

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Lietzén, Niko; Viitasaari, Lauri; Ilmonen, Pauliina

## Modeling temporally uncorrelated components of complex-valued stationary processes

*Published in:*  
Modern Stochastics: Theory and Applications

*DOI:*  
[10.15559/21-VMSTA190](https://doi.org/10.15559/21-VMSTA190)

Published: 01/11/2021

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
Lietzén, N., Viitasaari, L., & Ilmonen, P. (2021). Modeling temporally uncorrelated components of complex-valued stationary processes. *Modern Stochastics: Theory and Applications*, 8(4), 475-508.  
<https://doi.org/10.15559/21-VMSTA190>

# Modeling temporally uncorrelated components of complex-valued stationary processes

Niko Lietzén<sup>a,\*</sup>, Lauri Viitasaari<sup>b</sup>, Pauliina Ilmonen<sup>c</sup>

<sup>a</sup>*Department of Mathematics and Statistics, University of Turku, P.O. Box 772, 00074 CGI, Finland*

<sup>b</sup>*Department of Mathematics, Uppsala University, Sweden*

<sup>c</sup>*Department of Mathematics and Systems Analysis, Aalto University School of Science, Finland*

[niko.lietzen@utu.fi](mailto:niko.lietzen@utu.fi) (N. Lietzén), [lauri.viitasaari@math.uu.se](mailto:lauri.viitasaari@math.uu.se) (L. Viitasaari),  
[pauliina.ilmonen@aalto.fi](mailto:pauliina.ilmonen@aalto.fi) (P. Ilmonen)

Received: 4 June 2021, Revised: 6 October 2021, Accepted: 6 October 2021,  
Published online: 10 November 2021

**Abstract** A complex-valued linear mixture model is considered for discrete weakly stationary processes. Latent components of interest are recovered, which underwent a linear mixing. Asymptotic properties are studied of a classical unmixing estimator which is based on simultaneous diagonalization of the covariance matrix and an autocovariance matrix with lag  $\tau$ . The main contributions are asymptotic results that can be applied to a large class of processes. In related literature, the processes are typically assumed to have weak correlations. This class is extended, and the unmixing estimator is considered under stronger dependency structures. In particular, the asymptotic behavior of the unmixing estimator is estimated for both long- and short-range dependent complex-valued processes. Consequently, this theory covers unmixing estimators that converge slower than the usual  $\sqrt{T}$  and unmixing estimators that produce non-Gaussian asymptotic distributions. The presented methodology is a powerful preprocessing tool and highly applicable in several fields of statistics.

**Keywords** Asymptotic theory, blind source separation, long-range dependency, multivariate analysis, noncentral limit theorems

**2010 MSC** 62H12, 60F05, 60G15, 60G10, 94A12, 94A08

---

\*Corresponding author.

## 1 Introduction

In modern statistics, there is an increasing demand for theoretically solid methodology which can be applied beyond standard real-valued data, in the realm of more exotic data structures. In this paper, we consider a complex-valued linear mixture model based on temporally uncorrelated components. We consider the problem under discrete weakly stationary processes. We aim to find latent processes of interest when only linear mixtures of them are observable. The recovery of the latent processes is referred to as the unmixing procedure. The properties of the recovered processes themselves can be the source of interest, or alternatively, the approach can be used to reduce multivariate models into several univariate models, which can simplify the modeling burden.

In the context of linear mixture models, the assumption of uncorrelated components or stronger conditions that imply uncorrelated components is considered to be natural in several applications, for example, in finance and signal processing, see the article [14] and the book [10]. Applications where the observed time series are naturally complex-valued are frequent in, e.g., signal processing and biomedical applications, see [19, 41]. In such applications, the interest can lie in the shapes of the unobservable signals, such as the shapes of three-dimensional image valued signals, which correspond to different parts of the brain under different activities. In the signal processing literature, the problem is often referred to as the blind source separation (BSS) problem. In the BSS literature, it has been argued that discarding the special complex-valued structure can lead to loss of information, see [1]. Thus, it is often beneficial to natively work with complex-valued signals in such applications. We wish to emphasize that our  $\mathbb{C}^d$ -valued model does not directly correspond to existing  $\mathbb{R}^{2d}$  valued models, e.g., the one in [23]. In  $\mathbb{R}^{2d}$  valued BSS, it is assumed that all  $2d$  components are uncorrelated. In our model, the real part and the corresponding complex part are allowed to be correlated. For a collection of BSS applications, see [10].

In parallel to the signal processing community, linear mixture models, with similar model assumptions as in BSS literature, have recently received notable attention in finance, see for example [14, 22]. In financial applications, the term blind source separation is rarely used, and usually more descriptive model naming conventions are utilized. Note that our complex-valued approach is highly applicable in many real-valued financial applications. In our approach, we assume little concerning the relationship between the real and imaginary parts of a single component. Thus, from the real-valued perspective, the problem can be equivalently considered as modeling real-valued temporally uncorrelated pairs, where the elements contained in a single pair are not necessarily uncorrelated. However, transforming this complex-valued problem into a real-valued counterpart is often unfeasible, as it usually introduces additional constraints that would be otherwise embedded within the complex-valued formulation. Once the temporally uncorrelated components are recovered, one can then, for example, model volatilities bivariate (or univariate) or assess risk by modeling tail behavior with the tools of bivariate (or univariate) extreme value theory. Moreover, our approach is natural in applications where a single observation is vector valued, that is, the observations have both a magnitude and a direction, e.g., modeling the latent components of wind at an airport.

The main focus in this paper is on asymptotic behavior of a classical unmixing estimator. We consider an algorithm that is identical to the so-called Algorithm for Multiple Unknown Signals Extraction (AMUSE), [35]. In the financial side, asymptotic properties of an alternative approach, with slightly differing model assumptions compared to our approach, are given in, e.g., [14]. In the context of real-valued BSS, asymptotics have been considered in, e.g., [23–25]. Compared to the above mentioned approaches, we consider a substantially wider class of processes. In particular, we analyze the asymptotic behavior of the corresponding estimators under both long- and short-range dependent complex-valued processes. Furthermore, we take a semiparametric approach, that is, our theory is not limited to specific parametric family of distributions. As a pinnacle of the novel theory presented in this paper, we consider square-integrable processes without summable autocovariance structures, which causes the limiting distribution of the corresponding unmixing estimator to be non-Gaussian. Instead of using the classical central limit theorem, we take a modern approach and utilize general central limit theorem type results that are also applicable for processes with strong temporal dependencies. Moreover, we consider convergence rates that differ from the usual  $\sqrt{T}$ . We wish to emphasize that modeling long-range dependent processes, that is, processes without summable autocovariance structures, are of paramount importance in many financial applications.

The paper is structured as follows. In Section 2, we recall some basic theory regarding complex-valued random variables. In Section 3, we present the temporally uncorrelated components mixing model. In Section 4, we formulate the estimation procedure and study the asymptotic properties of the estimators. In Section 5, we consider a data example involving photographs, which can be presented as complex-valued time series. The proofs for the theoretical results are presented in the supplementary Appendix.

## 2 Random variables in the complex plane

Throughout, let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a common probability space and let  $\mathbf{z}_\cdot := (\mathbf{z}_t)_{t \in \mathbb{N}}$  be a collection of random variables  $\mathbf{z}_t : \Omega \rightarrow \mathbb{C}^d$ ,  $t \in \mathbb{N} = \{1, 2, \dots\}$ , where the dimension parameter  $d$  is a finite constant. Furthermore, let  $z_t^{(k)} = \text{pr}_k \circ \mathbf{z}_t$ , where  $\text{pr}_k$  is a projection to the  $k$ th complex coordinate. We refer to the process  $z_\cdot^{(k)} := (z_t^{(k)})_{t \in \mathbb{N}}$  as the  $k$ th component of the  $\mathbb{C}^d$ -valued stochastic process  $\mathbf{z}_\cdot$ . Note that the components can be expressed in the form  $z_t^{(k)} = a_t^{(k)} + ib_t^{(k)}$ , where  $a_t^{(k)}, b_t^{(k)}$  are real-valued  $\forall k \in \{1, 2, \dots, d\}$  and  $i$  is the imaginary unit. We denote the complex conjugate of  $\mathbf{z}_t$  as  $\mathbf{z}_t^*$  and we denote the conjugate transpose of  $\mathbf{z}_t$  as  $\mathbf{z}_t^H := (\mathbf{z}_t^*)^\top$ . Furthermore, we use  $\mathcal{B}(\mathbb{C}^d)$  and  $\mathcal{B}(\mathbb{R}^{2d})$  to denote the Borel-sigma algebras on  $\mathbb{C}^d$  and  $\mathbb{R}^{2d}$ , respectively.

We use the following notation for the multivariate mean and the (unsymmetrized) autocovariance matrix with lag  $\tau \in \{0\} \cup \mathbb{N}$ :

$$\boldsymbol{\mu}_{\mathbf{z}_t} = \mathbb{E}[\mathbf{z}_t] \in \mathbb{C}^d, \quad \ddot{\mathbf{S}}_\tau[\mathbf{z}_t] = \mathbb{E}\left[(\mathbf{z}_t - \boldsymbol{\mu}_{\mathbf{z}_t})(\mathbf{z}_{t+\tau} - \boldsymbol{\mu}_{\mathbf{z}_{t+\tau}})^H\right] \in \mathbb{C}^{d \times d}.$$

Furthermore, we use the notation  $\mathbf{S}_\tau[\mathbf{z}_t] = \frac{1}{2}[\ddot{\mathbf{S}}_\tau[\mathbf{z}_t] + (\ddot{\mathbf{S}}_\tau[\mathbf{z}_t])^H] \in \mathbb{C}^{d \times d}$  for the symmetrized autocovariance matrix with lag  $\tau \in \{0\} \cup \mathbb{N}$ .

In the case of univariate stochastic processes, we use  $\ddot{S}_\tau[x_t, y_s]$  to denote  $\ddot{S}_\tau[x_t, y_s] = \mathbb{E}[(x_t - \mu_{x_t})(y_s + \tau - \mu_{y_{s+\tau}})^*]$ . Note that the unsymmetrized autocovariance matrix  $\ddot{S}_\tau$  is not necessarily conjugate symmetric, i.e., Hermite symmetric or Hermitian, and it can have complex-valued diagonal entries. In contrast, the symmetrized autocovariance matrix  $S_\tau$  is always conjugate symmetric, that is,  $S_\tau = S_\tau^H$ , and the diagonal entries of the symmetrized autocovariance matrix are by definition real-valued.

In the literature, the definition of stationarity for complex-valued stochastic processes varies. In this paper, we use the following definition.

**Definition 1.** The  $\mathbb{C}^d$ -valued process  $\mathbf{z} := (\mathbf{z}_t)_{t \in \mathbb{N}}$  is weakly stationary if the components of  $\mathbf{z}$  are square-integrable and if for any  $\tau \in \{0\} \cup \mathbb{N}$  and for any pair  $(u, v) \in \mathbb{N} \times \mathbb{N}$ , we have that  $\mathbb{E}[\mathbf{z}_u] = \mathbb{E}[\mathbf{z}_v]$  and  $\ddot{S}_\tau[\mathbf{z}_u] = \ddot{S}_\tau[\mathbf{z}_v]$ .

Let  $\mathbf{Z} := (\mathbf{Z}_j)_{j \in \mathcal{T}}$ ,  $\mathcal{T} = \{1, 2, \dots, T\}$ , be a sampled process generated by  $\mathbf{z}$ , where  $\mathbf{Z}_j$  is  $\mathbb{C}^d$ -valued for every  $j \in \mathcal{T}$ . We use the following classical finite sample estimators for the mean vector and the autocovariance matrix with lag  $\tau \in \{0, 1, \dots, T-1\}$ :

$$\hat{\mu}[\mathbf{Z}] = \frac{1}{T} \sum_{j=1}^T \mathbf{Z}_j, \quad \tilde{S}_\tau[\mathbf{Z}] = \frac{1}{T-\tau} \sum_{j=1}^{T-\tau} (\mathbf{Z}_j - \hat{\mu}[\mathbf{Z}]) (\mathbf{Z}_{j+\tau} - \hat{\mu}[\mathbf{Z}])^H,$$

where for  $\tau = 0$ , the scaling used in  $\tilde{S}_\tau$  is  $1/(T-1)$ . Furthermore, the symmetrized autocovariance matrix estimator with lag  $\tau \in \{0, 1, \dots, T-1\}$  is defined as  $\hat{S}_\tau[\mathbf{Z}] = \frac{1}{2}[\tilde{S}_\tau[\mathbf{Z}] + (\tilde{S}_\tau[\mathbf{Z}])^H]$ . Note that the symmetrized autocovariance matrix estimator is always conjugate symmetric.

Let  $\mathbf{y}_1$  and  $\mathbf{y}_2$  be  $\mathbb{R}^d$ -valued random vectors such that the concatenated random vector, that is,  $(\mathbf{y}_1^\top \ \mathbf{y}_2^\top)^\top$ , follows a  $\mathbb{R}^{2d}$ -valued Gaussian distribution. Then, the  $\mathbb{C}^d$ -valued random vector  $\mathbf{y} = \mathbf{y}_1 + i\mathbf{y}_2$  follows the  $\mathbb{C}^d$ -valued Gaussian distribution with the location parameter  $\mu_y$ , the covariance matrix  $\Sigma_y$  and the relation matrix  $\mathbf{P}_y$ , defined as

$$\begin{aligned} \mu_y &= \mathbb{E}[\mathbf{y}] = \mathbb{E}[\mathbf{y}_1] + i\mathbb{E}[\mathbf{y}_2] \in \mathbb{C}^d, \quad \Sigma_y = \mathbb{E}[(\mathbf{y} - \mu_y)(\mathbf{y} - \mu_y)^H] \in \mathbb{C}^{d \times d}, \\ \mathbf{P}_y &= \mathbb{E}[(\mathbf{y} - \mu_y)(\mathbf{y} - \mu_y)^\top] \in \mathbb{C}^{d \times d}. \end{aligned}$$

We distinguish between complex- and real-valued Gaussian distributions with the number of given parameters —  $\mathcal{N}_d(\mu_y, \Sigma)$  denotes a  $d$ -variate real-valued Gaussian distribution and  $\mathcal{N}_d(\mu_y, \Sigma, \mathbf{P})$  denotes a  $d$ -variate complex-valued Gaussian distribution.

The classical central limit theorem (CLT) for independent and identically distributed complex-valued random variables is given as follows. Let  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$  be a collection of i.i.d.  $\mathbb{C}^d$ -vectors with square-integrable components, mean  $\mu_x$ , covariance matrix  $\Sigma$  and relation matrix  $\mathbf{P}$ . Then,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (\mathbf{x}_j - \mu_x) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbf{x} \sim \mathcal{N}_d(\mathbf{0}, \Sigma, \mathbf{P}),$$

where  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution. We can harness the rich literature on real-valued limiting theorems via the following lemma.

**Lemma 1.** *Let  $\{\mathbf{v}_n\}_{n \in \mathbb{N}}$  be a collection of  $\mathbb{R}^{2d}$ -valued random vectors  $\mathbf{v}_j^\top = \begin{pmatrix} \mathbf{x}_j^\top & \mathbf{y}_j^\top \end{pmatrix}$ , where  $\mathbf{x}_j, \mathbf{y}_j$  are  $\mathbb{R}^d$ -valued for every  $j \in \mathbb{N}$ , and let  $\mathbf{v}^\top = (\mathbf{x}^\top \ \mathbf{y}^\top)$ . Then,  $\mathbf{v}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbf{v}$  if and only if  $\mathbf{x}_n + i\mathbf{y}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbf{x} + i\mathbf{y}$ .*

Lemma 1 can be, for example, applied to Gaussian vectors in a straightforward manner, see Corollary 1 in the supplementary Appendix.

In our work, we utilize Lemma 1 in order to apply real-valued results for complex-valued scenarios, but we wish to emphasize that it also works in the other way — the complex-valued asymptotic distribution automatically grants the corresponding real-valued limiting distribution. For additional details regarding complex-valued statistics, see, e.g., [13, 15, 31].

### 3 Temporally uncorrelated components model

In this section, we consider a linear temporally uncorrelated components model for discrete time complex-valued processes. We define the following version of the model.

**Definition 2.** The  $\mathbb{C}^d$ -process  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  follows the temporally uncorrelated components mixing model, if

$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t + \boldsymbol{\mu}_{\mathbf{x}}, \quad \forall t \in \mathbb{N},$$

where  $\mathbf{A}$  is a constant nonsingular  $\mathbb{C}^{d \times d}$ -matrix,  $\boldsymbol{\mu}_{\mathbf{x}} \in \mathbb{C}^d$  is a constant location parameter and  $\mathbf{z}$  is a  $\mathbb{C}^d$ -process with components having continuous marginal distributions. In addition, the process  $\mathbf{z}$  satisfies the following four conditions for every  $t, \kappa \in \mathbb{N}$ ,

- (1)  $\boldsymbol{\mu}_{\mathbf{z}_t} = \mathbf{0}$ ,      (2)  $\mathbf{S}_0[\mathbf{z}_t] = \mathbf{I}_d$ ,
- (3)  $\check{\mathbf{S}}_\kappa[\mathbf{z}_t]$  is finite and depends only on  $\kappa$ ,
- (4)  $\exists \tau \in \mathbb{N} : \mathbf{S}_\tau[\mathbf{z}_t] = \boldsymbol{\Lambda}_\tau = \text{diag}\left(\lambda_\tau^{(1)}, \dots, \lambda_\tau^{(d)}\right)$ ,

such that  $+\infty > \lambda_\tau^{(1)} > \lambda_\tau^{(2)} > \dots > \lambda_\tau^{(d)} > -\infty$ .

