction rates is given in Section 7.2.1. Our main theorem, establishing contraction rates for the spectral density matrix in the Hellinger topology is presented in Section 7.2.2. The proof of the theorem will be developed in Sections 7.2.3-7.2.5 and the final Section 7.2.6 concludes with a discussion of our findings.

7.2.1. Introduction

Let (ε_n) be a sequence of positive numbers, typically fulfilling $\varepsilon_n \to 0$ as $n \to \infty$ (although the latter property is not needed technically for the definition). Let $(Z_1, \dots, Z_n) \sim P_\theta^n$, with $\theta \in \Theta$ and Θ being some parameter space endowed with a metric d. Then (ε_n) is a *contraction rate* at θ_0 with respect to d, if for every sequence (M_n) of positive numbers with $M_n \to \infty$ it holds

$$P\left(B^c_{M_n\varepsilon_n}(\theta_0)|Z_1, \dots, Z_n\right) \to 0, \quad \text{in } P^n_{\theta_0} \text{ probability as } n \to \infty, \tag{7.30}$$

with $B_\varepsilon(\theta_0) = \{\theta \in \Theta \colon d(\theta_0, \theta) < \varepsilon\}$. Here, *for every* sequence (M_n) in particular refers to sequences that diverge *arbitrarily slowly* to infinity as n grows. In fact, there are examples in which the sequence M_n can be replaced by a positive constant M (see e.g. Section 2 in Ghosal et al. (2000)), however for many proof techniques this is not possible. Still, for M_n tending to infinity arbitrarily slowly, the radius $M_n\varepsilon_n$ in (7.30) can be viewed as *close enough* to ε_n, such that ε_n allows to be interpreted as *order* of convergence speed. For a more detailed discussion, see e.g. Chapter 8 in Ghosal and van der Vaart (2017).

Theorem 7.1 is not suited to derive good contraction rates, since it is usually necessary to employ stronger assumptions on the model and the prior. Of course, it can be expected that model testability and prior positivity will also be needed and it is additionally required to quantify *how much* mass the prior distribution puts in certain regions and *how fast* the convergence of the model tests takes place – this will be made precise in the upcoming Theorem 7.17, which can be conceived as an extension of Theorem 7.1.

Rates of convergence for iid models have been derived in Ghosal et al. (2000), and – using a different entropy measure – also for certain regression models in Shen and Wasserman (2001). A general theory for non-iid observations has been developed in Ghosal and Van Der Vaart (2007), where the authors also established contraction rates for the spectral density of a univariate Gaussian time series for a truncated version of the Bernstein-Dirichlet prior from Choudhuri et al. (2004a) and Whittle's likelihood in the Hellinger topology. Furthermore, the authors also derived conditions for contraction rates in the $\mathbb{L}^2$-topology, using the full Gaussian likelihood and prior distributions that are confined to uniform Hölder classes.

Theorem 7.17. *Let $Z_{1,n}, \dots, Z_{n,n}$ be independently distributed having density $p_{j,n}(\cdot, \theta)$ with respect to a σ-finite measure on a Borel space $\mathcal{X}_j$ for $j = 1, \dots, n$, where $\theta \in \Theta$ and Θ is some measurable space. Let $\theta_0 \in \Theta$. Denote the joint distribution of $(Z_{1,n}, \dots, Z_{n,n})$ under $\theta \in \Theta$ by P_θ^n. Let the Kullback-Leibler (KL) divergences $K_{j,n}$ and variance terms $V_{j,n}$ be defined as in (7.1) and let*

$$K_n(\theta_0, \theta) := \frac{1}{n} \sum_{j=1}^{n} K_{j,n}(\theta_0, \theta), \quad V_n(\theta_0, \theta) := \frac{1}{n} \sum_{j=1}^{n} V_{j,n}(\theta_0, \theta).$$

Consider the KL-type neighborhoods

$$B_{n,2}(\theta_0, \delta) := \left\{\theta \in \Theta \colon K_n(\theta_0, \theta) < \delta^2, V_n(\theta_0, \theta) < \delta^2\right\}.$$

Let P be a prior on Θ and consider a sieve sequence (Θ_n) with $\Theta_n \subset \Theta$. Let (ε_n) be a sequence of positive numbers with $\varepsilon_n \to 0$ and $n\varepsilon_n^2 \to \infty$ as $n \to \infty$. Let d_n, e_n be two semimetrics (as defined in Definition B.30 in the Appendix) on Θ. Let the following assumptions be fulfilled:

- *Prior mass of neighborhoods: There exists $c > 0$ such that*

  (a) $P(B_{n,2}(\theta_0, \varepsilon_n)) \geq \exp(-cn\varepsilon_n^2)$.

- *Existence of tests: There exist $\xi, K > 0$ such that for all $\delta > 0$ and $\theta_1 \in \Theta$ with $d_n(\theta_1, \theta_0) > \delta$, there exists a test φ_n with*

  (b) $E_{\theta_0}^n \varphi_n \leq \exp(-Kn\delta^2)$,

(c) $\sup_{\theta\in\Theta:\, e_n(\theta,\theta_1)<\xi\delta} E_\theta^n(1-\varphi_n) \leq \exp(-Kn\delta^2)$,

where E_θ^n *refers to the expected value under* P_θ^n.

- *Sieve entropy and complement mass: With the* $\xi > 0$ *from (c) it holds*

  (d) $\sup_{\varepsilon>\varepsilon_n} N(\xi\varepsilon, \{\theta\in\Theta_n:\, d_n(\theta,\theta_0)<2\varepsilon\}, e_n) \leq \exp(n\varepsilon_n^2)$,

  (e) $\frac{P(\Theta_n^c)}{P(B_{n,2}(\theta_0,\varepsilon_n))} = o(\exp(-2n\varepsilon_n^2))$ *as* $n\to\infty$.

Then ε_n *is a posterior contraction rate with respect to* d_n *in the sense of* (7.30).

Proof. The result is Theorem 1 and Lemma 1 in Ghosal and Van Der Vaart (2007). Indeed, we observe that Assumption (a) implies

$$\frac{P(\theta\in\Theta_n:\, j\varepsilon_n < d_n(\theta,\theta_0)\leq 2j\varepsilon_n)}{P(B_{n,2}(\theta_0,\varepsilon_n))} \leq \exp(Kn\varepsilon_n^2 j^2/2)$$

for all sufficiently large $j\in\mathbb{N}$, since the numerator on the left hand side is smaller or equal to 1. Hence Assumption (2.5) in their Theorem 1 is fulfilled. Furthermore, Assumptions (b) and (c) correspond to Assumption (2.2) in Ghosal and Van Der Vaart (2007), and Assumption (d) to Assumption (2.4). Finally, Assumption (e) corresponds to the assumptions of their Lemma 1. □

When comparing the assumptions of the posterior contraction Theorem 7.17 with the posterior consistency Theorem 7.1, the following differences are prominent. The assumptions of prior positivity (assumptions (a)-(c) of Theorem 7.1) are strengthened to an actual prior mass quantification of neighborhoods (assumption (a) of Theorem 7.17). The assumption of uniformly exponentially powerful testing (assumptions (d) and (e) of Theorem 7.1) have been modified towards an assumption for local testing against fixed alternatives, at an exponential $n\delta^2$ convergence rate (see assumptions (b) and (c) of Theorem 7.17) and a stronger upper bound for the sieve entropy in terms of covering numbers (assumption (d)). The prior mass of the sieve complement (assumption (f) of Theorem 7.1) has been refined to assumption (e). The result from Theorem 7.17 can be stated for more general experiments (e.g. dependent observations) and under suitable conditions it is also possible to strengthen the posterior contraction rate from convergence in P_θ^n probability to P_θ^n almost sure convergence, see Section 8.3 in Ghosal and van der Vaart (2017) for further details.

In view of these differences, there are two main challenges when extending a consistency result towards posterior contraction rates. Namely, the choice of a suitable topology (semimetrics d_n and e_n fulfilling the testability assumptions (b) and (c) from Theorem 7.17) and to establish the prior mass condition (a) of Theorem 7.17. Regarding the first issue, the $\mathbb{L}^1$ topology (integrated Frobenius norm) is not suited, since it is too restrictive. In fact, an inspection of the proof of Lemma 7.15 reveals that an exponential $n\delta^2$ rate cannot be established by our proof technique. A *default choice* in the setting of independently, non-identically distributed random variables is the root average squared Hellinger distance

$$d_{n,H}(\theta,\theta_0) := \sqrt{\frac{1}{n}\sum_{j=1}^n d_H^2\left(p_{j,n}(\cdot,\theta), p_{j,n}(\cdot,\theta_0)\right)} \tag{7.31}$$

(see (3.1) in Ghosal and Van Der Vaart (2007)), with the Hellinger distance $d_H(p,q)$ between two probability densities p, q on a Borel space $\mathcal{X}$ being defined as

$$d_H^2(p,q) = \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{p(z)} - \sqrt{q(z)}\right)^2 dz = 1 - \int_{\mathcal{X}} \sqrt{p(z)q(z)}dz.$$

It is known that tests satisfying assumptions (b) and (c) from Theorem 7.17 exist for $d_n = e_n = d_{n,H}$ (with $K = \frac{1}{2}$ and $\xi = \frac{1}{18}$), as the following result summarizes.

Lemma 7.18. *Let Θ be a measurable space and let $Z_{1,n}, \dots, Z_{n,n}$ be independently distributed with densities $p_j(\cdot, \theta)$ with respect to a σ-finite measure on a Borel space for $j = 1, \dots, n$, where $\theta \in \Theta$. Let $\varepsilon > 0$ and $\theta_0, \theta_1 \in \Theta$ with $d_{n,H}(\theta, \theta_0) > \varepsilon$. Then there exists a test φ_n such that*

$$\mathrm{E}_{\theta_0}^n \varphi_n \leq \exp(-\frac{1}{2}n\varepsilon^2), \qquad \sup_{\theta \in \Theta:\, d_{n,H}(\theta,\theta_1) \leq \frac{\varepsilon}{18}} \mathrm{E}_\theta^n (1 - \varphi_n) \leq \exp(-\frac{1}{2}n\varepsilon^2),$$

where E_θ^n denotes the expected value under the joint distribution of $(Z_{1,n}, \dots, Z_{n,n}|\theta)$.

Proof. See Lemma 2 in Ghosal and Van Der Vaart (2007). □

It is known that a contraction rate of $\varepsilon_n = n^{-a/(2+2a)}(\log n)^{(1+2a)/(2+2a)}$ for an a-smooth spectral density (where we use the notion of smoothness in the sense of Assumption ***f***2) for $a \in (1,2)$ of a univariate Gaussian time series with respect to $d_{n,H}$ under Whittle's likelihood can be achieved with a Bernstein-Dirichlet prior (see Ghosal and Van Der Vaart (2007) and Example 9.19 in Ghosal and van der Vaart (2017), where the authors present a result for $a \in (0,2]$, however their argument is only valid for $a \in (1,2)$, since their proof technique relies on the mutual contiguity result from Choudhuri et al. (2004b), which is only valid for $a > 1$). This rate can be improved when using splines, which are known to have better approximation properties than Bernstein polynomials (c.f. Remark B.14 in the Appendix). It it thus of interest to investigate whether these rates can also be attained in the multivariate setting under the Bernstein-Hpd-Gamma prior. In the upcoming Section 7.2.2 (see Theorem 7.20), we will settle the answer to affirmative.

7.2.2. Contraction Rates in the Root Average Squared Hellinger Distance

As for posterior consistency (see Section 7.1.2), we will also consider a truncated prior. The truncation is conducted *two-sided*, keeping the eigenvalues of the spectral density matrices uniformly away from both zero and infinity a priori. This is in line with Assumption ***f***1 on the ground truth spectral density (c.f. the discussion after Assumption ***f***2 for a brief discussion of the statistical interpretation of these conditions). To elaborate the prior, let $0 < \tau_1 \leq \tau_2 < \infty$ and consider

$$\mathcal{C}_{\tau_0,\tau_1} := \{\boldsymbol{f} \in \mathcal{D}_{d\times d} \colon \lambda_{\min}(\boldsymbol{f}(\omega)) \geq \tau_0, \lambda_{\max}(\boldsymbol{f}(\omega)) \leq \tau_1 \text{ for all } 0 \leq \omega \leq \pi\}. \tag{7.32}$$

Denote by P the Bernstein-Hpd-Gamma prior as in (5.2) and assume that $P(\mathcal{C}_{\tau_0,\tau_1}) > 0$ holds (this is e.g. the case if the prior fulfills Assumption GP4 and Assumption *k*2 below, see the

upcoming Corollary 7.21). Define the two-sided truncation of P as

$$P_{\tau_0,\tau_1}(F) := \frac{P(F \cap \mathcal{C}_{\tau_0,\tau_1})}{P(\mathcal{C}_{\tau_0,\tau_1})}, \quad F \subset \mathcal{D}_{d\times d} \text{ measurable.} \tag{7.33}$$

We will employ the following assumption on the prior probability mass function of the polynomial degree k:

Assumption k2. *There exist positive constants A_1, A_2 and positive constant κ_1, κ_2 such that*

$$A_1 \exp(-\kappa_1 k \log k) \leq p(k) \leq A_2 \exp(-\kappa_2 k), \quad k \in \mathbb{N}.$$

We shall give a brief discussion of this assumption.

Remark 7.19. *(a) Any prior probability mass function of the form $p(k) = C\exp(-ck\log k)$ for $k \in \mathbb{N}$ and positive constants c, C clearly fulfills Assumption k2. Similarly, so does any prior with probability mass function of the form $p(k) = C\exp(-ck)$.*

(b) If $k \sim \mathrm{Poi}(C)$ a priori, i.e. with probability mass function $p(k) = \frac{C^k}{k!}\exp(-C)$ for $k \in \mathbb{N}$ and a positive constant C, then Assumption k2 is also fulfilled. This follows with the same argument as in Remark 7.2 (b).

Based on the general result from Theorem 7.17 and the existence of tests for $d_{n,H}$ from Lemma 7.18, we can now formulate the contraction rate results for the spectral density matrix with respect to $d_{n,H}$, under suitable assumptions on Bernstein Hpd Gamma prior.

Theorem 7.20. *Let $\{\underline{Z}_t\}$ be a Gaussian stationary time series in $\mathbb{R}^d$ with mean zero and spectral density matrix $\boldsymbol{f}_0$ fulfilling Assumptions f1-f2 with $1 < a \leq 2$. Let $\tau_0 \in (0, b_0)$ and $\tau_1 \in (b_1, \infty)$ (with b_0 and b_1 from Assumption f1). Denote by $P^n_{\boldsymbol{f}_0}$ the joint distribution of the Fourier coefficients $\tilde{\underline{Z}}_1, \ldots, \tilde{\underline{Z}}_N$ from (7.7). Let the prior on the spectral density matrix $\boldsymbol{f}$ be given by the truncated Bernstein-Hpd-Gamma prior P_{τ_0,τ_1} from (7.33) with prior on $\boldsymbol{\Phi}$ fulfilling Assumption GP4 and prior on k fulfilling Assumption k2. Then*

$$\varepsilon_n = n^{-a/(2+2a)}(\log n)^{(1+2a)/(2+2a)}$$

is a contraction rate with respect to the root average squared Hellinger distance $d_{n,H}$ under Whittle's likelihood $\tilde{P}^n_W$ (from (5.3)), i.e. for every positive sequence (M_n) with $M_n \to \infty$ for $n \to \infty$ it holds

$$P^n_{W;\tau_0,\tau_1}\left(\{\boldsymbol{f} \colon d_{n,H}(\boldsymbol{f}, \boldsymbol{f}_0) \geq M_n\varepsilon_n\} \,|\underline{Z}_1, \ldots, \underline{Z}_n\right) \to 0, \quad \text{in } P^n_{\boldsymbol{f}_0} \text{ probability as } n \to \infty,$$

where $P^n_{W;\tau_0,\tau_1}$ denotes the pseudo posterior distribution obtained by updating the P_{τ_0,τ_1} prior with Whittle's likelihood $\tilde{P}^n_W$.

We first observe that in the setting of Theorem 7.20, the average squared Hellinger distance $d^2_{n,H}$ is given by

$$d^2_{n,H}(\boldsymbol{f}, \boldsymbol{f}_0) = \frac{1}{N}\sum_{j=1}^{N} d^2_H\big(p_{j,N}(\cdot|\boldsymbol{f}_0), p_{j,N}(\cdot|\boldsymbol{f})\big) \tag{7.34}$$

with $p_{1,N}, \ldots, p_{N,N}$ from (7.12). We will prove Theorem 7.20 by verifying the assumptions of Theorem 7.17. Since the assumptions of the contiguity result from Corollary 4.5 are fulfilled, we can without loss of generality assume that $\tilde{\underline{Z}}_1, \ldots, \tilde{\underline{Z}}_n$ are independent and distributed as under $\tilde{P}^n_W(\cdot|\boldsymbol{f}_0)$.

7.2.3. Prior Mass of Neighborhoods

The two-sided truncation in (7.33) is well-defined under the above assumptions, as the following result summarizes.

Corollary 7.21. *Let the assumptions of Theorem 7.20 be fulfilled. Let* $\varepsilon \in (0, \min\{b_0 - \tau_0, \tau_1 - b_1\})$ *and*

$$B_\varepsilon := \{\boldsymbol{f} \in \mathcal{D}_{d\times d} \colon \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} < \varepsilon\}.$$

Then it holds $B_\varepsilon \subset \mathcal{C}_{\tau_0,\tau_1}$ *for the two-sided truncation set* $\mathcal{C}_{\tau_0,\tau_1}$ *from (7.32). In particular, it holds* $P(\mathcal{C}_{\tau_0,\tau_1}) > 0$.

Proof. First observe that from Lemma B.3, we obtain $\lambda_{\min}(\boldsymbol{A} + \boldsymbol{B}) \geq \lambda_{\min}(\boldsymbol{A}) + \lambda_{\min}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{S}_d$. Let $\boldsymbol{f} \in B_\varepsilon$. These considerations lead to

$$\lambda_{\min}(\boldsymbol{f}(\omega)) \geq \lambda_{\min}(\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)) + \lambda_{\min}(\boldsymbol{f}_0(\omega)) \geq \lambda_{\min}(\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)) + b_0$$

for $0 \leq \omega \leq \pi$. Since $\varepsilon^2 > \|\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)\|^2$, it follows from (B.8) in the Appendix that $|\lambda_j(\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega))| < \varepsilon$ for $j = 1, \ldots, d$, and in particular $\lambda_{\min}(\boldsymbol{f}(\omega) - \boldsymbol{f}_0(\omega)) > -\varepsilon$, yielding $\lambda_{\min}(\boldsymbol{f}(\omega)) \geq b_0 - \varepsilon \geq \tau_0$ for $0 < \varepsilon < b_0 - \tau_0$. With the same argument as in the proof of Lemma 7.4, we also obtain $\lambda_{\max}(\boldsymbol{f}(\omega)) \leq \tau_1$ for $0 < \varepsilon < \tau_1 - b_1$. This shows $\boldsymbol{f} \in \mathcal{C}_{\tau_0,\tau_1}$ and the rest of the proof is analogous to the proof of Lemma 7.4. □

Similar to (7.20), we define a sieve structure Θ_n on $\Theta = \mathcal{D}_{d\times d}$ by

$$\Theta_n := \bigcup_{k=1}^{k_n} \left\{\mathfrak{B}(k, \underline{\boldsymbol{W}}) \colon \underline{\boldsymbol{W}} \in \mathcal{S}_d^{+k}\right\} \cap \mathcal{C}_{\tau_0,\tau_1}, \quad k_n := \rho\varepsilon_n^{-2/a}, \tag{7.35}$$

with a from Assumption ***f*2** and ε_n as in Theorem 7.20 and ρ is a positive constant to be specified later. We start with quantifying the prior mass of neighborhoods of $\boldsymbol{\Phi}$.

Lemma 7.22. *Let* $\boldsymbol{f}_0$ *fulfill Assumptions* ***f*1-*f*2**. *Let* k_n *be as in (7.35). Let* $\boldsymbol{\Phi} \sim \mathrm{CRM}_{d\times d}(\nu)$ *with Poisson mean measure* ν *from (3.1) fulfilling Assumption GP4. Let*

$$\boldsymbol{F}_0(A) := \int_A \boldsymbol{f}_0(\omega) d\omega, \quad A \subset [0, \pi] \text{ measurable}, \tag{7.36}$$

denote the spectral Hpd measure corresponding to $\boldsymbol{f}_0$. *Denote by* $(I_{j,k_n} \colon j = 1, \ldots, k_n)$ *the partition of* $[0, \pi]$ *from (5.1). Then there exist positive constants* c_0, C_0 *not depending on* n *such that*

$$P\left(\sum_{j=1}^{k_n} \|\boldsymbol{\Phi}(I_{j,k_n}) - \boldsymbol{F}_0(I_{j,k_n})\| < \frac{\varepsilon_n}{k_n}\right) \geq C_0 \exp(c_0 k_n \log \varepsilon_n)$$

holds for all n *with* ε_n *as in Theorem 7.20.*

Proof. First we note that

$$P\left(\sum_{j=1}^{k_n}\|\boldsymbol{\Phi}(I_{j,k_n})-\boldsymbol{F}_0(I_{j,k_n})\|<\frac{\varepsilon_n}{k_n}\right)\geq P\left(\max_{j=1,\dots,k_n}\|\boldsymbol{\Phi}(I_{j,k_n})-\boldsymbol{F}_0(I_{j,k_n})\|<\frac{\varepsilon_n}{k_n^2}\right).$$

Since it holds $\frac{1}{k_n}\geq\varepsilon_n$ for n large enough, the probability on the right hand side is bounded from below by

$$P\left(\max_{j=1,\dots,k_n}\|\boldsymbol{\Phi}(I_{j,k_n})-\boldsymbol{F}_0(I_{j,k_n})\|<\varepsilon_n^3\right)=\prod_{j=1}^{k_n}P\left(\|\boldsymbol{\Phi}(I_{j,k_n})-\boldsymbol{F}_0(I_{j,k_n})\|<\varepsilon_n^3\right),$$

where the independence property of $\boldsymbol{\Phi}$ from Remark 3.1 was used in the last step. Since $\mathcal{X}=[0,\pi]$ fulfills Assumptions $\mathcal{X}1$ and $\mathcal{X}2$, the assumptions of Theorem 3.7 and Corollary 3.8 are fulfilled. With $C_{\alpha,j,k_n}:=\int_{I_{j,k_n}}\alpha(x,\mathbb{S}_d^+)dx$, an application of Corollary 3.8 yields, for $j=1,\dots,k_n$,

$$P\left(\|\boldsymbol{\Phi}(I_{j,k_n})-\boldsymbol{F}_0(I_{j,k_n})\|<\varepsilon_n^3\right)\geq C\kappa_{\alpha,j,k_n}\exp\left((d^2+C_{\alpha,j,k_n}+1)3\log\varepsilon_n\right)$$

with

$$\begin{aligned}\kappa_{\alpha,j,k_n}&=\exp(-cC_{\alpha,j,k_n})\left(1-\exp\left\{-\frac{1}{2}W\left(\frac{2}{C_{\alpha,j,k_n}}\right)\right\}\right)g_0\mathcal{L}(I_{j,k_n})\\&=\exp(-cC_{\alpha,j,k_n})\left(1-\exp\left\{-\frac{1}{2}W\left(\frac{2}{C_{\alpha,j,k_n}}\right)\right\}\right)\frac{g_0\pi}{k_n},\end{aligned}$$

with g_0 from Assumption GP4 and positive constants c,C only depending on β_0,β_1 as in Assumption GP4, d and b_1 as in Assumptions $\boldsymbol{f}1$. Using $\sum_{j=1}^{k_n}C_{\alpha,j,k_n}=\int_0^\pi\alpha(x,\mathbb{S}_d^+)dx=C_\alpha$, this leads to

$$\begin{aligned}&\prod_{j=1}^{k_n}P\left(\|\boldsymbol{\Phi}(I_{j,k_n})-\boldsymbol{F}_0(I_{j,k_n})\|<\varepsilon_n^3\right)\\&\geq C^{k_n}\left(\prod_{j=1}^{k_n}\kappa_{\alpha,j,k_n}\right)\exp\left((C_\alpha+k_n(d^2+1))3\log\varepsilon_n\right),\end{aligned}\tag{7.37}$$

with

$$\prod_{j=1}^{k_n}\kappa_{\alpha,j,k_n}=\exp(-cC_\alpha)\left(\frac{g_0\pi}{k_n}\right)^{k_n}\prod_{j=1}^{k_n}\left(1-\exp\left\{-\frac{1}{2}W\left(\frac{2}{C_{\alpha,j,k_n}}\right)\right\}\right).$$

Since, by Assumption GP4 and Lemma B.10 in the Appendix it holds

$$C_{\alpha,j,k_n}=\int_{I_{j,k_n}}\int_{\mathbb{S}_d^+}\alpha(x,d\boldsymbol{U})dx\leq g_1\int_{I_{j,k_n}}dx\int_{\mathbb{S}_d^+}d\boldsymbol{U}=\frac{g_1\pi}{k_n}\frac{\tilde{\Gamma}_d(d)}{\Gamma(d^2)}\leq C_1$$

for all $j=1,\dots,k_n$ and hence, recalling that the Lambert W function (as defined in (B.27) in the Appendix) is strictly monotonically increasing on $(0,\infty)$,

$$1-\exp\left(-\frac{1}{2}W\left(\frac{2}{C_{\alpha,j,k_n}}\right)\right)\geq 1-\exp\left(-\frac{1}{2}W\left(\frac{2}{C_1}\right)\right)=:\tilde{c}>0,$$

we get

$$\prod_{j=1}^{k_n} \kappa_{\alpha,j,k_n} \geq \exp(-cC_\alpha) \left(\frac{\tilde{c}\pi g_0}{k_n}\right)^{k_n} \geq \exp(-cC_\alpha) \left(\pi \tilde{c} g_0\right)^{k_n} \varepsilon_n^{k_n}.$$

With (7.37), this leads to

$$\prod_{j=1}^{k_n} P\left(\|\boldsymbol{\Phi}(I_{j,k_n}) - \boldsymbol{F}_0(I_{j,k_n})\| < \varepsilon_n^3\right) \geq \tilde{C}^{k_n} \exp(-cC_\alpha) \exp\left((C_\alpha + k_n(3d^2+4)) \log \varepsilon_n\right)$$

for a positive constant $\tilde{C}$. The right hand side is equal to

$$\exp(-cC_\alpha) \exp\left(k_n \log \varepsilon_n \left(3d^2 + 4 + \frac{C_\alpha}{k_n} + \frac{\log \tilde{C}}{\log \varepsilon_n}\right)\right) \geq C_0 \exp(c_0 k_n \log \varepsilon_n)$$

for all n, where $C_0 := \exp(-cC_\alpha) > 0$ and c_0 is a positive constant, since $3d^2 + 4 + \{C_\alpha/k_n + \log \tilde{C}/\log \varepsilon_n\}$ is bounded from above. □

The following result quantifies the prior mass of neighborhoods of $\boldsymbol{f}_0$ under the truncated Bernstein-Hpd-Gamma prior.

Lemma 7.23. *Let the assumptions of Theorem 7.20 be fulfilled. For $j = 1, \ldots, N$ let the Kullback Leibler terms $K_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}), K_N(\boldsymbol{f}_0, \boldsymbol{f})$ and $V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f})$ be as in (7.13) and (7.14) and define $V_N(\boldsymbol{f}_0, \boldsymbol{f}) := \frac{1}{N}\sum_{j=1}^N V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f})$. For $\varepsilon > 0$, let*

$$B_{N,2}(\boldsymbol{f}_0, \varepsilon) := \left\{\boldsymbol{f} \in \Theta_n \colon K_N(\boldsymbol{f}_0, \boldsymbol{f}) < \varepsilon^2, V_N(\boldsymbol{f}_0, \boldsymbol{f}) < \varepsilon^2\right\}, \tag{7.38}$$

with the sieve Θ_n from (7.35). Then there exists a positive constant c such that

$$P_{\tau_0,\tau_1}(B_{N,2}(\boldsymbol{f}_0, \varepsilon_n)) \geq \exp(-cn\varepsilon_n^2)$$

holds for all n with $\varepsilon_n = n^{-a/(2+2a)}(\log n)^{(1+2a)/(2+2a)}$ as in Theorem 7.20.

Proof. The proof is similar to Section 7.3 in Ghosal and Van Der Vaart (2007) and Section 9.3 in Ghosal and van der Vaart (2017), where a related result has been shown for the univariate case. Let $\tilde{\varepsilon}_n := \frac{b_0 \varepsilon_n}{2}$ and $B_{\tilde{\varepsilon}_n} := \{\boldsymbol{f} \in \Theta_n \colon \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} < \tilde{\varepsilon}_n\}$. Let $\boldsymbol{f} \in B_{\tilde{\varepsilon}_n}$. From Lemma 7.5 (a), we get

$$V_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) \leq \lambda_{\min}(\boldsymbol{f}(\omega_j))^{-2} \|\boldsymbol{f}_{0j} - \boldsymbol{f}(\omega_j)\|^2 \leq \varepsilon_n^2, \quad j = 1, \ldots, N,$$

where in the last step it was used that $\lambda_{\min}(\boldsymbol{f}(\omega_j)) \geq \frac{b_0}{2}$ by (7.17) for n large enough. By (7.18), the assumptions of part (b) of Lemma 7.5 are fulfilled, yielding $K_{j,N}(\boldsymbol{f}_0, \boldsymbol{f}) \leq \varepsilon_n^2$. We conclude that $B_{\tilde{\varepsilon}_n} \subset B_{N,2}(\boldsymbol{f}_0, \varepsilon_n)$ and proceed by bounding the prior mass of $B_{\tilde{\varepsilon}_n}$ from below.

For $\tilde{\varepsilon}_n$ small enough, it holds $B_{\tilde{\varepsilon}_n} \subset \mathcal{C}_{\tau_0,\tau_1}$ (with the two-sided truncation set $\mathcal{C}_{\tau_0,\tau_1}$ from (7.32)) by Corollary 7.21 and hence $P_{\tau_0,\tau_1}(B_{\tilde{\varepsilon}_n}) = \frac{P(B_{\tilde{\varepsilon}_n})}{P(\mathcal{C}_{\tau_0,\tau_1})} \geq P(B_{\tilde{\varepsilon}_n})$, where P denotes the Bernstein-Hpd-Gamma prior without truncation.