To improve the fluency of the paper, from hereon, the term mixing model is used to refer to the temporally uncorrelated components mixing model. The conditions (1)–(4) of Definition 2 imply that the latent process  $\mathbf{z}$  is weakly stationary in the sense of Definition 1. Note that under these model assumptions the concatenated  $\mathbb{R}^{2d}$ -process  $(\text{Re}(\mathbf{z}_t^\top) \ \text{Im}(\mathbf{z}_t^\top))^\top$  is not necessarily weakly stationary in the classical real-valued sense. One could also require that the concatenated vector is stationary, which can be done by adding to the model a condition involving the complex-valued relation matrix, defined in Section 2. However, adding the extra condition would also fix the relationship, up to heterogeneous sign-changes, between the real and imaginary parts inside a single component. In this paper, we do not wish to make assumptions regarding the complex phase of a single component. Consequently, many of the

following results involve a so-called phase-shift matrix  $\mathbf{J}$ , that is, a complex-valued diagonal matrix with diagonal entries of the form  $\exp(i\theta_1), \dots, \exp(i\theta_d)$ . A phase-shift matrix  $\mathbf{J}$  is by definition unitary, i.e.,  $\mathbf{J}\mathbf{J}^H = \mathbf{J}^H\mathbf{J} = \mathbf{I}_d$ . Note that all possible phase-shift matrices can be constructed by considering phases on some suitable  $2\pi$ -length interval, e.g., the usual  $[0, 2\pi)$ .

Condition (3) is included in Definition 2 to ensure that we can apply limiting results that require the traditional real-valued weak stationarity. Since Condition (3) is solely used to guarantee stationarity, the only requirements for  $\check{\mathbf{S}}_\kappa[\mathbf{z}_t]$  are that it has finite elements and that it is invariant with respect to the time parameter  $t$ .

For a process  $\mathbf{x}$ , that follows the mixing model, the corresponding unmixing problem is to uncover the latent process  $\mathbf{z}$ , by using only the information contained in the observable process  $\mathbf{x}$ .

**Definition 3.** Let  $\mathbf{x}_\cdot := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2, such that condition (4) holds for some fixed  $\tau$ . The affine functional  $g : \mathbf{a} \mapsto \mathbf{\Gamma}(\mathbf{a} - \boldsymbol{\mu}) : \mathbb{C}^d \rightarrow \mathbb{C}^d$  is a solution to the corresponding unmixing problem if the process  $g \circ \mathbf{x}_\cdot$  satisfies conditions (1)–(3) and condition (4) is satisfied with the fixed  $\tau$ .

Given a process  $\mathbf{x}_\cdot$  that follows the mixing model and a corresponding solution  $g : \mathbf{a} \mapsto \mathbf{\Gamma}(\mathbf{a} - \boldsymbol{\mu})$ , we refer to  $\mathbf{A}$ , see Definition 2, as the mixing matrix and we refer to  $\mathbf{\Gamma}$  as the unmixing matrix. The objective of the unmixing matrix is to reverse the effects of the mixing matrix. However, under our model assumptions, we cannot uniquely determine the relationship between the mixing and the unmixing matrix. The relationship between the two matrices is  $\mathbf{\Gamma}\mathbf{A} = \mathbf{J}$ , where  $\mathbf{J} \in \mathbb{C}^{d \times d}$  is some phase-shift matrix, see Lemma 3 in the supplementary Appendix.

Under our assumptions, we cannot distinguish between solutions that contain unmixing matrices that are a phase-shift away from each other. Thus, we say that two solutions  $g_1 : \mathbf{a} \mapsto \mathbf{\Gamma}_1(\mathbf{a} - \boldsymbol{\mu}_1)$  and  $g_2 : \mathbf{a} \mapsto \mathbf{\Gamma}_2(\mathbf{a} - \boldsymbol{\mu}_2)$  are equivalent, if there exists a phase-shift matrix  $\mathbf{J}$  such that  $\mathbf{\Gamma}_1 = \mathbf{\Gamma}_2\mathbf{J}$ .

We next provide a solution procedure for the unmixing problem. Recall that in the mixing model we assume that the mixing matrix  $\mathbf{A}$  and the latent process  $\mathbf{z}_\cdot$  are unobservable. Therefore, the solution procedure relies solely on statistics of the observable process  $\mathbf{x}_\cdot$ .

**Theorem 1.** Let  $\mathbf{x}_\cdot := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2. The functional  $g : \mathbf{a} \mapsto \mathbf{\Gamma}(\mathbf{a} - \boldsymbol{\mu}) : \mathbb{C}^d \rightarrow \mathbb{C}^d$  is a solution to the corresponding unmixing problem if and only if the eigenvector equation

$$(\mathbf{S}_0[\mathbf{x}_t])^{-1} \mathbf{S}_\tau[\mathbf{x}_t] \mathbf{\Gamma}^H = \mathbf{\Gamma}^H \boldsymbol{\Lambda}_\tau, \quad \forall t \in \mathbb{N}, \quad (1)$$

and the scaling equation

$$\mathbf{\Gamma} \mathbf{S}_0[\mathbf{x}_t] \mathbf{\Gamma}^H = \mathbf{I}_d, \quad \forall t \in \mathbb{N}, \quad (2)$$

are satisfied.

One can apply Theorem 1 to find solutions by first estimating the eigenvalues and eigenvectors of  $(\mathbf{S}_0[\mathbf{x}_t])^{-1} \mathbf{S}_\tau[\mathbf{x}_t]$  and then scaling the eigenvectors so that the scaling equation is satisfied. In addition, see Corollary 2 in the supplementary Appendix.

#### 4 Estimation and asymptotic properties

The algorithm for multiple unknown signals extraction (AMUSE), [35], is a widely applied blind source separation unmixing procedure. AMUSE and the corresponding asymptotic properties have been previously studied in the real-valued case, see, e.g., [23]. Here, we adopt the estimation part of the AMUSE algorithm. However, our underlying model assumptions are not identical to the ones in [23] and [35]. In this section, we consider the consistency and the limiting distribution of the corresponding unmixing estimator under complex-valued long-range and short-range dependent processes, significantly extending the results given in [23].

**Definition 4.** Let  $\mathbf{x}_\cdot := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 and let  $\mathbf{X}$  be a  $\mathbb{C}^{T \times d}$ -valued,  $1 \leq d < T < \infty$ , sampled stochastic process generated by  $\mathbf{x}_\cdot$ . Let  $\hat{\boldsymbol{\mu}} := \hat{\boldsymbol{\mu}}[\mathbf{X}]$  and let  $\mathbf{1}_T$  be a  $\mathbb{R}^T$ -vector full of ones. The mapping  $\hat{\mathbf{g}} : \mathbf{C} \mapsto (\mathbf{C} - \mathbf{1}_T \hat{\boldsymbol{\mu}}^\top) \hat{\mathbf{F}}^\top : \mathbb{C}^{T \times d} \rightarrow \mathbb{C}^{T \times d}$  is a solution to the finite sample unmixing problem, if  $\hat{\lambda}_\tau^{(1)} \geq \dots \geq \hat{\lambda}_\tau^{(d)}$ , such that

$$\hat{\mathbf{F}} \hat{\mathbf{S}}_0[\mathbf{X}] \hat{\mathbf{F}}^H = \mathbf{I}_d \quad \text{and} \quad \hat{\mathbf{F}} \hat{\mathbf{S}}_\tau[\mathbf{X}] \hat{\mathbf{F}}^H = \hat{\boldsymbol{\Lambda}}_\tau = \text{diag}(\hat{\lambda}_\tau^{(1)}, \dots, \hat{\lambda}_\tau^{(d)}).$$

Recall that  $\hat{\mathbf{S}}_\tau$  is the symmetrized autocovariance matrix, and it is, by definition, conjugate symmetric. In addition, recall that the diagonal entries and the eigenvalues of a conjugate symmetric matrix are, again by definition, real-valued. Hereby, the ordering of  $\hat{\lambda}_\tau^{(j)}$  is always well-defined. The diagonal elements (eigenvalues)  $\hat{\lambda}_\tau^{(j)}$  can be seen as finite sample estimates of  $\lambda_\tau^{(j)}$ . Even though the population eigenvalues  $\lambda_\tau^{(j)}$  are assumed to be distinct, it is possible, due to, e.g., limited computational accuracy, that ties occur among the estimates  $\hat{\lambda}_\tau^{(j)}$ . Hereby, in the finite sample case, strict inequalities are not imposed on the estimates  $\hat{\lambda}_\tau^{(j)}$  in Definition 4.

Using Lemma 4 of the supplementary Appendix, we can straightforwardly implement the unmixing procedure. The first step is to calculate  $\hat{\mathbf{S}}_0 := \hat{\mathbf{S}}_0[\mathbf{X}]$ , from a realization  $\mathbf{X}$ , and the corresponding conjugate symmetric inverse square root  $\hat{\boldsymbol{\Sigma}}_0 := \hat{\mathbf{S}}_0^{-1/2}$ . Recall that, by assumption, the components of  $\mathbf{z}_\cdot$  have continuous marginal distributions and the mixing matrix is nonsingular. Thus, the eigenvalues of covariance matrix estimates are always real-valued and almost surely positive. Hereby, the matrix  $\hat{\boldsymbol{\Sigma}}_0$  can be obtained by estimating the eigendecomposition of  $\hat{\mathbf{S}}_0$ . The next step is to choose a lag-parameter  $\tau$  and estimate the eigendecomposition  $\hat{\mathbf{S}}_\tau[\mathbf{X} \hat{\boldsymbol{\Sigma}}_0^\top] = \hat{\boldsymbol{\Sigma}}_0 \hat{\mathbf{S}}_\tau[\mathbf{X}] \hat{\boldsymbol{\Sigma}}_0 = \hat{\mathbf{V}} \hat{\boldsymbol{\Lambda}}_\tau \hat{\mathbf{V}}^H$ . The covariance matrix and autocovariance matrix estimators are affine equivariant in the sense that  $\hat{\mathbf{S}}_j[\mathbf{X} \mathbf{C}^\top] = \mathbf{C} \hat{\mathbf{S}}_j[\mathbf{X}] \mathbf{C}^H$ ,  $j \in \{0, \tau\}$ , for all nonsingular  $\mathbb{C}^{d \times d}$ -matrices  $\mathbf{C}$ . An estimate for the latent process can then be found via the mapping  $(\mathbf{X} - \mathbf{1}_T \hat{\boldsymbol{\mu}}^\top)(\hat{\mathbf{V}}^H \hat{\boldsymbol{\Sigma}}_0)^\top$ , where  $\hat{\boldsymbol{\mu}}$  is the sample mean vector of  $\mathbf{X}$ .

In practice, one should choose the lag parameter  $\tau \neq 0$  such that the diagonal elements of  $\hat{\boldsymbol{\Lambda}}_\tau$  are as distinct as possible. Strategies for choosing  $\tau$  are discussed, e.g., in [9]. From a computational perspective, the estimation procedure is relatively fast. Hereby, in practice, one should try several different values of  $\tau$  and study  $\hat{\boldsymbol{\Lambda}}_\tau$ .

We emphasize that one can apply the theory of this paper to other estimators as well. That is, under minor model assumptions, the estimators  $\hat{\mathbf{S}}_0$ ,  $\hat{\mathbf{S}}_\tau$  can be re-



placed with any matrix-valued estimators that have the so-called complex-valued affine equivariance property, see [16].

The conditions given in Definition 4 remain true if we replace  $\hat{\Gamma}$  with  $\mathbf{J}\hat{\Gamma}$ , where  $\mathbf{J} \in \mathbb{C}^{d \times d}$  can be any phase-shift matrix. Thus, as in the population level, we say that the estimates  $\hat{\Gamma}_1$  and  $\hat{\Gamma}_2$  are equivalent if  $\hat{\Gamma}_1 = \mathbf{J}\hat{\Gamma}_2$  for some phase-shift matrix  $\mathbf{J}$ .

Justified by the affine invariance property given in supplementary Appendix Lemmas 5 and 6, we can, without loss of generality, derive the rest of the theory under the assumption of trivial mixing, that is, in the case when the mixing matrix is  $\mathbf{A} = \mathbf{I}_d$ . See also the beginning of the Proof of Lemma 2.

#### 4.1 Limiting behavior of the AMUSE estimator

We next consider limiting properties of the finite sample solutions. Note that the finite sample statistics and the sampled stochastic process  $\mathbf{X}$  depend on the sample size  $T$ . In the framework of this paper, our usage of the notation is equivalent with the one presented in [38]. We use the short expression  $\mathbf{X}_T = \mathcal{O}_p(1)$  to denote that a sequence of  $\mathbb{C}^{d \times d}$ -matrices  $\mathbf{X}_1, \mathbf{X}_2, \dots$ , converges in probability to a zero matrix. In addition, we use  $\mathbf{Y}_\beta = \mathcal{O}_p(1)$  to denote that a collection of  $\mathbb{C}^{d \times d}$ -matrices  $\{\mathbf{Y}_\beta : \beta \in B\}$  is uniformly tight under some nonempty indexing set  $B$ . By saying that a matrix is uniformly tight, we mean that real part and imaginary part of every element in the matrix is uniformly tight.

**Lemma 2.** *Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 and let  $\mathbf{X}$  be a  $\mathbb{C}^{T \times d}$ -valued,  $1 \leq d < T < \infty$ , sampled stochastic process generated by  $\mathbf{x}$ , and let  $\hat{g} : \mathbf{C} \mapsto (\mathbf{C} - \mathbf{I}_T \hat{\boldsymbol{\mu}}^\top) \hat{\Gamma}^\top$  be a  $T$ -indexed sequence of corresponding finite sample solutions. Furthermore, let  $\alpha_T(\hat{\mathbf{S}}_0[\mathbf{X}] - \mathbf{S}_0[\mathbf{x}_t]) = \mathcal{O}_p(1)$  and  $\beta_T(\hat{\mathbf{S}}_\tau[\mathbf{X}] - \mathbf{S}_\tau[\mathbf{x}_t]) = \mathcal{O}_p(1)$ , for some real-valued sequences  $\alpha_T$  and  $\beta_T$ , which satisfy  $\alpha_T \uparrow \infty$ ,  $\beta_T \uparrow \infty$  as  $T \rightarrow \infty$ . Then, for  $\gamma_t = \min(\alpha_t, \beta_t)$ , there exists a sequence of  $T$ -indexed phase-shift matrices  $\hat{\mathbf{J}}$ , such that*

$$\gamma_T (\hat{\mathbf{J}} \hat{\Gamma} - \Gamma) = \mathcal{O}_p(1)$$

*holds asymptotically.*

Hereby, under the assumptions of Lemma 2, there exists a sequence of  $T$ -indexed matrices  $\hat{\mathbf{J}}$  such that  $\hat{\mathbf{J}} \hat{\Gamma}$  converges in probability to  $\Gamma$ .

**Theorem 2.** *Let  $\mathbf{X}$ ,  $g$  and  $\hat{g}$  be defined as in Lemma 2 and denote the element  $(j, k)$  of  $\hat{\mathbf{S}}_\tau[\mathbf{X}]$  as  $[\hat{\mathbf{S}}_\tau]_{jk}$ . Then, under the assumptions of Lemma 2 and the trivial mixing scenario  $\mathbf{A} = \mathbf{I}_d$ , there exists a sequence of  $T$ -indexed phase-shift matrices  $\hat{\mathbf{J}}$ , such that*

$$\begin{aligned} \gamma_T (\hat{\mathbf{J}}_{jj} \hat{\Gamma}_{jj} - 1) &= \frac{\gamma_T}{2} \left( [\hat{\mathbf{S}}_0]_{jj} - 1 \right) + \mathcal{O}_p(1/\gamma_T), \quad \forall j \in \{1, \dots, d\}, \quad \text{and} \\ \gamma_T (\lambda_\tau^{(k)} - \lambda_\tau^{(j)}) \hat{\mathbf{J}}_{jj} \hat{\Gamma}_{jk} &= \gamma_T \left( \lambda_\tau^{(j)} [\hat{\mathbf{S}}_0]_{jk} - [\hat{\mathbf{S}}_\tau]_{jk} \right) + \mathcal{O}_p(1/\gamma_T), \quad j \neq k. \end{aligned}$$

By Theorem 2, we can directly find the asymptotic distribution of the unmixing matrix estimator  $\hat{\Gamma}$ , if we have the asymptotic distributions and the convergence rates

of the estimators  $\hat{\mathbf{S}}_0$  and  $\hat{\mathbf{S}}_\tau$ . Note that, if the rates  $\alpha_T$  and  $\beta_T$  differ, the estimator with the faster convergence will tend to zero in probability in Theorem 2. Furthermore, note that the asymptotic distributions of the diagonal elements of  $\gamma_T(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d)$  only depend on the asymptotic distribution of the covariance matrix estimator  $\hat{\mathbf{S}}_0$ . However, the rate of convergence of the off-diagonal elements depends on both  $\hat{\mathbf{S}}_0$  and  $\hat{\mathbf{S}}_\tau$ , and consequently, if the covariance estimator  $\hat{\mathbf{S}}_0$  converges faster than the autocovariance estimator  $\hat{\mathbf{S}}_\tau$ , the diagonal elements of  $\gamma_T(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d)$  converge to zero in distribution and in probability. Conversely, it may happen that the autocovariance estimator  $\hat{\mathbf{S}}_\tau$  converges faster. In that case, the off-diagonal elements of  $\gamma_T(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d)$  converge to zero in distribution and in probability.

Theorem 2 can alternatively be presented in matrix form as

$$\begin{aligned}\gamma_T \left( \text{diag} \left[ \hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d \right] \right) &= \gamma_T \left( \frac{1}{2} \text{diag} \left[ \hat{\mathbf{S}}_0 - \mathbf{I}_d \right] \right) + \mathcal{O}_p(1/\gamma_T), \\ \gamma_T \left( \hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \text{diag} \left[ \hat{\mathbf{J}}\hat{\mathbf{\Gamma}} \right] \right) &= \gamma_T \left( \mathbf{H} \odot \left[ \mathbf{\Lambda}_\tau \hat{\mathbf{S}}_0 - \hat{\mathbf{S}}_\tau \right] \right) + \mathcal{O}_p(1/\gamma_T),\end{aligned}$$

where  $\mathbf{H}_{jj} = 0$  and  $\mathbf{H}_{jk} = (\lambda_\tau^{(k)} - \lambda_\tau^{(j)})^{-1}$ ,  $j \neq k$ ,  $\odot$  denotes the Hadamard (i.e., entrywise) product and  $\hat{\mathbf{J}}$  is the  $T$ -indexed sequence of phase-shift matrices that set the diagonal elements of  $\hat{\mathbf{\Gamma}}$  to be on the positive real-axis.

#### 4.2 Limiting behavior under summable covariance structures

In this section, we consider a class of stochastic processes that satisfy the Breuer–Major theorem for weakly stationary processes. The Breuer–Major theorem is considered in the context of Gaussian subordinated processes, see, e.g., [2, 8, 30]. A univariate real-valued Gaussian subordinated process  $z$ , defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , is a weakly stationary process that can be expressed in the form  $z = f \circ \mathbf{y}$ , where  $\mathbf{y}$  is a  $\mathbb{R}^\ell$ -variate Gaussian process and  $f : \mathbb{R}^\ell \rightarrow \mathbb{R}$ .

We emphasize that Gaussian subordinated processes form a very rich model class. For example, recently in [39], the authors showed that arbitrary one-dimensional marginal distributions and a rich class of covariance structures can be modeled by  $f \circ \mathbf{y}$  with simple univariate stationary Gaussian process  $\mathbf{y}$ . While the model class is very rich, underlying driving Gaussian processes still provide a large toolbox for analyzing limiting behavior of various estimators. Such limit theorems have been a topic of active research for decades. For recent relevant papers on the topic, we refer to [27–30] and the references therein.

We next give the definitions of univariate real-valued Hermite polynomials and Hermite ranks. We define the  $k$ th,  $k \in \{0\} \cup \mathbb{N}$ , (probabilistic) Hermite polynomial  $H_k$ , using Rodrigues’ formula

$$H_k(x) = (-1)^k \exp \left( x^2/2 \right) \frac{d^k}{dx^k} \exp \left( -x^2/2 \right).$$

The first four Hermite polynomials are hereby  $H_0(x) = 1$ ,  $H_1(x) = x$ ,  $H_2(x) = x^2 - 1$  and  $H_3(x) = x^3 - 3x$ . The set  $\{H_k/\sqrt{k!} : k \in \{0\} \cup \mathbb{N}\}$  forms an orthonormal basis on the Hilbert space  $\mathcal{L}^2(\mathbb{R}, P_y)$ , where  $P_y$  denotes the law of a univariate standard Gaussian random variable. Consequently, every function  $f \in \mathcal{L}^2(\mathbb{R}, P_y)$  can be

decomposed as

$$f(x) = \sum_{k=0}^{\infty} a_k H_k(x), \quad (3)$$

where  $a_k \in \mathbb{R}$  for every  $k \in \{0\} \cup \mathbb{N}$ . If  $x$  and  $y$  follow the univariate standard Gaussian distribution, the orthogonality of the Hermite polynomials and the decomposition of Equation (3) give

$$\mathbb{E}[f(x)f(y)] = \sum_{k=0}^{\infty} k! \alpha_k^2 (\mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])])^k = \sum_{k=0}^{\infty} k! \alpha_k^2 (\text{Cov}[x, y])^k.$$

In the following, we consider real-valued functions of several variables and apply some known results that are based on multivariate Hermite polynomials. However, our derivations (in Section 4.3) require the use of univariate Hermite polynomials only. Thus, we only define univariate Hermite polynomials here.

The Hermite rank for a function  $f$  is defined as follows.

**Definition 5** (Hermite rank). Let  $\mathbf{y}$  be a  $\mathbb{R}^\ell$ -valued Gaussian random vector,  $\ell \in \mathbb{N}$ , and let  $f : \mathbb{R}^\ell \rightarrow \mathbb{R}$  be a function such that  $f \circ \mathbf{y}$  is square-integrable. The function  $f$  has Hermite rank  $q$ , with respect to  $\mathbf{y}$ , if

$$\mathbb{E}[(f(\mathbf{y}) - \mathbb{E}[f(\mathbf{y})]) p_m(\mathbf{y})] = 0,$$

for all Hermite polynomials  $p_m : \mathbb{R}^d \rightarrow \mathbb{R}$  that are of degree  $m \leq q - 1$  and there exists a polynomial  $p_q$  of degree  $q$  such that

$$\mathbb{E}[(f(\mathbf{y}) - \mathbb{E}[f(\mathbf{y})]) p_q(\mathbf{y})] \neq 0.$$

Note that in the one-dimensional setting, the Hermite rank  $q$  of a function  $f$  is the smallest nonnegative integer in Equation (3), such that  $\alpha_q \neq 0$ . We next present the multivariate version of the well-known Breuer–Major theorem from [8].

**Theorem 3.** Let  $\mathbf{y}_\cdot := (\mathbf{y}_t)_{t \in \mathbb{N}}$  be a  $\mathbb{R}^\ell$ -valued centered stationary Gaussian process. Let  $f_1, f_2, \dots, f_d : \mathbb{R}^\ell \rightarrow \mathbb{R}$  be measurable functions that have the Hermite rank at least  $q \in \mathbb{N}$ , with respect to  $\mathbf{y}_\cdot$ , and let  $f_j \circ \mathbf{y}_\cdot$  be square-integrable for every  $j \in \{1, \dots, d\}$ . Additionally, let the series of covariances satisfy

$$\sum_{\tau=0}^{\infty} \left| \ddot{S}_\tau \left[ y_1^{(j)}, y_1^{(k)} \right] \right|^q = \sum_{\tau=0}^{\infty} \left| \mathbb{E} \left[ y_1^{(j)} y_{1+\tau}^{(k)} \right] \right|^q < +\infty, \quad \forall j, k \in \{1, \dots, \ell\}.$$

Then, the  $\mathbb{R}^d$ -sequence

$$\frac{1}{\sqrt{T}} \left( \sum_{t=1}^T (f_1(\mathbf{y}_t) - \mathbb{E}[f_1(\mathbf{y}_t)]) \quad \cdots \quad \sum_{t=1}^T (f_d(\mathbf{y}_t) - \mathbb{E}[f_d(\mathbf{y}_t)]) \right)^\top$$

converges in distribution, as  $T \rightarrow \infty$ , to a centered Gaussian vector  $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_d)$  with finite covariances  $\mathbb{E}[\rho_j \rho_k]$  equal to

$$S_0[f_j(\mathbf{y}_1), f_k(\mathbf{y}_1)] + \sum_{\tau=1}^{\infty} (\ddot{S}_\tau[f_j(\mathbf{y}_1), f_k(\mathbf{y}_1)] + \ddot{S}_\tau[f_k(\mathbf{y}_1), f_j(\mathbf{y}_1)]).$$

The proof of Theorem 3 is omitted here. Theorem 3 follows directly from the univariate Breuer–Major theorem, given in [8], and by using the usual Cramér–Wold device, see, e.g., [7, Theorem 29.4]. A similar version of Theorem 3 can also be found in Section 5 of [2].

We next present the following assumption which enables us to find the asymptotic behavior of the unmixing matrix estimator using the Breuer–Major theorem.

**Assumption 1.** Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 with the trivial mixing matrix  $\mathbf{A} = \mathbf{I}_d$  and let  $\mathbf{X}$  be a  $\mathbb{C}^{T \times d}$ -valued,  $1 \leq d < T < \infty$ , sampled stochastic process generated by  $\mathbf{x}$ . Denote  $\mathbf{z} = (\mathbf{b}_t + i\mathbf{c}_t)_{t \in \mathbb{N}} = \mathbf{b} + i\mathbf{c}$ . We assume that there exists  $\ell \in \mathbb{N}$  and a centered  $\mathbb{R}^\ell$ -valued stationary Gaussian process  $\boldsymbol{\eta}$ , such that, for all  $k \in \{1, \dots, d\}$ , the component  $\mathbf{b}^{(k)}$  has the same finite-dimensional distributions and, asymptotically, the same autocovariance function as  $\tilde{f}_k(\boldsymbol{\eta})$ , for some function  $\tilde{f}_k : \mathbb{R}^\ell \rightarrow \mathbb{R}$ . Similarly, we assume that, for all  $k \in \{1, \dots, d\}$ , the component  $\mathbf{c}^{(k)}$  has the same finite-dimensional distributions and, asymptotically, the same autocovariance function as  $\tilde{f}_{k+d}(\boldsymbol{\eta})$ , for some function  $\tilde{f}_{k+d} : \mathbb{R}^\ell \rightarrow \mathbb{R}$ . Furthermore, we assume that  $\mathbb{E}[|\tilde{f}_k(\boldsymbol{\eta}_1)|^4] < +\infty$ ,  $\forall k \in \{1, \dots, 2d\}$ , and that  $r_{\boldsymbol{\eta}}^{(j,k)}(t) = \mathbb{E} \left[ \boldsymbol{\eta}_{t+1}^{(j)} \boldsymbol{\eta}_1^{(k)} \right]$  satisfies  $r_{\boldsymbol{\eta}}^{(j,j)}(0) = 1$ ,  $\forall j \in \{1, \dots, \ell\}$ , and

$$\sum_{t=1}^{\infty} \left| r_{\boldsymbol{\eta}}^{(j,k)}(t) \right| < +\infty, \quad \forall j, k \in \{1, \dots, \ell\}.$$

Note that the finite-dimensional distributions of the  $\mathbb{R}$ -valued stochastic process  $\mathbf{c}$ , is the collection of probability measures defined as,

$$\left\{ \mathbb{P} \left[ \{c_{t_1} \in B_1, \dots, c_{t_m} \in B_m\} \mid m \in \mathbb{N} : \forall B_j \in \mathcal{B}(\mathbb{R}) \text{ and } \{t_1, \dots, t_m\} \subset \mathbb{N} \right] \right\}.$$

We again want to emphasize that a wide class of stochastic processes satisfy Assumption 1. Indeed, we allow an arbitrary dimension  $\ell$  for the driving Gaussian process  $\boldsymbol{\eta}$ , and arbitrary (summable) covariance structures. In comparison, it was shown in [39] that in many cases it would suffice to relate only one stationary Gaussian process  $\boldsymbol{\eta}^{(j)}$  to each function  $\tilde{f}_j$ .

We are now ready to consider the asymptotic distribution for the unmixing matrix estimator under Assumption 1.

**Theorem 4.** Let Assumption 1 be satisfied and let  $\hat{\mathbf{g}} : \mathbf{C} \mapsto (\mathbf{C} - \mathbf{I}_T \hat{\boldsymbol{\mu}}^\top) \hat{\boldsymbol{\Gamma}}^\top$  be a  $T$ -indexed sequence of finite sample solutions. Then, under the trivial mixing scenario  $\mathbf{A} = \mathbf{I}_d$ , there exists a  $T$ -indexed sequence  $\hat{\mathbf{J}}$  of phase-shift matrices such that

$$\sqrt{T} \cdot \text{vec} \left( \hat{\mathbf{J}} \hat{\boldsymbol{\Gamma}} - \mathbf{I}_d \right) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathbf{v} \sim \mathcal{N}_{d^2}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}}, \mathbf{P}_{\mathbf{v}}).$$

The exact forms for  $\boldsymbol{\Sigma}_{\mathbf{v}}$  and  $\mathbf{P}_{\mathbf{v}}$  are given in the supplementary Appendix.

Note that it is possible to present Theorem 4 with even weaker assumptions. In the current formulation, we require that the Gaussian process  $\mathbf{y}$  (equivalently,  $\boldsymbol{\eta}$ ) has summable covariances and cross-covariances. By studying the exact Hermite ranks

of the underlying functions, weaker summability conditions would suffice (cf. Theorem 3). However, given the Hermite ranks of the functions  $f_j$ , it is in general impossible to say anything about the Hermite ranks of transformations such as  $f_j^2$  arising from  $\hat{S}_0$ , see the proof of Theorem 4. On the other hand, assuming Hermite rank equal to one is a very natural assumption in many occasions. For example, the modeling approach given in [39] sets the rank of  $f_j$  to equal 1. Hermite rank 1 is stable in a sense that even small perturbations in a function with higher Hermite rank leads to rank 1 again. For further discussion on the stability of Hermite rank 1, we refer to [4].

### 4.3 Notes on noncentral limit theorems

In this section, we provide examples where the convergence rate of the unmixing estimator differs from the standard  $\sqrt{T}$  and where the limiting distribution is non-Gaussian. Such situations arise, especially, when the convergence summability condition of Theorem 3 does not hold.

**Assumption 2.** Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 with the trivial mixing matrix  $\mathbf{A} = \mathbf{I}_d$  and let  $\mathbf{X}$  be a  $\mathbb{C}^{T \times d}$ -valued,  $1 \leq d < T < +\infty$ , sampled stochastic process generated by  $\mathbf{x}$ . and  $\mathbf{z} = \mathbf{b} + i\mathbf{c}$ . We assume that there exists centered  $\mathbb{R}$ -valued stationary Gaussian processes  $\eta_*^{(k)}$ ,  $k \in \{1, 2, \dots, 2d\}$ , such that, for all  $k \in \{1, \dots, d\}$ , the component  $\mathbf{b}_*^{(k)}$  has the same finite-dimensional distributions and, asymptotically, the same autocovariance function as  $\tilde{f}_k(\eta_*^{(k)})$ , for some function  $\tilde{f}_k : \mathbb{R} \rightarrow \mathbb{R}$ . Similarly, we assume that, for all  $k \in \{1, \dots, d\}$ , the component  $\mathbf{c}_*^{(k)}$  has the same finite-dimensional distributions and, asymptotically, the same autocovariance function as  $\tilde{f}_{k+d}(\eta_*^{(k+d)})$ , for some function  $\tilde{f}_{k+d} : \mathbb{R} \rightarrow \mathbb{R}$ . Furthermore, we assume that  $\mathbb{E}[|\tilde{f}_k(\eta_*^{(k)})|^4] < +\infty$ ,  $\forall k \in \{1, \dots, 2d\}$ , and that the Gaussian processes  $\eta_*^{(1)}, \eta_*^{(2)}, \dots, \eta_*^{(2d)}$  are mutually independent.

In comparison to Assumption 1, we here assume that real and imaginary parts of each component are driven by a single Gaussian process. As discussed in Section 4.2, this is not a huge restriction. In addition, we assume independent components and that the real and imaginary parts of each component are independent. Although independence is a common assumption in the blind source separation literature, we note that our results can be extended to cover dependencies as long as, for every  $j, k \in \{1, 2, \dots, 2d\}$ , the cross-covariances between  $\eta^{(k)}$  and  $\eta^{(j)}$  are negligible compared to the decay of the autocovariance functions  $r_{\eta^{(j)}}(t)$ .

If the autocovariance functions  $r_{\eta^{(j)}}$  satisfy

$$\sum_{t=1}^{\infty} |r_{\eta^{(j)}}(t)| < +\infty, \quad \forall j \in \{1, \dots, 2d\}, \quad (4)$$

then we are in the situation of Theorem 4. More generally, a weakly stationary series is called short-range dependent if the corresponding autocovariance function  $r$  satisfies Equation (4). Hence, we assume that at least one of the functions  $r_{\eta^{(j)}}$  is not summable. For such components, we assume the following long-range dependence condition.

**Definition 6.** We say that a  $\mathbb{R}$ -valued weakly stationary process  $\eta$  is long-range dependent if the autocovariance function  $r_\eta$  satisfies

$$\lim_{k \rightarrow \infty} k^{2-2H} r_\eta(k) = C \quad (5)$$

for some  $H \in [1/2, 1)$  and  $C \in (0, \infty)$ .

Note that the definition of long-range dependence varies in the literature. For details on long-range dependent processes and their different definitions, we refer to [5, 6, 33].

There is a large literature on limit theorems under long-range dependence. See, e.g., [3, 11, 34] for a few central works on the topic. For an insightful and detailed presentation we refer to [18], and for recent advances on the topic, see, e.g., [21, 40]. In the case of long-range dependent processes, the rate of convergence of the normalized mean is slower than the usual  $\sqrt{T}$  and the limiting distribution depends on the corresponding Hermite rank. More precisely, the limiting distribution follows a so-called  $q$ -Hermite distribution, where  $q$  is the Hermite rank of the underlying function. Note that  $q$ -Hermite distributions can be fully characterized by the corresponding Hermite rank  $q$  and a self-similarity parameter  $H$ , i.e., the Hurst index. In the stable case  $q = 1$ , we obtain a Gaussian limit, and in the case  $q = 2$  we obtain the so-called *Rosenblatt distribution* that is not Gaussian. Similarly, for  $q \geq 3$  the limiting distribution is not Gaussian. For details on Hermite distributions and processes, we refer to [3, 12, 36]. In particular, we apply the following known result (see [11, Theorem 1]).