Let $\boldsymbol{f} = \sum_{j=1}^{k_n} \boldsymbol{\Phi}(I_{j,k_n}) b(\cdot/\pi | j, k_n - j + 1)$. Recall that the components of $\boldsymbol{f}_0$ are (by Assumption $\boldsymbol{f}2$ and Lemma B.15 in the Appendix) continuously differentiable with derivative being

Hölder of order $a - 1 > 0$. Denote by $\boldsymbol{F}_0$ the spectral (Hpd) measure corresponding to $\boldsymbol{f}_0$ as in (7.36). An application of Lemma B.13 in the Appendix yields $\|\boldsymbol{f}_0 - \mathfrak{B}(k_n, \boldsymbol{F}_0)\|_{F,\infty} \leq C k_n^{-a/2}$ and hence, using $\|b(\cdot|j, k-j+1)\|_\infty \leq k$ (see (B.21) in the Appendix),

$$\|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} \leq \|\boldsymbol{f} - \mathfrak{B}(k_n, \boldsymbol{F}_0)\|_{F,\infty} + C k_n^{-a/2} \leq k_n \sum_{j=1}^{k_n} \|\boldsymbol{\Phi}(I_{j,k_n}) - \boldsymbol{F}_0(I_{j,k_n})\| + C k_n^{-a/2}$$

(recall the definition of the equidistant partition of $[0, \pi]$ from (5.1)). Choosing ρ in (7.35) large enough yields $C k_n^{-a/2} = C\rho^{-a/2}\varepsilon_n \leq \frac{\tilde{\varepsilon}_n}{2}$ and thus

$$\begin{aligned} P(B_{\tilde{\varepsilon}_n}) &\geq P\left(\left\{(k, \boldsymbol{\Phi})\colon k = k_n, k_n \sum_{j=1}^{k_n} \|\boldsymbol{\Phi}(I_{j,k_n}) - \boldsymbol{F}_0(I_{j,k_n})\| < \frac{\tilde{\varepsilon}_n}{2}\right\}\right) \\ &= P(k = k_n) P\left(\left\{\boldsymbol{\Phi}\colon \sum_{j=1}^{k_n} \|\boldsymbol{\Phi}(I_{j,k_n}) - \boldsymbol{F}_0(I_{j,k_n})\| < \frac{\tilde{\varepsilon}_n}{2k_n}\right\}\right), \end{aligned} \tag{7.39}$$

where in the last step it was used that k and $\boldsymbol{\Phi}$ are independent under the prior. By the choice of k_n it also holds

$$k_n \log k_n = \frac{-2}{a} k_n \log \varepsilon_n + k_n \log \rho \leq \frac{-3}{a} k_n \log \varepsilon_n \tag{7.40}$$

for n large enough. Hence, by prior assumption *k2* it holds

$$P(k = k_n) = p(k_n) \geq A_1 \exp(-\kappa_1 k_n \log k_n) \geq A_1 \exp(\tilde{c}_1 k_n \log \varepsilon_n)$$

for a positive constant $\tilde{c}_1$. Furthermore, it follows with Lemma 7.22 that

$$P\left(\sum_{j=1}^{k_n} \|\boldsymbol{\Phi}(I_{j,k_n}) - \boldsymbol{F}_0(I_{j,k_n})\| < \frac{\tilde{\varepsilon}_n}{2k_n}\right) \geq C \exp(\tilde{c}_2 k_n \log \varepsilon_n)$$

for positive constants $C, \tilde{c}_2$. Combining these results into (7.39), we arrive at

$$P(B_{\tilde{\varepsilon}_n}) \geq C \exp(\tilde{c}_3 k_n \log \varepsilon_n) = C \exp(\rho \tilde{c}_3 \varepsilon_n^{-2/a} \log \varepsilon_n)$$

for a positive constant $\tilde{c}_3$. Furthermore, ε_n fulfills (for n large enough)

$$\begin{aligned} \varepsilon_n^{-2/a} \log \varepsilon_n &= n^{1/(1+a)} (\log n)^{-(1/a+2)/(1+a)} \left(\frac{-a}{2+2a} \log n + \frac{1+2a}{2+2a} \log \log n\right) \\ &\geq \frac{-a}{2+2a} n^{1/(1+a)} (\log n)^{1-(1/a+2)/(1+a)} \\ &= -c_4 n n^{-a/(1+a)} (\log n)^{(a-1/a-1)/(1+a)} \\ &= -c_4 n n^{-a/(1+a)} (\log n)^{(1+2a)/(1+a)} |o(1)|, \end{aligned}$$

i.e.

$$\varepsilon_n^{-2/a} \log \varepsilon_n \geq -c_4 n \varepsilon_n^2 |o(1)| \tag{7.41}$$

as $n\varepsilon_n^2 \to \infty$ for a positive constant c_4. Hence it follows

$$P(B_{\tilde{\varepsilon}_n}) \geq C \exp(-\rho \tilde{c}_3 c_4 n \varepsilon_n^2) = \exp(\log C - \rho \tilde{c}_3 c_4 n \varepsilon_n^2) \geq \exp(-2\rho \tilde{c}_3 c_4 n \varepsilon_n^2), \tag{7.42}$$

yielding the claim with $c = 2\rho \tilde{c}_3 c_4 > 0$. □

7.2.4. Sieve Entropy and Complement Mass

The next result quantifies the sieve entropy in terms of covering numbers with respect to $d_{n,H}$. The idea is to bound $d_{n,H}$ from above by means of the maximum Frobenius norm (see the upcoming Lemma 7.24), and then apply the existing result for the latter from Lemma 7.8.

Lemma 7.24. *Let $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathcal{S}_d^+$ with $\lambda_{\min}(\boldsymbol{\Sigma}_i) \geq \tau_0 > 0$ and $\lambda_{\max}(\boldsymbol{\Sigma}_i) \leq \tau_1 < \infty$ for $i = 1, 2$. Denote by p_i the density of the $CN_d(\underline{0}, \boldsymbol{\Sigma}_i)$ distribution for $i = 1, 2$. Then the squared Hellinger distance between p_1 and p_2 can be bounded from above by*

$$d_H^2(p_1, p_2) \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|$$

for proportionality constants depending only on τ_0, τ_1 and d.

Proof. The normalizing constant of the probability density of the $CN_d(0, \boldsymbol{\Sigma}^{-1})$ distribution is

$$\int_{\mathbb{C}^d} \exp(-\underline{z}^* \boldsymbol{\Sigma} \underline{z}) d\underline{z} = \frac{\pi^d}{|\boldsymbol{\Sigma}|}, \quad \boldsymbol{\Sigma} \in \mathcal{S}_d^+.$$

With this we compute, using the results from Lemma B.5 in the Appendix,

$$\begin{aligned} d_H^2(p_1, p_2) &= 1 - \int_{\mathbb{C}^d} \sqrt{p_1(\underline{z})} \sqrt{p_2(\underline{z})} d\underline{z} \\ &= 1 - \frac{1}{\pi^d |\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2|^{1/2}} \int_{\mathbb{C}^d} \exp\left(-\underline{z}^* \left[\frac{\boldsymbol{\Sigma}_1^{-1}}{2} + \frac{\boldsymbol{\Sigma}_2^{-1}}{2} \right] \underline{z} \right) d\underline{z} \\ &= 1 - \frac{2^d}{|\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2|^{1/2} |\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}|} = 1 - \frac{|2\boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_2^{1/2}|}{|\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2|}. \end{aligned}$$

From this, we compute

$$\begin{aligned} \frac{|2\boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_2^{1/2}|}{|\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2|} &= \frac{|2\boldsymbol{I}_d|}{|\boldsymbol{\Sigma}_1^{-1/2}||\boldsymbol{\Sigma}_2^{-1/2}||\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2|} = \frac{|2I_d|}{|\boldsymbol{\Sigma}_2^{-1/2}||\boldsymbol{\Sigma}_1^{1/2} + \boldsymbol{\Sigma}_1^{-1/4} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/4}|} \\ &= \frac{|2\boldsymbol{I}_d|}{|\boldsymbol{\Sigma}_2^{-1/4} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_2^{-1/4} + \boldsymbol{\Sigma}_2^{-1/4} \boldsymbol{\Sigma}_1^{-1/4} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/4} \boldsymbol{\Sigma}_2^{-1/4}|} = \frac{|2\boldsymbol{I}_d|}{|\boldsymbol{Q} + \tilde{\boldsymbol{Q}}|}, \end{aligned}$$

with $\boldsymbol{Q} = \boldsymbol{\Sigma}_2^{-1/4} \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_2^{-1/4} \in \mathcal{S}_d^+$ and $\tilde{\boldsymbol{Q}} = \boldsymbol{\Sigma}_2^{-1/4} \boldsymbol{\Sigma}_1^{-1/4} \boldsymbol{\Sigma}_2 \boldsymbol{\Sigma}_1^{-1/4} \boldsymbol{\Sigma}_2^{-1/4} \in \mathcal{S}_d^+$ and $\boldsymbol{\Sigma}^{1/4}$ denoting the Hpd square root of $\boldsymbol{\Sigma}^{1/2}$ for $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$. Note that the eigenvalues of $\boldsymbol{Q}$ and $\tilde{\boldsymbol{Q}}$ are bounded away from 0 (this follows from the Min-Max Theorem of Courant Fisher from Lemma B.3 in the Appendix, since the eigenvalues of $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ are bounded and bounded away from 0 by assumption) and hence $|\boldsymbol{Q} + \tilde{\boldsymbol{Q}}| \geq \prod_{i=1}^d (\lambda_i(\boldsymbol{Q}) + \lambda_i(\tilde{\boldsymbol{Q}}))$ (see Lemma B.5 in the Appendix) is bounded away from 0. This yields

$$d_H^2(p_1, p_2) = \frac{|\boldsymbol{Q} + \tilde{\boldsymbol{Q}}| - |2\boldsymbol{I}_d|}{|\boldsymbol{Q} + \tilde{\boldsymbol{Q}}|} \lesssim \left| \frac{\boldsymbol{Q} + \tilde{\boldsymbol{Q}}}{2} \right| - |\boldsymbol{I}_d|,$$

where the proportionality factor depends only on τ_0, τ_1 and d. Observe that the eigenvalues of $\boldsymbol{Q} + \tilde{\boldsymbol{Q}}$ are also bounded, i.e. $\lambda_{\max}(\frac{\boldsymbol{Q} + \tilde{\boldsymbol{Q}}}{2}) \leq \tau$ for some positive constant τ (depending only on τ_0, τ_1 and d). Using $|\boldsymbol{I}_d| = 1$ and $|\boldsymbol{\Sigma}| \leq \lambda_{\max}(\boldsymbol{\Sigma})^d$ for $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$ yields

$$0 \leq d_H^2(p_1, p_2) \lesssim \left(\lambda_{\max} \left(\frac{\boldsymbol{Q} + \tilde{\boldsymbol{Q}}}{2} \right) \right)^d - 1.$$

It follows that $\tau \geq \lambda_{\max}(\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2}) \geq 1$, which implies that the largest eigenvalue $\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2} - \boldsymbol{I}_d$ is nonnegative (note however that $\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2} - \boldsymbol{I}_d$ is not positive semidefinite in general, but may have eigenvalues smaller than zero as well). By Lemma B.26 in the Appendix, there exists a positive constant $\tilde{\tau}$ such that

$$\left(\lambda_{\max}\left(\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2}\right)\right)^d - 1 \leq \tilde{\tau}\left(\lambda_{\max}\left(\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2}\right) - 1\right) \lesssim \lambda_{\max}\left(\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2}\right) - 1.$$

By Lemma B.4 in the Appendix, we find

$$\begin{aligned}\lambda_{\max}\left(\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2}\right) - 1 = \lambda_{\max}\left(\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2} - \boldsymbol{I}_d\right) &\leq \left\|\frac{\boldsymbol{Q}+\tilde{\boldsymbol{Q}}}{2} - \boldsymbol{I}_d\right\|_2 \\ &\leq \frac{1}{2}\|\boldsymbol{Q} - \boldsymbol{I}_d\|_2 + \frac{1}{2}\|\tilde{\boldsymbol{Q}} - \boldsymbol{I}_d\|_2,\end{aligned}$$

and it follows

$$d_H^2(p_1, p_2) \lesssim \|\boldsymbol{Q} - \boldsymbol{I}_d\| + \|\tilde{\boldsymbol{Q}} - \boldsymbol{I}_d\|. \tag{7.43}$$

For the first summand in (7.43) we get, using Lemma B.6 in the appendix (and recalling $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|$)

$$\|\boldsymbol{Q} - \boldsymbol{I}_d\| = \|\boldsymbol{\Sigma}_2^{-1/2}\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{I}_d\| \leq \|\boldsymbol{\Sigma}_2^{-1/2}\|_2\|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\| \leq \tau_0^{-1/2}\|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\|.$$

Consider the mapping $\boldsymbol{\Psi}\colon \mathcal{S}_d^+ \to \mathcal{S}_d^+$ defined as the Hermitian positive definite matrix square root $\boldsymbol{\Sigma} \mapsto \boldsymbol{\Psi}(\boldsymbol{\Sigma}) := \boldsymbol{\Sigma}^{1/2}$. Then for $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ it holds

$$\frac{\|\boldsymbol{\Psi}(\boldsymbol{\Sigma}_1) - \boldsymbol{\Psi}(\boldsymbol{\Sigma}_2)\|}{\|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|} \leq \frac{\|\boldsymbol{\Psi}(\boldsymbol{\Sigma}_1) - \boldsymbol{\Psi}(\boldsymbol{\Sigma}_2) - \boldsymbol{\Psi}'(\boldsymbol{\Sigma}_1)[\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1]\|}{\|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|} + \frac{\|\boldsymbol{\Psi}'(\boldsymbol{\Sigma}_1)[\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1]\|}{\|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|}, \tag{7.44}$$

where $\boldsymbol{\Psi}'$ is as in Lemma B.11. The first summand on the right hand side of (7.44) is the difference quotient of $\boldsymbol{\Psi}(\boldsymbol{Z})$ at $\boldsymbol{Z} = \boldsymbol{\Sigma}_1$ and converges to 0 as $\|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\| \to 0$, since $\boldsymbol{\Psi}$ is Fréchet differentiable by Lemma B.11. In particular, it is bounded for all $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ with bounded eigenvalues. The second summand on the right hand side of (7.44) is smaller or equal to $\|\boldsymbol{\Psi}'(\boldsymbol{\Sigma}_1)\|$, which is bounded by a constant (only depending on τ_0 and d) by Lemma B.11. This yields $\|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\| \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|$ and thus

$$\|\boldsymbol{Q} - \boldsymbol{I}_d\| \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|. \tag{7.45}$$

For the second summand in (7.43), we first observe

$$\|\tilde{\boldsymbol{Q}} - \boldsymbol{I}_d\| \leq \|\tilde{\boldsymbol{Q}} - \boldsymbol{Q}^{-1}\| + \|\boldsymbol{Q}^{-1} - \boldsymbol{I}_d\|.$$

With the same argument as for (7.45), we readily obtain $\|\boldsymbol{Q}^{-1} - \boldsymbol{I}_d\| \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|$. Furthermore

$$\begin{aligned}\|\tilde{\boldsymbol{Q}} - \boldsymbol{Q}^{-1}\| &= \|\boldsymbol{\Sigma}_2^{-1/4}\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2^{-1/4} - \boldsymbol{\Sigma}_2^{1/4}\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{1/4}\| \\ &= \|\boldsymbol{\Sigma}_2^{-1/4}(\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/4} - \boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{1/2})\boldsymbol{\Sigma}_2^{-1/4}\| \\ &\leq \tau_0^{-1/2}\|\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/4} - \boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{1/2}\| \\ &\leq \tau_0^{-1/2}(\|\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/4} - \boldsymbol{\Sigma}_2^{1/2}\| + \|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\|\end{aligned}$$

and

$$\|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\| \leq \tau_1^{1/2}\|\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_2^{1/2} - \boldsymbol{I}_d\| \leq \left(\frac{\tau_1}{\tau_0}\right)^{1/2} \|\boldsymbol{\Sigma}_2^{1/2} - \boldsymbol{\Sigma}_1^{1/2}\| \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|$$

as well as

$$\begin{aligned}\|\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/4} - \boldsymbol{\Sigma}_2^{1/2}\| &\leq \|\boldsymbol{\Sigma}_1^{-1/4}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1/4} - \boldsymbol{\Sigma}_1^{1/2}\| + \|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\| \\ &\leq \tau_0^{-1/2}\|\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1\| + \|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\| \\ &\lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|\end{aligned}$$

yield $\|\tilde{\boldsymbol{Q}} - \boldsymbol{I}_d\| \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|$. Combining this with (7.45) and (7.43) yields the claim. □

Lemma 7.25. *Let the assumptions of Theorem 7.20 be fulfilled. Then for sieve Θ_n from (7.35), the ε-covering number in the root average squared Hellinger topology fulfills*

$$\log \sup_{\varepsilon > \varepsilon_n} N(\xi\varepsilon, \{\boldsymbol{f} \in \Theta_n \colon d_{n,H}(\boldsymbol{f}, \boldsymbol{f}_0) < 2\varepsilon\}, d_{n,H}) \leq n\varepsilon_n^2$$

with ε_n as in Theorem 7.20.

Proof. The proof follows the arguments of Section 9.5.2 in Ghosal and van der Vaart (2017). Recall the representation of $d_{n,H}(\boldsymbol{f}, \boldsymbol{f}_0)$ from (7.34). From Lemma 7.24 we conclude for every $\boldsymbol{f} \in \Theta_n$

$$d_{n,H}^2(\boldsymbol{f}, \boldsymbol{f}_0) \leq \max_{j=1,\dots,N} d_H^2\left(p_{j,N}(\cdot|\boldsymbol{f}_0), p_{j,N}(\cdot|\boldsymbol{f})\right) \lesssim \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty}.$$

From this, an application of Lemma B.32 in the Appendix yields

$$\begin{aligned}\log \sup_{\varepsilon > \varepsilon_n} N(\xi\varepsilon, \{\boldsymbol{f} \in \Theta_n \colon d_{n,H}(\boldsymbol{f}, \boldsymbol{f}_0) < 2\varepsilon\}, d_{n,H}) &\leq \log N(\xi\varepsilon_n, \Theta_n, d_{n,H}) \\ &\leq \log N(c\xi^2\varepsilon_n^2, \Theta_n, \|\cdot\|_{F,\infty}) \\ &\leq c_1 k_n \log k_n - c_2 k_n \log \varepsilon_n + c_3\end{aligned}$$

for positive constants c_1, c_2, c_3, where the result from Lemma 7.8 was used in the last step. Using (7.40) and (7.41), this is bounded from above

$$\left(\frac{-3}{a} - c_2\right) k_n \log \varepsilon_n + c_3 = -\rho\left(\frac{3}{a} + c_2\right) \varepsilon_n^{-2/a} \log \varepsilon_n + c_3 \leq n\varepsilon_n^2 |o(1)| + c_3$$

as $n\varepsilon_n^2 \to \infty$. Since it also holds $c_3 = n\varepsilon_n^2 c_3/(n\varepsilon_n^2) = |o(1)| n\varepsilon_n^2$, this leads to

$$\log N(c\xi^2\varepsilon_n^2, \Theta_n, \|\cdot\|_{F,\infty}) \leq n\varepsilon_n^2 |o(1)| \leq n\varepsilon_n^2 \tag{7.46}$$

for n large enough, concluding the proof. □

7.2.5. Proof of Contraction Rate Theorem

The last piece in the puzzle is to bound the prior mass of the sieve complement $\Theta_n^c = \mathcal{D}_{d\times d} \setminus \Theta_n$, which will be done in the following Lemma.

Lemma 7.26. *Let the assumptions of Theorem 7.20 be fulfilled. Then it holds*

$$\frac{P_{\tau_0,\tau_1}(\Theta_n^c)}{P_{\tau_0,\tau_1}(B_{n,2}(\boldsymbol{f}_0, \varepsilon_n))} = o\left(\exp(-2n\varepsilon_n^2)\right)$$

with Θ_n as in (7.35) and $B_{n,2}(\boldsymbol{f}_0, \varepsilon_n)$ from (7.38).

Proof. The proof follows the arguments of Section 9.5.2 in Ghosal and van der Vaart (2017). First note that – by the same argument as in the proof of Lemma 7.16 – there exist positive constants $C, \tilde{c}$ such that

$$P_{\tau_0,\tau_1}(\Theta_n^c) \leq \frac{C}{\log k_n} \exp(-\tilde{c} k_n \log k_n) \leq \exp(-\tilde{c} k_n \log k_n)$$

holds for n large enough. With k_n as in (7.35) and ε_n as in the assumptions, we get $\tilde{c} k_n \log k_n \geq \rho n \varepsilon_n^2$ with (7.40) and (7.41). This yields

$$P_{\tau_0,\tau_1}(\Theta_n^c) \leq \exp(-\rho n \varepsilon_n^2). \tag{7.47}$$

Using the result from Lemma 7.23, we arrive at

$$\frac{P_{\tau_0,\tau_1}(\Theta_n^c)}{P_{\tau_0,\tau_1}(B_{N,2}(f_0, \varepsilon_n))} \leq \exp\left(-(\rho - c) n \varepsilon_n^2\right)$$

and choosing ρ sufficiently large (such that $\rho - c > 2$ and in accordance with the choice before (7.39)) yields the result. □

Now we can show the main result of this section by putting all previous results together.

Proof (of Theorem 7.20). We will apply Theorem 7.17. By Lemma 7.18, assumptions (b) and (c) of Theorem 7.17 are fulfilled, by Lemma 7.23, also assumption (a). Furthermore, assumption (d) follows from Lemma 7.25. Finally, assumption (e) from Theorem 7.17 is fulfilled for the sieve Θ_n from (7.35) by Lemma 7.26. □

7.2.6. Discussion

We established posterior contraction rates for the spectral density matrix of a multivariate Gaussian time series in the Hellinger topology. The rates coincide with the univariate case, which has been considered in Ghosal and Van Der Vaart (2007). It is known that these rates can be improved when using a different polynomial basis than the Bernstein polynomials.

It may be argued that the Hellinger topology is not the most "natural" topology for a curve estimation problem and it will be of interest for future research to establish rates in different topologies (such as $\mathbb{L}^1$). As an example, in Ghosal and Van Der Vaart (2007) the authors derived

conditions for $\mathbb{L}^2$ contraction rates for the spectral density of a univariate time series. These conditions include prior assumptions that are related to an a priori bound of the derivative. One main challenge for rates in different topologies is to show the existence of appropriate tests (condition (b) and (c) of Theorem 7.17). As an example, the tests for the $\mathbb{L}^1$ topology from Lemma 7.15 do not lend themselves to this purpose, since they do not yield the needed $n\varepsilon_n^2$ rate. On the other hand, the technique from Theorem 7.17 is not the only feasible way of deriving posterior contraction rates. In fact, it has been argued in Hoffmann et al. (2015) (see Section 5.3 there for a detailed discussion) that this proof technique has limitations and may yield suboptimal rates for certain models and topologies.

Part III.

Semiparametric Model

8.

Bayesian Semiparametric Linear Model

In many cases, a stationary centered time series is not the only object of interest, but merely describes the error sequence in a model with an additional finite-dimensional parameter of interest. Examples of such a situation range from linear models (with model coefficients being parameters of interest) and change point analysis (with change points being parameters of interest) to nonlinear regression (with the regression function being parameter of interest). Such models have seen recent attention in the Bayesian nonparametric literature. For example, in Amewou-Atisso et al. (2003), a Bayesian linear regression model is analyzed, where the density of the iid noise is modeled nonparametrically with Polya Tree or Dirichlet Process mixture of normals. In Edwards et al. (2015), the noise time series of a parametric gravitational wave signal has been modeled nonparametrically with a Bernstein-Dirichlet prior on the spectral density. Despite the existence of many other Bayesian semiparametric approaches in this spirit (see Section 12.3 in Ghosal and van der Vaart (2017) and the references therein for further examples), only few Bayesian semiparametric models for time series analysis have been considered in the literature, in particular from an asymptotic perspective.

8.1. Model and Prior Specification

We consider the following multivariate Bayesian linear model:

$$\underline{Z}_t = \boldsymbol{X}_t \underline{B} + \underline{e}_t, \quad t = 1, \dots, n, \tag{8.1}$$

where $\{\underline{Z}_t\}$ is the observed time series in $\mathbb{R}^d$ and $\boldsymbol{X}_t \in \mathbb{R}^{d \times r}$ is a (fixed) design matrix and $\underline{B} \in \mathbb{R}^r$ the linear model coefficient. The noise $\{\underline{e}_t\}$ is assumed to be a stationary Gaussian time series in $\mathbb{R}^d$ with $\mathrm{E}\underline{e}_t = \underline{0}$. The model equation (8.1) can equivalently be written in a vectorized formulation

$$\underline{Z} = \boldsymbol{X}\underline{B} + \underline{e},$$

with the stacked observations $\underline{Z} = (\underline{Z}_1^T, \dots, \underline{Z}_n^T)^T \in \mathbb{R}^{nd}$ and $\underline{e} = (\underline{e}_1^T, \dots, \underline{e}_n^T)^T \in \mathbb{R}^{nd}$ as well as $\boldsymbol{X} = (\boldsymbol{X}_1^T, \dots, \boldsymbol{X}_n^T)^T \in \mathbb{R}^{nd \times r}$. In what follows, we assume that the design matrix $\boldsymbol{X}$ is of full rank such that $\boldsymbol{X}^T\boldsymbol{X}$ is invertible. Denoting by $\boldsymbol{F}_{nd} \in \mathbb{R}^{nd \times nd}$ the orthogonal Fourier

transformation matrix from Lemma 4.1, the above formulation can equivalently be written in the frequency domain as

$$\tilde{\underline{Z}} = \tilde{\boldsymbol{X}}\underline{B} + \tilde{\underline{e}}, \tag{8.2}$$

with $\tilde{\underline{Z}} = \boldsymbol{F}_{nd}\underline{Z} \in \mathbb{R}^{nd}$ and $\tilde{\underline{e}} = \boldsymbol{F}_{nd}\underline{e} \in \mathbb{R}^{nd}$ as well as $\tilde{\boldsymbol{X}} = \boldsymbol{F}_{nd}\boldsymbol{X} \in \mathbb{R}^{nd\times r}$. Recall from Section 4.2 that $\tilde{\underline{Z}}$ contains the stacked real- and imaginary parts of the frequency domain observations. This representation invites us to employ Whittle's Likelihood to conduct inference about $\underline{B}$ and $\boldsymbol{f}$ in the frequency domain. Under Whittle's Likelihood P_W^n from (4.16), the frequency domain noise vector $\tilde{\underline{e}}$ is multivariate normal with mean zero and block diagonal covariance matrix $\boldsymbol{D}_{nd}$ as in (4.16), i.e. $\tilde{\underline{e}} \sim N_{nd}(\underline{0}, \boldsymbol{D}_{nd})$. Thus Whittle's Likelihood for the model (8.1) is given by the Lebesgue density

$$p_W^n(\tilde{\underline{z}}|\underline{B}, \boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^{nd}|\boldsymbol{D}_{nd}|}} \exp\left(-\frac{1}{2}\left(\tilde{\underline{z}} - \tilde{\boldsymbol{X}}\underline{B}\right)^T \boldsymbol{D}_{nd}^{-1}\left(\tilde{\underline{z}} - \tilde{\boldsymbol{X}}\underline{B}\right)\right), \tag{8.3}$$

for $\tilde{\underline{z}} \in \mathbb{R}^{nd}$. The Bayesian model is completed by a prior specification for $\underline{B}$ and $\boldsymbol{f}$. If one does not want to rely on any parametric assumption about $\{\underline{e}_t\}$ (aside the Gaussianity), it seems natural to model it nonparametrically in terms of a nonparametric model for $\boldsymbol{f}$. In this case, the above model is a *semiparametric* model, consisting of a parametric (finite dimensional) parameter of interest $\underline{B}$ and a nonparametric (infinite-dimensional) component $\boldsymbol{f}$. We will consider independent priors of the form $P(d\underline{B}, d\boldsymbol{f}) = P(d\underline{B})P(d\boldsymbol{f})$. For the spectral density matrix $\boldsymbol{f}$, we employ the nonparametric Bernstein-Hpd-Gamma prior from Section 5.1. For the linear model parameter $\underline{B}$, many priors are possible and commonly used (see e.g. Section 9 in Christensen et al. (2011) for an overview). One particularly popular noninformative prior choice which we will also employ for the numerical illustration in the upcoming Section 8.3 is the *Standard Improper Reference (SIR) Prior* (where the name is due to Section 9 in Christensen et al. (2011)), given by $p(\underline{B}) \propto 1$.

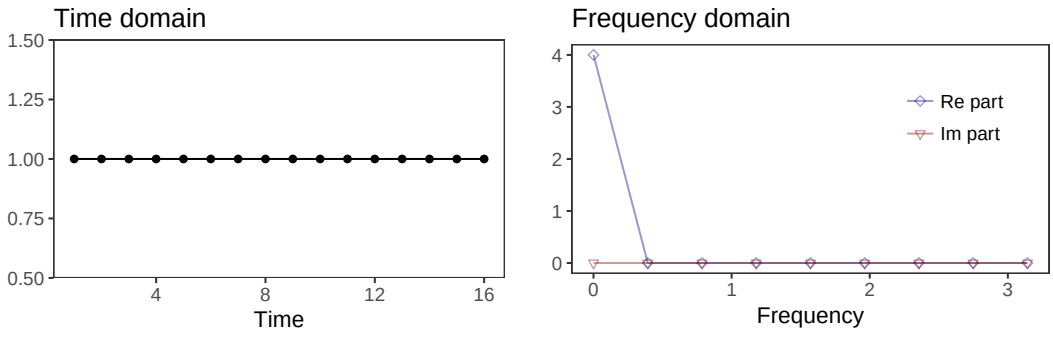
Examples

Let us consider some exemplary models that are accommodated in the model formulation (8.1):

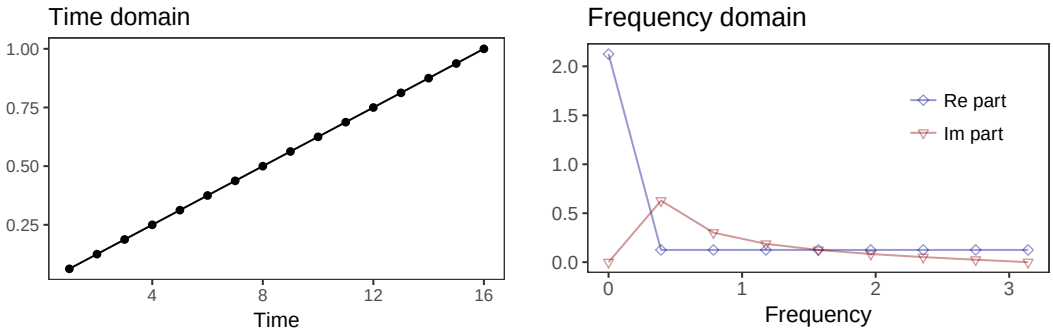
(a) Spectral density model: $\tilde{\underline{Z}} = \tilde{\underline{e}}$ with the spectral density matrix $\boldsymbol{f}$ being the only parameter. This is included in formulation (8.1) by letting $r = 0$.

(b) Mean model $\underline{Z}_t = \underline{\mu} + \underline{e}_t$: Letting $r = d$ and $\underline{B} = (\mu_1, \ldots, \mu_d)^T$ as well as $\boldsymbol{X}_t = \boldsymbol{I}_d$ for $t = 1, \ldots, n$ corresponds to each time series component having an individual mean (instead of assuming mean zero). It can readily be seen that the frequency domain representation is given by $\tilde{\underline{Z}} = \tilde{\underline{\mu}} + \tilde{\underline{e}}$ with $\tilde{\underline{\mu}} = (\sqrt{n}\underline{\mu}, 0, \ldots, 0) \in \mathbb{R}^{nd}$. Note that the only occurrence of the mean $\underline{\mu}$ from the time domain in the frequency domain is at $\omega = 0$. This is illustrated in Figure 8.1(a). As an important consequence, the full conditional of $\underline{\mu}$ under Whittle's Likelihood (8.3) only depends on $\boldsymbol{f}$ through $\boldsymbol{f}(0)$. We will have a more detailed look at this model from an asymptotic perspective in the upcoming Chapter 9.

(c) Linear trend model $\underline{Z}_t = \underline{\mu} + \underline{b}t + \underline{e}_t$: Letting $r = 2d$ and $\underline{B} = (\mu_1, \ldots, \mu_d, b_1, \ldots, b_d)$ as well as $\boldsymbol{X}_t = (\boldsymbol{I}_d, t\boldsymbol{I}_d) \in \mathbb{R}^{d\times 2d}$ corresponds to the inclusion of a linear trend for each component



(a)



(b)

Figure 8.1.: Time and frequency domain representation of (a) a mean signal $S_t = \mu$ for $t = 1, \dots, n$ and (b) a linear trend signal $S_t = bt$ for $t = 1, \dots, n$, for the exemplary signal length of $n = 16$ and $\mu = 1$ and $b = \frac{1}{16}$.

(with individual slopes and intercepts). The frequency domain representation of this model is more involved. In fact, a linear trend in the time domain yields a trigonometric signal in the frequency domain that is visible at all Fourier frequencies. To elaborate, with $S_t := t$ observe that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} S_t = \frac{\sqrt{n}(n+1)}{2}$ is the frequency domain representation at $\omega = 0$ and with $\omega_j = \frac{j2\pi}{n}$ for $j = 1, \dots, \lfloor n/2 \rfloor$ it holds (see Appendix A12 in Kammler (2000)) $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} S_t \exp(-it\omega_j) = \frac{\sqrt{n}}{2}(1 + i \cot \frac{\omega_j}{2})$ with the cotangent function $\cot \omega = \frac{\cos \omega}{\sin \omega}$ for $0 < \omega < \pi$. This is illustrated in Figure 8.1(b). As a consequence, the full conditional of $\underline{b}$ under Whittle's Likelihood (8.3) depends on $\boldsymbol{f}$ through $\boldsymbol{f}(0), \boldsymbol{f}(\omega_1), \dots, \boldsymbol{f}(\omega_{\lfloor n/2 \rfloor})$, i.e. at *all* Fourier frequencies. This is an important difference to the mean $\underline{\mu}$ model from (b).

(d) Linear trend model with common slope $\underline{Z}_t = \underline{\mu} + bt + \underline{e}_t$: Letting $r = d + 1$ and $\underline{B} = (\mu_1, \dots, \mu_d, b)$ as well as $\boldsymbol{X}_t = (\boldsymbol{I}_d | (t, \dots, t)^T) \in \mathbb{R}^{d \times (d+1)}$ corresponds to the inclusion of a linear trend for each component (with common slope and individual intercepts).

8.2. Numerical Simulations of Posterior Samples

We employ a Gibbs sampler to draw samples from the joint posterior $P_W^n(d\underline{B}, d\boldsymbol{f} | \tilde{\underline{Z}})$ of $(\underline{B}, \boldsymbol{f})$ in model (8.1) under Whittle's Likelihood $P_W^n(d\tilde{\underline{Z}} | \underline{B}, \boldsymbol{f})$ from (8.3). To draw from the full con-

ditional of $\boldsymbol{f}$, the Metropolis-within-Gibbs algorithm from Section 5.2.2 is used. We will show in the following that the full conditional of $\underline{B}$ is available, if the SIR prior $p(\underline{B}) \propto 1$ is employed. Indeed, recall that under $P_W^n(\cdot|\underline{B}, \boldsymbol{f})$, it holds that $\tilde{\underline{e}}$ is multivariate normal with mean $\underline{0}$ and covariance matrix $\boldsymbol{D}_{nd}$. The conditional likelihood of $\underline{B}$ in this case is (see Section 9.2.2 in Christensen et al. (2011))

$$\underline{B}|\underline{Z}, \boldsymbol{f} \sim N_r\left(\hat{\underline{B}}, (\tilde{\boldsymbol{X}}^T \boldsymbol{D}_{nd}^{-1} \tilde{\boldsymbol{X}})^{-1}\right), \tag{8.4}$$

where $\hat{\underline{B}} = (\tilde{\boldsymbol{X}}^T \boldsymbol{D}_{nd}^{-1} \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}} \boldsymbol{D}_{nd}^{-1} \tilde{\underline{Z}}$ is the maximum (conditional) likelihood estimator of $\underline{B}$. Under the SIR prior (being flat), the full conditional coincides with the conditional likelihood. Consider the mean model (b) from Section 8.1 as an example. The maximum likelihood estimator $\hat{\underline{\mu}}_n$ is the sample mean, which under P_W^n is distributed as the first Fourier coefficient $\tilde{\underline{Z}}(0)$ corresponding to $\omega = 0$. Hence conditional likelihood (and the full conditional under the SIR prior) of $\underline{\mu}$ is

$$\underline{\mu}|\underline{Z}, \boldsymbol{f} \sim N_d\left(\hat{\underline{\mu}}_n, 2\pi \boldsymbol{f}(0)\right), \quad \hat{\underline{\mu}}_n := \frac{1}{n}\sum_{t=1}^{n} \underline{Z}_t. \tag{8.5}$$

If another prior instead of the SIR prior is used, the distribution in (8.4) does not constitute the full conditional of $\underline{B}$. However, it can still be used as a proposal distribution in an MH step for $\underline{B}$. From our numerical experiments, we found that good mixing properties can be achieved for other prior choices than the SIR prior, when employing an MH step with the following random walk proposal scheme based on (8.4):

$$\underline{B}^{*,(i+1)} \sim N_r\left(\underline{B}^{(i)}, (\tilde{\boldsymbol{X}}^T \boldsymbol{D}_{nd}^{-1} \tilde{\boldsymbol{X}})^{-1}\right). \tag{8.6}$$

It is important to note that the Bernstein-Hpd-Gamma prior from Section 5.1 does not enforce $\boldsymbol{f}(0)$ to be symmetric positive definite (rather, it is Hermitian positive definite). The property of $\boldsymbol{f}(0)$ being spd is however desirable and in fact mandatory for a valid statistical inference in the semiparametric model: Unlike for the spectral density inference from Chapter 5, the boundary frequencies (corresponding to $\omega = 0$ and $\omega = \pi$) can *not* be left away from Whittle's Likelihood because otherwise the parameter $\underline{B}$ would not be identifiable in the model. Consider the mean model (b) from Section 8.1 as an example. We have seen that in fact the *only* occurrence of $\underline{\mu}$ in the frequency domain representation (8.2) is at the first Fourier coefficient $\tilde{\underline{Z}}(0)$ corresponding to $\omega = 0$. With a view on the full conditional of $\underline{\mu}$ from (8.5), This emphasizes the conceptual need of $\boldsymbol{f}(0)$ to be spd. Therefore, we will discard the imaginary part of $\boldsymbol{f}(0)$ (and similarly the imaginary part of $\boldsymbol{f}(\pi)$) in the Bernstein mixture parametrization (5.2) of $\boldsymbol{f}$.

8.3. Numerical Illustration

In this Section, we evaluate the performance of the proposed procedure for a multivariate linear model. We consider the linear trend model with common slope from (d) in Section 8, which can be written as

$$\underline{Z}_t = \underline{\mu} + bt + \underline{e}_t, \quad t = 1, \ldots, n. \tag{8.7}$$

Similar to the simulation study in Section 6.2, we compare the NP procedure with the VAR procedure. Additionally, we also consider the results of a procedure, in which the residual time series $\underline{e}_t$ is modeled as Gaussian White Noise. This procedure is denoted by the WN procedure (where NP stands for White Noise model). It corresponds to the VAR procedure, where the order is fixed to $p = 0$ and it is included to illustrate the necessity of taking an existing time dependence in the residuals into account.

The respective prior on the parametric part $\underline{B} = (\underline{\mu}^T, b)^T$ is given by the SIR prior $p(\underline{B}) \propto 1$ and the respective prior on the time series component is independent of $p(\underline{B})$ and as in Section 6.2 (i.e. Bernstein-Hpd-Gamma prior on $\boldsymbol{f}$ for NP and the Independent Normal-Inverse-Wishart prior for VAR) with the same prior parameters as in Section 6.2. The order p for the VAR procedure is determined in an AIC based model selection step, with the noise $\underline{e}_1, \ldots, \underline{e}_n$ directly (which is available for simulated data). The Gibbs sampling algorithms of NP and VAR are amended by an additional step for $\underline{B}$ to accommodate the linear model, with a conjugate Gibbs step drawn from (8.4) for NP and an MH step with proposal (8.6) for VAR. Each Markov Chain is run for a total length of 80,000 iterations, where the first 30,000 iterations are discarded as burn-in period and the remaining 50,000 are thinned by a factor of 5, yielding a respective posterior sample size of 10,000.

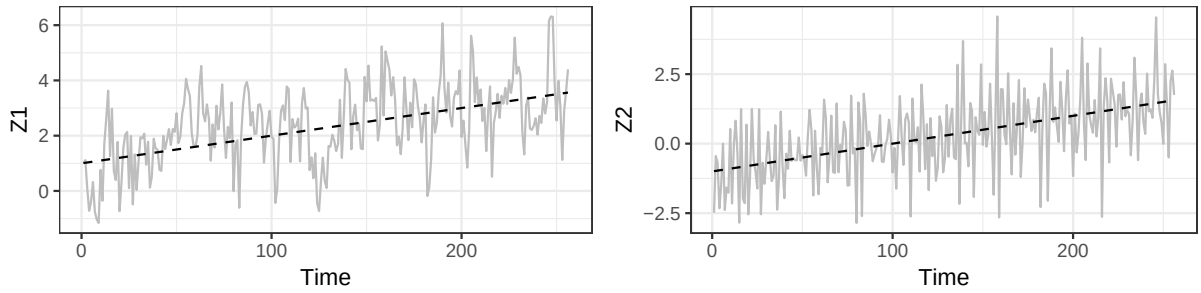
Simulated Data

We consider simulated data drawn from the linear model (8.7) for dimension $d = 2$ with $\mu_1 = 1$ and $\mu_2 = -1$ as well as $b = 0.01$. For the noise time series $\{\underline{e}_t\}$, we use the VAR(2) model from (6.9) and the VMA(1) model from (6.10). For each noise model, we generate $N = 200$ independent realizations for each sample size of $n = 256$ and $n = 512$ as well as $n = 1024$. Exemplary realizations are shown in Figure 8.2. For each component of $\underline{B} = (\mu_1, \mu_2, b)$, we consider the posterior median as Bayes estimator and compute 90% credibility regions with the 0.05 and 0.95 quantile of the respective empirical marginal posterior distribution. We then compare the average (among the N replications) Root Mean Squared Error (RMSE) of the Bayes estimators and the empirical coverage of the credibility regions. The median width of the credibility intervals is also considered. Furthermore, we will also consider $\mathbb{L}^1$-error and $\mathbb{L}^2$-error of the posterior median spectral density and the empirical coverage of uniform 90% regions for $\boldsymbol{f}$ as in Section 6.2.

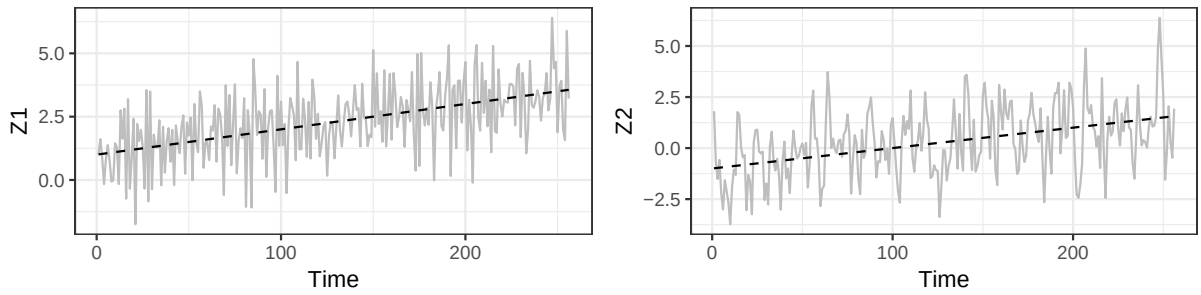
Results

The results can be found in Table 8.1. First of all, we observe that all results of the WN procedure are worse than for the VAR and NP procedure. Whereas this is not too surprising for spectral density inference results (in particular $\mathbb{L}^1$-error and $\mathbb{L}^2$-error), the results for μ_1, μ_2 and b illustrate the point that when the time dependence of the nuisance component is ignored in the model, the inference quality for the parameter of interest will suffer.

It can be seen that the $\mathbb{L}^1$ error and $\mathbb{L}^2$ error of the Bayes estimate of $\boldsymbol{f}$ is smaller for the VAR procedure for VAR(2) data, whereas for VMA(1) data, the NP procedure performs better.



(a)



(b)

Figure 8.2.: Realization of linear trend model (8.7) with (a) VAR(2) noise from (6.9) and (b) VMA(1) noise from (6.10) of respective length $n = 256$. The left panel depicts the first component and the right panel the second component of the time series $\underline{Z}_t$ in gray. The linear trend is visualized as dashed line.

	VAR(2) data								
	$n = 256$			$n = 512$			$n = 1024$		
	WN	NP	VAR	WN	NP	VAR	WN	NP	VAR
$\mathbb{L}^1$-error $\boldsymbol{f}$	0.277	0.104	0.072	0.276	0.082	0.052	0.276	0.063	0.035
$\mathbb{L}^2$-error $\boldsymbol{f}$	0.299	0.132	0.096	0.297	0.107	0.070	0.297	0.085	0.047
Coverage $\boldsymbol{f}$	0.00	0.48	0.91	0.00	0.35	0.88	0.00	0.26	0.88
RMSE μ_1	0.193	0.135	0.131	0.138	0.094	0.092	0.095	0.066	0.066
RMSE μ_2	0.147	0.055	0.049	0.107	0.037	0.035	0.073	0.026	0.025
$n \times$ RMSE b	0.285	0.085	0.069	0.209	0.057	0.049	0.142	0.038	0.035
Coverage μ_1	0.75	0.86	0.91	0.74	0.87	0.90	0.72	0.87	0.90
Coverage μ_2	0.87	0.88	0.92	0.86	0.87	0.91	0.84	0.89	0.89
Coverage b	0.80	0.88	0.91	0.78	0.88	0.89	0.78	0.89	0.88
Length μ_1	0.433	0.406	0.441	0.306	0.288	0.309	0.216	0.205	0.216
Length μ_2	0.438	0.178	0.167	0.309	0.120	0.116	0.218	0.083	0.081
$n \times$ Length b	0.721	0.276	0.232	0.511	0.179	0.163	0.362	0.120	0.114
	VMA(1) data								
	$n = 256$			$n = 512$			$n = 1024$		
	WN	NP	VAR	WN	NP	VAR	WN	NP	VAR
$\mathbb{L}^1$-error $\boldsymbol{f}$	0.311	0.093	0.155	0.310	0.070	0.121	0.309	0.054	0.093
$\mathbb{L}^2$-error $\boldsymbol{f}$	0.323	0.110	0.189	0.322	0.084	0.146	0.321	0.064	0.110
Coverage $\boldsymbol{f}$	0.00	0.62	0.98	0.00	0.45	0.96	0.00	0.27	0.96
RMSE μ_1	0.136	0.047	0.046	0.098	0.032	0.031	0.066	0.023	0.023
RMSE μ_2	0.186	0.132	0.132	0.131	0.092	0.092	0.092	0.067	0.067
$n \times$ RMSE b	0.257	0.040	0.034	0.187	0.023	0.020	0.125	0.016	0.014
Coverage μ_1	0.89	0.88	0.95	0.90	0.86	0.93	0.90	0.85	0.91
Coverage μ_2	0.81	0.81	0.91	0.81	0.83	0.90	0.80	0.84	0.89
Coverage b	0.85	0.96	1.00	0.84	0.95	0.99	0.85	0.87	0.98
Length μ_1	0.439	0.145	0.190	0.310	0.095	0.120	0.219	0.066	0.079
Length μ_2	0.472	0.354	0.480	0.334	0.255	0.322	0.236	0.186	0.221
$n \times$ Length b	0.721	0.170	0.205	0.511	0.089	0.113	0.361	0.052	0.067

Table 8.1.: Average $\mathbb{L}^1$-error, average $\mathbb{L}^2$-error, RMSE, empirical coverage and lengths of uniform 90% credibility regions of the NP procedure and the VAR procedure for 200 realizations of the linear trend model (8.7) with noise from (6.9) and (6.10).

Furthermore, the NP procedure produces less coverage for $\boldsymbol{f}$, whereas the VAR procedure yields approximately honest regions in the well specified case and larger regions for $\boldsymbol{f}$ than NP in both models. These observations are in line with the results from Section 6.2 (see Table 6.2).

The RMSEs of the parameters are very similar for both the NP and VAR procedure, whereas a tendency of the NP procedure to yield slightly more biased estimates (in terms of higher RMSE values) can be conceived, especially for the slope b. This bias effect is diminishing with increasing sample size, indicating that it may have been introduced by the Whittle Likelihood approximation, c.f. the discussion at the end of Section 4.1.

As for the coverage for the linear model coefficients, it can be observed that for the VAR(2) data, both NP and VAR procedures perform well and yield credibility regions that are close to being honest (i.e. with an empirical coverage close to 90%), where the length of the credibility intervals of NP are smaller than of VAR for μ_1 and larger for μ_2 and b. For the VMA(1) data on the other hand, the VAR procedure yields much wider intervals for all components of $\underline{B}$ and produces almost unit coverage for b, whereas the NP procedure produces coverage slightly below honesty for μ_1, μ_2 and slightly above honesty for b.

Discussion

If the main interest lies in inference about the linear model coefficients $\underline{B}$ (and the marginal posterior for $\boldsymbol{f}$ is not of interest), then there is no "clear winner" based on this simulation study. As for the joint posterior of $(\underline{B}, \boldsymbol{f})$, the NP procedure is more robust against misspecification as the parametric VAR method, as seen in terms of the $\mathbb{L}^1$ errors for $\boldsymbol{f}$ and the intervals lengths for b in Table 8.1. We found that the NP procedure yields slightly more biased Bayes estimates of $\underline{B}$ (particularly for small and moderate sample sizes) and whereas it is conjectured that this effect is introduced by the Whittle Likelihood approximation, it will be of interest for future research to investigate this effect in more detail. A further direction for future research constitutes the empirical comparison of the procedures for non-Gaussian data.

9.

Asymptotic properties

In this Section, we will have a closer look from an asymptotic perspective at the mean model

$$\underline{Z}_t = \underline{\mu} + \underline{e}_t, \quad t = 1, \ldots, n \tag{9.1}$$

where the mean vector $\underline{\mu} \in \mathbb{R}^d$ is the parameter of interest and $\underline{e}_t$ is a stationary Gaussian time series in $\mathbb{R}^d$ with $\mathrm{E}\underline{e}_t = 0$ and spectral density $\boldsymbol{f}$, constituting the noise in this model. Recall from (8.3) that Whittle's Likelihood for model (9.1) is given in terms of the Lebesgue density

$$p_W^n(\tilde{\underline{z}}|\mu, \boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^{nd}|\boldsymbol{D}_{nd}|}} \exp\left\{-\frac{1}{2}(\tilde{\underline{z}} - \tilde{\underline{\mu}})^T \boldsymbol{D}_{nd}^{-1}(\tilde{\underline{z}} - \tilde{\underline{\mu}})\right\} \tag{9.2}$$

for $\tilde{\underline{z}} \in \mathbb{R}^{nd}$ with the frequency domain mean vector $\tilde{\underline{\mu}} = (\sqrt{n}\underline{\mu}^T, 0, \ldots, 0)^T \in \mathbb{R}^{nd}$. We will also consider the complex-valued frequency domain formulation of Whittle's Likelihood, which (similar to (4.8)) is defined in terms of the Fourier coefficients $\tilde{\underline{Z}}_0, \ldots, \tilde{\underline{Z}}_{\lfloor n/2 \rfloor}$ from (4.5) as

$$p_W^n(\tilde{\underline{z}}_0, \ldots, \tilde{\underline{z}}_{\lfloor n/2 \rfloor}|\underline{\mu}, \boldsymbol{f}) = p_{0,n}(\tilde{\underline{z}}_0|\underline{\mu}, \boldsymbol{f}) \prod_{j=1}^{\lfloor n/2 \rfloor} p_{j,n}(\tilde{\underline{z}}_j|\boldsymbol{f}) \tag{9.3}$$

with

$$p_{0,n}(\tilde{\underline{z}}_0|\underline{\mu}, \boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^d|2\pi\boldsymbol{f}(0)|}} \exp\left(-\frac{1}{4\pi}\left(\tilde{\underline{z}}_0 - \sqrt{n}\underline{\mu}\right)^T \boldsymbol{f}(0)^{-1}\left(\tilde{\underline{z}}_0 - \sqrt{n}\underline{\mu}\right)\right)$$

and, for n even,

$$p_{n/2,n}(\tilde{\underline{z}}_{n/2}|\boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^d|2\pi\boldsymbol{f}(\pi)|}} \exp\left(-\frac{1}{4\pi}\tilde{\underline{z}}_{n/2}^T \boldsymbol{f}(\pi)^{-1}\tilde{\underline{z}}_{n/2}\right),$$

and $p_{j,n}$ for $j = 1, \ldots, \lceil n/2 \rceil - 1$ as in (7.12). We will consider a prior of form $P(d\mu, d\boldsymbol{f}) = P(d\mu)P(d\boldsymbol{f})$, i.e. μ and $\boldsymbol{f}$ are a priori independent. The question arises which conditions on $P(d\underline{\mu})$ and $P(d\boldsymbol{f})$ are sufficient to ensure that the joint posterior of $(\underline{\mu}, \boldsymbol{f})$ behaves "well" in an asymptotic sense. Furthermore, if the primary interest lies in inference about $\underline{\mu}$, the question is how the marginal posterior of $\underline{\mu}$ behaves asymptotically. As an example, in the parametric iid setting it is well known that (under regularity assumptions) the posterior for $\underline{\mu}$ contracts at rate $n^{-1/2}$ around the true value and that the asymptotic limiting distribution of the posterior

is Gaussian, a result that actually holds true in much wider generality and is well-known as *Bernstein-von-Mises* phenomenon (see Section 10.2 in van der Vaart (2000) for a formulation in the parametric iid setting or Theorem 2.1 in Kleijn and Van der Vaart (2012) for a more general formulation that includes the non-iid case and a possible model misspecification). Thus, we also wish to investigate: Is it possible in the model (9.1) to get the same rate of contraction for $\underline{\mu}$ as if $\boldsymbol{f}$ was actually known (i.e. $n^{-1/2}$)? And does the marginal posterior of $\underline{\mu}$ converge to a known distribution (e.g. Gaussian), i.e. can we establish a Bernstein-von-Mises type result?

We will investigate these questions in the following sections. In Section 9.1, we establish contraction rates for the joint posterior distribution of $(\underline{\mu}, \boldsymbol{f})$. It will be shown there that the same rate of contraction in the Hellinger topology can be achieved as for $\boldsymbol{f}$ in Section 7.2.2 – under the same prior assumptions for $\boldsymbol{f}$ and under prior positivity and regularity assumptions on $\underline{\mu}$. This has the clear interpretation that the "bottleneck" for the rate of contraction is the nonparametric part $\boldsymbol{f}$. Indeed, we will see in Section 7.2.2 that the posterior contraction is highly anisotropic: The marginal posterior of $\underline{\mu}$ is shown to contract at rate $n^{-1/2}$. It is rather surprising that the conditions on $P(d\boldsymbol{f})$ to achieve this are even much weaker than the ones employed for the joint contraction rate theorem. The question of whether a Bernstein-von-Mises theorem can be established remains open. It is indicated by considerations in the final Section 9.3 that additional assumptions have to be employed to achieve this goal. The reason is that for marginal asymptotic normality of $\underline{\mu}$, it is needed that the marginal posterior of $\boldsymbol{f}(0)$ is consistent. However, consistency in $\mathbb{L}^1$ as shown in Section 7.1.2 is not strong enough for this purpose.

9.1. Joint Posterior Contraction Rates

We will work with the following prior assumption for $\underline{\mu}$:

Assumption μ1. *The prior $P(d\underline{\mu})$ on $\underline{\mu}$ is such that all moments exist. Furthermore, it is assumed that in a neighborhood of $\underline{\mu}_0$, it possesses a Lebesgue density that is continuous and bounded away from 0.*

We can now formulate the main result of this Section, stating that the same contraction rate can be achieved for the joint posterior of $(\underline{\mu}, \boldsymbol{f})$ as in Section 7.2.2.

Theorem 9.1. *Let $\{\underline{Z}_t\}$ be a Gaussian stationary time series in $\mathbb{R}^d$ with mean $\underline{\mu}_0$ and spectral density matrix $\boldsymbol{f}_0$ fulfilling Assumptions **f**1-**f**2 with $1 < a \leq 2$. Let $\tau_0 \in (0, b_0)$ and $\tau_1 \in (b_1, \infty)$ (with b_0 and b_1 from Assumption **f**1). Let the prior on $(\underline{\mu}, \boldsymbol{f})$ be given as $P_{\tau_0,\tau_1}(d\underline{\mu}, d\boldsymbol{f}) := P(d\underline{\mu})P_{\tau_0,\tau_1}(d\boldsymbol{f})$, with $P(d\underline{\mu})$ fulfilling Assumption μ1 and the truncated Bernstein-Hpd-Gamma prior P_{τ_0,τ_1} from (7.33) fulfilling Assumptions GP4 and k2. Then*

$$\varepsilon_n = n^{-a/(2+2a)}(\log n)^{(1+2a)/(2+2a)}$$

is a posterior contraction rate for $(\underline{\mu}, \boldsymbol{f})$ with respect to the root average squared Hellinger distance $d_{n,H}$, i.e. for every positive sequence (M_n) with $M_n \to \infty$ for $n \to \infty$ it holds

$$P^n_{W;\tau_0,\tau_1}\Big(\big\{(\mu, \boldsymbol{f})\colon d_{n,H}\big((\underline{\mu}, \boldsymbol{f}), (\underline{\mu}_0, \boldsymbol{f}_0)\big) \geq M_n\varepsilon_n\big\}|Z_1, \ldots, Z_n\Big) \to 0, \quad \text{in } P^n_{(\underline{\mu}_0,\boldsymbol{f}_0)} \text{ probability}$$

as $n \to \infty$*, where* $P^n_{W;\tau_0,\tau_1}$ *denotes the pseudo posterior distribution obtained by updating the prior with Whittle's likelihood* P^n_W *from* (9.2) *and* $P^n_{(\underline{\mu}_0,\boldsymbol{f}_0)}$ *denotes the joint distribution of* $\underline{Z}_1,\dots,\underline{Z}_n$ *under* $(\underline{\mu}_0,\boldsymbol{f}_0)$*.*

In the setting of Theorem 9.1, the average squared Hellinger distance is given by, with $N := \lfloor n/2 \rfloor$,

$$d^2_{n,H}((\mu,\boldsymbol{f}),(\mu_0,\boldsymbol{f})) = \frac{1}{N+1}\left(d^2_H\left(p_{0,n}(\cdot|\underline{\mu},\boldsymbol{f}),p_{0,n}(\cdot|\underline{\mu}_0,\boldsymbol{f}_0)\right) + \sum_{j=1}^{N} d^2_H\left(p_{j,n}(\cdot|\boldsymbol{f}),p_{j,n}(\cdot|\boldsymbol{f}_0)\right)\right) \tag{9.4}$$

with $p_{0,n},\dots,p_{N,n}$ from (9.3). To show Theorem 9.1, we will proceed similar to Section 7.2.2 by showing that the assumptions of the general contraction rates Theorem 7.17 are fulfilled. The proof will rely on the contiguity result from Theorem 4.4. We will start with the prior mass of neighborhoods in Section 9.1.1, continue with the sieve entropy in Section 9.1.2 and collect the result for the proof of Theorem 9.1 in Section 9.1.3.

9.1.1. Prior Mass of Neighborhoods

The following result bounds the KL terms between multivariate normals from above in terms of distances of the respective matrices and mean vectors. It is similar in spirit to Lemma 7.5 and will be important for translating the KL prior mass conditions into prior mass conditions for μ and $\boldsymbol{f}$.