**Proposition 1.** Let  $\eta$  be a  $\mathbb{R}$ -valued stationary Gaussian process with autocovariance function  $r_\eta$  such that  $r_\eta(0) = 1$  and  $r_\eta$  satisfies Equation (5) for some  $H \in [1/2, 1)$ . Let  $f$  be a function such that  $\mathbb{E}[(f(\eta_1))^2] < +\infty$  and the Hermite rank of  $f$  equals  $q$ . If  $q(2H - 2) > -1$ , then, for some constant  $C = C_{f,\eta} > 0$ , we have

$$T^{q(1-H)-1} \sum_{t=1}^T (f(\eta_t) - \mathbb{E}[f(\eta_t)]) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} C Z_q,$$

where  $Z_q$  follows a  $q$ -Hermite distribution.

In view of Proposition 1, in the long-range dependent case the Hermite rank plays a crucial role. Thus, one cannot pose general results without any a priori knowledge on the ranks. Indeed, it can be shown — see, e.g., [6, Equation 4.26] with  $d = H - 1/2$  — that if  $f$  has Hermite rank  $q$  and  $r$  satisfies Equation (5) with some  $H$  such that  $q(2H - 2) > -1$ , then

$$\lim_{T \rightarrow \infty} T^{q(2-2H)} \mathbb{E} \left[ \left( \frac{1}{T} \sum_{t=1}^T (f(\eta_t) - \mathbb{E}[f(\eta_t)]) \right)^2 \right] = C. \quad (6)$$

This justifies the normalization in Proposition 1 and also gives the constant explicitly, as  $\mathbb{E}[Z_q] = 0$  and  $\mathbb{E}[(Z_q)^2] = 1$ . We pose the following assumption on the autocovariances  $r_{\eta(j)}$  and Hermite ranks.

**Assumption 3.** We assume that for every  $j \in \{1, 2, \dots, 2d\}$  the autocovariance functions  $r_{\eta^{(j)}}$  with  $r_{\eta^{(j)}}(0) = 1$  satisfy either Equation (4) or Equation (5), and that Equation (5) is satisfied for at least one index. Let  $I \subset \{1, 2, \dots, 2d\}$  denote the set of indices for which Equation (5) holds, and for every  $i \in I$ , denote the Hermite rank of  $x \mapsto \tilde{f}_i(x)$  by  $q_{1,i}$  and the Hermite rank of  $x \mapsto [\tilde{f}_i(x) - \mathbb{E}(\tilde{f}_i(\eta^{(i)}))]^2$  by  $q_{2,i}$ . We assume that, for any indices  $j, k \in I$ ,  $j \neq k$ ,

$$\max_{i \in I} q_{2,i}(2H_i - 2) > \max_{j,k \in I} \{q_{1,k}(2H_k - 2) + q_{1,j}(2H_j - 2), -1\}, \quad (7)$$

and, for any  $k \in I$ ,

$$\max_{i \in I} q_{2,i}(2H_i - 2) \geq q_{1,k}(4H_k - 4). \quad (8)$$

We stress that Assumption 3 is not very restrictive. We allow a combination of short- and long-range dependent processes  $\eta^{(j)}$  and we assume that at least one of the processes is long-range dependent, since otherwise we may apply Theorem 4. In this respect, Condition (7) guarantees that at least one of the  $[\tilde{f}_i(x) - \mathbb{E}(\tilde{f}_i(\eta^{(i)}))]^2$  is long-range dependent. Indeed, if  $\max_{i \in I} q_i(2H_i - 2) < -1$ , then Theorem 4 applies again. Moreover, one could also study the limiting case  $\max_{i \in I} q_i(2H_i - 2) = -1$ . Then, one usually obtains a Gaussian limit, but with a rate given by  $\sqrt{T/\log(T)}$ . Our more restrictive Assumptions (7) and (8) are posed in order to apply known results on limit theorems in the long-range dependent case. Indeed, if these conditions were violated, then we would need to study the convergence of complicated dependent random vectors towards some combinations of Hermite distributions. Unfortunately, to the best of our knowledge, there are no studies in this topic that would fit into our general setting (on some results in special cases, see, e.g., [3]).

In view of Proposition 1, the rate of convergence,  $T^\gamma$ , of our unmixing estimator arises from Equation (7). We set

$$\gamma := -\frac{1}{2} \max_{i \in I} q_{2,i}(2H_i - 2). \quad (9)$$

By Equation (7), we have that  $\gamma < \frac{1}{2}$ , meaning that our rate is slower than the usual  $\sqrt{T}$ . The following results show that cross-terms do not contribute to the limit.

**Proposition 2.** Let Assumptions 2 and 3 be satisfied. Then for every  $j \neq k$ , we have that

$$T^{\gamma-1} \sum_{t=1}^T \left( \tilde{f}_k(\eta_t^{(k)}) - \mathbb{E}[\tilde{f}_k(\eta_t^{(k)})] \right) \left( \tilde{f}_j(\eta_t^{(j)}) - \mathbb{E}[\tilde{f}_j(\eta_t^{(j)})] \right) \xrightarrow[T \rightarrow \infty]{\mathcal{L}_2} 0 \quad (10)$$

and

$$T^{\gamma-1} \sum_{t=1}^T \left( \tilde{f}_k(\eta_t^{(k)}) - \mathbb{E}[\tilde{f}_k(\eta_t^{(k)})] \right) \left( \tilde{f}_j(\eta_{t+\tau}^{(j)}) - \mathbb{E}[\tilde{f}_j(\eta_t^{(j)})] \right) \xrightarrow[T \rightarrow \infty]{\mathcal{L}_2} 0 \quad (11)$$

where  $\xrightarrow{\mathcal{L}_2}$  denotes convergence in the space  $\mathcal{L}_2(\Omega, \mathbb{P})$ . Consequently, the convergence also holds in probability.

In the presence of long-range dependency, the mean estimator  $\hat{\boldsymbol{\mu}}[\mathbf{X}]$  in the definition of  $\tilde{\mathbf{S}}_\tau[\mathbf{X}]$  has to be studied separately as it may contribute to the limit.

In order to obtain the limiting distribution for  $T^\gamma(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d)$ , we introduce some notation. For each  $k \in \{1, 2, \dots, 2d\}$  we define Hermite processes  $Z_{q_{1,k}}$  and  $Z_{q_{2,k}}$ , such that for a given  $k$ , we allow  $Z_{q_{1,k}}$  and  $Z_{q_{2,k}}$  to be dependent on each other, while both are independent of  $Z_{q_{1,j}}$ ,  $Z_{q_{2,j}}$  for  $j \neq k$ . We also introduce constants  $C_{1,k}$ ,  $C_{2,k}$  in the following way: if  $k \in I$  is such that equality holds in (8), i.e.,  $\max_{i \in I} q_{2,i}(2H_i - 2) = 2[q_{1,k}(2H_k - 2)]$ , then we let  $C_{1,k}$  to be the constant given in Proposition 1 associated to the limit

$$T^{q_{1,k}(1-H_k)-1} \sum_{t=1}^T \left( \tilde{f}_k(\boldsymbol{\eta}_t^{(k)}) - \mathbb{E}[\tilde{f}_k(\boldsymbol{\eta}_t^{(k)})] \right) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} C_{1,k} Z_{q_{1,k}}.$$

Otherwise we set  $C_{1,k} = 0$ . Similarly, if  $k \in I$  is such that

$$q_{2,k}(2H_k - 2) = \max_{i \in I} q_{2,i}(2H_i - 2), \quad (12)$$

we let  $C_{2,k}$  to be the constant given in Proposition 1 associated to the limit

$$T^{q_{2,k}(1-H_k)-1} \sum_{t=1}^T \left( h_k(\boldsymbol{\eta}_t^{(k)}) - \mathbb{E}[h_k(\boldsymbol{\eta}_t^{(k)})] \right) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} C_{2,k} Z_{q_{2,k}},$$

where

$$h_k(\boldsymbol{\eta}_t^{(k)}) = \left( \tilde{f}_k(\boldsymbol{\eta}_t^{(k)}) - \mathbb{E}[\tilde{f}_k(\boldsymbol{\eta}_t^{(k)})] \right)^2.$$

Otherwise we set  $C_{2,k} = 0$ . The following theorem is our main result of this subsection.

**Theorem 5.** *Let Assumptions 2 and 3 hold. Then,*

$$T^\gamma (\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \frac{1}{2} \boldsymbol{\Upsilon},$$

where  $\boldsymbol{\Upsilon}$  is a  $\mathbb{R}^{d \times d}$ -valued diagonal matrix with elements given by

$$(\boldsymbol{\Upsilon})_{jj} \sim C_{2,j} Z_{q_{2,j}} + C_{1,j}^2 Z_{q_{1,j}}^2 + C_{2,j+d} Z_{q_{2,j+d}} + C_{1,j+d}^2 Z_{q_{1,j+d}}^2.$$

We remark that it might be that most of the elements vanish. Indeed, the coefficient  $C_{2,j}$  ( $C_{2,j+d}$ , respectively) is nonzero only if the real part (imaginary part, respectively) of  $[f_j(x) - \mathbb{E}(f_j(\boldsymbol{\eta}))]^2$  converges with the correct rate  $T^\gamma$ . In this case,  $C_{1,j}$  ( $C_{1,j+d}$ , respectively) is nonzero only if the corresponding element in  $\tilde{\boldsymbol{\mu}}$  converges with the rate  $T^{\gamma/2}$ .

## 5 Image source separation

The restoration of images, which have been mixed together by some transformation, is a classical example in blind source separation (BSS), see, e.g., [26]. The image





**Fig. 1.** Original images

source separation is usually only considered for black and white images, where the colors can be presented using a single color parameter. Since there is only a single color parameter associated with a single pixel, the black and white image source problem can be straightforwardly formulated as a real-valued BSS problem. This does not remain true for colored photographs.

In our example, we consider colored photographs and conduct the example using the statistical software R ([32]). We start with three photographs that are of size  $1536 \times 2048$ ,  $1200 \times 1600$  and  $960 \times 720$ , the figures are imported to R using the package JPEG, [37]. As a first step, we resize all of the photographs to be of equal size  $960 \times 720$ . The resized images are stored in a  $960 \times 720 \times 3$  array format, such that a red-green-blue color parameter triplet, denoted as  $(R, G, B)$ , is assigned to every pixel.

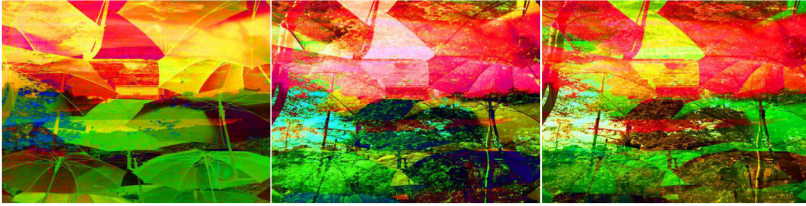
We apply a bijective cube to unit sphere transformation for every color triplet. Then, we use the well-known stereographic projection, which is an almost bijective transformation between the unit sphere and the complex plane. The stereographic projection is bijective everywhere except for the north pole of the unit sphere. For almost all photographs, we can choose the color coordinate system such that no pixel has a color triplet located on the north pole. This holds for our example and we can apply the inverse mappings in order to get back to the color cube surface.

We then have a single complex-number corresponding to every pixel of the color corrected images and we can present the images in a  $\mathbb{C}^{960 \times 720}$ -matrix form. We denote the images with complex-valued elements as  $\mathbf{Z}_1$ ,  $\mathbf{Z}_2$  and  $\mathbf{Z}_3$  and apply the following transformation,

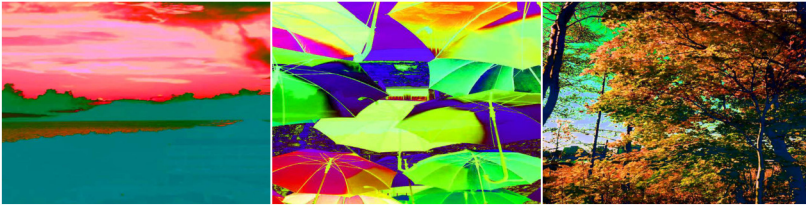
$$\mathbf{X}^\top = \begin{pmatrix} \text{vec}(\mathbf{X}_1)^\top \\ \text{vec}(\mathbf{X}_2)^\top \\ \text{vec}(\mathbf{X}_3)^\top \end{pmatrix}^\top = \begin{pmatrix} \text{vec}(\mathbf{Z}_1)^\top \\ \text{vec}(\mathbf{Z}_2)^\top \\ \text{vec}(\mathbf{Z}_3)^\top \end{pmatrix}^\top (\mathbf{I}_3 + \mathbf{E} + i\mathbf{F}),$$

where  $\mathbf{I}_3$  is the identity matrix and the elements of  $\mathbf{E}$  and  $\mathbf{F}$  were generated independently from the univariate uniform distribution  $\text{Unif}(-1/2, 1/2)$ . The complex-valued mixed images  $\mathbf{X}_1$ ,  $\mathbf{X}_2$  and  $\mathbf{X}_3$  can then be plotted by performing the chain of inverse mappings in reverse order. The corresponding mixed images are presented in Figure 2.

We then apply the unmixing procedure presented in Section 4 to the matrix  $\mathbf{X}$ , using several different lag parameters  $\tau$ . In this example,  $\mathbf{X}$  has three columns and 691200 rows, such that every element of the matrix is complex-valued. In controlled settings, where the true mixing matrix is known, the minimum distance index (MD)



**Fig. 2.** Mixed images

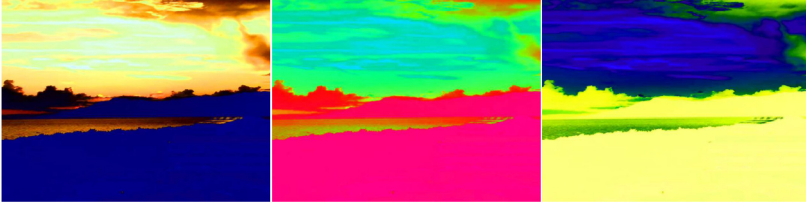


**Fig. 3.** Unmixed images using the AMUSE procedure with  $\tau = 1$

index is a straightforward way to compare the performance of different estimators, see [20] for the complex-valued formulation. The MD index is a performance index, defined between zero and one, where zero corresponds to the separation being perfect. In addition, the MD index ensures that comparison between different estimators is fair, this is especially important in this paper since under our model assumptions, different estimators do not necessarily estimate the same population quantities. The minimum distance index is based on the minimum distance between the known mixing matrix and the unmixing matrix estimate with respect to permutations, scaling and phase-shifts. While applying the AMUSE procedure, if we can find a parameter  $\tau$  such that the diagonal elements of  $\hat{\mathbf{A}}_\tau$  are sufficiently distinct (from a computational perspective), we can recover the unmixing matrix up to phase-shifts. Hereby, the choice of the mixing matrix in this example is inconsequential from the viewpoint of minimum distance index, identical values are obtained for a fixed  $\tau$  with any finite mixing matrix. Consequently, in this example, minimum distance index will produce the same value if the mixing matrix was replaced with, e.g., the identity matrix.

We tried several different lag parameters and the best performance, in the light of the MD index and visual inspection, was attained with  $\tau = 1$  and the corresponding MD index value was approximately 0.174. The unmixed images obtained using the unmixing procedure with  $\tau = 1$  are presented in Figure 3. In addition, for example, lag parameter choices  $\tau = 2, 3, 961, 962, 963$  provided only slightly worse results, the corresponding MD index values were between 0.177 and 0.183. Note that the autocovariances are calculated from the vectorized images, and for example the first and the 961st entries are neighboring pixels in the unvectorized images.

Comparing the original color corrected images in Figure 1 and the unmixed images in Figure 3, the shapes seem to match almost perfectly. The color schemes vary since the complex phase is not uniquely fixed in our model. In addition, recall that under our model, solutions that are a phase-shift away from each other are consid-



**Fig. 4.** Unmixed images produced by equivalent solutions

ered to be equivalent. In Figure 4, we present the first unmixed image, produced by the unmixing procedure with  $\tau = 1$ , under three equivalent solutions. The images in Figure 4 are obtained such that we first find an unmixing estimate  $\hat{\Gamma}$  and then right-multiply the obtained estimate with a diagonal matrix with diagonal entries  $\exp(i\theta)$ . In Figure 4, the left image is obtained with  $\theta = \pi/4$ , the middle one is obtained with  $\theta = 3\pi/4$  and the right image is obtained with  $\theta = 5\pi/4$ .

## A Appendix

**Proof of Lemma 1.** Both directions of the claim follow directly from the multivariate version of the continuous mapping theorem. Note that the mapping  $f : (r_1, r_2, \dots, r_{2d})^\top \mapsto (r_1, r_2, \dots, r_d)^\top + i(r_{d+1}, r_{d+2}, \dots, r_{2d})^\top : \mathbb{R}^{2d} \rightarrow \mathbb{C}^d$  is homeomorphic, that is,  $f$  is continuous and bijective, and the preimage  $f^{-1}$  is also continuous.  $\square$

**Corollary 1.** Let  $\{v_n\}_{n \in \mathbb{N}}$  be a collection of  $\mathbb{R}^{2d}$ -valued random vectors  $v_j^\top = \begin{pmatrix} x_j^\top & y_j^\top \end{pmatrix}$ , where  $x_j, y_j$  are  $\mathbb{R}^d$ -valued for every  $j \in \mathbb{N}$ . Let

$$\alpha_n \sum_{j=1}^n (v_j - \mu_v) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} v \sim \mathcal{N}_{2d}(\theta, \Sigma_v),$$

where  $\alpha_n \uparrow \infty$  as  $n \rightarrow \infty$  and

$$\mu_v = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad \text{and} \quad \Sigma_v = \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^\top & \Sigma_y \end{pmatrix}.$$

Then, for the sequence of  $\mathbb{C}^d$ -valued random vectors  $z_j = x_j + iy_j$ , we have that

$$\alpha_n \sum_{j=1}^n (z_j - \mu_z) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} z \sim \mathcal{N}_d(\theta, \Sigma_z, P_z),$$

where  $\mu_z = \mu_x + i\mu_y$ ,  $\Sigma_z = \Sigma_x + \Sigma_y + i(\Sigma_{xy}^\top - \Sigma_{xy})$  and  $P_z = \Sigma_x - \Sigma_y + i(\Sigma_{xy}^\top + \Sigma_{xy})$ .