Lemma 9.2. *Let* $\underline{\mu}_0,\underline{\mu}_1 \in \mathbb{R}^d$ *and* $\boldsymbol{\Sigma}_0,\boldsymbol{\Sigma}_1 \in \mathcal{S}^+_d(\mathbb{R})$*, where* $\mathcal{S}^+_d(\mathbb{R})$ *denotes the cone of symmetric positive definite matrices in* $\mathbb{R}^{d\times d}$*. with* $\lambda_{\min}(\boldsymbol{\Sigma}_i) \geq \tau_0$ *and* $\lambda_{\max}(\boldsymbol{\Sigma}_i) \leq \tau_1$ *for* $i = 0,1$ *and some positive constants* τ_0,τ_1*. Assume that* $\lambda_{\min}(\boldsymbol{\Sigma}_1^{-1/2}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1/2}) \geq \frac{1}{2}$*. Let* p_i *denote the density of the* $N_d(\underline{\mu}_i,\boldsymbol{\Sigma}_i)$ *distribution for* $i = 0,1$*. Denote by*

$$K := \int_{\mathbb{R}^d} \log\frac{p_0(\underline{z})}{p_1(\underline{z})} p_0(\underline{z})d\underline{z}, \quad V := \int_{\mathbb{R}^d}\left(\log\frac{p_0(\underline{z})}{p_1(\underline{z})} - K\right)^2 p_0(\underline{z})d\underline{z}$$

the Kullback-Leibler divergence and associated variance term from p_0 *to* p_1*. Then it holds*

$$K \lesssim \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|^2 + \|\underline{\mu}_0 - \underline{\mu}_1\|^2 \tag{9.5}$$

and

$$V \lesssim \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\|^2 + \|\underline{\mu}_0 - \underline{\mu}_1\| + \|\underline{\mu}_0 - \underline{\mu}_1\|^2 \tag{9.6}$$

with proportionality constants only depending on τ_0,τ_1 *and* d*.*

Proof. We start by computing for $\underline{z} \in \mathbb{R}^d$

$$2\log\frac{p_0(\underline{z})}{p_1(\underline{z})} = \log|\boldsymbol{\Sigma}_1| - \log|\boldsymbol{\Sigma}_0| + (\underline{z}-\underline{\mu}_1)^T\boldsymbol{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_1) - (\underline{z}-\underline{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\underline{z}-\underline{\mu}_0).$$

The third summand can be decomposed as

$$\begin{aligned}(\underline{z}-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_1) &= (\underline{z}-\underline{\mu}_0+\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_0+\underline{\mu}_0-\underline{\mu}_1)\\ &= (\underline{z}-\underline{\mu}_0)^T\mathbf{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_0)+2(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_0)\\ &\quad+(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{\mu}_0-\underline{\mu}_1),\end{aligned}$$

where the symmetry of $\mathbf{\Sigma}_1^{-1}$ was used in the last step. This yields

$$\begin{aligned}2\log\frac{p_0(z)}{p_1(z)} &= \log|\mathbf{\Sigma}_1|-\log|\mathbf{\Sigma}_0|+(\underline{z}-\underline{\mu}_0)^T\left(\mathbf{\Sigma}_1^{-1}-\mathbf{\Sigma}_0^{-1}\right)(\underline{z}-\underline{\mu}_0)\\ &\quad+2(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_0)+(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{\mu}_0-\underline{\mu}_1).\end{aligned} \tag{9.7}$$

Letting $l_1(\underline{z}) := \log|\mathbf{\Sigma}_1|-\log|\mathbf{\Sigma}_0|+(\underline{z}-\underline{\mu}_0)^T\left(\mathbf{\Sigma}_1^{-1}-\mathbf{\Sigma}_0^{-1}\right)(\underline{z}-\underline{\mu}_0)$ we observe that $\frac{1}{2}\int l_1(\underline{z})p_0(\underline{z})dz$ is the Kullback-Leibler divergence between an $N_d(\underline{\mu}_0,\mathbf{\Sigma}_0)$ and an $N_d(\underline{\mu}_0,\mathbf{\Sigma}_1)$ distribution. Since by assumption it holds $\lambda_{\min}(\mathbf{\Sigma}_1^{-1/2}\mathbf{\Sigma}_0\mathbf{\Sigma}_1^{-1/2})\geq\frac{1}{2}$, we find that the assumptions of Lemma 7.5 (b) are fulfilled. While Lemma 7.5 is actually formulated for the complex multivariate normal CN_d distribution, the results and the proof are valid completely analogous for the real multivariate normal N_d distribution. Thus we can conclude

$$\int l_1(\underline{z})p_0(\underline{z})dz \lesssim \frac{1}{\lambda_{\min}(\mathbf{\Sigma}_1)^2}\|\mathbf{\Sigma}_0-\mathbf{\Sigma}_1\|^2 \lesssim \|\mathbf{\Sigma}_0-\mathbf{\Sigma}_1\|^2.$$

Letting $l_2(\underline{z}) := 2(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{z}-\underline{\mu}_0)$ it follows that $\int l_2(\underline{z})p_0(\underline{z})d\underline{z}=0$, since μ_0 is the mean under p_0. Since it also holds

$$(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}(\underline{\mu}_0-\underline{\mu}_1)\leq\frac{1}{\tau_0}\|\underline{\mu}_0-\underline{\mu}_1\|^2\lesssim\|\underline{\mu}_0-\underline{\mu}_1\|^2$$

by the Minmax theorem of Courant-Fisher, assertion (9.5) follows.

To show (9.6), let $\underline{Z}\sim N_d(\underline{0},\mathbf{\Sigma}_0)$. Then $V=\mathrm{Var}[\log\frac{p_0(\underline{Z}+\mu_0)}{p_1(\underline{Z}+\mu_0)}]$ and from (9.7) this yields

$$4V=\mathrm{Var}\left[\underline{Z}^T\left(\mathbf{\Sigma}_1^{-1}-\mathbf{\Sigma}_0^{-1}\right)\underline{Z}+2(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}\underline{Z}\right]=:\mathrm{Var}\left[\eta_1+\eta_2\right],$$

which can also be written as

$$\mathrm{Var}[\eta_1+\eta_2]=\mathrm{Var}[\eta_1]+\mathrm{Var}[\eta_2]+2\mathrm{Cov}(\eta_1,\eta_2).$$

Note that $\mathrm{Var}[\eta_1]$ is the KL variance term between an $N_d(\underline{0},\mathbf{\Sigma}_0)$ and an $N_d(\underline{0},\mathbf{\Sigma}_1)$ distribution. From part (a) of Lemma 7.5 (recalling that the result there is formulated for the CN_d distribution but valid and proven analogously for the N_d distribution), we get

$$\mathrm{Var}[\eta_1]\lesssim\frac{1}{\lambda_{\min}(\mathbf{\Sigma}_1)^2}\|\mathbf{\Sigma}_0-\mathbf{\Sigma}_1\|^2\lesssim\|\mathbf{\Sigma}_0-\mathbf{\Sigma}_1\|^2.$$

As $\mathrm{E}\underline{Z}=0$, it follows $\mathrm{E}[(\underline{\mu}_0-\underline{\mu}_1)^T\mathbf{\Sigma}_1^{-1}\underline{Z}]=0$ and an application of Lemma B.24 in the Appendix

yields

$$\begin{aligned}\frac{1}{4}\mathrm{Var}[\eta_2] = \mathrm{Var}\left[(\underline{\mu}_0 - \underline{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} \underline{Z}\right] &= \mathrm{E}\left[(\underline{\mu}_0 - \underline{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} \underline{Z}\underline{Z}^T \boldsymbol{\Sigma}_1^{-1} (\underline{\mu}_0 - \underline{\mu}_1)\right] \\ &= \mathrm{E}\left[\underline{Z}^T \boldsymbol{\Sigma}_1^{-1} (\underline{\mu}_0 - \underline{\mu}_1)(\underline{\mu}_0 - \underline{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} \underline{Z}\right] \\ &= \mathrm{tr}\left(\boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1} (\underline{\mu}_0 - \underline{\mu}_1)(\underline{\mu}_0 - \underline{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1}\right) \\ &= \mathrm{tr}\left((\underline{\mu}_0 - \underline{\mu}_1)^T \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1} (\underline{\mu}_0 - \underline{\mu}_1)\right) \\ &\leq \lambda_{\max}\left(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{-1}\right) \|\underline{\mu}_0 - \underline{\mu}_1\|^2 \\ &\lesssim \|\underline{\mu}_0 - \underline{\mu}_1\|^2,\end{aligned}$$

where the cyclic property of the trace from Lemma B.2 in the Appendix was used. For the covariance term, we use the Cauchy-Schwartz inequality to obtain

$$|\mathrm{Cov}(\eta_1, \eta_2)| \leq \big(\mathrm{Var}[\eta_1]\mathrm{Var}[\eta_0]\big)^{1/2} \lesssim \|\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1\| \|\underline{\mu}_0 - \underline{\mu}_1\| \lesssim \|\underline{\mu}_0 - \underline{\mu}_1\|,$$

yielding (9.6). □

Recall the definition of the Bernstein polynomial expansion operator $\mathfrak{B}$ from (7.4) and of the two-sided prior truncation set $\mathcal{C}_{\tau_0,\tau_1}$ from (7.32). We will consider the following sieve structure Θ_n on the parameter space $\Theta = \mathbb{R}^d \times \mathcal{D}_{d\times d}$:

$$\begin{aligned}\Theta_n &:= \Theta_n^{(1)} \times \Theta_n^{(2)}, \\ \Theta_n^{(1)} &:= \left\{\underline{\mu} \in \mathbb{R}^d \colon \|\underline{\mu}\| < M_n\right\}, \quad M_n := \exp(\rho_1 n \varepsilon_n^2), \\ \Theta_n^{(2)} &:= \bigcup_{k=1}^{k_n} \left\{\mathfrak{B}(k, \underline{\boldsymbol{W}}) \colon \underline{\boldsymbol{W}} \in \mathcal{S}_d^{+k}\right\} \cap \mathcal{C}_{\tau_0,\tau_1}, \quad k_n := \rho_2 \varepsilon_n^{-2/a},\end{aligned} \tag{9.8}$$

with positive constants ρ_1 and ρ_2 that are to be specified later. In fact, ρ_2 will be chosen sufficiently large and ρ_1 sufficiently small in later considerations. The next result quantifies the prior mass of shrinking Kullback-Leibler neighborhoods. It is similar in spirit to Lemma 7.23.

Lemma 9.3. *Let the assumptions of Theorem 9.1 be fulfilled. Let $N := \lfloor n/2 \rfloor$. Consider the KL divergence*

$$K_n\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) := \frac{1}{N+1}\left(K_{0,n}\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) + \sum_{j=1}^{N} K_{j,n}(\boldsymbol{f}_0, \boldsymbol{f})\right)$$

with

$$\begin{aligned}K_{0,n}\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) &= \mathrm{E}_{(\underline{\mu}_0, \boldsymbol{f}_0)} \log \frac{p_{0,n}(\tilde{\underline{Z}}_0 | \underline{\mu}_0, \boldsymbol{f}_0)}{p_{0,n}(\tilde{\underline{Z}}_0 | \underline{\mu}, \boldsymbol{f})}, \\ K_{j,n}(\boldsymbol{f}_0, \boldsymbol{f})) &= \mathrm{E}_{(\underline{\mu}_0, \boldsymbol{f}_0)} \log \frac{p_{j,n}(\tilde{\underline{Z}}_j | \boldsymbol{f}_0)}{p_{j,n}(\tilde{\underline{Z}}_j | \boldsymbol{f})}, \quad j = 1, \ldots, N,\end{aligned}$$

with $p_{0,n}, \ldots, p_{N,n}$ as in (9.3), and the corresponding KL variance term

$$V_n\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) := \frac{1}{N+1}\left(V_{0,n}\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) + \sum_{j=1}^{N} V_{j,n}(\boldsymbol{f}_0, \boldsymbol{f})\right).$$

For $\varepsilon > 0$, let

$$\begin{aligned} &B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon) \\ &:= \Big\{(\underline{\mu}, \boldsymbol{f}) \in \Theta_n \colon K_n\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) < \varepsilon^2, V_n\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) < \varepsilon^2\Big\} \end{aligned} \tag{9.9}$$

with the sieve Θ_n from (9.8). Then there exists a positive constant c such that

$$P_{\tau_0,\tau_1}\big(B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n)\big) \geq \exp(-cn\varepsilon_n^2)$$

holds for all $n \in \mathbb{N}$ large enough, with $\varepsilon_n = n^{-a/(2+2a)}(\log n)^{(1+2a)/(2+2a)}$.

Proof. Let $(\underline{\mu}, \boldsymbol{f}) \in B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n)$. First recall from (7.17) and (7.18) that for ε_n small enough, it holds

$$\min_{0\leq\omega\leq\pi} \lambda_{\min}(\boldsymbol{f}(\omega)) \geq \frac{b_0}{2}, \quad \min_{0\leq\omega\leq\pi} \lambda_{\min}\left(\boldsymbol{f}(\omega)^{-1/2}\boldsymbol{f}_0(\omega)\boldsymbol{f}(\omega)^{-1/2}\right) \geq \frac{1}{2},$$

with b_0 from Assumption $\boldsymbol{f}1$. Thus an application of Lemma 7.5 yields

$$K_{j,n}(\boldsymbol{f}_0, \boldsymbol{f}) \lesssim \|\boldsymbol{f}(\omega_j) - \boldsymbol{f}_0(\omega_j)\|^2, \quad V_{j,n}(\boldsymbol{f}_0, \boldsymbol{f}) \lesssim \|\boldsymbol{f}(\omega_j) - \boldsymbol{f}_0(\omega_j)\|^2$$

for $j = 1, \dots, N$. Similarly, an application of Lemma 9.2 to $K_{0,n}$ (recalling that $p_{0,n}(\cdot|\underline{\mu}, \boldsymbol{f})$ is the density of the $N_d(\sqrt{n}\underline{\mu}, 2\pi\boldsymbol{f}(0))$ distribution) yields

$$K_{0,n}\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) \lesssim n\|\underline{\mu} - \underline{\mu}_0\|^2 + \|\boldsymbol{f}(0) - \boldsymbol{f}_0(0)\|^2,$$

and

$$\begin{aligned} V_{j,n}\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) &\lesssim n^{1/2}\|\underline{\mu} - \underline{\mu}_0\| + n\|\underline{\mu} - \underline{\mu}_0\|^2 + \|\boldsymbol{f}(0) - \boldsymbol{f}_0(0)\|^2 \\ &\lesssim n\left(\|\underline{\mu} - \underline{\mu}_0\| + \|\underline{\mu} - \underline{\mu}_0\|^2\right) + \|\boldsymbol{f}(0) - \boldsymbol{f}_0(0)\|^2. \end{aligned}$$

This yields

$$\begin{aligned} K_n\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) &\lesssim \frac{1}{N+1}\left(n\|\underline{\mu} - \underline{\mu}_0\|^2 + \sum_{j=0}^{N} \|\boldsymbol{f}(\omega_j) - \boldsymbol{f}_0(\omega_j)\|^2\right) \\ &\lesssim \|\underline{\mu} - \underline{\mu}_0\|^2 + \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty}^2 \end{aligned}$$

with the maximum Frobenius norm $\|\cdot\|_{F,\infty}$ from (7.3) and similarly

$$V_n\big((\underline{\mu}_0, \boldsymbol{f}_0), (\underline{\mu}, \boldsymbol{f})\big) \lesssim \|\underline{\mu} - \underline{\mu}_0\| + \|\underline{\mu} - \underline{\mu}_0\|^2 + \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty}^2.$$

Thus, there exists a positive constant c such that $B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0) \times B_{c\varepsilon_n}(\boldsymbol{f}_0) \subset B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n)$ with $\tilde{\varepsilon} = \min\{\varepsilon_n, \varepsilon_n^2\}$ and $B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0)$ denoting the ball (in $\mathbb{R}^d$) of radius $c\tilde{\varepsilon}_n$ and $B_{c\varepsilon_n}(\boldsymbol{f}_0) := \{\boldsymbol{f} \in \mathcal{D}_{d\times d} \colon \|\boldsymbol{f} - \boldsymbol{f}_0\|_{F,\infty} < c\varepsilon_n\}$ the ball in $\mathcal{D}_{d\times d}$ with respect to $\|\cdot\|_{F,\infty}$.

We continue by bounding the prior mass of $B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0) \times B_{c\varepsilon_n}(\boldsymbol{f}_0)$ from below. First we recall from Assumption $\mu 1$ that there exists a positive constant C such that the Lebesgue density of $P(d\mu)$ fulfills $p(\underline{\mu}) \geq C$ for all $\underline{\mu} \in B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0)$ for n large enough. Since $\tilde{\varepsilon}_n = \varepsilon_n^2$ for n large enough, this yields (with $\mathcal{L}$ denoting the Lebesgue measure in $\mathbb{R}^d$)

$$P(B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0)) \geq C\mathcal{L}(B_{c\varepsilon_n^2}(\underline{\mu}_0)) \geq \tilde{C}\varepsilon_n^{2d} = \exp(\log\tilde{C} + 2d\log\varepsilon_n) \geq \exp(-c_1 n\varepsilon_n^2)$$

for a positive constants $\tilde{C}$, where in the last step it was used that ε_n fulfills

$$\log \tilde{C} + d \log \varepsilon_n \geq \varepsilon_n^{-2/a} \log \varepsilon_n \geq -c_1 n \varepsilon_n^2$$

for some positive constant c_1 for n large enough by (7.41). From (7.42), it follows that the Bernstein-Hpd-Gamma prior fulfills $P_{\tau_0,\tau_1}(B_{c\varepsilon_n}(\boldsymbol{f}_0)) \geq \exp(-c_2 n \varepsilon_n^2)$ for a positive constant c_2, yielding that the joint prior for $(\mu, \boldsymbol{f})$ fulfills

$$P_{\tau_0,\tau_1}\Big(B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0) \times B_{c\varepsilon_n}(\boldsymbol{f}_0)\Big) = P\big(B_{c\tilde{\varepsilon}_n}(\underline{\mu}_0)\big) P_{\tau_0,\tau_1}\big(B_{c\varepsilon_n}(\boldsymbol{f}_0)\big) \geq \exp\big(-(c_1+c_2) n \varepsilon_n^2\big),$$

which concludes the proof. □

9.1.2. Sieve Entropy and Complement Mass

The following result bounds the Hellinger distance between two multivariate normals from above in terms of distances of the respective matrices and mean vectors. It is similar in spirit to Lemma 7.24 and will be important for quantifying the sieve entropy.

Lemma 9.4. *Let $\underline{\mu}_1, \underline{\mu}_2 \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathcal{S}_d^+(\mathbb{R})$ with $\lambda_{\min}(\boldsymbol{\Sigma}_i) \geq \tau_0$ and $\lambda_{\max}(\boldsymbol{\Sigma}_i) \leq \tau_1$ for $i = 1, 2$ and some positive constants τ_0, τ_1. Let p_i denote the density of the $N_d(\underline{\mu}_i, \boldsymbol{\Sigma}_i)$ distribution for $i = 1, 2$. Then it holds*

$$d_H^2(p_1, p_2) \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\| + \|\underline{\mu}_1 - \underline{\mu}_2\|$$

with proportionality constants depending only on τ_0, τ_1 and d.

Proof. We start by noting the representation (see Pardo (2005), p.51)

$$\begin{aligned} d_H^2(p_1, p_2) &= 1 - \frac{2^{d/2}}{|\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2|^{1/4} |\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}|^{1/2}} \exp\left(-\frac{1}{4}(\underline{\mu}_1 - \underline{\mu}_2)^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\underline{\mu}_1 - \underline{\mu}_2)\right) \\ &= 1 - \frac{2^{d/2}}{|\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2|^{1/4} |\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}|^{1/2}} + \frac{2^{d/2}}{|\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2|^{1/4} |\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}|^{1/2}} \times \\ &\quad \left(1 - \exp\left(-\frac{1}{4}(\underline{\mu}_1 - \underline{\mu}_2)^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1} (\underline{\mu}_1 - \underline{\mu}_2)\right)\right) \end{aligned}$$

By analogous arguments as in the proof of Lemma 7.24, it holds

$$0 \leq 1 - \frac{2^{d/2}}{|\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2|^{1/4} |\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}|^{1/2}} \lesssim \|\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2\|.$$

Using the boundedness assumption of the eigenvalues, we obtain from Lemma B.5 in the Appendix

$$|\boldsymbol{\Sigma}_i| \geq \tau_0^d, \quad |\boldsymbol{\Sigma}_i^{-1}| \geq \tau_1^{-d}, \quad |\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}| \geq (2\tau_1)^{-d},$$

hence

$$\begin{aligned}
&\frac{2^{d/2}}{|\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2|^{1/4}|\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1}|^{1/2}}\left(1-\exp\left(-\frac{1}{4}(\underline{\mu}_1-\underline{\mu}_2)^T\left(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right)^{-1}(\underline{\mu}_1-\underline{\mu}_2)\right)\right)\\
&\lesssim 1-\exp\left(-\frac{1}{4}(\underline{\mu}_1-\underline{\mu}_2)^T\left(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right)^{-1}(\underline{\mu}_1-\underline{\mu}_2)\right)\\
&\leq 1-\exp\left(-\frac{1}{4}\lambda_{\max}\left(\left(\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2\right)^{-1}\right)\|\underline{\mu}_1-\underline{\mu}_2\|^2\right)\\
&= 1-\exp(-c\|\underline{\mu}_1-\underline{\mu}_2\|^2),
\end{aligned}$$

where the Min-Max theorem of Courant-Fisher (see Lemma B.3 in the Appendix) was employed. The function $h_c\colon [0,\infty)\to[0,\infty), h_c(x) := 1-\exp(-cx^2)$ has a point of inflection at $x^* = (2c)^{-1/2}$, where the maximal derivative is attained, with value $h_c'(x^*) = \sqrt{2c}\exp(-1/2) > 0$. Thus we compute

$$1-\exp(-c\|\underline{\mu}_1-\underline{\mu}_2\|^2) \leq h_c'(x^*)\|\underline{\mu}_1-\underline{\mu}_2\| \lesssim \|\underline{\mu}_1-\underline{\mu}_2\|,$$

which concludes the proof. □

With the previous result, we can derive an upper bound for the metric entropy of the sieve Θ_n.

Lemma 9.5. *Let the assumptions of Theorem 9.1 be fulfilled. Then for the sieve Θ_n from (9.8), the ε-covering number in the root average squared Hellinger topology fulfills*

$$\log \sup_{\varepsilon>\varepsilon_n} N\left(\xi\varepsilon, \{(\underline{\mu},\boldsymbol{f})\in\Theta_n\colon d_{n,H}^2\big((\underline{\mu}_0,\boldsymbol{f}_0),(\underline{\mu},\boldsymbol{f})\big)<2\varepsilon\}, d_{n,H}\right) \leq n\varepsilon_n^2.$$

Proof. Recall the representation of $d_{n,H}^2$ from (9.4). Let $N = \lfloor n/2\rfloor$. From Lemma 9.4, we obtain

$$d_H^2\Big(p_{j,n}(\cdot|\boldsymbol{f}_0), p_{j,n}(\cdot|\boldsymbol{f})\Big) \lesssim \|\boldsymbol{f}_0(\omega_j)-\boldsymbol{f}(\omega_j)\|, \quad j=1,\dots,N,$$

and from Lemma 9.4, recalling that $p_{0,n}(\cdot|\underline{\mu},\boldsymbol{f})$ is the density of the $N_d(\sqrt{n}\underline{\mu}, 2\pi\boldsymbol{f}(0))$ distribution, we obtain

$$d_H^2\Big(p_{0,n}(\cdot|\underline{\mu}_0,\boldsymbol{f}_0), p_{0,n}(\cdot|\underline{\mu},\boldsymbol{f})\Big) \lesssim n^{1/2}\|\underline{\mu}_0-\underline{\mu}\| + \|\boldsymbol{f}_0(0)-\boldsymbol{f}(0)\|$$

yielding

$$d_{n,H}^2\big((\underline{\mu}_0,\boldsymbol{f}_0),(\underline{\mu},\boldsymbol{f})\big) \lesssim n^{-1/2}\|\underline{\mu}_0-\underline{\mu}\| + \|\boldsymbol{f}_0-\boldsymbol{f}\|_{F,\infty}.$$

Thus, by Lemma B.32 in the Appendix, there exists a positive constant c such that

$$\log N(\varepsilon,\Theta_n,d_{n,H}) \leq \log N(cn^{1/2}\varepsilon^2,\Theta_n^{(1)},\|\cdot\|) + \log N(c\varepsilon^2,\Theta_n^{(2)},\|\cdot\|_{F,\infty}).$$

In particular, this yields

$$\begin{aligned}
&\log \sup_{\varepsilon>\varepsilon_n} N\left(\xi\varepsilon, \{(\underline{\mu},\boldsymbol{f})\in\Theta_n\colon d_{n,H}^2\big((\underline{\mu}_0,\boldsymbol{f}_0),(\underline{\mu},\boldsymbol{f})\big)<2\varepsilon\}, d_{n,H}\right)\\
&\leq \log N(\xi\varepsilon_n,\Theta_n,d_{n,H})\\
&\leq \log N(c\xi^2 n^{1/2}\varepsilon_n^2,\Theta_n^{(1)},\|\cdot\|) + \log N(c\xi^2\varepsilon_n^2,\Theta_n^{(2)},\|\cdot\|_{F,\infty}).
\end{aligned} \tag{9.10}$$

Recalling the definition of $M_n = \exp(\rho_1 n\varepsilon_n^2)$ from (9.8), an application of Lemma B.31 in the Appendix reveals

$$\begin{aligned}\log N(c\xi^2 n^{1/2}\varepsilon_n^2, \Theta_n^{(1)}, \|\cdot\|) &\leq d(\log(3M_n) - \log(c\xi^2 n^{1/2}\varepsilon_n^2)) \\ &= d\log 3 + d\rho_1 n\varepsilon_n^2 - d\log(c\xi^2 n^{1/2}\varepsilon_n^2) \leq 2d\rho_1 n\varepsilon_n^2\end{aligned}$$

for n large enough. Choosing ρ_1 small enough, this yields

$$\log N(c\xi^2 n^{1/2}\varepsilon_n^2, \Theta_n^{(1)}, \|\cdot\|) \leq \frac{1}{2} n\varepsilon_n^2.$$

From (7.46), we obtain

$$\log N(c\xi^2\varepsilon_n^2, \Theta_n^{(2)}, \|\cdot\|_{F,\infty}) \leq \frac{1}{2} n\varepsilon_n^2,$$

and in view of (9.10), this concludes the proof. □

9.1.3. Proof of Joint Contraction Rate Theorem

It remains to bound the prior mass of the sieve complement, which is the following result.

Lemma 9.6. *Let the assumptions of Theorem 9.1 be fulfilled. Then it holds*

$$\frac{P_{\tau_0,\tau_1}(\Theta_n^c)}{P_{\tau_0,\tau_1}(B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n))} = o\left(\exp(-2n\varepsilon_n^2)\right)$$

with Θ_n as in (9.8) and $B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n)$ from (9.9).

Proof. First observe that

$$\Theta_n^c = \left(\Theta_n^{(1)} \times \Theta_n^{(2)}\right)^c = \left(\Theta_n^{(1)c} \times \Theta^{(2)}\right) \cup \left(\Theta^{(1)} \times \Theta_n^{(2)c}\right),$$

with $\Theta_n^{(i)c}$ denoting the complement of $\Theta_n^{(i)}$ in $\Theta^{(i)}$ for $i = 1, 2$, where $\Theta^{(1)} = \mathbb{R}^d$ and $\Theta^{(2)} = \mathcal{D}_{d\times d}$. Since $\underline{\mu}$ and $\boldsymbol{f}$ are a priori independent, this yields

$$P_{\tau_0,\tau_1}(\Theta_n^c) = P(\Theta_n^{(1)c}) + P_{\tau_0,\tau_1}(\Theta_n^{(2)c}).$$

By Assumption $\mu 1$, all prior moments of $\underline{\mu}$ exist. Thus we compute with $M_n = \exp(\rho_1 n\varepsilon_n^2)$ for $\delta > 0$ with Chebyshevs inequality

$$P(\Theta_n^{(1)c}) = P(\|\underline{\mu}\| > M_n) \leq M_n^{-\delta/\rho_1} \int_{\|\underline{\mu}\|>M_n} \|\underline{\mu}\|^{\delta/\rho_1} P(d\underline{\mu}) \lesssim M_n^{-\delta/\rho_1}$$

with proportionality constants depending only on δ, ρ_1 and $P(d\underline{\mu})$. Since $M_n^{-\delta/\rho_1} = \exp(-\delta n\varepsilon_n^2)$, it follows from Lemma 9.3 that

$$\frac{P(\Theta_n^{(1)c})}{P_{\tau_0,\tau_1}(B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n))} \lesssim \exp(-(\delta - c)n\varepsilon_n^2) = o\left(\exp(-2n\varepsilon_n^2)\right)$$

for $\delta > 2 + c$. On the other hand, it also holds

$$\frac{P_{\tau_0,\tau_1}(\Theta_n^{(2)c})}{P_{\tau_0,\tau_1}(B_{n,2}(\underline{\mu}_0, \boldsymbol{f}_0, \varepsilon_n))} = o\left(\exp(-2n\varepsilon_n^2)\right)$$

by (7.47) and Lemma 9.3, hence the result follows. □

Now we are able to present the proof of the main result.

Proof (of Theorem 9.1). We will apply Theorem 7.17. By Lemma 7.18, assumptions (b) and (c) of Theorem 7.17 are fulfilled and by Lemma 9.3, also assumption (a). Assumption (d) is fulfilled by Lemma 9.5 and assumption (e) for the sieve Θ_n from (9.8) by Lemma 9.6. □

9.2. Marginal Contraction at Parametric Rate

The following result gives sufficient conditions for the marginal posterior to contract at parametric rate and motivates the following considerations in this section.

Theorem 9.7. *Let $X_1, \dots, X_n$ be random variables in some Euclidean space with likelihood function $P(dX_1, \dots, dX_n|\theta, \eta)$ where the parameter is partitioned as (θ, η) with $\theta \in \Theta$ and $\eta \in \mathcal{H}$ and $\Theta, \mathcal{H}$ being (jointly) measurable spaces. Let $(\theta_0, \eta_0) \in \Theta \times \mathcal{H}$. Assume that some prior $P(d\theta, d\eta)$ is specified on $\Theta \times \mathcal{H}$. Denote by P_0^n the joint distribution of $X_1, \dots, X_n$ under (θ_0, η_0) and by E_0^n the corresponding expected value operator. Let (ε_n) be a sequence of positive numbers. Assume that there exists a sequence $(\mathcal{H}_n)$ of subsets of $\mathcal{H}$ such that*

(a) the marginal posterior of η concentrates on $\mathcal{H}_n$ asymptotically, i.e.

$$P(\eta \in \mathcal{H}_n|X_1, \dots, X_n) \to 1, \quad \text{in } P_0^n \text{ probability as } n \to \infty,$$

(b) the conditional posterior of θ contracts uniformly at rate ε_n, i.e. for every sequence (M_n) of positive numbers with $M_n \to \infty$ it holds

$$\mathrm{E}_0^n \sup_{\eta \in \mathcal{H}_n} P(\|\theta - \theta_0\| > \varepsilon_n M_n|\eta, X_1, \dots, X_n) \to 0.$$

Then the marginal posterior of θ contracts at rate ε_n, i.e.

$$P(\|\theta - \theta_0\| > \varepsilon_n M_n|X_1, \dots, X_n) \to 0, \quad \text{in } P_0^n \text{ probability as } n \to \infty$$

for every positive sequence (M_n) with $M_n \to \infty$.

Proof. See Theorem 6.2 in Bickel and Kleijn (2012). □

The goal of this Section is to apply Theorem 9.7 to show that the marginal posterior of $\underline{\mu}$ contracts at rate $n^{-1/2}$. This result will be presented in Theorem 9.9. Consider the following space of Hpd functions that are bounded and bounded away from zero:

$$\mathcal{H} := \left\{ \boldsymbol{f} \colon [0, \pi] \to \mathcal{S}_d^+ \colon \sup_{0 \le \omega \le \pi} \|\boldsymbol{f}(\omega)\| < \infty, \sup_{0 \le \omega \le \pi} \|\boldsymbol{f}^{-1}(\omega)\| < \infty \right\}$$

and, for constants $0 < \tau_0, \tau_1 < \infty$, the following subset:

$$\mathcal{H}_{\tau_0, \tau_1} := \{ \boldsymbol{f} \in \mathcal{H} \colon \lambda_{\min}(\boldsymbol{f}(0)) \ge \tau_0, \lambda_{\max}(\boldsymbol{f}(0)) \le \tau_1 \}. \tag{9.11}$$

Observe that in contrast to the prior restriction set from (7.32) as considered previously, the restriction by τ_0 and τ_1 in (9.11) only applies to $\boldsymbol{f}(0)$. We will work with the following assumptions on the prior on $\underline{\mu}$ and $\boldsymbol{f}$:

Assumption μ2. *The prior on $(\underline{\mu}, \boldsymbol{f})$ is of the form $P(d\underline{\mu}, d\boldsymbol{f}) \propto P(d\boldsymbol{f})$, i.e. $\underline{\mu}$ and $\boldsymbol{f}$ are a priori independent and the prior on $\underline{\mu}$ is improper with $p(\underline{\mu}) \propto 1$. The prior on $\boldsymbol{f}$ is assumed to fulfill $P(\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}) = 1$ for some $0 < \tau_0 < \tau_1 < \infty$ with $\mathcal{H}_{\tau_0,\tau_1}$ from (9.11).*

We will use the following notational shorthand for $\underline{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma} \in \mathcal{S}_d^+(\mathbb{R})$:

$$N_d(\underline{\mu}, \boldsymbol{\Sigma})[B] := \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \int_B \exp\left(-\frac{1}{2}(\underline{z}-\underline{\mu})^T \boldsymbol{\Sigma}^{-1}(\underline{z}-\underline{\mu})\right) d\underline{z},$$

for $B \subset \mathbb{R}^d$ measurable. Furthermore, for $r > 0$, denote by $B_r(\underline{\mu}_0) = \{\underline{\mu} \in \mathbb{R}^d \colon \|\underline{\mu} - \underline{\mu}_0\| < r\}$ the Euclidean ball of radius r. We will need the following Lemma.

Lemma 9.8. *Let $r > 0$ and $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 \in \mathcal{S}_d^+(\mathbb{R})$ such that $\lambda_{\min}(\boldsymbol{\Sigma}_2) \geq \lambda_{\max}(\boldsymbol{\Sigma}_1)$. Then*

$$N_d(\underline{0}, \boldsymbol{\Sigma}_1)\,[B_r(\underline{0})] \geq N_d(\underline{0}, \boldsymbol{\Sigma}_2)\,[B_r(\underline{0})]\,.$$

Proof. Denote by $\boldsymbol{\Sigma}_i^{1/2}$ the spd square root of $\boldsymbol{\Sigma}_i$ for $i = 1, 2$. First we compute, using the substitution $\underline{z} = \boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\underline{y}$, $d\underline{z} = |\boldsymbol{\Sigma}_2|^{1/2}|\boldsymbol{\Sigma}_1|^{-1/2}d\underline{y}$:

$$\begin{aligned} N_d(\underline{0}, \boldsymbol{\Sigma}_2)\,[B_r(\underline{0})] &= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_2|^{1/2}} \int_{B_r(\underline{0})} \exp\left(-\frac{1}{2}\underline{z}^T\boldsymbol{\Sigma}_2^{-1}\underline{z}\right) d\underline{z} \\ &= \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_1|^{1/2}} \int_{\tilde{B}_r(\underline{0})} \exp\left(-\frac{1}{2}\underline{y}^T\boldsymbol{\Sigma}_1^{-1}\underline{y}\right) d\underline{y} \\ &= N_d(\underline{0}, \boldsymbol{\Sigma}_1)\left[\tilde{B}_r(\underline{0})\right] \end{aligned}$$

with $\tilde{B}_r(\underline{0}) := \{\underline{y} \in \mathbb{R}^d \colon \|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\underline{y}\|^2 < r^2\}$. It suffices to show $\tilde{B}_r(\underline{0}) \subset B_r(\underline{0})$. To do so, we employ the Min-Max theorem of Courant-Fisher and get

$$\|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\underline{y}\|^2 = \underline{y}^T\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2^{1/2}\underline{y} \geq \lambda_{\min}\left(\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2^{1/2}\right)\|\underline{y}\|^2.$$

Since by Lemma B.3 in the Appendix it holds

$$\lambda_{\min}\left(\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2^{1/2}\right) = \frac{1}{\lambda_{\max}\left(\boldsymbol{\Sigma}_2^{-1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1/2}\right)} \geq \frac{1}{\lambda_{\max}(\boldsymbol{\Sigma}_2^{-1})\lambda_{\max}(\boldsymbol{\Sigma}_1)} = \frac{\lambda_{\min}(\boldsymbol{\Sigma}_2)}{\lambda_{\max}(\boldsymbol{\Sigma}_1)} \geq 1$$

by assumption, we get $\|\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1^{-1/2}\underline{y}\| \geq \|\underline{y}\|$ and thus $\tilde{B}_r(\underline{0}) \subset B_r(\underline{0})$, concluding the proof. □

Now we can formulate the main result of this section.

Theorem 9.9. *Let $\{\underline{Z}_n\}$ be a stationary Gaussian time series in $\mathbb{R}^d$ with mean $\underline{\mu}_0$ and spectral density matrix $\boldsymbol{f}_0$ fulfilling Assumptions f1 and f2. Let the prior fulfill Assumption μ2 with $0 < \tau_0 < \lambda_{\min}(\boldsymbol{f}_0(0))$ and $\lambda_{\max}(\boldsymbol{f}_0(0)) < \tau_1 < \infty$. Denote by $P_W^n(d\underline{\mu}, d\boldsymbol{f}|\underline{Z}_1, \ldots, \underline{Z}_n)$ the joint posterior of $(\underline{\mu}, \boldsymbol{f})$ when updating $P(d\underline{\mu}, d\boldsymbol{f})$ with Whittle's likelihood (9.2). Then the marginal posterior of $\underline{\mu}$ contracts at rate $n^{-1/2}$, i.e.*

$$P_W^n(n^{1/2}\|\underline{\mu} - \underline{\mu}_0\| > M_n|\underline{Z}_1, \ldots, \underline{Z}_n) \to 0, \quad \text{in } P_0^n \text{ probability as } n \to \infty$$

for every sequence (M_n) of positive numbers with $M_n \to \infty$, where P_0^n denotes the joint distribution of $(\underline{Z}_1, \ldots, \underline{Z}_n)$ under $(\underline{\mu}_0, \boldsymbol{f}_0)$.

The Assumption $\boldsymbol{f}2$ on $\boldsymbol{f}_0$ in Theorem 9.9 is only needed to justify the Whittle approximation (see Theorem 4.4). If the full Gaussian likelihood is used instead of Whittle's Likelihood, it can be relaxed. In fact, an inspection of the upcoming proof reveals that it is sufficient to assume that $\boldsymbol{f}_0$ is continuous and $\boldsymbol{f}_0 \in \mathcal{H}_{\tau_0,\tau_1}$ (with $\mathcal{H}_{\tau_0,\tau_1}$ from (9.11)) in this case.

Proof of Theorem 9.9. Due to the contiguity result from Theorem 4.4, we can assume that the joint distribution of $(\underline{Z}_1, \ldots, \underline{Z}_n)$ is given by $P_W^n(\cdot|\underline{\mu}_0, \boldsymbol{f}_0)$. By Assumption $\mu 2$, the prior on $\underline{\mu}$ is flat and hence the conditional posterior of $\underline{\mu}$ is equal to the conditional likelihood. Hence it follows from (8.5) that $P_W^n(d\underline{\mu}|\boldsymbol{f}, \underline{Z}_1, \ldots, \underline{Z}_n)$ is a d-variate normal distribution with mean $\hat{\underline{\mu}}_n := \frac{1}{n}\sum_{t=1}^n \underline{Z}_t$ and covariance matrix $n^{-1}2\pi\boldsymbol{f}(0)$. Since it holds

$$P_W^n(\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}|\underline{Z}_1, \ldots, \underline{Z}_n) = 1, \quad \text{for all } (\underline{Z}_1, \ldots, \underline{Z}_n).$$

by Assumption $\mu 2$, it follows that assumption (a) of Theorem 9.7 is satisfied and we will continue to show that assumption (b) is also satisfied, which is equivalent to

$$\mathrm{E}_0^n \sup_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} P_W^n(n^{1/2}\|\underline{\mu} - \underline{\mu}_0\| > M_n|\boldsymbol{f}, \underline{Z}_1, \ldots, \underline{Z}_n) \to 0. \tag{9.12}$$

The conditional probability on the left hand side of (9.12) is equal to

$$\begin{aligned} P_W^n(n^{1/2}\|\underline{\mu} - \underline{\mu}_0\| > M_n|\boldsymbol{f}, \underline{Z}_1, \ldots, \underline{Z}_n) &= P_W^n(\underline{\mu} \notin B_{n^{-1/2}M_n}(\underline{\mu}_0)|\boldsymbol{f}, \underline{Z}_1, \ldots, \underline{Z}_n) \\ &= 1 - N_d\left(\hat{\underline{\mu}}_n, n^{-1}2\pi\boldsymbol{f}(0)\right)\left[B_{n^{-1/2}M_n}(\underline{\mu}_0)\right] \end{aligned}$$

and (9.12) is equivalent to

$$\mathrm{E}_0^n \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d\left(\hat{\underline{\mu}}_n, n^{-1}2\pi\boldsymbol{f}(0)\right)\left[B_{n^{-1/2}M_n}(\underline{\mu}_0)\right] \to 1. \tag{9.13}$$

Under $P_W^n(\cdot|\underline{\mu}_0, \boldsymbol{f}_0)$, it holds $\hat{\underline{\mu}}_n = \underline{\mu}_0 + \underline{r}_n$ with $\underline{r}_n \sim N_d\left(\underline{0}, n^{-1}2\pi\boldsymbol{f}_0(0)\right)$. This follows from (4.8), noting that $\hat{\underline{\mu}}_n$ is equal to the Fourier coefficient $\tilde{\underline{Z}}_0$ at the zero frequency. Using translation by $\underline{\mu}$, we obtain

$$N_d\left(\hat{\underline{\mu}}_n, n^{-1}2\pi\boldsymbol{f}(0)\right)\left[B_{n^{-1/2}M_n}(\underline{\mu}_0)\right] = N_d\left(\underline{r}_n, n^{-1}2\pi\boldsymbol{f}(0)\right)\left[B_{n^{-1/2}M_n}(\underline{0})\right].$$

Let $\tilde{\underline{r}} := \sqrt{n}(2\pi)^{-1/2}\underline{r}_n$. Then $\tilde{\underline{r}} \sim N_d(\underline{0}, \boldsymbol{f}_0(0))$ and, using $|a\boldsymbol{A}| = |a|^d|\boldsymbol{A}|$ for $a \in \mathbb{R}$ and $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, we compute

$$\begin{aligned} &N_d\left(\underline{r}_n, n^{-1}2\pi\boldsymbol{f}(0)\right)\left[B_{n^{-1/2}M_n}(\underline{0})\right] \\ &= \frac{n^{d/2}}{(2\pi)^d|\boldsymbol{f}(0)|^{1/2}}\int_{|\underline{z}|<n^{-1/2}M_n} \exp\left(-\frac{n}{4\pi}(\underline{z}-\underline{r}_n)^T\boldsymbol{f}(0)^{-1}(\underline{z}-\underline{r}_n)\right)d\underline{z} \\ &= \frac{n^{d/2}}{(2\pi)^d|\boldsymbol{f}(0)|^{1/2}}\int_{|\underline{z}|<n^{-1/2}M_n} \exp\left(-\frac{1}{2}\left(\frac{\sqrt{n}\underline{z}}{\sqrt{2\pi}} - \tilde{\underline{r}}\right)^T\boldsymbol{f}(0)^{-1}\left(\frac{\sqrt{n}\underline{z}}{\sqrt{2\pi}} - \tilde{\underline{r}}\right)\right)d\underline{z}. \end{aligned}$$

Employing the transformation $\underline{y} = \sqrt{n}(2\pi)^{-1/2}\underline{z}$ with $d\underline{z} = (2\pi)^{d/2}n^{-d/2}d\underline{y}$, this reveals

$$\begin{aligned} &N_d\left(\underline{r}_n, n^{-1}2\pi\boldsymbol{f}(0)\right)\left[B_{n^{-1/2}M_n}(\underline{0})\right] \\ &= \frac{1}{(2\pi)^{d/2}|\boldsymbol{f}(0)|^{1/2}}\int_{|\underline{y}|<\tilde{M}_n} \exp\left(-\frac{1}{2}\left(\underline{y}-\tilde{\underline{r}}\right)^T\boldsymbol{f}(0)^{-1}\left(\underline{y}-\tilde{\underline{r}}\right)\right)d\underline{y} \\ &= N_d\left(\tilde{\underline{r}}, \boldsymbol{f}(0)\right)\left[B_{\tilde{M}_n}(\underline{0})\right], \end{aligned}$$

with $\tilde{M}_n = (2\pi)^{-1/2} M_n$. Thus the left hand side of (9.13) is equal to

$$\begin{aligned}
&\mathrm{E}_0^n \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{\tilde{r}}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n}(\underline{0}) \right] \\
&= \frac{1}{(2\pi)^{d/2} |\boldsymbol{f}_0(0)|^{1/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2} \underline{r}^T \boldsymbol{f}_0(0)^{-1} \underline{r} \right) \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{r}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n}(\underline{0}) \right] d\underline{r} \\
&\geq \frac{1}{(2\pi)^{d/2} |\boldsymbol{f}_0(0)|^{1/2}} \int_{B_{\tilde{M}_n/2}(0)} \exp\left(-\frac{1}{2} \underline{r}^T \boldsymbol{f}_0(0)^{-1} \underline{r} \right) \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{r}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n}(\underline{0}) \right] d\underline{r}.
\end{aligned}$$

For $\|\underline{r}\| < \tilde{M}_n/2$ it holds $B_{\tilde{M}_n/2}(\underline{0}) \subset B_{\tilde{M}_n}(-\underline{r})$ and thus

$$N_d(\underline{r}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n}(\underline{0}) \right] = N_d(\underline{0}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n}(-\underline{r}) \right] \geq N_d(\underline{0}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right]$$

for any $\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}$, yielding

$$\begin{aligned}
&\mathrm{E}_0^n \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{\tilde{r}}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n}(\underline{0}) \right] \\
&\geq \frac{1}{(2\pi)^{d/2} |\boldsymbol{f}_0(0)|^{1/2}} \int_{\tilde{B}_{M_n/2}(\underline{0})} \exp\left(-\frac{1}{2} \underline{r}^T \boldsymbol{f}_0(0)^{-1} \underline{r} \right) \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{0}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right] d\underline{r} \\
&= N_d(\underline{0}, \boldsymbol{f}_0(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right] \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{0}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right].
\end{aligned}$$

Now consider the matrix $\boldsymbol{T} := \mathrm{diag}(\tau_1, \ldots, \tau_1) \in \mathbb{R}^{d \times d}$. By assumption, it holds $\boldsymbol{f}(0) \leq \boldsymbol{T}$ for all $\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}$ and an application of Lemma 9.8 yields

$$\begin{aligned}
&N_d(\underline{0}, \boldsymbol{f}_0(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right] \inf_{\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}} N_d(\underline{0}, \boldsymbol{f}(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right] \\
&\geq N_d(\underline{0}, \boldsymbol{f}_0(0)) \left[B_{\tilde{M}_n/2}(\underline{0}) \right] N_d(\underline{0}, \boldsymbol{T}) \left[B_{\tilde{M}_n/2}(\underline{0}) \right] \to 1,
\end{aligned}$$

because $\tilde{M}_n = (2\pi)^{-1/2} M_n \to \infty$ as $n \to \infty$. We have shown that the assumptions of Theorem 9.7 are fulfilled, and an application thereof yields the claim. □

9.3. Outlook: Bernstein-von-Mises

In Section 9.2, we have seen that an $n^{-1/2}$ rate of contraction can be achieved for the marginal posterior of $\underline{\mu}$. The question arises whether the limiting distribution of the posterior distribution is Gaussian, i.e. whether a Bernstein-von-Mises type theorem can be established. In a parametric setting, it is necessary that the likelihood is asymptotically equal to a properly scaled Gaussian location model, a property known as *Local Asymptotic Normality*. It can be expected that this property is also needed in the semiparametric context. The following general theorem gives sufficient conditions for scalar parameters of interest.

Theorem 9.10 (Semiparametric Bernstein-von-Mises)**.** *Let $Z_1, \ldots, Z_n$ be random variables in some Euclidean space with likelihood having Lebesgue density $p(Z_1, \ldots, Z_n | \theta, \eta)$, where the parameter is partitioned as (θ, η) with $\theta \in \Theta \subset \mathbb{R}$ and $\eta \in \mathcal{H}$ and $\mathcal{H}$ is a measurable space. Let θ and η be independent under the prior, i.e. $P(d\theta, d\eta) = P(d\theta)P(d\eta)$. Let (Θ_n) and $(\mathcal{H}_n)$ be sequences of measurable subsets of Θ and $\mathcal{H}$. Assume that the following properties hold:*

(a) *Local Asymptotic Normality: There exists a tight sequence* (G_n) *of random variables (i.e. the sequence of distribution functions of* G_n *is tight) and a sequence of positive numbers* (a_n)*, bounded away from 0, such that the log likelihood ratio can be written as*

$$\log \frac{p(Z_1,\dots,Z_n|\theta,\eta)}{p(Z_1,\dots,Z_n|\theta_0,\eta)} = \sqrt{n}(\theta-\theta_0)G_n - \frac{1}{2}na_n|\theta-\theta_0|^2 + R_n(\theta,\eta)$$

for all $(\theta,\eta) \in \Theta \times \mathcal{H}$ *with a remainder term* $R_n(\theta,\eta)$ *fulfilling*

$$\sup_{\theta\in\Theta_n,\eta\in\mathcal{H}_n} \frac{R_n(\theta,\eta)}{1+n|\theta-\theta_0|^2} \to 0, \quad \text{as } n\to\infty. \tag{9.14}$$

The sequences (G_n) *and* (a_n) *may depend on* (θ_0,η_0)*, but not on* (θ,η)*.*

(b) *Prior positivity: In an open neighborhood of* θ_0*, the prior on* θ *possesses a continuous Lebesgue density that is positive and bounded away from 0.*

(c) *Joint and uniform conditional nuisance posterior concentration: It holds*

$$P(\theta\in\Theta_n,\eta\in\mathcal{H}_n|Z_1,\dots,Z_n)\to 1$$

as well as

$$\inf_{\theta\in\Theta_n} P(\eta\in\mathcal{H}_n|Z_1,\dots,Z_n,\theta)\to 1.$$

Then for the marginal posterior $P(d\theta|Z_1,\dots,Z_n)$ *of* θ *it holds*

$$\mathrm{E}_0\delta_{\mathrm{TV}}\left[P(\theta\in\cdot|Z_1,\dots,Z_n), N(\theta_0+a_n^{-1}G_n, n^{-1}a_n^{-1})\right]\to 0,$$

with E_0 *denoting the expected value under* (θ_0,η_0) *and* δ_{TV} *denoting the total variation distance between two measures.*

Proof. See Theorem 12.9 in Ghosal and van der Vaart (2017). □

It may be noted that Theorem 9.10 can be formulated in more generality, employing so-called *least favorable transformations* of the nuisance parameter, see Section 12.3.2 in Ghosal and van der Vaart (2017). However, the above formulation is sufficient for our considerations. Let us restrict our attention to the univariate case $d=1$ and investigate how the assumptions of Theorem 9.10 translate to the mean $\theta=\mu$ of a time series with nuisance parameter $\eta=f$ being the spectral density. With Whittle's Likelihood (9.3), the log likelihood ratio is

$$\log\frac{p_W(Z_1,\dots,Z_n|\mu,f)}{p_W(Z_1,\dots,Z_n|\mu_0,f)} = \frac{1}{2}\left(\frac{(\tilde{Z}_0-\sqrt{n}\mu_0)^2}{2\pi f(0)} - \frac{(\tilde{Z}_0-\sqrt{n}\mu)^2}{2\pi f(0)}\right) \tag{9.15}$$

with the Fourier coefficient at $\omega=0$ given by $\tilde{Z}_0 = \frac{1}{\sqrt{n}}\sum_{t=1}^n Z_t$. Let $G_n := \tilde{Z}_0 - \sqrt{n}\mu_0$. Under $P_W(\cdot|\mu_0,f_0)$, it holds $G_n \sim N(0,2\pi f_0(0))$. Hence the sequence (G_n) is tight. Furthermore, the right hand side of (9.15) is equal to

$$\begin{aligned}\frac{1}{4\pi f(0)}\left(G_n^2 - \left(G_n-\sqrt{n}(\mu-\mu_0)\right)^2\right) &= \frac{1}{4\pi f(0)}\left(2\sqrt{n}G_n(\mu-\mu_0) - n(\mu-\mu_0)^2\right)\\ &= \frac{1}{2\pi f_0(0)}\left(\sqrt{n}G_n(\mu-\mu_0) - \frac{1}{2}n(\mu-\mu_0)^2\right)\\ &\quad + R_n(\mu,f)\end{aligned}$$

with remainder term

$$\begin{aligned} R_n(\mu, f) &= \frac{1}{2\pi}\left(\sqrt{n}G_n(\mu-\mu_0) - \frac{1}{2}n(\mu-\mu_0)^2\right)\left(\frac{1}{f(0)} - \frac{1}{f_0(0)}\right) \\ &=: \left(R_n^{(1)}(\mu) + R_n^{(2)}(\mu)\right)\left(\frac{1}{f(0)} - \frac{1}{f_0(0)}\right). \end{aligned}$$

From these considerations, it follows that assumption (a) of Theorem 9.10 is fulfilled if the remainder fulfills the decay condition (9.14). Thus we compute

$$\frac{\left|R_n^{(1)}(\mu)\left(\frac{1}{f(0)} - \frac{1}{f_0(0)}\right)\right|}{1+n(\mu-\mu_0)^2} \lesssim \frac{\sqrt{n}|\mu-\mu_0|\,|G_n|}{1+n(\mu-\mu_0)^2}\left|\frac{1}{f(0)} - \frac{1}{f_0(0)}\right|. \tag{9.16}$$

Recall the set $\mathcal{H}_{\tau_0,\tau_1}$ from (9.11). It follows from (9.16) that

$$\sup_{\mu\in\mathbb{R}, f\in\mathcal{H}_{\tau_0,\tau_1}} \frac{\left|R_n^{(1)}(\mu)\left(\frac{1}{f(0)} - \frac{1}{f_0(0)}\right)\right|}{1+n(\mu-\mu_0)^2} \to 0, \quad n\to\infty.$$

Furthermore, we compute

$$\frac{\left|R_n^{(2)}(\mu)\left(\frac{1}{f(0)} - \frac{1}{f_0(0)}\right)\right|}{1+n(\mu-\mu_0)^2} \lesssim \frac{n(\mu-\mu_0)^2}{1+n(\mu-\mu_0)^2}\left|\frac{1}{f(0)} - \frac{1}{f_0(0)}\right|.$$

Thus, for condition (9.14) to be fulfilled, we need to find posterior concentration set sequences (Θ_n) and $(\mathcal{H}_n)$ such that either $\sup_{\mu\in\Theta_n} n(\mu-\mu_0)^2 \to 0$ holds or

$$\sup_{f\in\mathcal{H}_n} |f(0) - f_0(0)| \to 0, \quad \text{as } n\to\infty. \tag{9.17}$$

The first condition can in general not be expected to hold, since it is known from the parametric case that the posterior of μ does not contract faster than with rate $n^{-1/2}$ (see Section 10.2 in van der Vaart (2000)). Thus it follows from Theorem 9.10 that a Bernstein-von-Mises theorem for μ can be established if a posterior concentration (in the sense of condition (c) of Theorem 9.10) set sequence $(\mathcal{H}_n)$ can be found such that condition (9.17) is fulfilled. The condition (9.17) is of different nature than the previous results of this work ($\mathbb{L}^1$ posterior consistency in Section 7.1.2 and Hellinger contraction rates in Sections 7.2.2 and 9.1), since it involves pointwise consistency at $\omega = 0$. The question under which conditions this can be established remains open. For future work, it will be of great interest to investigate if consistency or contraction rates employing pointwise conditions can as well be established and this could be used to establish (9.17), implying a Bernstein-von-Mises result for μ. One possible approach to tackle this issue would be to establish consistency in the supremum norm, a road that has been taken in the recent literature for some special models, see Castillo (2014) and Yoo and Ghosal (2016).

In the previous Sections, we restricted our attention to the mean model. It is conjectured that the results and proof ideas can readily be carried over to more general models, as e.g. the linear model (8.1). A possible difficulty in this context will be that the Fourier Transform of the design matrix induces a frequency domain signal not only on $\omega = 0$ but on all Fourier Frequencies (see the examples in Section 8.1).

Appendix

A.

Assumptions

Chapter 1 and Chapter 2

Assumption α1. *The integral* $\int_{\bar{\mathbb{S}}_d^+} \log\left(1+\frac{1}{\beta(\boldsymbol{U})}\right)\alpha(d\boldsymbol{U})$ *is finite.*

Assumption $\mathcal{X}$1. *Let $\mathcal{X}$ be a Polish space, i.e. a topological space that is homeomorphic to a complete metric space having a countable dense subset. Let $\mathcal{X}$ be equipped with a locally compact and σ-finite and non-trivial Borel measure denoted by dx.*

Chapter 3

Assumption $A\Gamma$1. *Let $\eta \equiv \eta_0 > d-1$ and $\omega\colon \mathcal{X} \to (0,\infty)$ be measurable such that $\int_{\mathcal{X}} \omega(x)dx < \infty$. Let $\boldsymbol{\Sigma}\colon \mathcal{X} \to \mathcal{S}_d^+$ be measurable such that $\sup_{x\in\mathcal{X}} \lambda_{\max}(\boldsymbol{\Sigma}(x)) \leq \tau < \infty$.*

Assumption GP1. *(a) Let $\alpha\colon \mathcal{X} \times \mathbb{B}(\bar{\mathbb{S}}_d^+) \to [0,\infty)$ such that $\{\alpha(x,\cdot)\}_{x\in\mathcal{X}}$ is a family of finite measures on $\bar{\mathcal{S}}_d^+$ and such that for all $B \in \mathbb{B}(\bar{\mathbb{S}}_d^+)$ the mapping $\mathcal{X} \ni x \mapsto \alpha(x,B)$ is measurable.*

(b) Let $\beta\colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \to (0,\infty)$ be measurable.

Assumption GP2. *The integral $\int_{\mathcal{X}\times\bar{\mathbb{S}}_d^+\times[0,\infty)} \min(1,r)\nu(dx,d\boldsymbol{U},dr)$ is finite for ν as in (3.1).*

Assumption GP2'. *The integral $C_\alpha := \int_{\mathcal{X}} \alpha(x,\bar{\mathbb{S}}_d^+)dx$ is finite and there exists a constant $\beta_0 > 0$ such that $\beta(x,\boldsymbol{U}) \geq \beta_0$ holds for all $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$ and all $x \in \mathcal{X}\setminus N$, where $N \subset \mathcal{X}$ is a null set.*

Assumption GP3. *(a) There exists $N \subset \mathcal{X}$ with $\int_N dx = 0$ such that $\operatorname{supp}(\alpha(x,\cdot)) = \bar{\mathbb{S}}_d^+$ holds for all $x \in \mathcal{X}\setminus N$.*

(b) The function β is locally bounded, i.e. for every $x_0 \in \mathcal{X}$ there exists a neighborhood $\mathcal{U}_0$ of x_0 such that $\sup_{x\in\mathcal{U}_0,\boldsymbol{U}\in\bar{\mathbb{S}}_d^+} \beta(x,\boldsymbol{U}) < \infty$.

Assumption GP4. *(a) It holds $\alpha(x,d\boldsymbol{U}) = g(x,\boldsymbol{U})d\boldsymbol{U}$ for a measurable function $g\colon \mathcal{X} \times \bar{\mathbb{S}}_d^+ \to (0,\infty)$. Furthermore, there exist positive constants g_0, g_1 such that $g_0 \leq g(x,\boldsymbol{U}) \leq g_1$ for all $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$ and all $x \in \mathcal{X}\setminus N$, where $N \subset \mathcal{X}$ is a null set, i.e. $\int_N dx = 0$.*

(b) There exist constants $0 < \beta_0 \leq \beta_1 < \infty$ *and a null set* $N \subset \mathcal{X}$ *such that* $\beta_0 \leq \beta(x, \boldsymbol{U}) \leq \beta_1$ *holds for all* $\boldsymbol{U} \in \bar{\mathbb{S}}_d^+$ *and all* $x \in \mathcal{X} \setminus N$.

Assumption $\mathcal{X}$2. *It holds* $\int_{\mathcal{X}} dx < \infty$.

Chapter 4

Assumption $\boldsymbol{f}$1. *The eigenvalues of* $\boldsymbol{f}(\omega)$ *are uniformly bounded and uniformly bounded away from 0. That is, there exist positive constants* b_0, b_1 *such that*

$$\lambda_{\min}(\boldsymbol{f}(\omega)) \geq b_0, \quad \lambda_{\max}(\boldsymbol{f}(\omega)) \leq b_1, \quad 0 \leq \omega \leq \pi.$$

Assumption $\boldsymbol{f}$2. *The autocovariance function* $\boldsymbol{\Gamma}(h) = \int_0^{2\pi} \boldsymbol{f}(\omega) \exp(ih\omega) d\omega$ *of* $\boldsymbol{f}$ *fulfills*

$$\sum_{h \in \mathbb{Z}} \|\boldsymbol{\Gamma}(h)\| |h|^a < \infty,$$

for some $a > 1$.