**Proof of Corollary 1.** The corollary follows directly by applying Lemma 1 and by calculating the mean, the covariance and the relation matrix of  $z$ .  $\square$

**Lemma 3.** Let  $\mathbf{x}_\cdot := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2. The functional  $g : \mathbf{a} \mapsto \mathbf{\Gamma}(\mathbf{a} - \boldsymbol{\mu}) : \mathbb{C}^d \rightarrow \mathbb{C}^d$  is a solution to the corresponding unmixing problem if and only if

$$\boldsymbol{\mu} = \boldsymbol{\mu}_\mathbf{x} \quad \text{and} \quad \mathbf{\Gamma}\mathbf{A} = \mathbf{J},$$

where  $\mathbf{J} \in \mathbb{C}^{d \times d}$  is a phase-shift matrix.

**Proof of Lemma 3.** Let  $\mathbf{y}_\cdot := g \circ \mathbf{x}_\cdot$ . First, let  $\mathbf{y}_\cdot$  be a solution, that is,  $\mathbf{y}_\cdot$  satisfies conditions (1)–(4) of Definition 2. Recall that  $\mathbf{\Gamma}$  is assumed to be nonsingular. Using condition (1), we get that

$$\mathbb{E}[\mathbf{y}_t] = \mathbb{E}[\mathbf{\Gamma}(\mathbf{x}_t - \boldsymbol{\mu})] = \mathbb{E}[\mathbf{\Gamma}(\mathbf{A}\mathbf{z}_t + \boldsymbol{\mu}_\mathbf{x} - \boldsymbol{\mu})] = \mathbf{\Gamma}\mathbb{E}[\boldsymbol{\mu}_\mathbf{x} - \boldsymbol{\mu}] = \mathbf{0},$$

and thus  $\boldsymbol{\mu}_\mathbf{x} = \boldsymbol{\mu}$ , since  $\mathbf{\Gamma}$  is nonsingular. Next, we can rewrite condition (2) as

$$\mathbf{S}_0[\mathbf{y}_t] = \mathbb{E}[\mathbf{y}_t \mathbf{y}_t^H] = \mathbb{E}[\mathbf{\Gamma}\mathbf{A}\mathbf{z}_t (\mathbf{\Gamma}\mathbf{A}\mathbf{z}_t)^H] = \mathbf{\Gamma}\mathbf{A}\mathbf{A}^H \mathbf{\Gamma}^H = \mathbf{I}_d,$$

which implies that  $\mathbf{\Gamma}\mathbf{A}$  is a unitary matrix. Similarly, we can rewrite condition (4) as

$$\mathbf{S}_\tau[\mathbf{y}_t] = \frac{1}{2} \left( \mathbb{E}[\mathbf{y}_t \mathbf{y}_{t+\tau}^H] + \mathbb{E}[\mathbf{y}_{t+\tau} \mathbf{y}_t^H] \right) = \mathbf{\Gamma}\mathbf{A}\boldsymbol{\Lambda}_\tau \mathbf{A}^H \mathbf{\Gamma}^H = \boldsymbol{\Lambda}_\tau,$$

which is equivalent to

$$\boldsymbol{\Lambda}_\tau \mathbf{A}^H \mathbf{\Gamma}^H = \mathbf{A}^H \mathbf{\Gamma}^H \boldsymbol{\Lambda}_\tau.$$

Since  $\boldsymbol{\Lambda}_\tau$  has real-valued distinct diagonal entries and since  $\boldsymbol{\Lambda}_\tau$  and  $\mathbf{A}^H \mathbf{\Gamma}^H$  commute, we get that  $\mathbf{\Gamma}\mathbf{A}$  is also a diagonal matrix. Consequently,  $\mathbf{\Gamma}\mathbf{A}$  is a unitary diagonal matrix, which implies that the diagonal elements of the matrix product are of the form  $\exp(i\theta_1), \dots, \exp(i\theta_d)$ .

For the second part of the proof, let  $\boldsymbol{\mu} = \boldsymbol{\mu}_\mathbf{x}$  and  $\mathbf{\Gamma}\mathbf{A} = \mathbf{J}$ , where  $\mathbf{J}$  is some phase-shift matrix. We next verify that  $\mathbf{y}_\cdot := g \circ \mathbf{x}_\cdot$  is a solution, that is, we verify that  $\mathbf{y}_\cdot$  satisfies conditions (1)–(4) of Definition 2. Condition (1) is clearly satisfied, since  $\mathbb{E}[\mathbf{x}_t] = \boldsymbol{\mu}_\mathbf{x}$ . Furthermore, we have that, for every  $\tau, t \in \mathbb{N}$ ,

$$\ddot{\mathbf{S}}_\tau[\mathbf{y}_t] = \mathbb{E}[\mathbf{\Gamma}\mathbf{A}\mathbf{z}_t (\mathbf{\Gamma}\mathbf{A}\mathbf{z}_{t+\tau})^H] = \mathbf{J}\ddot{\mathbf{A}}_\tau \mathbf{J}^H,$$

where  $\ddot{\mathbf{A}}_0 = \mathbf{I}_d$ . Thus, conditions (2)–(3) are satisfied. Finally, we have for a fixed  $\tau$  and for every  $t \in \mathbb{N}$  that

$$\mathbf{S}_\tau[\mathbf{y}_t] = \mathbf{J}\mathbf{S}_\tau[\mathbf{z}_t]\mathbf{J}^H = \mathbf{J}\boldsymbol{\Lambda}_\tau \mathbf{J}^H = \boldsymbol{\Lambda}_\tau,$$

since diagonal matrices commute. Thus, condition (4) is satisfied and the proof is complete.  $\square$

**Proof of Theorem 1.** In order to simplify the notation, we denote  $\mathbf{S}_0 := \mathbf{S}_0[\mathbf{x}_t]$  and  $\mathbf{S}_\tau := \mathbf{S}_\tau[\mathbf{x}_t]$ . First, let  $g$  be a solution. Then, since  $\mathbf{x}_\cdot$  follows the mixing model, Lemma 3 gives

$$\mathbf{S}_0 \mathbf{\Gamma}^H = \mathbf{A} \mathbf{A}^H \mathbf{\Gamma}^H = \mathbf{A} \mathbf{J}^H \implies \mathbf{A} = \mathbf{S}_0 \mathbf{\Gamma}^H \mathbf{J}.$$

Using the above expressions, we get that

$$\mathbf{S}_\tau \mathbf{\Gamma}^H = \mathbf{A} \mathbf{A}_\tau \mathbf{A}^H \mathbf{\Gamma}^H = \mathbf{S}_0 \mathbf{\Gamma}^H \mathbf{J} \mathbf{A}_\tau \mathbf{J}^H = \mathbf{S}_0 \mathbf{\Gamma}^H \mathbf{A}_\tau,$$

and left-multiplying by  $\mathbf{S}_0^{-1}$  yields Equation (1). In addition, Lemma 3 directly gives Equation (2). Since the process  $\mathbf{x}$  is weakly stationary, the previous is true for every  $t \in \mathbb{N}$ .

Next, let Equations (1) and (2) be satisfied for every  $t \in \mathbb{N}$ . Left-multiplying by  $\mathbf{S}_0$ , Equation (1) can be reformulated as

$$\mathbf{A} \mathbf{A}_\tau \mathbf{A}^H \mathbf{\Gamma}^H = \mathbf{A} \mathbf{A}^H \mathbf{\Gamma}^H \mathbf{A}_\tau,$$

which, after left-multiplication by  $\mathbf{A}^{-1}$ , gives that  $\mathbf{A}^H \mathbf{\Gamma}^H$  and  $\mathbf{A}_\tau$  commute. Since the diagonal matrix  $\mathbf{A}_\tau$  has real-valued distinct diagonal elements, we get that  $\mathbf{A}^H \mathbf{\Gamma}^H$  is a diagonal matrix. Then the scaling equation gives that  $\mathbf{\Gamma} \mathbf{A} \mathbf{A}^H \mathbf{\Gamma}^H = \mathbf{I}_d$ , i.e.,  $\mathbf{\Gamma} \mathbf{A}$  is a unitary matrix. Consequently,  $\mathbf{\Gamma} \mathbf{A} = \mathbf{J}$ , where  $\mathbf{J}$  is some phase-shift matrix. By Lemma 3, the functional  $g$  is hereby a solution to the corresponding unmixing problem, and the proof is complete.  $\square$

**Corollary 2.** *Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 and denote  $\mathbf{S}_j := \mathbf{S}_j[\mathbf{x}_t]$ . Let  $\mathbf{S}_0^{-1/2}$  be a conjugate symmetric matrix, such that  $\mathbf{S}_0^{-1/2} \mathbf{S}_0 \mathbf{S}_0^{-1/2} = \mathbf{I}_d$ . Then the eigendecomposition  $\mathbf{S}_0^{-1/2} \mathbf{S}_\tau \mathbf{S}_0^{-1/2} = \mathbf{V} \mathbf{A}_\tau \mathbf{V}^H$  is satisfied for some unitary  $\mathbf{V}$ , and the functional*

$$g : \mathbf{a} \mapsto \mathbf{V}^H \mathbf{S}_0^{-\frac{1}{2}} (\mathbf{a} - \boldsymbol{\mu}_x) : \mathbb{C}^d \rightarrow \mathbb{C}^d$$

*is a solution to the corresponding unmixing problem.*

**Proof of Corollary 2.** Under the assumptions of Definition 2, we have that  $\mathbf{S}_0$  is a positive-definite matrix and that  $\mathbf{S}_\tau$  is a conjugate symmetric matrix. Thus, the matrix-square root of  $\mathbf{S}_0$  exists and the conjugate symmetric matrix-square root is unique, and consequently similar arguments as in the real-valued counterpart presented in [17] can be applied here. Hereby,  $\mathbf{S}_0^{-1/2} \mathbf{S}_\tau \mathbf{S}_0^{-1/2}$  can always be eigendecomposed.

The second part of the proof follows directly from Theorem 1 by verifying that  $\mathbf{\Gamma} = \mathbf{V}^H \mathbf{S}_0^{-1/2}$  satisfies Equations (1) and (2).  $\square$

**Lemma 4.** *Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 and let  $\mathbf{X}$  be a  $\mathbb{C}^{T \times d}$ -valued,  $1 \leq d < T < \infty$ , sampled stochastic process generated by  $\mathbf{x}$ . Denote  $\hat{\boldsymbol{\mu}} := \hat{\boldsymbol{\mu}}[\mathbf{X}]$  and  $\hat{\mathbf{S}}_j := \hat{\mathbf{S}}_j[\mathbf{X}]$ ,  $j \in \{0, \tau\}$ . Let  $\hat{\boldsymbol{\Sigma}}_0$  be a conjugate symmetric matrix that satisfies  $\hat{\boldsymbol{\Sigma}}_0 \hat{\mathbf{S}}_0 \hat{\boldsymbol{\Sigma}}_0 = \mathbf{I}_d$  and let  $\hat{\boldsymbol{\Sigma}}_0 \hat{\mathbf{S}}_\tau \hat{\boldsymbol{\Sigma}}_0 = \hat{\mathbf{V}} \hat{\mathbf{A}}_\tau \hat{\mathbf{V}}^H$ , where  $\hat{\mathbf{V}}$  is unitary. Such matrices  $\hat{\boldsymbol{\Sigma}}_0$ ,  $\hat{\mathbf{V}}$  exist, and the mapping*

$$\hat{g} : \mathbf{C} \mapsto (\mathbf{C} - \mathbf{I}_T \hat{\boldsymbol{\mu}}^\top) \left( \hat{\mathbf{V}}^H \hat{\boldsymbol{\Sigma}}_0 \right)^\top : \mathbb{C}^{T \times d} \rightarrow \mathbb{C}^{T \times d}$$

*is a solution to the finite sample unmixing problem.*

**Proof of Lemma 4.** Note that Lemma 4 can be seen as a finite sample version of Corollary 2. By assumption, the components of  $\mathbf{z}$  have continuous marginals, which implies that covariance matrix estimates are almost surely nonsingular. Thus, we can almost surely find a unique matrix  $\hat{\Sigma}_0$  which is the conjugate symmetric inverse square root of  $\hat{S}_0$ . We proceed to verify that  $\hat{\Gamma} = \hat{\mathbf{V}}^H \hat{\Sigma}_0$  satisfies the conditions of Definition 4. First,

$$\hat{\mathbf{V}}^H \hat{\Sigma}_0 \hat{S}_0 \left( \hat{\mathbf{V}}^H \hat{\Sigma}_0 \right)^H = \hat{\mathbf{V}}^H \hat{\mathbf{V}} = \mathbf{I}_d,$$

and similarly,

$$\hat{\mathbf{V}}^H \hat{\Sigma}_0 \hat{S}_\tau \left( \hat{\mathbf{V}}^H \hat{\Sigma}_0 \right)^H = \hat{\mathbf{V}}^H \hat{\mathbf{V}} \hat{\Lambda}_\tau \hat{\mathbf{V}}^H \hat{\mathbf{V}} = \hat{\Lambda}_\tau. \quad \square$$

**Lemma 5.** Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 and let  $g : a \mapsto \Gamma(\mathbf{a} - \boldsymbol{\mu})$  be a corresponding solution. Furthermore, let  $\tilde{\mathbf{x}} := (\mathbf{C}\mathbf{x}_t + \mathbf{b})_{t \in \mathbb{N}}$ , where  $\mathbf{C} \in \mathbb{C}^{d \times d}$  is nonsingular,  $\mathbf{b} \in \mathbb{C}^d$  and let  $\tilde{g} : a \mapsto \tilde{\Gamma}(\mathbf{a} - \tilde{\boldsymbol{\mu}})$  be a solution for the process  $\tilde{\mathbf{x}}$ . Then,

$$\tilde{\Gamma} = \mathbf{J}\Gamma\mathbf{C}^{-1}$$

for some phase-shift matrix  $\mathbf{J}$ .

**Proof of Lemma 5.** By Lemma 3, we have that  $\Gamma\mathbf{A} = \mathbf{J}_1$  and  $\tilde{\Gamma}\mathbf{C}\mathbf{A} = \mathbf{J}_2$ , where  $\mathbf{J}_1$  and  $\mathbf{J}_2$  are some phase-shift matrices. Recall that the mixing matrix  $\mathbf{A}$  is nonsingular. Hereby, we get that  $\Gamma = \mathbf{J}_1\mathbf{A}^{-1}$  and  $\tilde{\Gamma} = \mathbf{J}_2(\mathbf{C}\mathbf{A})^{-1}$ . By using the obtained expressions for  $\Gamma$  and  $\tilde{\Gamma}$ , we get that

$$\tilde{\Gamma} = \mathbf{J}_2\mathbf{I}_d\mathbf{A}^{-1}\mathbf{C}^{-1} = \mathbf{J}_2\mathbf{J}_1^H\mathbf{J}_1\mathbf{A}^{-1}\mathbf{C}^{-1} = \mathbf{J}_3\Gamma\mathbf{C}^{-1},$$

where  $\mathbf{J}_3 = \mathbf{J}_2\mathbf{J}_1^H$  is a phase-shift matrix as the set of phase-shift matrices is stable under matrix multiplication.  $\square$

**Lemma 6.** Let  $\mathbf{x} := (\mathbf{x}_t)_{t \in \mathbb{N}}$  be a process that satisfies Definition 2 and let  $\mathbf{X}$  be a  $\mathbb{C}^{T \times d}$ -valued,  $1 \leq d < T < \infty$ , sampled stochastic process generated by  $\mathbf{x}$ , and let  $\hat{g} : \mathbf{C} \mapsto (\mathbf{C} - \mathbf{I}_T \hat{\boldsymbol{\mu}}^\top) \hat{\Gamma}^\top$  be a corresponding finite sample solution defined in Lemma 4. Let  $\mathbf{B} \in \mathbb{C}^{d \times d}$  be nonsingular, let  $\mathbf{b} \in \mathbb{C}^d$ , let  $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{B}^\top + \mathbf{I}_T \mathbf{b}^\top$  and let  $\tilde{g} : \mathbf{C} \mapsto (\mathbf{C} - \mathbf{I}_T \tilde{\boldsymbol{\mu}}^\top) \tilde{\Gamma}^\top$  be a corresponding finite sample solution for  $\tilde{\mathbf{X}}$ . Then,  $\tilde{\Gamma} = \mathbf{J}\hat{\Gamma}\mathbf{B}^{-1}$  is a finite sample solution for  $\tilde{\mathbf{X}}$ , where  $\mathbf{J}$  is some phase-shift matrix.