Chapter 7

Assumption k1. *There exist positive constants* c, C *such the prior probability mass function of* k *fulfills*

$$0 < p(k) \leq C \exp(-ck \log k), \quad k \in \mathbb{N}.$$

Assumption k2. *There exist positive constants* A_1, A_2 *and positive constant* κ_1, κ_2 *such that*

$$A_1 \exp(-\kappa_1 k \log k) \leq p(k) \leq A_2 \exp(-\kappa_2 k), \quad k \in \mathbb{N}.$$

Chapter 9

Assumption μ1. *The prior* $P(d\underline{\mu})$ *on* $\underline{\mu}$ *is such that all moments exist. Furthermore, it is assumed that in a neighborhood of* $\underline{\mu}_0$, *it possesses a Lebesgue density that is continuous and bounded away from 0.*

Assumption μ2. *The prior on* $(\underline{\mu}, \boldsymbol{f})$ *is of the form* $P(d\underline{\mu}, d\boldsymbol{f}) \propto P(d\boldsymbol{f})$, *i.e.* $\underline{\mu}$ *and* $\boldsymbol{f}$ *are a priori independent and the prior on* $\underline{\mu}$ *is improper with* $p(\underline{\mu}) \propto 1$. *The prior on* $\boldsymbol{f}$ *is assumed to fulfill* $P(\boldsymbol{f} \in \mathcal{H}_{\tau_0,\tau_1}) = 1$ *for some* $0 < \tau_0 < \tau_1 < \infty$ *with* $\mathcal{H}_{\tau_0,\tau_1}$ *from* (9.11).

B.

Mathematical Appendix

B.1. Matrix Algebra

For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{d\times d}$, the *Kronecker product* of $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B}$ is defined as

$$\boldsymbol{A} \otimes \boldsymbol{B} := \begin{pmatrix} a_{11}\boldsymbol{B} & \dots & a_{1d}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{d1}\boldsymbol{B} & \dots & a_{dd}\boldsymbol{B} \end{pmatrix} \in \mathbb{C}^{d^2\times d^2}. \tag{B.1}$$

The following calculation rules apply.

Lemma B.1. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D} \in \mathbb{C}^{d\times d}$. Then it holds*

$$(\boldsymbol{A} \otimes \boldsymbol{B})^* = \boldsymbol{A}^* \otimes \boldsymbol{B}^*, \quad (\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD}).$$

Proof. See Section 2 in Chapter 2 of Magnus and Neudecker (2007). □

Let $m > d-1$ and $\boldsymbol{S} \in \mathcal{S}_d^+(\mathbb{R})$ with $\mathcal{S}_d^+(\mathbb{R})$ denoting the set of symmetric positive definite $\mathbb{R}^{d\times d}$ matrices. The *Inverse Wishart* distribution $\text{Wish}_{d\times d}^{-1}(m, \boldsymbol{S})$ with m degrees of freedom and scale matrix $\boldsymbol{S}$ is defined in terms of the Lebesgue density

$$p(\boldsymbol{Z}) = \frac{|\boldsymbol{S}|^{m/2}}{2^{md/2}\Gamma_d(m/2)} |\boldsymbol{Z}|^{-(m+d+1)/2} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{S}\boldsymbol{X}^{-1}\right), \quad \boldsymbol{Z} \in \mathcal{S}_d^+(\mathbb{R}), \tag{B.2}$$

where Γ_d denotes the multivariate Gamma function (see Chapter 1.4 in Gupta and Nagar (1999)).

Norms and Eigenvalues

For a matrix $\boldsymbol{A} \in \mathcal{S}_d$, denote the eigenvalues of $\boldsymbol{A}$ by $\lambda_1(\boldsymbol{A}), \dots, \lambda_d(\boldsymbol{A}) \in \mathbb{R}$. Furthermore, denote by $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ the smallest and largest eigenvalue. The trace is defined as

$$\operatorname{tr}(\boldsymbol{A}) := \sum_{i=1}^{d} a_{ii} = \sum_{i=1}^{d} \lambda_i(\boldsymbol{A}).$$

The following *cyclic property* of the trace applies:

Lemma B.2. *Let* $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{C}^{d\times d}$. *Then*

$$\operatorname{tr}(\boldsymbol{ABC}) = \operatorname{tr}(\boldsymbol{BCA}) = \operatorname{tr}(\boldsymbol{CAB}).$$

Proof. See Section 10 in Chapter 1 of Magnus and Neudecker (2007). □

The following result is well-known as the Min-Max principle of Courant-Fisher.

Lemma B.3 (A.1 in Marshall et al. (2011))**.** *Let* $\boldsymbol{A} \in \mathcal{S}_d$. *Then*

$$\lambda_{\min}(\boldsymbol{A}) = \min_{0\neq \underline{z}\in\mathbb{C}^d} \frac{\underline{z}^*\boldsymbol{A}\underline{z}}{\|\underline{z}\|^2}, \quad \lambda_{\max}(\boldsymbol{A}) = \max_{0\neq \underline{z}\in\mathbb{C}^d} \frac{\underline{z}^*\boldsymbol{A}\underline{z}}{\|\underline{z}\|^2}.$$

In this work, we consider the following matrix norms. Let $\boldsymbol{A} = (a_{ij}) \in \mathbb{C}^{d\times d}$. Then the *Frobenius Norm* is defined as

$$\|\boldsymbol{A}\| := \sqrt{\operatorname{tr}(\boldsymbol{AA}^*)} = \sqrt{\sum_{i,j=1}^{d} |a_{ij}|^2}. \tag{B.3}$$

The 1-*Norm* is defined as

$$\|\boldsymbol{A}\|_1 := \sum_{i,j=1}^{d} |a_{ij}|, \tag{B.4}$$

whereas the *Euclidean Norm* is defined as

$$\|\boldsymbol{A}\|_2 := \sqrt{\lambda_{\max}(\boldsymbol{AA}^*)} \tag{B.5}$$

and the *Max Norm* as

$$\|\boldsymbol{A}\|_\infty := \max_{i,j=1,\dots,d} |a_{ij}|. \tag{B.6}$$

The *Trace Norm* is defined as

$$\|\boldsymbol{A}\|_T := \operatorname{tr}\left((\boldsymbol{AA}^*)^{1/2}\right) = \sum_{j=1}^{d} \sqrt{(\lambda_j(\boldsymbol{AA}^*)} \tag{B.7}$$

where $(\boldsymbol{AA}^*)^{1/2}$ denotes the Hermitian positive semidefinite matrix square root of $\boldsymbol{AA}^* \in \bar{\mathcal{S}}_d^+$, see (B.15). For $\boldsymbol{A} \in \mathcal{S}_d$ the Frobenius and trace norm simplify to

$$\|\boldsymbol{A}\| = \sqrt{\sum_{j=1}^{d} \lambda_j(\boldsymbol{A})^2}, \quad \|\boldsymbol{A}\|_T = \sum_{j=1}^{d} |\lambda_j(\boldsymbol{A})| \tag{B.8}$$

and for $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$, we have the particularly convenient representations

$$\|\boldsymbol{A}\|_T = \operatorname{tr}(\boldsymbol{A}), \quad \|\boldsymbol{A}\|_2 = \lambda_{\max}(\boldsymbol{A}). \tag{B.9}$$

We will frequently need the following inequalities.

Lemma B.4. *(a) For $\boldsymbol{A} \in \mathbb{C}^{d\times d}$ it holds*

$$\frac{1}{d}\|\boldsymbol{A}\|_1 \leq \|\boldsymbol{A}\| \leq \|\boldsymbol{A}\|_1, \quad \|\boldsymbol{A}\|_2 \leq \|\boldsymbol{A}\| \leq \sqrt{d}\|\boldsymbol{A}\|_2, \quad \|\boldsymbol{A}\| \leq d\|\boldsymbol{A}\|_\infty,$$

and

$$\|\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^*\| = \|\boldsymbol{A}\|$$

for all unitary $\boldsymbol{U} \in \mathbb{C}^{d\times d}$.

(b) For $\boldsymbol{A} \in \mathcal{S}_d$ it holds

$$\lambda_{\max}(\boldsymbol{A}) \leq \|\boldsymbol{A}\|_2, \quad |\lambda_{\min}(\boldsymbol{A})| \leq \|\boldsymbol{A}\|_2.$$

(c) For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{d\times d}$ it holds

$$|\operatorname{tr}(\boldsymbol{A}\boldsymbol{B})| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|, \quad \|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|, \quad \|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|.$$

(d) For $\boldsymbol{A}, \boldsymbol{B} \in \bar{\mathcal{S}}_d^+$ and $\boldsymbol{C} \in \mathcal{S}_d$ it holds

$$\lambda_{\min}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B}) \leq \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \leq \lambda_{\max}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B}) \tag{B.10}$$

and

$$\lambda_{\min}(\boldsymbol{A})\|\boldsymbol{C}\| \leq \|\boldsymbol{A}^{1/2}\boldsymbol{C}\boldsymbol{A}^{1/2}\| \leq \lambda_{\max}(\boldsymbol{A})\|\boldsymbol{C}\|. \tag{B.11}$$

Proof. Part (a) and (b) and (c) are standard results, see e.g. Appendix II of Davies (1973). Part (c) can be found in Appendix I of Dzhaparidze and Kotz (2012). For a proof of (B.10), see Sections H.1.g and H.1.h in Chapter 9 of Marshall et al. (2011). To show (B.11), we use the cyclic property of the trace from Lemma B.2 to compute

$$\|\boldsymbol{A}^{1/2}\boldsymbol{C}\boldsymbol{A}^{1/2}\|^2 = \operatorname{tr}\left(\boldsymbol{A}^{1/2}\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}\boldsymbol{A}^{1/2}\right) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}) \geq \lambda_{\min}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}),$$

where (B.10) was used in the last step for $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$ and $\boldsymbol{C}\boldsymbol{A}\boldsymbol{C} \in \bar{\mathcal{S}}_d^+$. By the same argument, it also holds

$$\operatorname{tr}(\boldsymbol{C}\boldsymbol{A}\boldsymbol{C}) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{C}\boldsymbol{C}) \geq \lambda_{\min}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{C}\boldsymbol{C}) = \lambda_{\min}(\boldsymbol{A})\|\boldsymbol{C}\|^2,$$

yielding the lower bound in (B.11) and the proof for the upper bound follows along the same lines. □

The absolute value of the determinant of $\boldsymbol{A} \in \mathbb{C}^{d\times d}$ is abbreviated by

$$|\boldsymbol{A}| := |\det(\boldsymbol{A})|. \tag{B.12}$$

The following results for the determinant are used frequently.

Lemma B.5. *(a) For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{d\times d}$ it holds*

$$\det \boldsymbol{A} = \prod_{j=1}^{n} \lambda_j(\boldsymbol{A}), \quad \det(\boldsymbol{A}\boldsymbol{B}) = \det \boldsymbol{A} \det \boldsymbol{B},$$

with $\lambda_1(\boldsymbol{A}), \ldots, \lambda_d(\boldsymbol{A}) \in \mathbb{C}$ denoting the eigenvalues of $\boldsymbol{A}$.

(b) For $\boldsymbol{A} \in \mathcal{S}_d^+$ it holds

$$|\boldsymbol{A}| = \det \boldsymbol{A}, \quad |\boldsymbol{A}^{1/2}| = |\boldsymbol{A}|^{1/2}, \quad |\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1},$$

with $\boldsymbol{A}^{1/2}$ denoting the Hpd matrix square root of $\boldsymbol{A}$ from (B.15).

(c) For $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{S}_d^+$ it holds

$$|\boldsymbol{A} + \boldsymbol{B}| \geq \prod_{i=1}^{d} \left(\lambda_i(\boldsymbol{A}) + \lambda_i(\boldsymbol{B})\right),$$

with $\lambda_1(\boldsymbol{A}) \leq \ldots \leq \lambda_d(\boldsymbol{A})$ and $\lambda_1(\boldsymbol{B}) \leq \ldots \leq \lambda_d(\boldsymbol{B})$ denoting the eigenvalues in nondecreasing order.

Proof. Part (a) and (b) are standard results an can e.g. be found in Section 9 in Chapter 1 of Magnus and Neudecker (2007). For part (c), see Section G.2.a in Chapter 9 of Marshall et al. (2011). □

The following result can be useful when working with the Frobenius norm.

Lemma B.6. *Let $\boldsymbol{A}, \boldsymbol{B} \in \bar{\mathcal{S}}_d^+$. Then*

$$\|\boldsymbol{B}^{1/2}\boldsymbol{A}\boldsymbol{B}^{1/2} - \boldsymbol{I}_d\| = \|\boldsymbol{B}\boldsymbol{A} - \boldsymbol{I}_d\| = \|\boldsymbol{A}\boldsymbol{B} - \boldsymbol{I}_d\| = \|\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2} - \boldsymbol{I}_d\|,$$

where $\boldsymbol{A}^{1/2} \in \bar{\mathcal{S}}_d^+$ denotes the Hermitian positive semidefinite square root.

Proof. By the cyclic property of the trace from Lemma B.2 it holds

$$\operatorname{tr}(\boldsymbol{B}^{1/2}\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^{1/2}) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^{1/2}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}\boldsymbol{A})$$

and the result follows from

$$\|\boldsymbol{C} - \boldsymbol{I}_d\|^2 = \operatorname{tr}\left((\boldsymbol{C} - \boldsymbol{I}_d)(\boldsymbol{C} - \boldsymbol{I}_d)^*\right) = \operatorname{tr}(\boldsymbol{C}\boldsymbol{C}^*) - 2\operatorname{tr}(\boldsymbol{C}) + d$$

for $\boldsymbol{C} = \boldsymbol{B}^{1/2}\boldsymbol{A}\boldsymbol{B}^{1/2}$ or $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ or $\boldsymbol{C} = \boldsymbol{A}^{1/2}\boldsymbol{B}\boldsymbol{A}^{1/2}$ or $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{A}$. □

The following result gives some insight into the structure of the cone $\bar{\mathcal{S}}_d^+$ of Hermitian positive semidefinite matrices.

Lemma B.7. *Let $\boldsymbol{A} \in \mathcal{S}_d$. Then $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$ if and only if $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \geq 0$ holds for all $\boldsymbol{B} \in \bar{\mathcal{S}}_d^+$.*

Proof. By Lemma B.4, it holds $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \geq 0$ for all $\boldsymbol{A}, \boldsymbol{B} \in \bar{\mathcal{S}}_d^+$. Assume that $\boldsymbol{A} \in \mathcal{S}_d \setminus \bar{\mathcal{S}}_d^+$. Denote by $\lambda_1, \ldots, \lambda_d$ the eigenvalues of $\boldsymbol{A}$. Then it holds $\lambda_i < 0$ for some i (because otherwise, it would hold $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$). Without loss of generality, let $\lambda_1 < 0$. Let $\boldsymbol{U} \in \mathbb{C}^{d \times d}$ be unitary such that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$, with $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$. Let $\boldsymbol{B} := \boldsymbol{U}\operatorname{Diag}(1, 0, \ldots, 0)\boldsymbol{U}^*$. Then it clearly holds $\boldsymbol{B} \in \bar{\mathcal{S}}_d^+$ and $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \lambda_1 < 0$. □

Jacobians of Transformations

Recall the definition of the Lebesgue measure $d\boldsymbol{Z}$ on $\mathcal{S}_d^+$ and $d\boldsymbol{U}$ on $\mathbb{S}_d^+$ from (2.5) and (2.6). We need the following results.

Lemma B.8. *Let $\boldsymbol{Z} \in \mathcal{S}_d^+$ and let $\boldsymbol{U} := \frac{1}{r}\boldsymbol{Z} \in \mathbb{S}_d^+$ with $r = \operatorname{tr} \boldsymbol{Z} > 0$. Then $d\boldsymbol{Z} = r^{d^2-1} dr d\boldsymbol{U}$.*

Proof. To show the result, we re-parametrize the complex matrices $\boldsymbol{Z}$ and $\boldsymbol{U}$ as real vectors and derive the Jacobian determinant of the vector-to-vector transformation. To elaborate, for $\boldsymbol{Z} = (z_{ij})_{i,j=1}^d$ and $\boldsymbol{U} = (u_{ij})_{i,j=1}^d$ and $r = \operatorname{tr} \boldsymbol{Z}$ we will consider the transformation $\underline{z} \mapsto \underline{\varphi}$, where

$$\underline{z} := (z_{11}, \Re z_{12}, \Im z_{12}, \ldots, \Im z_{d-1,d}, z_{dd}) \in \mathbb{R}^{d^2},$$
$$\underline{\varphi} := (r, u_{11}, \Re u_{12}, \Im u_{12}, \ldots, \Im u_{d-1,d}) \in \mathbb{R}^{d^2}.$$

This parametrization is in line with the definition of the Lebesgue measure $d\boldsymbol{Z}$ and $d\boldsymbol{U}$ from (2.5) and (2.6). We will compute the determinant of the Jacobian $\boldsymbol{J} = (\frac{\partial \underline{z}_i}{\partial \underline{\varphi}_j})_{i,j=1}^{d^2}$ by the following cofactor expansion along the first column of $\boldsymbol{J}$:

$$\det \boldsymbol{J} = \sum_{i=1}^{d^2} C_{i1} J_{i1}, \quad C_{i1} = (-1)^{i+1} \det \boldsymbol{A}_{i1} \tag{B.13}$$

and $\boldsymbol{A}_{ij}$ denoting the (i,j)-minor matrix of $\boldsymbol{J}$, i.e. the $(d^2-1) \times (d^2-1)$ matrix obtained by deleting the i-th row and j-th column from $\boldsymbol{J}$. The values C_{ij} are called the *cofactors* of $\boldsymbol{J}$. The matrix $\boldsymbol{J}$ is of the following structure:

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial \underline{z}}{\partial r} & \multicolumn{2}{c}{r\boldsymbol{I}_{d^2-1}} \\ & \multicolumn{2}{c}{} \\ & \multicolumn{2}{c}{\frac{\partial \underline{z}_{d^2}}{\partial \underline{\varphi}}} \end{pmatrix}.$$

First observe that, since $\boldsymbol{Z} = r\boldsymbol{U}$ and $u_{dd} = 1 - \sum_{j=1}^{d-1} u_{jj}$, the first column of $\boldsymbol{J}$ is given by

$$\frac{\partial \underline{z}}{\partial r} = \left(u_{11}, \Re u_{12}, \Im u_{12}, \ldots, \Im u_{d-1,d}, 1 - \sum_{j=1}^{d-1} u_{jj} \right)^T.$$

Furthermore, for $i = 1, \ldots, d^2 - 1$, the row vector $(\partial \underline{z}_i)/(\partial \underline{\varphi})$ contains exactly one non-zero element aside from the its first entry, namely $(\partial \underline{z}_i)/(\partial \underline{\varphi}_{i+1}) = r$. The last row $(\partial \underline{z}_{d^2})/(\partial \underline{\varphi}) = (\partial z_{dd})/(\partial \underline{\varphi})$ contains exactly $d-1$ non-zero entries aside from its first entry, namely the entries $(\partial z_{dd})/(\partial u_{jj}) = -r$ for $j = 1, \ldots, d-1$. Thus, all cofactors C_{i1} of $\boldsymbol{J}$ along the first column are zero, except for the ones corresponding to z_i's fulfilling $z_i = z_{kk}$ for some k (because for all the other z_i's, the $(i-1)$-th column of the corresponding cofactor-submatrix is zero). Denote these non-zero cofactors by C_k and the corresponding cofactor-submatrices by $\boldsymbol{A}_k \in \mathbb{R}^{(d^2-1)\times(d^2-1)}$, $k = 1, \ldots, d$. Since $\boldsymbol{A}_d$ is diagonal with entries r, we have $\det \boldsymbol{A}_d = r^{d^2-1}$ and hence

$$C_d = (-1)^{d^2+1} \det \boldsymbol{A}_d = (-1)^{d^2+1} r^{d^2-1}.$$

For $k = 1, \dots, d-1$, the matrix $\boldsymbol{A}_k$ is of the following structure:

$$\boldsymbol{A}_k = \left(\begin{array}{c|c|c} & 0 & \\ r\boldsymbol{I}_a & \vdots & \boldsymbol{0}_{a\times b} \\ & & \\ \hline & & \\ r\boldsymbol{0}_{b\times a} & \vdots & \boldsymbol{I}_b \\ & 0 & \\ \hline * & -r & * \end{array}\right),$$

where $a + b = d^2 - 2$. A cofactor expansion along the last row of $\boldsymbol{A}_k$ yields

$$\det \boldsymbol{A}_k = (-1)^{k+d^2-1}(-r)r^{d^2-2} = (-1)^{d^2+k}r^{d^2-1}$$

and hence

$$C_k = (-1)^{1+k} \det \boldsymbol{A}_k = (-1)^{d^2+1}r^{d^2-1}, \quad k = 1, \dots, d-1.$$

Combining these insights with (B.13) reveals

$$\begin{aligned} \det \boldsymbol{J} &= \sum_{k=1}^{d-1} u_{kk}C_k + \left(1 - \sum_{k=1}^{d-1} u_{kk}\right) C_d \\ &= (-1)^{d^2+1} \sum_{k=1}^{d-1} u_{kk} r^{d^2-1} + (-1)^{d^2+1} \left(1 - \sum_{k=1}^{d-1} u_{kk}\right) r^{d^2-1} \\ &= (-1)^{d^2+1} r^{d^2-1}, \end{aligned}$$

hence $|\boldsymbol{J}| = r^{d^2-1}$, concluding the proof. □

Lemma B.9. *Let $\boldsymbol{Z} \in \mathcal{S}_d^+$ and let $\boldsymbol{\Sigma} \in \mathbb{C}^{d\times d}$ be invertible. Let $\boldsymbol{Y} = \boldsymbol{\Sigma}\boldsymbol{Z}\boldsymbol{\Sigma}^*$. Then $d\boldsymbol{Y} = |\boldsymbol{\Sigma}\boldsymbol{\Sigma}^*|^d d\boldsymbol{Z}$.*

Proof. See Mathai (1997), Theorem 3.5. □

Let $\eta > d - 1$. The complex multivariate Gamma function is defined as

$$\tilde{\Gamma}_d(\eta) = \int_{\mathcal{S}_d^+} \operatorname{etr}(-\boldsymbol{Z})|\boldsymbol{Z}|^{\eta-d} d\boldsymbol{Z} = \pi^{d(d-1)/2} \prod_{j=1}^{d} \Gamma(\eta - j + 1), \tag{B.14}$$

where $\Gamma(a) = \int_0^\infty z^{a-1} \exp(-z) dz$ is the Gamma function, being defined for $a > 0$.

Lemma B.10. *Let $\eta > d - 1$. Then*

$$\int_{\mathbb{S}_d^+} |\boldsymbol{U}|^{\eta-d} d\boldsymbol{U} = \frac{\tilde{\Gamma}_d(\eta)}{\Gamma(\eta d)}.$$

Proof. The proof is an adaption of Proposition 5.5 in Pérez-Abreu and Stelzer (2014) and uses the same idea of transforming the spherical integral to an integral over the whole cone $\mathcal{S}_d^+$, by employing a one-dimensional Gamma radial component. Observe that for any $a > 0$ we have

$$\begin{aligned}\int_{\mathbb{S}_d^+} |\boldsymbol{U}|^{\eta-d} d\boldsymbol{U} &= \frac{1}{\Gamma(a)} \int_0^\infty \int_{\mathbb{S}_d^+} r^{a-\eta d+d^2-1} \exp(-r)|r\boldsymbol{U}|^{\eta-d} dr d\boldsymbol{U} \\ &= \frac{1}{\Gamma(a)} \int_{\mathcal{S}_d^+} \operatorname{tr}(\boldsymbol{Z})^{a-\eta d} \operatorname{etr}(-\boldsymbol{Z})|\boldsymbol{Z}|^{\eta-d} d\boldsymbol{Z},\end{aligned}$$

where the transformation

$$\boldsymbol{Z} = r\boldsymbol{U}, \quad r = \operatorname{tr} \boldsymbol{Z}, \quad dr d\boldsymbol{U} = r^{-d^2+1} d\boldsymbol{Z}$$

from Lemma B.8 was used in the last step. Now choosing $a = \eta d$ and using the definition (B.14) of the complex multivariate Gamma function concludes the proof. □

Hpd Matrix Square Root

Any matrix $\boldsymbol{A} \in \mathcal{S}_d$ can be written as $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$ with $\boldsymbol{U}$ being a unitary matrix containing the eigenvectors of $\boldsymbol{A}$ as columns and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$, being the diagonal matrix of the eigenvalues $\lambda_1, \dots, \lambda_d \in \mathbb{R}$ of $\boldsymbol{A}$. If $\boldsymbol{A} \in \bar{\mathcal{S}}_d^+$, it holds $\lambda_j \geq 0$ for $j = 1, \dots, d$ and we can define in this case

$$\boldsymbol{A}^{1/2} := \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^*, \quad \boldsymbol{\Lambda}^{1/2} := \operatorname{diag}(\lambda_1^{1/2}, \dots, \lambda_d^{1/2}). \tag{B.15}$$

We call $\boldsymbol{A}$ the *Hermitian positive semidefinite square root* of $\boldsymbol{A}$ (or Hpd square root if $\boldsymbol{A} \in \mathcal{S}_d^+$). It is clear that $\boldsymbol{A}^{1/2} \in \bar{\mathcal{S}}_d^+$ (and $\boldsymbol{A}^{1/2} \in \mathcal{S}_d^+$ if $\boldsymbol{A} \in \mathcal{S}_d^+$) and that $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$.

Similar to the one-dimensional case, the matrix square root is differentiable and uniformly Lipschitz, if the domain of definition is uniformly bounded away from 0 (in terms of eigenvalues), as we will show in the following. Recall that a function $\boldsymbol{\Psi} \colon \mathcal{S}_d \to \mathcal{S}_d$ is called Fréchet differentiable at $\boldsymbol{Z} \in \mathcal{S}_d$, if there exists a linear function $\boldsymbol{\Psi}'(\boldsymbol{Z}) \colon \mathcal{S}_d \to \mathcal{S}_d$ and a positive constant C such that $\|\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}]\| \leq C\|\boldsymbol{A}\|$ holds for all $\boldsymbol{A} \in \mathcal{S}_d$ and

$$\frac{\|\boldsymbol{\Psi}(\boldsymbol{Z}+\boldsymbol{X}) - \boldsymbol{\Psi}(\boldsymbol{Z}) - \boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{X}]\|}{\|\boldsymbol{X}\|} \to 0, \quad \text{as } \|\boldsymbol{X}\| \to 0. \tag{B.16}$$

Lemma B.11. *Consider the mapping $\boldsymbol{\Psi} \colon \mathcal{S}_d^+ \to \mathcal{S}_d^+$ defined as the Hermitian positive definite matrix square root $\boldsymbol{\Psi}(\boldsymbol{Z}) = \boldsymbol{Z}^{1/2}$. Then $\boldsymbol{\Psi}$ is Fréchet differentiable. For $\boldsymbol{Z} \in \mathcal{S}_d^+$, the Fréchet derivative $\boldsymbol{\Psi}'(\boldsymbol{Z}) \colon \mathcal{S}_d \to \mathcal{S}_d$ is given by*

$$\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}] = \boldsymbol{U}\boldsymbol{\phi}(\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U})\boldsymbol{U}^*, \quad \boldsymbol{A} \in \mathcal{S}_d \tag{B.17}$$

with

$$\phi(\tilde{\boldsymbol{A}})_{ij} := \frac{\tilde{a}_{ij}}{\sqrt{\lambda_i} + \sqrt{\lambda_j}}, \quad \tilde{\boldsymbol{A}} = (\tilde{a}_{ij}) \in \mathcal{S}_d,$$

where $\lambda_1, \dots, \lambda_d > 0$ are the eigenvalues of $\boldsymbol{Z}$ and $\boldsymbol{U}$ is unitary such that $\boldsymbol{Z} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^*$, with $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$. Furthermore, for all $\tau > 0$ and $\mathcal{S}_{d|\tau}^+ := \{\boldsymbol{Z} \in \mathcal{S}_d^+ \colon \lambda_{\min}(\boldsymbol{Z}) \geq \tau\}$ there exists a positive constant C only depending on τ and d such that

$$\|\boldsymbol{\Psi}'(\boldsymbol{Z})\| \leq C, \quad \text{for all } \boldsymbol{Z} \in \mathcal{S}_{d|\tau}^+, \tag{B.18}$$

with $\|\boldsymbol{\Psi}'(\boldsymbol{Z})\| := \sup_{\boldsymbol{0}\neq\boldsymbol{A}\in\mathcal{S}_d} \frac{\|\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}]\|}{\|\boldsymbol{A}\|}$ *denoting the operator norm of* $\boldsymbol{\Psi}'(\boldsymbol{Z})$.

Proof. Applying the Fréchet differential operator to the left and right hand side of $\boldsymbol{\Psi}(\boldsymbol{Z})\boldsymbol{\Psi}(\boldsymbol{Z}) = \boldsymbol{Z}$ and using the chain rule (see Chapter 8, Section 2, equation (15) in Magnus and Neudecker (2007)) yields

$$\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}]\boldsymbol{\Psi}(\boldsymbol{Z}) + \boldsymbol{\Psi}(\boldsymbol{Z})\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}] = \boldsymbol{A}, \quad \boldsymbol{A}\in\mathcal{S}_d. \tag{B.19}$$

First assume that $\boldsymbol{U} = \boldsymbol{I}_d$, i.e. $\boldsymbol{Z} = \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1,\dots,\lambda_d)$. Let $\boldsymbol{B} = \boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}]$. Then (B.19) becomes

$$\boldsymbol{B}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{\Lambda}^{1/2}\boldsymbol{B} = \boldsymbol{A}, \quad \boldsymbol{\Lambda}^{1/2} = \operatorname{diag}(\sqrt{\lambda}_1,\dots,\sqrt{\lambda}_d),$$

from which $a_{ij} = (\sqrt{\lambda}_i + \sqrt{\lambda}_j)b_{ij}$ and hence $b_{ij} = \frac{a_{ij}}{\sqrt{\lambda}_i+\sqrt{\lambda}_j}$ can directly be concluded. For the case $\boldsymbol{U}\neq\boldsymbol{I}_d$, let $\boldsymbol{X}\in\mathcal{S}_d$ such that $\boldsymbol{\Lambda}+\boldsymbol{X}\in\mathcal{S}_d^+$. Recalling $\boldsymbol{\Psi}(\boldsymbol{Z}) = \boldsymbol{Z}^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^*$ and the unitary invariance of the Frobenius norm (i.e. $\|\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^*\| = \|\boldsymbol{A}\|$ for all unitary matrices $\boldsymbol{U}$ and all $\boldsymbol{A}\in\mathbb{C}^{d\times d}$, see Lemma B.4) we consider the difference quotient

$$\begin{aligned}
&\frac{\|(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^* + \tilde{\boldsymbol{X}})^{1/2} - \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^* - \boldsymbol{U}\boldsymbol{\Psi}'(\boldsymbol{\Lambda})[\boldsymbol{U}^*\tilde{\boldsymbol{X}}\boldsymbol{U}]\boldsymbol{U}^*\|}{\|\tilde{\boldsymbol{X}}\|}\\
&= \frac{\|\boldsymbol{U}((\boldsymbol{\Lambda}+\boldsymbol{U}^*\tilde{\boldsymbol{X}}\boldsymbol{U})^{1/2} - \boldsymbol{\Lambda}^{1/2} - \boldsymbol{\Psi}'(\boldsymbol{\Lambda})[\boldsymbol{U}^*\tilde{\boldsymbol{X}}\boldsymbol{U}])\boldsymbol{U}^*\|}{\|\tilde{\boldsymbol{X}}\|}\\
&= \frac{\|(\boldsymbol{\Lambda}+\tilde{\boldsymbol{X}})^{1/2} - \boldsymbol{\Lambda}^{1/2} - \boldsymbol{\Psi}'(\boldsymbol{\Lambda})[\boldsymbol{U}^*\tilde{\boldsymbol{X}}\boldsymbol{U}]\|}{\|\boldsymbol{U}^*\tilde{\boldsymbol{X}}\boldsymbol{U}\|} \to 0,
\end{aligned}$$

as $\|\tilde{\boldsymbol{X}}\|\to 0$ by definition of $\boldsymbol{\Psi}'(\boldsymbol{\Lambda})$. This shows that $\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}] = \boldsymbol{U}\boldsymbol{\Psi}'(\boldsymbol{\Lambda})[\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}]\boldsymbol{U}^*$, concluding the proof of (B.17).

Consider the max norm $\|\cdot\|_\infty$ from (B.6). By the equivalence of matrix norms, there exist positive constants c_1, c_2 only depending on d such that $c_1\|\boldsymbol{A}\|_\infty \le \|\boldsymbol{A}\| \le c_2\|\boldsymbol{A}\|_\infty$ holds for all $\boldsymbol{A}\in\mathbb{C}^{d\times d}$. For $\boldsymbol{Z}\in\mathcal{S}_{d|\tau}^+$ and $\boldsymbol{A}\in\mathcal{S}_d$, we get

$$\|\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}]\| = \|\boldsymbol{U}\boldsymbol{\phi}(\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U})\boldsymbol{U}^*\| = \|\boldsymbol{\phi}(\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U})\| \le c_2\|\boldsymbol{\phi}(\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U})\|_\infty,$$

where the unitary invariance of the Frobenius norm $\|\cdot\|$ was used in the second step. Furthermore it clearly holds

$$\|\boldsymbol{\phi}(\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U})\|_\infty \le \frac{1}{2\sqrt{\lambda_{\min}(\boldsymbol{Z})}}\|\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}\|_\infty \le \frac{1}{2\sqrt{\tau}}\|\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}\|_\infty,$$

and finally

$$\|\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}\|_\infty \le \frac{1}{c_1}\|\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}\| = \frac{1}{c_1}\|\boldsymbol{A}\|$$

yields

$$\|\boldsymbol{\Psi}'(\boldsymbol{Z})[\boldsymbol{A}]\| \le \frac{c_2}{2c_1\sqrt{\tau}}\|\boldsymbol{A}\|,$$

which concludes (B.18) with $C := \frac{c_2}{2c_1\sqrt{\tau}}$. □

B.2. Bernstein Polynomials

The Bernstein polynomial basis of degree $k \in \mathbb{N}$ on $[0,1]$ consists of the following k Beta density functions:

$$b(x|j, k-j+1) = \frac{\Gamma(k+1)}{\Gamma(j)\Gamma(k-j+1)} x^{j-1}(1-x)^{k-j}, \quad 0 \leq x \leq 1, \quad j = 1, \ldots, k \tag{B.20}$$

with $\Gamma(j) = \int_0^\infty r^{j-1} \exp(-j) dr$ denoting the Gamma function. Given some mixture weights $\underline{w} := (w_1, \ldots, w_k) \in \mathbb{R}^k$, the Bernstein polynomial mixture of degree k is defined as

$$\mathfrak{B}(k, \underline{w})[x] := \sum_{j=1}^{k} w_j b(x|j, k-j+1), \quad 0 \leq x \leq 1.$$

By Section 1 in Petrone (1999), it holds $\int_0^1 b(x|j, k-j+1)dx = 1$ and $\max_{0\leq x\leq 1} |b(x|j, k-1+1)| \leq k$ for all $k \in \mathbb{N}$ and all $j = 1, \ldots, k$. This implies

$$\int_0^1 \mathfrak{B}(k, \underline{w})[x]dx = \sum_{j=1}^{k} w_j, \quad \max_{0\leq x\leq 1} |\mathfrak{B}(k, \underline{w})[x]| \leq k \max_{j=1,\ldots,k} |w_j|. \tag{B.21}$$

We first present some approximation properties of Bernstein polynomials.

Lemma B.12. *Let $\boldsymbol{f}\colon [0,1] \to \mathbb{C}^{d\times d}$ be continuous. For $k \in \mathbb{N}$, consider the Bernstein polynomial*

$$\boldsymbol{f}_k(x) = \sum_{j=1}^{k} \left(\int_{\frac{j-1}{k}}^{\frac{j}{k}} \boldsymbol{f}(t)dt \right) b(x|j, k-j+1), \quad 0 \leq x \leq 1, \tag{B.22}$$

where the integral is understood component-wise. Then it holds

$$\max_{0\leq x\leq 1} \|\boldsymbol{f}(x) - \boldsymbol{f}_k(x)\| \to 0, \quad \text{as } k \to \infty.$$

Furthermore, if $\boldsymbol{f}$ takes only values in $\bar{\mathcal{S}}_d^+$ (or $\mathcal{S}_d^+$), then so does $\boldsymbol{f}_k$.

Proof. Let $\varepsilon > 0$. Consider the components f_{rs} for $r, s = 1, \ldots, d$ of $\boldsymbol{f}$. By the uniform approximation property of Bernstein polynomials (see Theorem 1.6.1 in Lorentz (2012) or Section 1 in Petrone (1999)), there exists for every $\delta > 0$ a positive integer $k_{0,rs}$ such that for all $k \geq k_{0,rs}$ it holds $\max_{0\leq x\leq 1} |f_{rs}(x) - f_{k,rs}(x)| < \delta$, with

$$f_{k,rs}(x) = \sum_{j=1}^{k} \left(\int_{\frac{j-1}{k}}^{\frac{j}{k}} f_{rs}(t)dt \right) b(x|j, k-j+1), \quad 0 \leq x \leq 1$$

denoting the r,s-th component of $\boldsymbol{f}_k$. Since $\int_{(j-1)/k}^{j/k} \boldsymbol{f}(t)dt \in \bar{\mathcal{S}}_d^+$ for $\boldsymbol{f}\colon [0,1] \to \bar{\mathcal{S}}_d^+$, it follows $\boldsymbol{f}_k(x) \in \bar{\mathcal{S}}_d^+$ for $0 \leq x \leq 1$ in this case. Similarly, if $\boldsymbol{f}$ takes values in $\mathcal{S}_d^+$, then it holds

$$\underline{z}^* \left(\int_{(j-1)/k}^{j/k} \boldsymbol{f}(t)dt \right) \underline{z} = \int_{(j-1)/k}^{j/k} \underline{z}^* \boldsymbol{f}(t) \underline{z} dt > 0, \quad j = 1, \ldots, k,$$

for all $\underline{z} \in \mathbb{C}^d$, which shows that $\int_{(j-1)/k}^{j/k} \boldsymbol{f}(t)dt \in \mathcal{S}_d^+$ Hence $\boldsymbol{f}_k$ only takes values in $\mathcal{S}_d^+$ in this case. With $k \geq k_0 := \max_{r,s=1,\dots,d} k_{0,rs}$ it holds $\max_{0\leq x\leq 1} \|\boldsymbol{f}(x) - \boldsymbol{f}_k(x)\|_\infty < \delta$. Note that $\max_{r,s=1,\dots,d} |A_{rs}| \geq \frac{1}{d^2}\|\boldsymbol{A}\|$ for all $\boldsymbol{A} \in \mathbb{C}^{d\times d}$. Choosing $\delta < \frac{\varepsilon}{d}$ we obtain for $k \geq k_0$ from Lemma B.4

$$\max_{0\leq x\leq 1} \|\boldsymbol{f}(x) - \boldsymbol{f}_k(x)\| \leq d \max_{0\leq x\leq 1} \|\boldsymbol{f}(x) - \boldsymbol{f}_k(x)\|_\infty < \varepsilon,$$

concluding the proof. □

The preceding result can be strengthened to get a rate of convergence, under additional regularity assumptions on $\boldsymbol{f}$, as the following result shows.

Lemma B.13. *Let $\boldsymbol{f}\colon [0,1] \to \mathbb{C}^{d\times d}$ and $\boldsymbol{f}_k$ be defined as in (B.22).*

(a) If the components of $\boldsymbol{f}$ are Hölder of order $0 < a \leq 1$, then there exists a positive constant C such that

$$\max_{0\leq x\leq 1} \|\boldsymbol{f}(x) - \boldsymbol{f}_k(x)\| \leq Ck^{-a/2}$$

holds for all $k \in \mathbb{N}$.

(b) If the components of $\boldsymbol{f}$ are continuously differentiable with derivatives being Hölder of order $0 < a \leq 1$, then there exists a positive constant C such that

$$\max_{0\leq x\leq 1} \|\boldsymbol{f}(x) - \boldsymbol{f}_k(x)\| \leq Ck^{-(1+a)/2}$$

holds for all $k \in \mathbb{N}$.

Proof. The rates of approximation for the components of $\boldsymbol{f}$ is well-known, see e.g. Theorem 1.6.1 and Theorem 1.6.2 in Lorentz (2012), or Lemma E.3 in Ghosal and van der Vaart (2017). The rate in the matrix norm can be obtained with the same argument as in Lemma B.12. □

Remark B.14. *The approximation rates achieved by Bernstein polynomials (see Lemma B.13) are not optimal. It is known that for a function f on $[0,1]$ of smoothness a, there exists a polynomial f_k of degree k such that $\|f - f_k\|_\infty < Ck^{-a}$ holds for all $k \in \mathbb{N}$ and some constant $C > 0$, see Section 1.6 in Lorentz (2012). These rates can e.g. be attained by splines, see Lemma E.5 in Ghosal and van der Vaart (2017).*

B.3. Fourier Series

Let $(\boldsymbol{b}_k)_{k\in\mathbb{Z}}$ be a sequence of matrices in $\mathbb{C}^{d\times d}$ such that $\sum_{k\in\mathbb{Z}} \|\boldsymbol{b}_k\|^2 < \infty$. Let

$$\boldsymbol{f}(\omega) := \sum_{k\in\mathbb{Z}} \boldsymbol{b}_k \exp(-ik\omega), \quad \omega \in \mathbb{R}. \tag{B.23}$$

Clearly, $\boldsymbol{f}$ is 2π-periodic and by Parceval's Theorem, the components of $\boldsymbol{f}$ are in $\mathbb{L}^2([0,2\pi])$. For $0 < a \leq 1$, we denote by $\mathcal{C}_{d\times d}^a$ the space of continuous 2π-periodic functions g which are Hölder of order a, i.e.

$$\mathcal{C}_{d\times d}^a := \left\{ \boldsymbol{g}\colon \mathbb{R} \to \mathbb{C}^{d\times d} \text{ } 2\pi\text{-periodic}\colon \sup_{\omega\in\mathbb{R}} \|\boldsymbol{g}(\omega+\delta) - \boldsymbol{g}(\omega)\| \leq C|\delta|^a \text{ for some } C > 0 \right\}.$$

The following Lemma summarizes some regularity conditions for $\boldsymbol{f}$ from (B.23), in terms of decay of the Fourier coefficients $\boldsymbol{b}_k$.

Lemma B.15. *Let $(\boldsymbol{b}_k)_{k\in\mathbb{Z}}$ be a sequence of matrices in $\mathbb{C}^{d\times d}$.*

(a) If $\sum_{k\in\mathbb{Z}} \|\boldsymbol{b}_k\| < \infty$, then the series in (B.23) converges uniformly for $0 \le \omega \le 2\pi$. In particular, $\boldsymbol{f}$ is continuous.

(b) If $\sum_{k\in\mathbb{Z}} |k|^a \|\boldsymbol{b}_k\| < \infty$ for some $0 < a \le 1$, then $\boldsymbol{f} \in \mathcal{C}^a_{d\times d}$.

(c) If $\sum_{k\in\mathbb{Z}} |k|^s \|\boldsymbol{b}_k\| < \infty$ for some positive integer s, then $\boldsymbol{f}$ is s times continuously differentiable with derivative $\boldsymbol{f}^{(s)}(\omega) = \sum_{k\in\mathbb{Z}} (-k)^s \boldsymbol{b}_k \exp(-ik\omega)$ for $\omega \in \mathbb{R}$, where the series converges uniformly for $0 \le \omega \le 2\pi$.

(d) If $\sum_{k\in\mathbb{Z}} |k|^{s+a} \|\boldsymbol{b}_k\| < \infty$ for some $0 < a < 1$ and some positive integer s, then $\boldsymbol{f}^{(s)} \in \mathcal{C}^a_{d\times d}$.

Proof. Let $N \ge 0$ and consider the trigonometric polynomial $\boldsymbol{f}_N := \sum_{k=0}^N \boldsymbol{b}_k \exp(-ik\omega)$. For $N > M \ge 0$, we get

$$\|\boldsymbol{f}_N - \boldsymbol{f}_M\|_{F,\infty} = \left\| \sum_{k=M+1}^{N} \boldsymbol{b}_k \exp(-ik\cdot) \right\|_{F,\infty} \le \sum_{k=M+1}^{N} \|\boldsymbol{b}_k\|.$$

Since $(\boldsymbol{b}_k)_{k\ge 0}$ (being convergent) is a Cauchy sequence in $\ell^1(\mathbb{C}^{d\times d})$, so are the components of $(\boldsymbol{f}_N)_{N\ge 0}$ a Cauchy sequence in the Banach space $C([0,2\pi], \|\cdot\|_\infty)$, thus convergent. Similarly, also $(\boldsymbol{f}_N)_{N\le 0}$ is convergent, yielding (a).

To show (b), let $\delta \in \mathbb{R}$ and define $D_\delta \boldsymbol{f}(\omega) := \boldsymbol{f}(\omega+\delta) - \boldsymbol{f}(\omega)$. It suffices to show $\|D_\delta \boldsymbol{f}(\omega)\| \le C|\delta|^a$ for some constant C that does not depend on δ or ω. With $\xi_k(\omega) := \exp(-ik\omega)$, we compute

$$\frac{|D_\delta \xi_k(\omega)|}{|\delta|^a} = \frac{|D_{k\delta}\xi_1(k\omega)|}{|\delta|^a} \le \frac{C|k\delta|^a}{|\delta|^a} = C|k|^a,$$

where the constant C does not depend on k, ω or δ, since $\xi_1 \in \mathcal{C}^a$. Thus

$$\frac{\|D_\delta \boldsymbol{f}(\omega)\|}{|\delta|^a} \le \sum_{k\in\mathbb{Z}} \|\boldsymbol{b}_k\| \frac{|D_\delta \xi_k(\omega)|}{|\delta|^a} \le C \sum_{k\in\mathbb{Z}} \|\boldsymbol{b}_k\| |k|^a.$$

This yields (b), since the right hand side is finite and does not depend on δ.

Part (c) follows readily from the dominated convergence theorem, using part (a). Part (d) follows by a combination of part (b) and (c). □

We will also encounter the following regularity class, which is defined in terms of decay of the Fourier coefficients. For $a \ge 0$, let

$$\mathcal{A}^a_{d\times d} := \left\{ \boldsymbol{f} = \sum_{k\in\mathbb{Z}} \boldsymbol{b}_k \exp(-ik\cdot) \colon \sum_{k\in\mathbb{Z}} \|\boldsymbol{b}_k\| |k|^a < \infty \right\}.$$

It is known that $\mathcal{A}^a_{d\times d}$ with the pointwise matrix multiplication is a Banach algebra (see Section 13.1 in Gröchenig (2013)). The next result is a generalization of Wiener's Lemma, stating that the invertible elements in $\mathcal{A}^a_{d\times d}$ are characterized by pointwise invertibility.

Lemma B.16. *For $a \geq 0$, let $\boldsymbol{f} \in \mathcal{A}^a_{d\times d}$ with $|\boldsymbol{f}(\omega)| \neq 0$ for $0 \leq \omega \leq \pi$. Then $\boldsymbol{f}^{-1} \in \mathcal{A}^a_{d\times d}$.*

Proof. The proof can be found in Lemma 13.3.2 in Gröchenig (2013). □

B.4. Poisson Processes

Let $\mathcal{Y}$ be a Borel space, i.e. a topological space endowed with the Borel σ-algebra. A *Poisson process* Π on $\mathcal{Y}$ is a random countable subset of $\mathcal{Y}$, such that for all $k > 0$ and all disjoint subsets $A_1, \ldots, A_k$ of $\mathcal{Y}$, the random variables $N(A_1), \ldots, N(A_k)$ are independent with $N(A_i) \sim \mathrm{Poi}(\nu(A_i))$, where $N(A_i) = \#\{\Pi \cap A_i\}$. The measure ν on $\mathcal{Y}$ is called the *mean measure* of Π. We write $\Pi \sim \mathrm{PP}(\nu)$. The proofs of the next four results can all be found in Kingman (1992) (see Sections 2.2, 2.5, 3.2, 4.1 and 5 there).

Theorem B.17 (Existence Theorem)**.** *Let ν be a σ-finite measure on a Borel space $\mathcal{Y}$ such that for all $y \in \mathcal{Y}$, the set $\{y\}$ is measurable with $\nu(\{y\}) = 0$. Then there exists a Poisson process Π on $\mathcal{Y}$ with mean measure ν.*

Theorem B.18 (Superposition Theorem)**.** *Let $\{\Pi_j \colon j \in \mathbb{N}\}$ be a countable collection of independent Poisson processes having respective mean measure ν_j on a Borel space $\mathcal{Y}$. Denote by $\Pi := \cup_{j=1}^{\infty} \Pi_j$ their superposition. Then it holds $\Pi \sim \mathrm{Poi}(\nu)$ with $\nu := \sum_{j=1}^{\infty} \nu_j$.*

Theorem B.19 (Campbell's Theorem)**.** *Let Π be a Poisson process with mean measure ν on a Borel space $\mathcal{Y}$ and let $g \colon \mathcal{Y} \to [0, \infty)$ be measurable. Then the sum $\Phi := \sum_{y\in\Pi} g(y)$ is absolutely convergent in probability if and only if $\int_{\mathcal{Y}} \min\{|g(y)|, 1\}\nu(dy)$ is finite. In this case, the distribution of Φ is given in terms of the Laplace transform*

$$\mathrm{E}\exp(-t\Phi) = \exp\left(-\int_{\mathcal{Y}} (1 - \exp(-tg(y)))\nu(dy)\right), \quad t \geq 0.$$

Theorem B.20 (Interval Theorem)**.** *Let Π be a Poisson process on $[0, \infty)$ with mean measure $\nu(dy) = \lambda dy$, where dy denotes the Lebesgue measure and λ is a positive constant. Let $y_1 := \inf \Pi$ and $y_{j+1} := \inf\{\Pi \setminus \{y_1, \ldots, y_j\}\}$ for $j \geq 1$. Then the increments*

$$z_1 := y_1, \quad z_{j+1} := y_{j+1} - y_j \quad (j \geq 1),$$

are independent with $z_j \sim \mathrm{Exp}(\lambda)$ for $j = 1, 2, \ldots$.

Theorem B.21 (Marking Theorem)**.** *Let $\mathcal{Y}, \mathcal{Z}$ be Borel spaces and let Π be a Poisson process on $\mathcal{Y}$ with mean measure ν. Consider a mapping $\alpha \colon \mathcal{Y} \times \mathbb{B}(\mathcal{Z}) \to [0, \infty)$ (with $\mathbb{B}(\mathcal{Z})$ denoting the Borel σ-algebra in $\mathcal{Z}$) such that for every $y \in \mathcal{Y}$, $\alpha(y, \cdot)$ is a probability measure on $\mathcal{Z}$ and that for every measurable $B \subset \mathcal{Z}$, the mapping $\mathcal{Y} \ni y \mapsto \alpha(y, B)$ is measurable. Consider the following random subset of $\mathcal{Y} \times \mathcal{Z}$:*

$$\Pi_0 := \{(y, m_y) \colon y \in \Pi\}, \quad m_y | y \overset{ind.}{\sim} \alpha(y, \cdot),$$

consisting of each element $y \in \mathcal{Y}$ of Π and a respective marking *$m_y \in \mathcal{Z}$, drawn independently from $\alpha(y, \cdot)$. Then Π_0 is a Poisson process on $\mathcal{Y} \times \mathcal{Z}$ with mean measure*

$$\nu_0(A) = \int_{\mathcal{Y}\times\mathcal{Z}} \mathbb{1}_A(y, z)\alpha(y, dz)\nu(dy), \quad A \subset \mathcal{X} \times \mathcal{Y} \text{ measurable.}$$

Theorem B.22 (Almost sure mapping Theorem)**.** *Let $\mathcal{Y}, \mathcal{Z}$ be Borel spaces and let $\Pi, \tilde{\Pi}$ be Poisson processes on $\mathcal{Y}$ and $\mathcal{Z}$ with mean measure ν and $\tilde{\nu}$ respectively. Assume that there exists a measurable function $h\colon \mathcal{Y} \to \mathcal{Z}$ such that $\tilde{\nu}(B) = \nu(h^{-1}(B))$ for every measurable $B \subset \mathcal{Z}$. Assume further that the underlying probability space of $\tilde{\Pi}$ is rich enough to carry a* $\mathrm{Unif}([0,1])$ *random variable (i.e. having the continuous uniform distribution on $(0,1)$) which is independent of $\tilde{\Pi}$. Let $\{z_j\}_{j\geq 1} \stackrel{d}{=} \Pi$. Then there exists a sequence $\{\tilde{z}_j\}_{j\geq 1}$ of random variables in $\mathcal{Y}$ with $\{\tilde{z}_j\}_{j\geq 1} \stackrel{d}{=} \{z_j\}_{j\geq 1}$ such that the almost sure representation $\tilde{\Pi} \stackrel{a.s.}{=} \{h(\tilde{z}_j)\}_{j\geq 1}$ holds.*

Proof. The conclusion $\tilde{\Pi} \stackrel{d}{=} \{h(z_j)\}_{j\geq 1}$ is standard, see e.g. Section 2.3 in Kingman (1992). The argument for the almost sure representation is given in Proposition 2.1 in Rosiński (2001). □

B.5. Miscellaneous

Let $k > 1$ and $a_1, \ldots, a_k > 0$. The Dirichlet distribution $\mathrm{Dir}(a_1, \ldots, a_k)$ is defined on the standard $(k-1)$-simplex, in terms of the probability density

$$p(x_1, \ldots, x_k) = \frac{\Gamma\left(\sum_{j=1}^k a_j\right)}{\prod_{j=1}^k \Gamma(a_j)} \prod_{j=1}^k x_j^{a_j - 1}, \quad x_1, \ldots, x_k \in (0,1), \quad \sum_{j=1}^k x_j = 1. \tag{B.24}$$

For $(X_1, \ldots, X_k) \sim \mathrm{Dir}(a_1, \ldots, a_k)$ with $a_0 := \sum_{j=1}^k a_j$ it holds

$$\mathrm{E}X_j = \frac{a_j}{a_0}, \quad \mathrm{Var}X_j = \frac{a_j(a_0 - a_j)}{a_0^2(a_0+1)}, \quad j = 1, \ldots, k. \tag{B.25}$$

The Gamma distribution $\mathrm{Ga}(a, b)$ with shape parameter $a > 0$ and rate parameter $b > 0$ is defined in terms of the Lebesgue density

$$p(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \quad x > 0. \tag{B.26}$$

The following result of Riemann sum convergence speed is well known in the literature. However, since a reference of the proof was not readily available, the proof is included for the sake of completeness.

Lemma B.23. *Let $g\colon [0,1] \to \mathbb{R}$ be continuously differentiable. Then, with $x_j := j/n$ for $j = 0, \ldots, n-1$ it holds*

$$\left| \int_0^1 g(x)dx - \frac{1}{n} \sum_{j=0}^{n-1} g(x_j) \right| \leq \frac{\max_{0\leq x\leq 1} |g'(x)|}{n}.$$

Proof. First observe that

$$\int_0^1 g(x)dx = \sum_{j=1}^{n-1} \int_{x_j}^{x_{j+1}} g(x)dx, \quad \frac{1}{n} \sum_{j=0}^{n-1} g(x_j) = \sum_{j=0}^{n-1} \int_{x_j}^{x_{j+1}} g(x_j)dx.$$

This leads to

$$\left| \int_0^1 g(x)dx - \frac{1}{n}\sum_{j=0}^{n-1} g(x_j) \right| \leq \sum_{j=0}^{n-1} \int_{x_j}^{x_{j+1}} |g(x) - g(x_j)|dx.$$

By the mean value theorem, there exist $\tilde{x}_j \in [j/n, (j+1)/n]$ for $j = 0, \ldots, n-1$ such that $g(x) - g(x_j) = \frac{g'(\tilde{x}_j)}{n}$. This yields $|g(x) - g(x_j)| \leq \frac{\max_{0\leq x\leq 1}|g'(x)|}{n}$ and hence

$$\sum_{j=0}^{n-1} \int_{x_j}^{x_{j+1}} |g(x) - g(x_j)|dx \leq \sum_{j=0}^{n-1} \frac{\max_{0\leq x\leq 1}|g'(x)|}{n^2} = \frac{\max_{0\leq x\leq 1}|g'(x)|}{n},$$

concluding the proof. □

The following result describes the distribution of random quadratic forms for the multivariate normal and complex multivariate normal distribution.

Lemma B.24. *(a) Let $\boldsymbol{\Sigma} \in \mathcal{S}_d^+(\mathbb{R})$ and $\underline{Z} \sim N_d(\underline{0}, \boldsymbol{\Sigma})$. Let $\boldsymbol{A} \in \mathcal{S}_d(\mathbb{R})$. Consider the random quadratic form $\Psi(\underline{Z}) := \underline{Z}^T \boldsymbol{A}\underline{Z}$. Then*

$$\Psi(\underline{Z}) \stackrel{d}{=} \sum_{i=1}^{d} \lambda_i |X_i|^2,$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\boldsymbol{\Sigma}\boldsymbol{A}$ and $X_1, \ldots, X_d \overset{iid}{\sim} N(0,1)$. In particular, it holds

$$\mathrm{E}\Psi(\boldsymbol{Z}) = \sum_{i=1}^{d} \lambda_i = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}), \quad \mathrm{Var}\Psi(\boldsymbol{Z}) = \sum_{i=1}^{d} 2\lambda_i^2 = 2\,\mathrm{tr}\left((\boldsymbol{\Sigma}\boldsymbol{A})^2\right).$$

(b) Let $\boldsymbol{\Sigma} \in \mathcal{S}_d^+$ and $\underline{Z} \sim CN_d(\underline{0}, \boldsymbol{\Sigma})$. Let $\boldsymbol{A} \in \mathcal{S}_d$. Consider the random quadratic form $\Psi(\underline{Z}) := \underline{Z}^* \boldsymbol{A}\underline{Z}$. Then

$$\Psi(\underline{Z}) \stackrel{d}{=} \sum_{i=1}^{d} \lambda_i |X_i|^2,$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of $\boldsymbol{\Sigma}\boldsymbol{A}$ and $X_1, \ldots, X_d \overset{iid}{\sim} CN(0,1)$. In particular, it holds

$$\mathrm{E}\Psi(\boldsymbol{Z}) = \sum_{i=1}^{d} \lambda_i = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}), \quad \mathrm{Var}\Psi(\boldsymbol{Z}) = \sum_{i=1}^{d} \lambda_i^2 = \mathrm{tr}\left((\boldsymbol{\Sigma}\boldsymbol{A})^2\right).$$

Proof. See the Appendix in Ibragimov (1963). □

The Lambert W function is defined as the inverse function of $\mathbb{R} \ni x \mapsto x\exp(x)$, such that

$$W(z)\exp(W(z)) = z, \quad \text{for } z \geq -\exp(-1). \tag{B.27}$$

Since $x \mapsto x\exp(x)$ is not injective, W consists of two branches for $-\exp(-1) \leq z < 0$, see Section 1 and Figure 1 in Corless et al. (1996). It is evident that W is strictly monotonically

increasing on $(0,\infty)$. Furthermore, W is concave on $[0,\infty)$ (see Section 3 in Corless et al. (1996)).

We will also need the following results.

Lemma B.25. *Let $\alpha > 0$. Then*

$$\frac{\alpha^2}{2}\exp\left(\frac{\alpha x}{2}\right) \leq \frac{(\alpha x - 1)\exp(\alpha x) + 1}{x^2} \leq \frac{\alpha^2}{2}\exp(\alpha x), \quad x \geq 0$$

and

$$\frac{(\alpha x - 1)\exp(\alpha x) + 1}{x^2} \geq \frac{\alpha^2}{2}\exp(\alpha x), \quad x \leq 0.$$

Proof. Using basic algebra, we find that the upper bound $\frac{(\alpha x-1)\exp(\alpha x)+1}{x^2} \leq \frac{\alpha^2}{2}\exp(\alpha x)$ for $x \geq 0$ is equivalent to

$$(\alpha x - 1)^2 + 1 - 2\exp(-\alpha x) \geq 0, \quad x \geq 0,$$

or equivalently $h(z) \geq 0$ for $z \geq 0$ with $h(z) := (z-1)^2 + 1 - 2\exp(-z)$. Since $h(0) = 0$ and the derivative h' of h fulfills

$$h'(z) = 2(z - 1 + \exp(-z)) \geq 0, \quad z \geq 0,$$

the upper bound follows. The proofs for the lower bound for $x \geq 0$ and the results for $x \leq 0$ are analogous. □

Lemma B.26. *Let $\tau \geq 1$ and $a > 0$. Then there exists a positive constant C, only depending on τ and a, such that*

$$z^a - 1 \leq C(z-1), \quad 1 \leq z \leq \tau.$$

Proof. Consider the function $g(z) := z^a - 1$. Assume first that $a \leq 1$. In this case, g is concave, yielding $g(z) \leq g'(z)(z-1) \leq a(z-1)$ for $z \geq 1$. If $a > 1$, then g is strictly convex, and the function graph of g is below the straight line between $g(1)$ and $g(\tau)$, which yields $g(z) \leq \frac{\tau^a - 1}{\tau - 1}(z-1)$ for $1 \leq z \leq \tau$ in this case. □

Lemma B.27. *Let $\boldsymbol{X}, \boldsymbol{Y}$ be independent random elements in $\mathcal{S}_d$. Then the support of $\boldsymbol{X} + \boldsymbol{Y}$ is the closure of $\{\boldsymbol{A} + \boldsymbol{B} \colon \boldsymbol{A} \in \operatorname{supp}(\boldsymbol{X}), \boldsymbol{B} \in \operatorname{supp}(\boldsymbol{Y})\}$.*

Proof. The result is Lemma 24.1 in Sato (1999), by identifying $\mathcal{S}_d$ with $\mathbb{R}^{d^2}$. □

Lemma B.28. *Let Z be a nonnegative random variable with distribution given by its Laplace transform*

$$\mathrm{E}\exp(-tZ) = \exp\left(-\int_0^\infty (1 - \exp(-rt))\nu(dr)\right), \quad t \geq 0,$$

and Lévy measure ν on $[0,\infty)$ with $\nu([0,\infty)) = \infty$ and $\int_0^1 r\nu(dr) < \infty$. Then $\min\operatorname{supp}(Z) = 0$.