**Proof of Lemma 6.** By Lemma 4, we can write  $\hat{\Gamma} = \hat{\mathbf{V}}^H \hat{\Sigma}_0$ , where  $\hat{\Sigma}_0$  denotes the conjugate symmetric inverse square root of  $\hat{S}_0$ . Note that the matrix-valued estimators are affine equivariant in the sense that

$$\hat{S}_j[\mathbf{X}\mathbf{B}^\top - \mathbf{1}_T \mathbf{b}^\top] = \mathbf{B}\hat{S}_j[\mathbf{X}]\mathbf{B}^H,$$

$j \in \{0, \tau\}$ , for all nonsingular  $\mathbb{C}^{d \times d}$ -matrices  $\mathbf{B}$  and all  $\mathbb{C}^d$ -vectors  $\mathbf{b}$ . We proceed to verify that  $\mathbf{J}\hat{\Gamma}\mathbf{B}^{-1}$  satisfies the criteria of a finite sample solution, given in Definition 4. Using the affine equivariance property, we get that

$$\mathbf{J}\hat{\Gamma}\mathbf{B}^{-1}\hat{S}_0[\tilde{\mathbf{X}}] \left( \mathbf{J}\hat{\Gamma}\mathbf{B}^{-1} \right)^H = \mathbf{J}\hat{\mathbf{V}}^H \hat{\Sigma}_0 \hat{S}_0[\mathbf{X}] \hat{\Sigma}_0 \hat{\mathbf{V}}^H = \mathbf{J}\mathbf{J}^H = \mathbf{I}_d$$

and

$$\mathbf{J}\hat{\mathbf{\Gamma}}\mathbf{B}^{-1}\hat{\mathbf{S}}_{\tau}[\tilde{\mathbf{X}}]\left(\mathbf{J}\hat{\mathbf{\Gamma}}\mathbf{B}^{-1}\right)^{\mathbf{H}} = \mathbf{J}\hat{\mathbf{V}}^{\mathbf{H}}\hat{\mathbf{\Sigma}}_0\hat{\mathbf{S}}_{\tau}[\mathbf{X}]\hat{\mathbf{\Sigma}}_0\hat{\mathbf{V}}\mathbf{J}^{\mathbf{H}} = \mathbf{J}\hat{\mathbf{\Lambda}}_{\tau}\mathbf{J}^{\mathbf{H}} = \hat{\mathbf{\Lambda}}_{\tau}.$$

Hereby,  $\mathbf{J}\hat{\mathbf{\Gamma}}\mathbf{B}^{-1}$  is a finite sample solution for  $\tilde{\mathbf{X}}$  and the proof is complete.  $\square$

**Lemma 7.** *Let  $\hat{\mathbf{\Sigma}}$  be a  $T$ -indexed sequence of  $\mathbb{C}^{d \times d}$ -valued nonsingular estimates and let  $\alpha_T$  be a real-valued sequence that satisfies  $\alpha_T \uparrow \infty$  as  $T \rightarrow \infty$ . Furthermore, let  $\alpha_T(\hat{\mathbf{\Sigma}} - \mathbf{I}_d) = \mathcal{O}_p(1)$ . Then the following two statements hold.*

$$(i) \quad \hat{\mathbf{\Sigma}}^a \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbf{I}_d, \text{ where } a \in \left\{-1, -\frac{1}{2}, 1\right\}.$$

$$(ii) \quad \alpha_T(\hat{\mathbf{\Sigma}}^a - \mathbf{I}_d) = \mathcal{O}_p(1), \text{ where } a \in \left\{-1, -\frac{1}{2}\right\}.$$

**Proof of Lemma 7.** The assumption  $\alpha_T(\hat{\mathbf{\Sigma}} - \mathbf{I}_d) = \mathcal{O}_p(1)$  implies that  $\hat{\mathbf{\Sigma}} - \mathbf{I}_d = (1/\alpha_T)\mathcal{O}_p(1) = o_p(1)\mathcal{O}_p(1) = o_p(1)$ , that is,  $\hat{\mathbf{\Sigma}} \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbf{I}_d$ . Note that the matrix inversion and the conjugate symmetric square root of the inversion are continuous transformations in the neighborhood of  $\mathbf{I}_d$ . Thus, we can apply the continuous mapping theorem which gives us that  $\hat{\mathbf{\Sigma}}^a \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbf{I}_d$ , as  $T \rightarrow \infty$ , for  $a \in \{-1, -\frac{1}{2}\}$ .

Using part (i) and Slutsky's lemma, we get that the inverse is uniformly tight, since

$$\begin{aligned} \alpha_T(\hat{\mathbf{\Sigma}}^{-1} - \mathbf{I}_d) &= (\hat{\mathbf{\Sigma}}^{-1} - \mathbf{I}_d)\alpha_T(\mathbf{I}_d - \hat{\mathbf{\Sigma}}) + \alpha_T(\mathbf{I}_d - \hat{\mathbf{\Sigma}}) \\ &= o_p(1)\mathcal{O}_p(1) + \mathcal{O}_p(1) = \mathcal{O}_p(1). \end{aligned}$$

For the final part, first note that

$$(\hat{\mathbf{\Sigma}}^{-1} - \mathbf{I}_d) = (\hat{\mathbf{\Sigma}}^{-\frac{1}{2}} - \mathbf{I}_d)(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}} + \mathbf{I}_d).$$

Since  $\hat{\mathbf{\Sigma}}^{-\frac{1}{2}} + \mathbf{I}_d$  converges in probability to  $2\mathbf{I}_d$  and the matrix inversion is a continuous transformation in the neighborhood of  $2\mathbf{I}_d$ , the continuous mapping theorem gives us that  $(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}} + \mathbf{I}_d)^{-1} \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \frac{1}{2}\mathbf{I}_d$ . Thus by Slutsky's lemma,

$$\alpha_T(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}} - \mathbf{I}_d) = \alpha_T(\hat{\mathbf{\Sigma}}^{-1} - \mathbf{I}_d)(\hat{\mathbf{\Sigma}}^{-\frac{1}{2}} + \mathbf{I}_d)^{-1} = \mathcal{O}_p(1). \quad \square$$

**Proof of Lemma 2.** Denote  $\hat{\mathbf{S}}_j := \hat{\mathbf{S}}_j[\mathbf{X}]$  and  $\mathbf{S}_j := \mathbf{S}_j[\mathbf{x}_t]$ , where  $j \in \{0, \tau\}$ . By Lemmas 5 and 6, it is sufficient to consider the trivial mixing scenario  $\mathbf{A} = \mathbf{I}_d$ . Note that  $\gamma_T(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}) = \gamma_T(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}}\mathbf{\Gamma}^{-1} - \mathbf{I}_d)\mathbf{\Gamma}$  and by Lemma 6 we have that  $\hat{\mathbf{J}}\hat{\mathbf{\Gamma}}\mathbf{\Gamma}^{-1}$  is a finite sample solution for  $\mathbf{X}\mathbf{\Gamma}^{\top}$ . Hereby, the trivial mixing can be generalized to any full-rank mixing matrix  $\mathbf{A}$ .

Under trivial mixing, we have that  $\alpha_T(\hat{\mathbf{S}}_0 - \mathbf{I}_d) = \mathcal{O}_p(1)$  and  $\beta_T(\hat{\mathbf{S}}_{\tau} - \mathbf{A}_{\tau}) = \mathcal{O}_p(1)$ . Furthermore, by Corollary 2 and Lemma 4, we have  $\hat{\mathbf{\Gamma}} = \hat{\mathbf{V}}^{\mathbf{H}}\hat{\mathbf{\Sigma}}_0$  and  $\mathbf{\Gamma} = \mathbf{V}^{\mathbf{H}}\mathbf{\Sigma}_0$ , where  $\hat{\mathbf{\Sigma}}_0 = \hat{\mathbf{S}}_0^{-1/2}$  and  $\mathbf{\Sigma}_0 = \mathbf{S}_0^{-1/2}$  are conjugate symmetric. Note that under trivial mixing we have  $\mathbf{S}_0^{-1/2} = \mathbf{I}_d$  and  $\mathbf{V} = \mathbf{J}$ , where  $\mathbf{J}$  is some phase-shift matrix.



We next denote  $\hat{\Sigma}_\tau = \hat{\Sigma}_0 \hat{S}_\tau \hat{\Sigma}_0$  and show that  $\gamma_T(\hat{\Sigma}_\tau - \Lambda_\tau) = \mathcal{O}_p(1)$ . The uniform tightness follows from Lemma 7, Slutsky's lemma and the factorization

$$\begin{aligned} \gamma_T(\hat{\Sigma}_\tau - \Lambda_\tau) &= \gamma_T \left[ (\hat{\Sigma}_0 - \mathbf{I}_d)(\hat{S}_\tau - \Lambda_\tau)(\hat{\Sigma}_0 - \mathbf{I}_d) + (\hat{\Sigma}_0 - \mathbf{I}_d)(\hat{S}_\tau - \Lambda_\tau) \right. \\ &\quad + (\hat{\Sigma}_0 - \mathbf{I}_d)\Lambda_\tau(\hat{\Sigma}_0 - \mathbf{I}_d) + (\hat{\Sigma}_0 - \mathbf{I}_d)\Lambda_\tau + (\hat{S}_\tau - \Lambda_\tau) \\ &\quad \left. + (\hat{S}_\tau - \Lambda_\tau)(\hat{\Sigma}_0 - \mathbf{I}_d) + \Lambda_\tau(\hat{\Sigma}_0 - \mathbf{I}_d) \right] \\ &= o_p(1) + \gamma_T \left[ (\hat{\Sigma}_0 - \mathbf{I}_d)\Lambda_\tau + (\hat{S}_\tau - \Lambda_\tau) + \Lambda_\tau(\hat{\Sigma}_0 - \mathbf{I}_d) \right] \\ &= \mathcal{O}_p(1). \end{aligned}$$

Next, recall the eigendecomposition  $\hat{\Sigma}_\tau = \hat{\mathbf{V}} \hat{\Lambda}_\tau \hat{\mathbf{V}}^H$  and that the space of unitary matrices is compact. Consequently,  $\mathbf{U} = \mathcal{O}_p(1)$  for any unitary  $\mathbf{U}$ . By right-multiplying both sides of the eigendecomposition with  $\hat{\mathbf{V}}$  and by subtracting  $\Lambda_\tau$  from both sides, we get

$$\hat{\Sigma}_\tau \hat{\mathbf{V}} - \Lambda_\tau = \hat{\mathbf{V}} \hat{\Lambda}_\tau - \Lambda_\tau,$$

where both sides of the equation can be further factorized as

$$(\hat{\Sigma}_\tau - \Lambda_\tau) \hat{\mathbf{V}} + \Lambda_\tau (\hat{\mathbf{V}} - \mathbf{I}_d) = (\hat{\mathbf{V}} - \mathbf{I}_d) \hat{\Lambda}_\tau + (\hat{\Lambda}_\tau - \Lambda_\tau),$$

and multiplying by  $\gamma_T$  and rearranging the terms yield

$$\gamma_T \left[ \Lambda_\tau (\hat{\mathbf{V}} - \mathbf{I}_d) - (\hat{\mathbf{V}} - \mathbf{I}_d) \hat{\Lambda}_\tau - (\hat{\Lambda}_\tau - \Lambda_\tau) \right] = \mathcal{O}_p(1). \quad (13)$$

By assumption, the diagonal matrix  $\Lambda_\tau$  has distinct real-valued diagonal elements. Furthermore, since the matrix  $\hat{\Lambda}_\tau$  is obtained by estimating an eigendecomposition, it is also diagonal, with diagonal elements denoted as  $\hat{\lambda}_\tau^{(1)}, \dots, \hat{\lambda}_\tau^{(d)}$ . Then, consider the element  $(j, k)$ ,  $j \neq k$ , of Equation (13):

$$\gamma_T \hat{\mathbf{V}}_{jk} (\lambda_\tau^{(j)} - \hat{\lambda}_\tau^{(k)}) = \mathcal{O}_p(1). \quad (14)$$

Since  $\gamma_T(\hat{\Sigma}_\tau - \Lambda_\tau) = \mathcal{O}_p(1)$ , we get that  $\hat{\lambda}_\tau^{(k)}$  converges in probability to  $\lambda_\tau^{(k)}$ . Furthermore, since the diagonal elements of  $\Lambda_\tau$  are distinct, we can divide both sides of Equation (14) by  $\lambda_\tau^{(j)} - \hat{\lambda}_\tau^{(k)}$ , which gives us that  $\gamma_T[\hat{\mathbf{V}}]_{jk} = \mathcal{O}_p(1)$  holds asymptotically when  $j \neq k$ .

Next, let  $\hat{\mathbf{J}}$  be a  $T$ -indexed sequence of phase-shift matrices, such that the diagonal entries of the product  $\hat{\mathbf{V}} \hat{\mathbf{J}}^H$  are in the positive real axis for every  $T \in \mathbb{N}$ . Note that any complex number can be expressed in the form  $r \exp(i\theta)$ , where  $\theta$  is the phase and  $r$  is the modulus, i.e., the length of the complex number. In the matrix case, we can similarly express  $\hat{\mathbf{V}}$  so that  $\hat{\mathbf{V}} = \hat{\mathbf{V}}_r \hat{\mathbf{V}}_\theta$ , where  $\hat{\mathbf{V}}_\theta$  is a phase-shift matrix that contains the phases of the diagonal elements of  $\hat{\mathbf{V}}$  and consequently  $\hat{\mathbf{V}}_r$  is a matrix with real-valued diagonal entries. Then, given a phase-shift matrix  $\hat{\mathbf{V}}_\theta = \text{diag}(\exp(i\theta_1), \dots, \exp(i\theta_d))$ , we can always construct a matrix  $\hat{\mathbf{J}}^H =$



$\text{diag}(\exp(-i\theta_1), \dots, \exp(-i\theta_d))$  such that the product  $\hat{\mathbf{V}}\hat{\mathbf{J}}^H$  has real-valued diagonal entries.

After the diagonal entries have been rotated to the positive real axis, the rotated diagonal elements are equal to the corresponding moduli, that is,  $\hat{\mathbf{V}}_{kk}\hat{\mathbf{J}}_{kk}^H = |\hat{\mathbf{V}}_{kk}|$ . Then, since  $\hat{\mathbf{V}}$  is unitary, each of its row vectors has length one and the absolute value of a single element is at most one, which gives us

$$\left|1 - \hat{\mathbf{V}}_{kk}\hat{\mathbf{J}}_{kk}^H\right| = 1 - \hat{\mathbf{V}}_{kk}\hat{\mathbf{J}}_{kk}^H = 1 - |\hat{\mathbf{V}}_{kk}| = 1 - \sqrt{1 - \sum_{\substack{h=1 \\ h \neq k}}^d |\hat{\mathbf{V}}_{hk}|^2}.$$

By unitarity of  $\hat{\mathbf{V}}$ , the above square root is always between zero and one, and thus squaring the square root produces a smaller or equal number. Hereby, using the (asymptotic) uniform tightness of the off-diagonal elements, we get that asymptotically

$$\left|1 - \hat{\mathbf{V}}_{kk}\hat{\mathbf{J}}_{kk}^H\right| \leq 1 - 1 + \sum_{\substack{h=1 \\ h \neq k}}^d |\hat{\mathbf{V}}_{hk}|^2 = \sum_{\substack{h=1 \\ h \neq k}}^d |\hat{\mathbf{V}}_{hk}|^2 = \mathcal{O}_p(1/\gamma_T^2).$$

Hereby, by combining the results for the diagonal and off-diagonal elements, we get that there exists a sequence of phase-shift matrices  $\hat{\mathbf{J}}$  such that  $\gamma_T(\hat{\mathbf{V}}\hat{\mathbf{J}}^H - \mathbf{I}_d) = \mathcal{O}_p(1)$  holds asymptotically.

The claim of the lemma can then be written as

$$\begin{aligned} \gamma_T [\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d] &= \gamma_T \left[ \left( (\hat{\mathbf{V}}\hat{\mathbf{J}}^H)^H \hat{\mathbf{\Sigma}}_0 - \mathbf{I}_d \right) \right] \\ &= \left( \hat{\mathbf{V}}\hat{\mathbf{J}}^H \right)^H \gamma_T (\hat{\mathbf{\Sigma}}_0 - \mathbf{I}_d) + \gamma_T \left( \hat{\mathbf{V}}\hat{\mathbf{J}}^H - \mathbf{I}_d \right)^H \\ &= \mathcal{O}_p(1)\mathcal{O}_p(1) + \mathcal{O}_p(1) = \mathcal{O}_p(1). \end{aligned} \quad \square$$

**Proof of Theorem 2.** Let  $\hat{\mathbf{S}}_j := \hat{\mathbf{S}}_j[\mathbf{X}]$ ,  $\tau \in \{0, \tau\}$ , and recall the equations from Definition 4

$$\hat{\mathbf{\Gamma}}\hat{\mathbf{S}}_0\hat{\mathbf{\Gamma}}^H = \mathbf{I}_d \quad \text{and} \quad \hat{\mathbf{\Gamma}}\hat{\mathbf{S}}_\tau\hat{\mathbf{\Gamma}}^H = \hat{\mathbf{\Lambda}}_\tau. \quad (15)$$

In order to simplify the notation, we denote  $\hat{\mathbf{G}} := \hat{\mathbf{J}}\hat{\mathbf{\Gamma}}$ , where  $\hat{\mathbf{J}}$  is a  $T$ -indexed sequence of phase-shift matrices such that the diagonal elements of  $\hat{\mathbf{J}}\hat{\mathbf{\Gamma}}$  are on the positive real axis. Under Lemma 2, we have that  $\gamma_T(\hat{\mathbf{G}} - \mathbf{I}_d) = \mathcal{O}_p(1)$  and note that both parts of Equation (15) also hold for  $\hat{\mathbf{G}}$ .