Proof. See Corollary 24.8 in Sato (1999). □

The following result gives an explicit formula for mean and covariance of infinitely divisible random Hpd matrices.

Lemma B.29. *Let $\boldsymbol{X}$ be a random element in $\bar{\mathcal{S}}_d^+$ with infinitely divisible distribution given in terms of the Lévy Khinchine representation (see Theorem 2.1) of its Laplace transform*

$$L(\boldsymbol{\Theta}) = \mathrm{E}\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{X}) = \exp\left(-\int_{\bar{\mathcal{S}}_d^+}(1-\operatorname{etr}(-\boldsymbol{\Theta}\boldsymbol{Z}))\nu(d\boldsymbol{Z})\right), \quad \boldsymbol{\Theta}\in\bar{\mathcal{S}}_d^+$$

with Lévy measure ν.

(a) If $\int_{\bar{\mathcal{S}}_d^+}\|\boldsymbol{Z}\|\nu(d\boldsymbol{Z})<\infty$, then $\mathrm{E}\boldsymbol{X}=\int_{\bar{\mathcal{S}}_d^+}\boldsymbol{Z}\nu(d\boldsymbol{Z})$.

(b) If $\int_{\bar{\mathcal{S}}_d^+}\|\boldsymbol{Z}\|^2\nu(d\boldsymbol{Z})<\infty$, then $\operatorname{Cov}\boldsymbol{X} := \mathrm{E}[\boldsymbol{X}^{\otimes 2}]-(\mathrm{E}\boldsymbol{X})^{\otimes 2}$ is well-defined and it holds $\operatorname{Cov}\boldsymbol{X}=\int_{\bar{\mathcal{S}}_d^+}\boldsymbol{Z}\otimes\boldsymbol{Z}\nu(d\boldsymbol{Z})$.

Proof. The cumulant generating function of $\boldsymbol{X}$ is defined as $K(\boldsymbol{\Theta}) := \log L(-\boldsymbol{\Theta})$ for $\boldsymbol{\Theta}\in\bar{\mathcal{S}}_d^-$ with $\bar{\mathcal{S}}_d^-$ denoting the cone of Hermitian negative semidefinite matrices. Differentiating K and evaluation at $\boldsymbol{\Theta}=\boldsymbol{0}$ yields the cumulants. The functional derivative of K in direction $\boldsymbol{T}\in\mathcal{S}_d$ is given by

$$\frac{\partial}{\partial\boldsymbol{\Theta}}K_{\boldsymbol{T}}(\boldsymbol{\Theta}) = \lim_{\varepsilon\to 0}\frac{K(\boldsymbol{\Theta}+\varepsilon\boldsymbol{T})-K(\boldsymbol{\Theta})}{\varepsilon} = \lim_{\varepsilon\to 0}\int_{\bar{\mathcal{S}}_d^+}\frac{\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})-\operatorname{etr}((\boldsymbol{\Theta}+\varepsilon\boldsymbol{T})\boldsymbol{Z})}{\varepsilon}\nu(d\boldsymbol{Z}).$$

By Theorem 1.2 in Mathai (1997) it holds $\frac{\partial}{\partial\boldsymbol{\Theta}}[\operatorname{tr}(\boldsymbol{\Theta}\boldsymbol{Z})]=\boldsymbol{Z}$, which implies $\frac{\partial}{\partial\boldsymbol{\Theta}}[\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})] = \operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\boldsymbol{Z}$ and hence

$$\lim_{\varepsilon\to 0}\frac{\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})-\operatorname{etr}((\boldsymbol{\Theta}+\varepsilon\boldsymbol{T})\boldsymbol{Z})}{\varepsilon} = \operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\operatorname{tr}(\boldsymbol{T}\boldsymbol{Z}).$$

An application of Lebesgue's dominated convergence theorem reveals

$$\frac{\partial}{\partial\boldsymbol{\Theta}}K_{\boldsymbol{T}}(\boldsymbol{\Theta}) = \int_{\bar{\mathcal{S}}_d^+}\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\operatorname{tr}(\boldsymbol{T}\boldsymbol{Z})\nu(d\boldsymbol{Z}),$$

where the integral on the right hand side is well-defined since $0<\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\leq 1$ holds and since $|\operatorname{tr}(\boldsymbol{T}\boldsymbol{Z})|\leq\|\boldsymbol{T}\|\|\boldsymbol{Z}\|$, which is ν-integrable by assumption. This shows

$$\frac{\partial}{\partial\boldsymbol{\Theta}}K(\boldsymbol{\Theta}) = \int_{\bar{\mathcal{S}}_d^+}\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\boldsymbol{Z}\nu(d\boldsymbol{Z})$$

and letting $\boldsymbol{\Theta}\to\boldsymbol{0}$ concludes (a). Similarly, for the second cumulant we use $\frac{\partial}{\partial\boldsymbol{\Theta}}[\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\boldsymbol{Z}] = \operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\boldsymbol{Z}\otimes\boldsymbol{Z}$ (see Magnus and Neudecker (2007), (24), p. 208) to get

$$\frac{\partial^2}{\partial\boldsymbol{\Theta}^2}K(\boldsymbol{\Theta}) = \int_{\bar{\mathcal{S}}_d^+}\operatorname{etr}(\boldsymbol{\Theta}\boldsymbol{Z})\boldsymbol{Z}\otimes\boldsymbol{Z}\nu(d\boldsymbol{Z})$$

and letting $\boldsymbol{\Theta}\to\boldsymbol{0}$ concludes (b). □

Metric Entropy

We will need the notion of a semimetric.

Definition B.30. *Let $\mathcal{X}$ be a set and $d\colon \mathcal{X}\times\mathcal{X}\to[0,\infty)$ be a symmetric function satisfying the triangle inequality $d(x,z)\le d(x,y)+d(y,z)$ for all $x,y,z\in\mathcal{X}$ and $d(x,x)=0$ for all $x\in\mathcal{X}$. Then d is called a* semimetric.

Semimetrics generalize the notion of a metric and they lack the identity of indiscernibles (that is, $d(x,y)=0$ to imply $x=y$). For $\varepsilon>0$, the ε-covering number of a set $\mathcal{X}$ with respect to a semimetric d on $\mathcal{X}$ is defined as

$$N(\varepsilon,\mathcal{X},d):=\inf\left\{N\in\mathbb{N}\colon \exists x_1,\dots,x_N\in\mathcal{X} \text{ with } \mathcal{X}=\bigcup_{j=1}^{N}B_\varepsilon(x_j)\right\}, \tag{B.28}$$

where $B_\varepsilon(x)=\{y\in\mathcal{X}\colon d(x,y)<\varepsilon\}$. The following lemma gives a bound for the covering number of closed balls in Euclidean spaces.

Lemma B.31. *For $\underline{x}=(x_1,\dots,x_d)\in\mathbb{R}^d$, denote for $p\ge 1$ the p-norm of $\underline{x}$ by $\|\underline{x}\|_p=(\sum_{i=1}^d|x_i|^p)^{1/p}$. Then for $M>0$ and $0<\epsilon<M$ it holds*

$$N\big(\epsilon,\{\underline{x}\in\mathbb{R}^d\colon\|\underline{x}\|_p\le M\},\|\cdot\|_p\big)\le\left(\frac{3M}{\epsilon}\right)^d.$$

Proof. See (A.9) in Ghosal and Van Der Vaart (2007). □

The following calculation rules for covering numbers apply.

Lemma B.32. *(a) Let d_1,d_2 be semimetrics on a set $\mathcal{X}$ such that there exists $a,b>0$ with*

$$d_1^a(x,y)\le b d_2(x,y),\quad \text{for all } x,y\in\mathcal{X}.$$

Then $N(\varepsilon,\mathcal{X},d_1)\le N(\frac{\varepsilon^a}{b},\mathcal{X},d_2)$.

(b) Let $\mathcal{X},\mathcal{Y}$ be sets and d be a semimetric on $\mathcal{X}$ and e be a semimetric on $\mathcal{Y}$. Let $f\colon\mathcal{X}\to\mathcal{Y}$ be surjective such that there exists $a,b>0$ with

$$e^a(f(x),f(y))\le b d(x,y),\quad \text{for all } x,y\in\mathcal{X}.$$

Then $N(\varepsilon,\mathcal{Y},e)\le N(\frac{\varepsilon^a}{b},\mathcal{X},d)$.

Proof. We only show part (b), because part (a) is actually a special case thereof (with $\mathcal{X}=\mathcal{Y}$ and f being the identity). Let $B_1,\dots,B_N$ be a covering of $\mathcal{X}$ with respect to d with radius $\frac{\varepsilon^a}{b}$, i.e.

$$B_j=B_{\frac{\varepsilon^a}{b},d}(x_j)=\left\{x\in\mathcal{X}\colon d(x,x_j)<\frac{\varepsilon^a}{b}\right\},\quad j=1,\dots,N$$

for some $x_1,\dots,x_N\in\mathcal{X}$ such that $\mathcal{X}=\bigcup_{j=1}^N B_j$. Let $f(B_j):=\{f(x)\colon x\in B_j\}$. Since f is surjective, it also holds $\mathcal{Y}=\bigcup_{j=1}^N f(B_j)$. Consider $\tilde{B}_j:=B_{\varepsilon,e}(f(x_j))$ for $j=1,\dots,N$. By

assumption, it holds $e(f(x), f(x_j)) \leq b^{1/a} d^{1/a}(x, x_j) < \varepsilon$ for all $x \in B_j$, yielding $f(B_j) \subset \tilde{B}_j$. Thus it also holds $\mathcal{Y} = \bigcup_{j=1}^{N} \tilde{B}_j$ and it follows that $\tilde{B}_1, \ldots, \tilde{B}_N$ is an ε-cover with respect to e of $\mathcal{Y}$. This argument shows that *any* $\frac{\varepsilon^a}{b}$-cover (with respect to d) of $\mathcal{X}$ induces an ε-cover (with respect to e) of $\mathcal{Y}$ the same size, concluding the proof.

□

C.

Notation

Throughout this work, matrices are written in bold as $\boldsymbol{A}$ and vectors underlined as $\underline{Z}$. Measures in the Bayesian context (prior, posterior or likelihood) are usually denoted by capital letters, e.g. $P(d\theta|\underline{Z})$ and (in case of existence), the corresponding Lebesgue density in lower case, e.g. $p(\theta|\underline{Z})$. The used symbols are summarized in the following.

Functions and Operators

$\mathbb{1}_A$	indicator function for set A
$\|\cdot\|_{F,\infty}$	maximum Frobenius norm, see (7.3)
$\mathfrak{B}(k, \boldsymbol{\Phi})$	Bernstein polynomial expansion operator, see (7.4)
$b(\cdot\|j, k-j+1)$	density of Beta$(j, k-j+1)$ distribution, see (B.2)
$b_{\tau_l}^{\tau_r}(\cdot\|j, k-j+1)$	truncated Bernstein polynomial basis function, see (5.4)
δ_x	Dirac Delta function at x
E_1	exponential integral function, see (3.11)
$\mathrm{E}[X]$	expected value of random variable X
$\boldsymbol{f}(\omega), \boldsymbol{F}(d\omega)$	spectral density and spectral measure matrix, see (0.1)
$\Gamma(a), \tilde{\Gamma}_d(\eta)$	Gamma and complex multivariate Gamma function, see (B.14)
$\boldsymbol{\Gamma}(h)$	covariance function of stationary time series
$L_\mu(\boldsymbol{\Theta})$	Laplace transform of measure μ, see Section 2.1.1
$\mathrm{Var}[X]$	variance of random variable X
$\mathrm{vec}_{\mathbb{R}}$	stacking operator, see (4.9)

Matrices

$\|\boldsymbol{A}\|$	absolute value of determinant
$\boldsymbol{A} < \boldsymbol{B}$	$\boldsymbol{B} - \boldsymbol{A} \in \mathcal{S}_d^+$
$\|\boldsymbol{A}\|$	Frobenius norm, see (B.3)
$\|\boldsymbol{A}\|_\#$	for $\# \in \{1, 2, \infty, T\}$: matrix norms, see (B.4)–(B.7)
$\boldsymbol{A}^{1/2}$	Hermitian positive definite square root of $\boldsymbol{A} \in \mathcal{S}_d^+$, see (B.15)

$\boldsymbol{A}^T, \boldsymbol{A}^*$	transpose and Hermitian conjugate
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product, see (B.1)
$\mathcal{B}\boldsymbol{A}$	isomorphism from $\mathbb{C}^{d\times d}$ to $\mathbb{R}^{2d\times 2d}$ from (4.15)
$\det \boldsymbol{A}$	determinant
$\operatorname{etr} \boldsymbol{A}$	exponential trace, i.e. $\exp(\operatorname{tr} \boldsymbol{A})$
$\boldsymbol{F}_{nd}$	Fourier Transformation matrix, see Lemma 4.1
$\boldsymbol{I}_d$	$d \times d$ identity matrix
$\lambda_{\min}(\boldsymbol{A}), \lambda_{\max}(\boldsymbol{A})$	smallest and largest eigenvalue of Hermitian matrix $\boldsymbol{A}$
$\operatorname{tr} \boldsymbol{A}$	trace

Probability Distributions

$A\Gamma(\eta, \omega, \boldsymbol{\Sigma})$	$A\Gamma$ distribution, see (2.13)
$CN_d(\underline{0}, \boldsymbol{\Sigma})$	complex d-variate normal, see (4.6)
$\mathrm{CWish}_{d\times d}(\eta, \boldsymbol{\Sigma})$	complex Wishart distribution, see (1.12)
$\mathrm{CPoi}(C, \nu^*)$	Compound Poisson distribution, see Theorem 2.2
$\mathrm{CRM}(\nu)$	Completely Random Measure with Poisson mean measure ν, see (1.9)
$\mathrm{CRM}_{d\times d}(\nu)$	Completely Random Hpd Measure with Poisson mean measure ν, see (3.2)
$\mathrm{DP}(G)$	Dirichlet Process with base measure G
$\mathrm{Ga}(\alpha, \beta)$	Gamma distribution with $\alpha, \beta > 0$
$\mathrm{Ga}_{d\times d}(\alpha, \beta)$	Hpd Gamma distribution, see (2.7)
$N_d(\underline{\mu}, \boldsymbol{\Sigma})$	d-variate normal with mean $\underline{\mu} \in \mathbb{R}^d$ and variance $\boldsymbol{\Sigma} \in \mathcal{S}_d^+(\mathbb{R})$
$\mathrm{Poi}(C)$	Poisson distribution
$\mathrm{Unif}(\mathcal{X})$	Uniform distribution on $\mathcal{X}$
$\mathrm{Wish}^{-1}_{d\times d}(\eta, \boldsymbol{\Sigma})$	Inverse Wishart distribution, see (B.2)

Spaces and Sets

$\mathbb{B}(\mathcal{X})$	Borel sets on a topological space $\mathcal{X}$
$\mathbb{C}$	complex numbers
$\mathcal{D}_{d\times d}$	set of continuous $\mathcal{S}_d^+$-valued functions on $\mathcal{X}$
$\mathcal{L}_{\mathcal{X}}$	Lebesgue measure on Euclidean space $\mathcal{X}$
$\mathbb{L}^p$	measurable functions with Lebesgue integrable absolute pth power $(1 \leq p < \infty)$
$\mathcal{M}_{d\times d}$	set of finite Hpd measures on $\mathcal{X}$
$\mathbb{N}$	positive integers
$\mathbb{R}$	real numbers
$\mathcal{S}_d$	space of Hermitian $d \times d$ matrices
$\mathcal{S}_d^+$	cone of Hermitian positive definite $d \times d$ matrices
$\mathcal{S}_d^+(\mathbb{R})$	cone of symmetric positive definite $d \times d$ matrices
$\mathbb{S}_d$	set of Hermitian $d \times d$ matrices with unit trace, see (2.3)
$\mathbb{S}_d^+$	set of Hermitian positive definite $d \times d$ matrices with unit trace
$\mathbb{Z}$	integers

Miscellaneous

$a_n \lesssim b_n$	there exists a positive constant C such that $a_n \leq C b_n$ for all n
$d_{n,H}$	root average squared Hellinger distance, see (7.31)
$\bar{E}$	closure of a subset $E \subset \mathcal{X}$ of a topological space $\mathcal{X}$
$\#E$	cardinality of set E
$\Im z$	imaginary part of $z \in \mathbb{C}$
$N(\varepsilon, \mathcal{X}, d)$	ε-covering number, see (B.28)
$p(z) \propto f(z)$	there exists a positive constant C such that $p(z) = C f(z)$
$\Re z$	real part of $z \in \mathbb{C}$
$\operatorname{supp}(\mu)$	support of measure μ, see (2.11)

Bibliography

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation.

Akaike, H. (1973). Block Toeplitz matrix inversion. *SIAM Journal on Applied Mathematics*, 24(2):234–241.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (2003). Posterior consistency for semi-parametric regression problems. *Bernoulli*, 9(2):291–312.

Bai, J. and Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1):49–60.

Barron, A., Schervish, M. J., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2):536–561.

Barron, A. R. (1989). Uniformly powerful goodness of fit tests. *The Annals of Statistics*, pages 107–124.

Belitser, E., Serra, P., and van Zanten, H. (2015). Rate-optimal Bayesian intensity smoothing for inhomogeneous Poisson processes. *Journal of statistical planning and inference*, 166:24–35.

Bickel, P. and Kleijn, B. (2012). The semiparametric Bernstein–von Mises theorem. *The Annals of Statistics*, 40(1):206–237.

Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag New York.

Blackwell, D. (1973). Discreteness of Ferguson selections. *The Annals of Statistics*, 1(2):356–358.

Bobecka, K., Emilion, R., and Wesolowski, J. (2010). On matrix variate Dirichlet vectors. Technical report, Université d'Orléans.

Boehmke, B., Greenwell, B., Freels, J., and Gutierrez, R. (2018). *anomalyDetection: Implementation of Augmented Network Log Anomaly Detection Procedures*. R package version 0.2.5.

Böttcher, A. (2006). *Analysis of Toeplitz Operators*. Springer Monographs in Mathematics. Springer-Verlag Berlin Heidelberg, second edition.

Bradley, R. C. (2002). On positive spectral density functions. *Bernoulli*, 8(2):175–193.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*. Springer Series in

Statistics. Springer New York.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274.

Bru, M.-F. (1991). Wishart processes. *Journal of Theoretical Probability*, 4(4):725–751.

Bureau of Meteorology of the Australian Government (2018a). Climate glossary – Southern Oscillation Index (SOI). http://www.bom.gov.au/climate/glossary/soi.shtml. Accessed: 2018-05-15.

Bureau of Meteorology of the Australian Government (2018b). The three phases of the El Niño Southern Oscillation (ENSO). http://www.bom.gov.au/climate/enso/history/ln-2010-12/three-phases-of-ENSO.shtml. Accessed: 2018-05-15.

Cadonna, A., Kottas, A., and Prado, R. (2016). Bayesian spectral modeling for multiple time series. Technical report, Jack Baskin School of Engineering, Santa Cruz.

Cai, T. T., Ren, Z., and Zhou, H. H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probability Theory and Related Fields*, 156(1-2):101–143.

Castillo, I. (2014). On Bayesian supremum norm contraction rates. *The Annals of Statistics*, 42(5):2058–2091.

Choi, T. and Ramamoorthi, R. (2008). Remarks on consistency of posterior distributions. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, pages 170–186. Institute of Mathematical Statistics Collections.

Choi, T. and Schervish, M. J. (2007). On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis*, 98(10):1969–1987.

Choudhuri, N., Ghosal, S., and Roy, A. (2004a). Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059.

Choudhuri, N., Ghosal, S., and Roy, A. (2004b). Contiguity of the Whittle measure for a Gaussian time series. *Biometrika*, 91(1):211–218.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011). *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. CRC Press.

Contreras-Cristán, A., Gutiérrez-Peña, E., and Walker, S. G. (2006). A note on Whittle's likelihood. *Communications in Statistics Simulation and Computation*, 35(4):857–875.

Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., and Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational mathematics*, 5(1):329–359.

Cui, X., Gupta, A. K., and Nagar, D. K. (2005). Wilks' factorization of the complex matrix variate Dirichlet distributions. *Revista Matemática Complutense*, 18(2):315–328.

Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series. *Advances in Applied Probability*, 5(3):469–497.

Dettweiler, E. (1976). Infinitely divisible measures on the cone of an ordered locally convex vector space. *Annales scientifiques de l'Université de Clermont*, 61(14):11–17.

Devroye, L. (1986). Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265. ACM.

Diaconis, P. and Freedman, D. (1986). On the consistency of bayes estimates. *The Annals of Statistics*, pages 1–26.

Dunson, D. B., Xue, Y., and Carin, L. (2008). The matrix stick-breaking process: flexible Bayes meta-analysis. *Journal of the American Statistical Association*, 103(481):317–327.

Dzhaparidze, K. and Kotz, S. (2012). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer Series in Statistics. Springer New York.

Eddelbuettel, D. and Francois, R. (2011). Rcpp: seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Edwards, M. C., Meyer, R., and Christensen, N. (2015). Bayesian semiparametric power spectral density estimation with applications in gravitational wave data analysis. *Physical Review D*, 92(6).

Edwards, M. C., Meyer, R., and Christensen, N. (2017). Bayesian nonparametric spectral density estimation using B-spline priors. *Statistics and Computing*, pages 1–12.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230.

Freedman, D. (1999). Wald lecture: On the Bernstein-von Mises theorem with infinite-dimensional parameters. *The Annals of Statistics*, 27(4):1119–1141.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 27(1):143–158.

Ghosal, S., Ghosh, J. K., and Van Der Vaart, A. (2000). Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531.

Ghosal, S. and Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *The Annals of Statistics*, pages 2413–2429.

Ghosal, S. and Van Der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–223.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press.

Goodman, N. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of Mathematical Statistics*, 34(1):152–177.

Graczyk, P., Letac, G., and Massam, H. (2003). The complex Wishart distribution and the symmetric group. *The Annals of Statistics*, pages 287–309.

Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundations and Trends in*

Communications and Information Theory, 2(3):155–239.

Gröchenig, K. (2013). *Foundations of time-frequency analysis*. Springer Science & Business Media.

Gupta, A. and Nagar, D. (1999). *Matrix Variate Distributions*. Monographs and Surveys in Pure and Applied Mathematics. Taylor & Francis.

Gupta, Arjun K., D. K. N. and Bedoya, E. (2007). Properties of the complex matrix variate Dirichlet distribution. *Scientiae Mathematicae Japonicae*, 66(1):53–62.

Häfner, F. and Kirch, C. (2017). Moving Fourier analysis for locally stationary processes with the bootstrap in view. *Journal of Time Series Analysis*, 38(6):895–922.

Hannan, E. (1970). *Multiple Time Series*. A Wiley publication in applied statistics. Wiley.

Hannan, E. and Wahlberg, B. (1989). Convergence rates for inverse Toeplitz matrix forms. *Journal of Multivariate Analysis*, 31(1):127–135.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian nonparametrics*. Cambridge University Press.

Hoffmann, M., Rousseau, J., and Schmidt-Hieber, J. (2015). On adaptive posterior concentration rates. *The Annals of Statistics*, 43(5):2259–2295.

Ibragimov, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory of Probability & Its Applications*, 8(4):366–401.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.

James, L. F., Lijoi, A., and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, 36(1):76–97.

Kallenberg, O. (1997). *Foundations of modern probability*. Springer.

Kammler, D. W. (2000). *A First Course in Fourier Analysis*. Prentice Hall.

Kingman, J. (1992). *Poisson Processes*. Oxford Studies in Probability. Clarendon Press.

Kirch, C., Edwards, M. C., Meier, A., and Meyer, R. (2017). Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis. *Preprint (arXiv:1701.04846)*.

Kleijn, B. and Van der Vaart, A. (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6:354–381.

Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358.

LePage, R. (1981). Multidimensional infinitely divisidle variables and processes part II. In *Probability in Banach Spaces III*, pages 279–284. Springer.

Lévy, P. (1948). The arithmetical character of the wishart distribution. *Mathematical Proceedings*

of the Cambridge Philosophical Society, 44(2):295–297.

Li, Z. and Krafty, R. T. (2018). Adaptive Bayesian time-frequency analysis of multivariate time series. *Journal of the American Statistical Association*.

Lijoi, A. (2010). Models beyond the Dirichlet process. In Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G., editors, *Bayesian Nonparametrics*. Camb. Ser. Stat. Probab. Math.

Lorentz, G. G. (2012). *Bernstein polynomials*. AMS Chelsea Publishing.

Magnus, J. R. and Neudecker, H. (2007). *Matrix differential calculus with applications in statistics and econometrics*. Wiley, third edition.

Maiwald, D. and Kraus, D. (2000). Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices. *IEE Proceedings-Radar, Sonar and Navigation*, 147(4):162–168.

Marshall, A. W., Olkin, I., and Arnold, B. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer Series in Statistics. Springer-Verlag New York, second edition.

Mathai, A. (1997). *Jacobians of Matrix Transformations and Functions of Matrix Argument*. World Scientific Publishing.

McCulloch, R. E. and Tsay, R. S. (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis*, 15(2):235–250.

Mittelbach, M., Matthiesen, B., and Jorswieck, E. A. (2012). Sampling uniformly from the set of positive definite matrices with trace constraint. *IEEE Transactions on Signal Processing*, 60(5):2167–2179.

Muliere, P. and Tardella, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *Canadian Journal of Statistics*, 26(2):283–297.

Nagar, D. K. and Gupta, A. K. (2009). *Contributions to complex matrix variate distributions theory*. Universidad de Antioquia.

Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9(4):307–333.

Pardo, L. (2005). *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC.

Pérez-Abreu, V. and Rosiński, J. (2007). Representation of infinitely divisible distributions on cones. *Journal of Theoretical Probability*, 20(3):535–544.

Pérez-Abreu, V. and Stelzer, R. (2014). Infinitely divisible multivariate and matrix gamma distributions. *Journal of Multivariate Analysis*, 130:155–175.

Petrone, S. (1999). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26(3):373–393.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Regazzini, E., Lijoi, A., and Prünster, I. (2003). Distributional results for means of normalized

random measures with independent increments. *The Annals of Statistics*, pages 560–585.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Robertson, J. B. and Rosenberg, M. (1968). The decomposition of matrix-valued measures. *The Michigan Mathematical Journal*, 15(3):353–368.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The annals of Statistics*, pages 1048–1072.

Rosen, O. and Stoffer, D. S. (2007). Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*, 94(2).

Rosenberg, M. (1974). Operators as spectral integrals of operator-valued functions from the study of multivariate stationary stochastic processes. *Journal of Multivariate Analysis*, 4(2):166–209.

Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Springer.

Rousseau, J. (2016). On the frequentist properties of bayesian nonparametric methods. *Annual Review of Statistics and Its Application*, 3:211–231.

Rousseau, J., Chopin, N., and Liseo, B. (2012). Bayesian nonparametric estimation of the spectral density of a long or intermediate memory gaussian process. *The Annals of Statistics*, 40(2):964–995.

Roychowdhury, A. and Kulis, B. (2015). Gamma processes, stick-breaking and variational inference. In *Proceedings of The 18th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 800–808.

Rudin, W. (1962). *Fourier analysis on groups*. Interscience Publishers (a division of John Wiley and Sons), New York–London.

Salehi, H. (1968). The Hellinger square-integrability of matrix-valued measures with respect to a non-negative hermitian measure. *Arkiv för Matematik*, 7(4):299–303.

Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge university press.

Schwartz, L. (1965). On Bayes procedures. *Probability Theory and Related Fields*, 4(1):10–26.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica sinica*, pages 639–650.

Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *The Annals of Statistics*, pages 687–714.

Shumway, R. H. and Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

Stoffer, D. (2017). *astsa: Applied Statistical Time Series Analysis*. R package version 1.8.

Sykulski, A. M., Olhede, S. C., and Lilly, J. M. (2016). The de-biased Whittle likelihood for second-order stationary stochastic processes. *Preprint (arXiv:1605.06718)*.

Szabó, B., van der Vaart, A., and van Zanten, J. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *The Annals of Statistics*, 43(4):1391–1428.

Taylor, S. and Letham, B. (2017). *prophet: Automatic Forecasting Procedure*. R package version 0.2.1.

Troskie, C. (1967). Noncentral multivariate Dirichlet distributions. *South African Statistical Journal*, 1(1):21–32.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Vaart, A. W. and van Zanten, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *The Annals of Statistics*, 37(5B):2655–2675.

Widom, H. (1974). Asymptotic behavior of block Toeplitz matrices and determinants. *Advances in Mathematics*, 13(3):284–322.

Widom, H. (1975). On the limit of block Toeplitz determinants. *Proceedings of the American Mathematical Society*, pages 167–173.

Widom, H. (1976). Asymptotic behavior of block Toeplitz matrices and determinants II. *Advances in Mathematics*, 21(1):1–29.

Wolpert, R. L., Clyde, M. A., and Tu, C. (2011). Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels. *The Annals of Statistics*, 39(4):1916–1962.

Wolpert, R. L. and Ickstadt, K. (1998). Simulation of Lévy random fields. In *Practical nonparametric and semiparametric Bayesian statistics*, pages 227–242. Springer.

Wu, Y. and Ghosal, S. (2010). The L1-consistency of Dirichlet mixtures in multivariate Bayesian density estimation. *Journal of Multivariate Analysis*, 101(10):2411–2419.

Yoo, W. W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *The Annals of Statistics*, 44(3):1069–1102.

Zhang, S. (2016). Adaptive spectral estimation for nonstationary multivariate time series. *Computational Statistics & Data Analysis*, 103:330–349.

Zhang, Z., Wang, D., Dai, G., and Jordan, M. I. (2014). Matrix-variate Dirichlet process priors with applications. *Bayesian Analysis*, 9(2):259–286.

Zheng, Y., Zhu, J., and Roy, A. (2009). Nonparametric Bayesian inference for the spectral density function of a random field. *Biometrika*, 97(1):238–245.