The left part of Equation (15) can then be expanded as

$$(\hat{\mathbf{G}} - \mathbf{I}_d)\hat{\mathbf{S}}_0\hat{\mathbf{G}}^H + (\hat{\mathbf{S}}_0 - \mathbf{I}_d)\hat{\mathbf{G}}^H + (\hat{\mathbf{G}}^H - \mathbf{I}_d) = 0,$$

where the left-hand side can be further expanded as

$$(\hat{\mathbf{G}} - \mathbf{I}_d)(\hat{\mathbf{S}}_0 - \mathbf{I}_d)(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\hat{\mathbf{G}} - \mathbf{I}_d)(\hat{\mathbf{S}}_0 - \mathbf{I}_d) + (\hat{\mathbf{G}} - \mathbf{I}_d)(\hat{\mathbf{G}}^H - \mathbf{I}_d)$$

$$\begin{aligned}
& + (\hat{\mathbf{G}} - \mathbf{I}_d) + (\hat{\mathbf{S}}_0 - \mathbf{I}_d)(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\hat{\mathbf{S}}_0 - \mathbf{I}_d) + (\hat{\mathbf{G}}^H - \mathbf{I}_d) \\
& = (\hat{\mathbf{G}} - \mathbf{I}_d) + (\hat{\mathbf{S}}_0 - \mathbf{I}_d) + (\hat{\mathbf{G}}^H - \mathbf{I}_d) + \mathcal{O}_p(1/\gamma_T^2).
\end{aligned}$$

By rearranging the terms, we get the form

$$(\hat{\mathbf{G}}^H - \mathbf{I}_d) = (\mathbf{I}_d - \hat{\mathbf{G}}) + (\mathbf{I}_d - \hat{\mathbf{S}}_0) + \mathcal{O}_p(1/\gamma_T^2). \quad (16)$$

Similarly, the right part of Equation (15) can be expanded as

$$(\hat{\mathbf{G}} - \mathbf{I}_d)\hat{\mathbf{S}}_\tau\hat{\mathbf{G}}^H + (\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau)\hat{\mathbf{G}}^H + \mathbf{\Lambda}_\tau(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\mathbf{\Lambda}_\tau - \hat{\mathbf{\Lambda}}_\tau) = 0,$$

where the left-hand side can be further expanded as

$$\begin{aligned}
& (\hat{\mathbf{G}} - \mathbf{I}_d)(\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau)(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\hat{\mathbf{G}} - \mathbf{I}_d)(\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau) \\
& + (\hat{\mathbf{G}} - \mathbf{I}_d)\mathbf{\Lambda}_\tau(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\hat{\mathbf{G}} - \mathbf{I}_d)\mathbf{\Lambda}_\tau + (\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau)(\hat{\mathbf{G}}^H - \mathbf{I}_d) \\
& + (\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau) + \mathbf{\Lambda}_\tau(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\mathbf{\Lambda}_\tau - \hat{\mathbf{\Lambda}}_\tau) \\
& = (\hat{\mathbf{G}} - \mathbf{I}_d)\mathbf{\Lambda}_\tau + (\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau) + \mathbf{\Lambda}_\tau(\hat{\mathbf{G}}^H - \mathbf{I}_d) + (\mathbf{\Lambda}_\tau - \hat{\mathbf{\Lambda}}_\tau) + \mathcal{O}_p(1/\gamma_T^2).
\end{aligned}$$

Hereby we obtain

$$(\hat{\mathbf{G}} - \mathbf{I}_d)\mathbf{\Lambda}_\tau + (\hat{\mathbf{S}}_\tau - \mathbf{\Lambda}_\tau) + \mathbf{\Lambda}_\tau(\hat{\mathbf{G}}^H - \mathbf{I}_d) = (\hat{\mathbf{\Lambda}}_\tau - \mathbf{\Lambda}_\tau) + \mathcal{O}_p(1/\gamma_T^2),$$

which is, by using the expression for  $\hat{\mathbf{G}}^H - \mathbf{I}_d$  given by Equation (16), equivalent to

$$\hat{\mathbf{G}}\mathbf{\Lambda}_\tau - \mathbf{\Lambda}_\tau\hat{\mathbf{G}} = (\mathbf{\Lambda}_\tau - \hat{\mathbf{S}}_\tau) + \mathbf{\Lambda}_\tau(\hat{\mathbf{S}}_0 - \mathbf{I}_d) + (\hat{\mathbf{\Lambda}}_\tau - \mathbf{\Lambda}_\tau) + \mathcal{O}_p(1/\gamma_T^2). \quad (17)$$

Recall that  $\hat{\mathbf{G}} = \hat{\mathbf{J}}\hat{\mathbf{\Gamma}}$  and that  $\mathbf{\Lambda}_\tau, \hat{\mathbf{\Lambda}}_\tau$  are both diagonal matrices with real-valued diagonal elements. Then, by rearranging the terms and by considering the element  $(j, j)$  of Equation (16), we get that

$$\hat{\mathbf{J}}_{jj}\hat{\mathbf{\Gamma}}_{jj} - 1 = \frac{1}{2} \left( 1 - [\hat{\mathbf{S}}_0]_{jj} \right) + \mathcal{O}_p(1/\gamma_T^2).$$

Similarly, by considering the element  $(j, k)$ ,  $j \neq k$ , of Equation (17), we get that

$$(\lambda_\tau^{(k)} - \lambda_\tau^{(j)})\hat{\mathbf{J}}_{jj}\hat{\mathbf{\Gamma}}_{jk} = \lambda_\tau^{(j)} [\hat{\mathbf{S}}_0]_{jk} - [\hat{\mathbf{S}}_\tau]_{jk} + \mathcal{O}_p(1/\gamma_T^2).$$

The theorem then follows by multiplying both sides by  $\gamma_T$ .  $\square$

**Proof of Theorem 4.** In order to incorporate the shift  $\tau$  given in Definition 2, we first introduce some notation. For  $\boldsymbol{\eta}_\cdot$  given in Assumption 1, we set  $\mathbf{y}_\cdot^\top = (\boldsymbol{\eta}_\cdot^\top \quad \boldsymbol{\eta}_{\cdot+\tau}^\top)$ . We also introduce functions  $g_j(\mathbf{y}_\cdot) : \mathbb{R}^{2\ell} \rightarrow \mathbb{R}$ ,  $j \in \{1, 2, \dots, 4d\}$ , through relations  $g_j(\mathbf{y}_\cdot) = \tilde{f}_j(\boldsymbol{\eta}_\cdot)$ ,  $j \in \{1, 2, \dots, 2d\}$  and  $g_{j+2d}(\mathbf{y}_\cdot) = \tilde{f}_j(\boldsymbol{\eta}_{\cdot+\tau})$ ,  $j \in \{1, 2, \dots, 2d\}$ . That is, for  $j \geq 2d + 1$  the function  $g_j$  corresponds to  $\tilde{f}_j$  evaluated at shift  $\boldsymbol{\eta}_{\cdot+\tau}$ .

The unsymmetrized autocovariance matrix estimator with lag  $\tau$  is defined as

$$\tilde{\mathbf{S}}_\tau[\mathbf{X}] = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} (\mathbf{X}_t - \hat{\boldsymbol{\mu}}) (\mathbf{X}_{t+\tau} - \hat{\boldsymbol{\mu}})^H,$$

where  $\hat{\boldsymbol{\mu}} := \hat{\boldsymbol{\mu}}[\mathbf{X}]$ . Let  $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_T \boldsymbol{\mu}^\top$  and  $\tilde{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ . Under Assumption 1, we have that the  $k$ th component  $\tilde{X}_t^{(k)}$  has the same asymptotic autocovariance function as  $g_k(\mathbf{y}_t) + i g_{k+d}(\mathbf{y}_t) - \mathbb{E}[g_k(\mathbf{y}_t) + i g_{k+d}(\mathbf{y}_t)]$ . To improve the fluency of the proof, we denote  $f_k(\mathbf{y}_t) := g_k(\mathbf{y}_t) - \mathbb{E}[g_k(\mathbf{y}_t)]$ .

We can reformulate the estimator as

$$\tilde{\mathbf{S}}_\tau[\mathbf{X}] = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} \left[ \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t+\tau}^H - \tilde{\mathbf{X}}_t \tilde{\boldsymbol{\mu}}^H - \tilde{\boldsymbol{\mu}} \tilde{\mathbf{X}}_{t+\tau}^H + \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H \right],$$

where the last three terms of the sum are equal to

$$\frac{1}{T - \tau} \left[ \sum_{t=T-\tau+1}^T \tilde{\mathbf{X}}_t \tilde{\boldsymbol{\mu}}^H + \sum_{t=1}^{\tau} \tilde{\boldsymbol{\mu}} \tilde{\mathbf{X}}_t^H - (T + \tau) \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H \right] = \mathcal{O}_p(1/T). \quad (18)$$

Equation (18) holds, since the first two terms are finite sums. Furthermore, since  $\mathbb{E}[\tilde{\boldsymbol{\mu}}] = \mathbf{0}$ , we get that under Assumption 1 the  $k$ th component  $\tilde{\mu}^{(k)}$  is asymptotically equivalent to  $(1/T) \sum_{t=1}^T [f_k(\mathbf{y}_t) - \mathbb{E}[f_k(\mathbf{y}_t)] + i(f_{k+d}(\mathbf{y}_t) - \mathbb{E}[f_{k+d}(\mathbf{y}_t)])]$ . We can then directly apply Theorem 3, which in combination with Prohorov's theorem and Corollary 1 gives that  $\tilde{\boldsymbol{\mu}} = \mathcal{O}_p(1/\sqrt{T})$ , and consequently the last term of Equation (18) is  $\mathcal{O}_p(1/T)$ .

The symmetrized autocovariance estimator can then be expressed as

$$\hat{\mathbf{S}}_\tau[\mathbf{X}] = \frac{1}{2(T - \tau)} \sum_{t=1}^{T-\tau} \left( \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t+\tau}^H + \tilde{\mathbf{X}}_{t+\tau} \tilde{\mathbf{X}}_t^H \right) + \mathcal{O}_p(1/T),$$

and the existence of fourth moments and the model assumptions given in Definition 2 give us that

$$\mathbb{E}[\hat{\mathbf{S}}_\tau[\mathbf{X}]] = \frac{1}{2} \left( \ddot{\mathbf{\Lambda}}_\tau + \ddot{\mathbf{\Lambda}}_\tau^H \right) + \mathcal{O}(1/T),$$

where  $\ddot{\mathbf{\Lambda}}_0 = \mathbf{I}_d$ . Hereby, we have that

$$\sqrt{T} \left( \hat{\mathbf{S}}_0[\mathbf{X}] - \mathbf{I}_d \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H - \mathbb{E}[\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H] \right) + o_p(1).$$

Note that asymptotically it is indifferent whether we scale the autocovariance estimators with  $1/T$  or  $1/(T - \tau)$ . Thus, for the fixed  $\tau$ , that satisfies Definition 2, we have that  $\mathbb{E}[\mathbf{\Lambda}_\tau \hat{\mathbf{S}}_0[\mathbf{X}] - \hat{\mathbf{S}}_\tau[\mathbf{X}]] = \mathcal{O}(1/T)$ , which gives that

$$\sqrt{T} \left( \mathbf{\Lambda}_\tau \hat{\mathbf{S}}_0[\mathbf{X}] - \hat{\mathbf{S}}_\tau[\mathbf{X}] \right)$$

is equal to

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{1}{2} \left( 2\mathbf{\Lambda}_\tau \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H - \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t+\tau}^H - \tilde{\mathbf{X}}_{t+\tau} \tilde{\mathbf{X}}_t^H \right. \\ \left. - \mathbb{E} \left[ 2\mathbf{\Lambda}_\tau \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H - \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t+\tau}^H - \tilde{\mathbf{X}}_{t+\tau} \tilde{\mathbf{X}}_t^H \right] \right) + o_p(1),$$

such that in the above expression the terms  $\tilde{\mathbf{X}}_{k+\tau} = \mathbf{0}$  when  $k + \tau > T$ .

Let  $\hat{\mathbf{J}}$  be the  $T$ -indexed sequence that sets the diagonal elements of  $\hat{\mathbf{\Gamma}}$  to the positive real axis. By Theorem 2 and Assumption 1, we have that the diagonal element  $(j, j)$  of  $\sqrt{T}(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d)$  is asymptotically equivalent to  $\mathbf{H}_{jj} + o_p(1)$  defined as

$$\mathbf{H}_{jj} = \frac{1}{2\sqrt{T}} \sum_{t=1}^T \left( h_{j,j}(\mathbf{y}_t) - \mathbb{E}[h_{j,j}(\mathbf{y}_t)] + i \left( \tilde{h}_{j,j}(\mathbf{y}_t) - \mathbb{E}[\tilde{h}_{j,j}(\mathbf{y}_t)] \right) \right),$$

such that for  $j \in \{1, \dots, d\}$  we have  $h_{j,j}(\mathbf{y}_t) = (f_j(\mathbf{y}_t))^2 + (f_{j+d}(\mathbf{y}_t))^2$  and  $\tilde{h}_{j,j}(\mathbf{y}_t) = 0$ .

The off-diagonal element  $(j, k)$ ,  $j \neq k$ , of  $\sqrt{T}(\hat{\mathbf{J}}\hat{\mathbf{\Gamma}} - \mathbf{I}_d)$  is asymptotically equivalent to  $\mathbf{H}_{jk} + o_p(1)$ , where  $\mathbf{H}_{jk}$  is equal to

$$\frac{1}{\sqrt{T}(\lambda_\tau^{(k)} - \lambda_\tau^{(j)})} \sum_{t=1}^T \left( h_{j,k}(\mathbf{y}_t) - \mathbb{E}[h_{j,k}(\mathbf{y}_t)] + i \left( \tilde{h}_{j,k}(\mathbf{y}_t) - \mathbb{E}[\tilde{h}_{j,k}(\mathbf{y}_t)] \right) \right),$$

where

$$h_{j,k}(\mathbf{y}_t) = \lambda_\tau^{(j)} [f_j(\mathbf{y}_t) f_k(\mathbf{y}_t) + f_{j+d}(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t)] - \frac{1}{2} [f_j(\mathbf{y}_t) f_{k+2d}(\mathbf{y}_t) \\ + f_{j+d}(\mathbf{y}_t) f_{k+3d}(\mathbf{y}_t) + f_{j+2d}(\mathbf{y}_t) f_k(\mathbf{y}_t) + f_{j+3d}(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t)]$$

and

$$\tilde{h}_{j,k}(\mathbf{y}_t) = \lambda_\tau^{(j)} [f_{j+d}(\mathbf{y}_t) f_k(\mathbf{y}_t) - f_j(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t)] - \frac{1}{2} [f_{j+d}(\mathbf{y}_t) f_{k+2d}(\mathbf{y}_t) \\ + f_{j+3d}(\mathbf{y}_t) f_k(\mathbf{y}_t) - f_j(\mathbf{y}_t) f_{k+3d}(\mathbf{y}_t) - f_{j+2d}(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t)].$$

Let  $\Re[\mathbf{H}]$  be the real part and  $\Im[\mathbf{H}]$  be the imaginary part of the matrix  $\mathbf{H}$ , and let  $\mathbf{v} = \text{vec}(\Re[\mathbf{H}] \quad \Im[\mathbf{H}])$ , such that the  $\mathbb{C}^{2d^2}$ -valued vector  $\mathbf{v}$  first contains the columns of the real part and then the columns of the imaginary part.

Under Assumption 1, we have that  $f_j(\mathbf{y}_t)$  has finite fourth moments for every  $j \in \{1, \dots, 4d\}$  and every  $t \in \mathbb{N}$ . Hereby, the Cauchy–Schwarz inequality gives that every  $h_{j,k}(\mathbf{y}_t)$  and  $\tilde{h}_{j,k}(\mathbf{y}_t)$  are square-integrable for every  $t \in \mathbb{N}$ . Furthermore, covariances and cross-covariances of the Gaussian process  $\mathbf{y}$ , are summable. Hereby, we can apply Theorem 3 for  $\mathbf{v}$ , which gives that

$$\mathbf{v} \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \boldsymbol{\rho} \sim \mathcal{N}_{2d^2}(\mathbf{0}, \boldsymbol{\Sigma}_\rho), \quad \text{where} \quad \boldsymbol{\Sigma}_\rho = \begin{pmatrix} \boldsymbol{\Sigma}_{\rho_1} & \boldsymbol{\Sigma}_{\rho_{12}} \\ \boldsymbol{\Sigma}_{\rho_{12}}^\top & \boldsymbol{\Sigma}_{\rho_2} \end{pmatrix}.$$

Corollary 1 then gives us

$$\sqrt{T} \cdot \text{vec} \left( \hat{\mathbf{J}} \hat{\mathbf{\Gamma}} - \mathbf{I}_d \right) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathbf{v} \sim \mathcal{N}_{d^2}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{v}}, \mathbf{P}_{\mathbf{v}}),$$

where  $\mathbf{\Sigma}_{\mathbf{v}} = \mathbf{\Sigma}_{\rho_1} + \mathbf{\Sigma}_{\rho_2} + i(\mathbf{\Sigma}_{\rho_{12}}^\top - \mathbf{\Sigma}_{\rho_{12}})$  and  $\mathbf{P}_{\mathbf{v}} = \mathbf{\Sigma}_{\rho_1} - \mathbf{\Sigma}_{\rho_2} + i(\mathbf{\Sigma}_{\rho_{12}}^\top + \mathbf{\Sigma}_{\rho_{12}})$ . Denoting  $\mathbf{v}^\top = (v_{1,1} \ v_{2,1} \cdots v_{d,d})$ , we get that  $\mathbf{\Sigma}_{\mathbf{v}}$  has entries of the form

$$\begin{aligned} \mathbb{E} [v_{j,k} v_{l,m}^*] &= S_0 [h_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] + S_0 [\tilde{h}_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \\ &\quad + \sum_{\tau=1}^{\infty} \left( R_\tau [h_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] + R_\tau [\tilde{h}_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \right) \\ &\quad + i \left( S_0 [\tilde{h}_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] - S_0 [h_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \right) \\ &\quad + i \sum_{\tau=1}^{\infty} \left( R_\tau [\tilde{h}_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] - R_\tau [h_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \right), \end{aligned}$$

where  $S_0$  and  $R_\tau$  are

$$\begin{aligned} \ddot{S}_\tau [x_t, y_s] &= \mathbb{E} [(x_t - \mu_{x_t})(y_{s+\tau} - \mu_{y_{s+\tau}})^*] \quad \text{and} \\ R_\tau [x_t, y_s] &= \ddot{S}_\tau [x_t, y_s] + \ddot{S}_\tau [y_s, x_t]. \end{aligned}$$

The elements of the relation matrix  $\mathbf{P}_{\mathbf{v}}$  have the form

$$\begin{aligned} \mathbb{E} [v_{j,k} v_{l,m}] &= S_0 [h_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] - S_0 [\tilde{h}_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \\ &\quad + \sum_{\tau=1}^{\infty} \left( R_\tau [h_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] - R_\tau [\tilde{h}_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \right) \\ &\quad + i \left( S_0 [\tilde{h}_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] + S_0 [h_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \right) \\ &\quad + i \sum_{\tau=1}^{\infty} \left( R_\tau [\tilde{h}_{j,k}(\mathbf{y}_1), h_{l,m}(\mathbf{y}_1)] + R_\tau [h_{j,k}(\mathbf{y}_1), \tilde{h}_{l,m}(\mathbf{y}_1)] \right). \end{aligned}$$

Note that the diagonal elements of  $\mathbf{\Sigma}_{\mathbf{v}}$  are real-valued and recall that  $\tilde{h}_{j,j} = 0$  for every  $j \in \{1, \dots, d\}$ .  $\square$

**Lemma 8.** *Let Assumptions 2 and 3 be satisfied. Let  $\hat{\boldsymbol{\mu}} := \hat{\boldsymbol{\mu}}[X]$  and  $\tilde{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}$ . Then for every element  $(j, k)$ , such that  $j, k \in \{1, 2, \dots, d\}$  and  $j \neq k$ , we have  $[T^\gamma \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H]_{jk} \xrightarrow[T \rightarrow \infty]{\mathcal{L}_1} 0$ , where  $\xrightarrow{\mathcal{L}_1}$  denotes convergence in the space  $\mathcal{L}_1(\Omega, \mathbb{P})$ . Consequently, the convergence also holds in probability.*

**Proof of Lemma 8.** Recall that the real part of the  $k$ th component  $\tilde{\boldsymbol{\mu}}^{(k)}$  is given by

$$\frac{1}{T} \sum_{k=1}^T \left[ \tilde{f}_k(\boldsymbol{\eta}_t^{(k)}) - \mathbb{E} \left( \tilde{f}_k(\boldsymbol{\eta}_t^{(k)}) \right) \right].$$

As in the proof of Proposition 2, we may assume  $k \in I$  and that  $q_{1,k}(2H_k - 2) > -1$ , in which case we have

$$\begin{aligned} \mathbb{E} \left[ \left( \Re(\tilde{\mu}^{(k)}) \right)^2 \right] &\leq \frac{c_1}{T} \sum_{t=1}^T \left| S_0 \left[ \tilde{f}_k(\eta_t^{(k)}), \tilde{f}_k(\eta_1^{(k)}) \right] \right| \\ &\leq \frac{c_2}{T} \sum_{t=1}^T t^{q_{1,k}(2H_k-2)} \leq c_3 T^{q_{1,k}(2H_k-2)}, \end{aligned}$$

where  $c_1, c_2, c_3 \in (0, \infty)$ . Similarly, for  $j \in I$  and  $c_4 \in (0, \infty)$ , we obtain

$$\mathbb{E} \left[ \left( \Re(\tilde{\mu}^{(j)}) \right)^2 \right] \leq c_4 T^{q_{1,j}(2H_j-2)}.$$

Hereby, the Cauchy–Schwartz inequality yields

$$\begin{aligned} \mathbb{E} \left[ \left| T^\gamma \Re(\tilde{\mu}^{(k)}) \Re(\tilde{\mu}^{(j)}) \right| \right] &\leq T^\gamma \sqrt{\mathbb{E} \left[ \left( \Re(\tilde{\mu}^{(k)}) \right)^2 \right] \mathbb{E} \left[ \left( \Re(\tilde{\mu}^{(j)}) \right)^2 \right]} \\ &\leq c_5 T^\gamma T^{q_{1,k}(H_k-1)+q_{1,j}(H_j-1)}, \end{aligned}$$

which converges to zero by Equation (19) and  $c_5 \in (0, \infty)$ . Note that identical arguments remain valid for the imaginary parts. Finally, the short-range dependency cases and the boundary cases  $q_{1,k}(2H_k - 2) = -1$  can be treated similarly as in the proof of Proposition 2.  $\square$

**Proof of Proposition 2.** With  $\gamma$  given by Equation (9), Condition (7) translates into

$$\gamma < \min_{j,k \in I} \{q_{1,k}(1 - H_k) + q_{1,j}(1 - H_j), 1/2\}. \quad (19)$$

Using independence of the processes  $\eta^{(k)}$  and  $\eta^{(j)}$  together with straightforward computations, we obtain

$$\begin{aligned} &\mathbb{E} \left[ T^{\gamma-1} \sum_{t=1}^T \left[ \left( \tilde{f}_k(\eta_t^{(k)}) - \mathbb{E} \left( \tilde{f}_k(\eta_t^{(k)}) \right) \right) \left( \tilde{f}_j(\eta_t^{(j)}) - \mathbb{E} \left( \tilde{f}_j(\eta_t^{(j)}) \right) \right) \right] \right]^2 \\ &= T^{2\gamma-2} \sum_{t,s=1}^T S_0 \left[ \tilde{f}_k(\eta_t^{(k)}), \tilde{f}_k(\eta_s^{(k)}) \right] S_0 \left[ \tilde{f}_j(\eta_t^{(j)}), \tilde{f}_j(\eta_s^{(j)}) \right] \\ &\leq c_1 T^{2\gamma-1} \sum_{t=1}^T \left| S_0 \left[ \tilde{f}_k(\eta_t^{(k)}), \tilde{f}_k(\eta_1^{(k)}) \right] S_0 \left[ \tilde{f}_j(\eta_t^{(j)}), \tilde{f}_j(\eta_1^{(j)}) \right] \right|, \end{aligned}$$

where  $c_1 \in (0, \infty)$  and the inequality follows from change of variables and the Fubini–Tonelli theorem. Note that if

$$\sum_{t=1}^{\infty} \left| S_0 \left( \tilde{f}_k(\eta_t^{(k)}), \tilde{f}_k(\eta_1^{(k)}) \right) S_0 \left( \tilde{f}_j(\eta_t^{(j)}), \tilde{f}_j(\eta_1^{(j)}) \right) \right| < +\infty,$$

the claim follows, since by Equation (19) we have that  $\gamma < 1/2$ . In particular, this is the case if  $f_k(\boldsymbol{\eta}^{(k)})$  or  $f_j(\boldsymbol{\eta}^{(j)})$  is short-range dependent. Hence we may assume that  $j, k \in I$  and that  $q_{1,k}(2H_k - 2) + q_{1,j}(2H_j - 2) \geq -1$ . Let first  $q_{1,k}(2H_k - 2) + q_{1,j}(2H_j - 2) > -1$ . In this case we get

$$\begin{aligned} & T^{2\gamma-1} \sum_{t=1}^T \left| S_0 \left( \tilde{f}_k(\boldsymbol{\eta}_t^{(k)}), \tilde{f}_k(\boldsymbol{\eta}_1^{(k)}) \right) S_0 \left( \tilde{f}_j(\boldsymbol{\eta}_t^{(j)}), \tilde{f}_j(\boldsymbol{\eta}_1^{(j)}) \right) \right| \\ & \leq c_2 T^{2\gamma-1} \sum_{t=1}^T t^{q_{1,k}(2H_k-2)+q_{1,j}(2H_j-2)} \\ & \leq c_2 T^{2\gamma+q_{1,k}(2H_k-2)+q_{1,j}(2H_j-2)}, \end{aligned}$$

which converges to zero by Equation (19) and  $c_2 \in (0, \infty)$ . Similarly, if  $q_{1,k}(2H_k - 2) + q_{1,j}(2H_j - 2) = -1$ , we obtain

$$\begin{aligned} & T^{2\gamma-1} \sum_{t=1}^T \left| S_0 \left( \tilde{f}_k(\boldsymbol{\eta}_t^{(k)}), \tilde{f}_k(\boldsymbol{\eta}_1^{(k)}) \right) S_0 \left( \tilde{f}_j(\boldsymbol{\eta}_t^{(j)}), \tilde{f}_j(\boldsymbol{\eta}_1^{(j)}) \right) \right| \\ & \leq c_3 T^{2\gamma-1} \sum_{t=1}^T t^{-1} \leq c_4 T^{2\gamma-1} \log T, \end{aligned}$$

which converges to zero, since  $\gamma < \frac{1}{2}$  and  $c_3, c_4 \in (0, \infty)$ . This verifies Equation (10). Treating Equation (11) similarly concludes the proof.  $\square$

**Proof of Theorem 5.** Similarly as in the proof of Theorem 4, we set

$$\mathbf{y}^\top = \left( \boldsymbol{\eta}^{(1)} \quad \boldsymbol{\eta}^{(2)} \quad \dots \quad \boldsymbol{\eta}^{(2d)} \quad \boldsymbol{\eta}_{+\tau}^{(1)} \quad \boldsymbol{\eta}_{+\tau}^{(2)} \quad \dots \quad \boldsymbol{\eta}_{+\tau}^{(2d)} \right).$$

We also introduce functions  $g_j(\mathbf{y}) : \mathbb{R}^{4d} \rightarrow \mathbb{R}$ ,  $j \in \{1, 2, \dots, 4d\}$ , through relations  $g_j(\mathbf{y}) = \tilde{f}_j(\boldsymbol{\eta}^{(j)})$ ,  $j \in \{1, 2, \dots, 2d\}$ , and  $g_{j+2d}(\mathbf{y}) = \tilde{f}_j(\boldsymbol{\eta}_{+\tau}^{(j)})$ ,  $j \in \{1, 2, \dots, 2d\}$ . Moreover, we denote  $f_k(\mathbf{y}_t) := g_k(\mathbf{y}_t) - \mathbb{E}[g_k(\mathbf{y}_t)]$ , which gives that  $\mathbb{E}[f_k(\mathbf{y}_t)] = 0$  for all  $k \in \{1, 2, \dots, 4d\}$ .

Following the proof of Theorem 4, we can express the symmetrized autocovariance matrix estimator as

$$\hat{\mathbf{S}}_\tau[\mathbf{X}] = \frac{1}{2(T-\tau)} \sum_{t=1}^{T-\tau} \left( \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t+\tau}^H + \tilde{\mathbf{X}}_{t+\tau} \tilde{\mathbf{X}}_t^H \right) + \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H + \mathcal{O}_p(1/T).$$

We begin by showing that the off-diagonal terms vanish. Following the proof of Theorem 4, we obtain

$$T^\gamma \left( \boldsymbol{\Lambda}_\tau \hat{\mathbf{S}}_0[\mathbf{X}] - \hat{\mathbf{S}}_\tau[\mathbf{X}] \right),$$

which is equal to

$$T^{\gamma-1} \sum_{t=1}^T \frac{1}{2} \left( 2\boldsymbol{\Lambda}_\tau \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H - \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t+\tau}^H - \tilde{\mathbf{X}}_{t+\tau} \tilde{\mathbf{X}}_t^H \right) + T^\gamma (\boldsymbol{\Lambda}_\tau - \mathbf{I}_d) \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H + o_p(1),$$

with the convention  $\tilde{\mathbf{X}}_{k+\tau} = \mathbf{0}$ , when  $k + \tau > T$ . By Lemma 8, the off-diagonal elements of  $T^\gamma \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H$  vanish. Thus, the off-diagonal element  $(j, k)$ ,  $j \neq k$ , of  $T^\gamma (\hat{\mathbf{J}}\hat{\boldsymbol{\Gamma}} - \mathbf{I}_d)$  is asymptotically equivalent to  $\mathbf{H}_{jk} + o_p(1)$ , where independence and the proof of Theorem 4 give that

$$\mathbf{H}_{jk} = \frac{T^{\gamma-1}}{\lambda_\tau^{(k)} - \lambda_\tau^{(j)}} \sum_{t=1}^T \left( h_{j,k}(\mathbf{y}_t) + i \tilde{h}_{j,k}(\mathbf{y}_t) \right),$$

with

$$\begin{aligned} h_{j,k}(\mathbf{y}_t) = & \lambda_\tau^{(j)} \left[ f_j(\mathbf{y}_t) f_k(\mathbf{y}_t) + f_{j+d}(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t) \right] - \frac{1}{2} \left[ f_j(\mathbf{y}_t) f_{k+2d}(\mathbf{y}_t) \right. \\ & \left. + f_{j+d}(\mathbf{y}_t) f_{k+3d}(\mathbf{y}_t) + f_{j+2d}(\mathbf{y}_t) f_k(\mathbf{y}_t) + f_{j+3d}(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t) \right] \end{aligned}$$

and

$$\begin{aligned} \tilde{h}_{j,k}(\mathbf{y}_t) = & \lambda_\tau^{(j)} \left[ f_{j+d}(\mathbf{y}_t) f_k(\mathbf{y}_t) - f_j(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t) \right] - \frac{1}{2} \left[ f_{j+d}(\mathbf{y}_t) f_{k+2d}(\mathbf{y}_t) \right. \\ & \left. + f_{j+3d}(\mathbf{y}_t) f_k(\mathbf{y}_t) - f_j(\mathbf{y}_t) f_{k+3d}(\mathbf{y}_t) - f_{j+2d}(\mathbf{y}_t) f_{k+d}(\mathbf{y}_t) \right]. \end{aligned}$$

Thus, by Proposition 2,  $\mathbf{H}_{jk}$  converges to zero in probability. Hence it remains to prove the convergence of the diagonal elements. We can write

$$T^\gamma (\hat{\mathbf{S}}_0[\mathbf{X}] - \mathbf{I}_d) = T^{\gamma-1} \sum_{t=1}^T \left( \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H - \mathbb{E} \left[ \tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t^H \right] \right) + T^\gamma \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H + o_p(1).$$

Thus, the diagonal element  $(j, j)$  of  $T^\gamma (\hat{\mathbf{J}}\hat{\boldsymbol{\Gamma}} - \mathbf{I}_d)$  is asymptotically equivalent to  $\frac{1}{2} \mathbf{H}_{jj} + \frac{1}{2} T^\gamma (\tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H)_{jj} + o_p(1)$ , where

$$\mathbf{H}_{jj} = T^{\gamma-1} \sum_{t=1}^T \left( h_{j,j}(\mathbf{y}_t) - \mathbb{E} [h_{j,j}(\mathbf{y}_t)] \right)$$

with  $h_{j,j}(\mathbf{y}_t) = (f_j(\mathbf{y}_t))^2 + (f_{j+d}(\mathbf{y}_t))^2$ , and

$$\left[ \tilde{\boldsymbol{\mu}} \tilde{\boldsymbol{\mu}}^H \right]_{jj} = \left( \Re[\tilde{\boldsymbol{\mu}}^{(j)}] \right)^2 + \left( \Im[\tilde{\boldsymbol{\mu}}^{(j)}] \right)^2.$$

Recall that  $\mathbb{E}[f_j(\mathbf{y}_t)] = 0$ ,

$$\Re \left( \tilde{\boldsymbol{\mu}}^{(j)} \right) = \frac{1}{T} \sum_{t=1}^T f_j(\mathbf{y}_t) \quad \text{and} \quad \Im \left( \tilde{\boldsymbol{\mu}}^{(j)} \right) = \frac{1}{T} \sum_{t=1}^T f_{j+d}(\mathbf{y}_t).$$

Hence, due to independence, it suffices to prove that, for every  $j \in \{1, \dots, d\}$ ,

$$T^{\gamma-1} \sum_{t=1}^T \left( h_{j,j}(\mathbf{y}_t) - \mathbb{E} [h_{j,j}(\mathbf{y}_t)] \right) + T^{\gamma-2} \left( \left[ \sum_{t=1}^T f_j(\mathbf{y}_t) \right]^2 + \left[ \sum_{t=1}^T f_{j+d}(\mathbf{y}_t) \right]^2 \right)$$



$$\xrightarrow[T \rightarrow \infty]{\mathcal{D}} C_{2,j} Z_{q_{2,j}} + C_{1,j}^2 Z_{q_{1,j}}^2 + C_{2,j+d} Z_{q_{2,j+d}} + C_{1,j+d}^2 Z_{q_{1,j+d}}^2.$$

The convergence follows from independence and the continuous mapping theorem, if,  $\forall j \in \{1, \dots, 2d\}$ , the following two-dimensional vector converges:

$$\left( \frac{T^{\gamma-1} \sum_{t=1}^T ((f_j(\mathbf{y}_t))^2 - \mathbb{E}[(f_j(\mathbf{y}_t))^2])}{T^{\frac{\gamma}{2}-1} \sum_{t=1}^T f_j(\mathbf{y}_t)} \right) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \begin{pmatrix} C_{2,j} Z_{q_{2,j}} \\ C_{1,j} Z_{q_{1,j}} \end{pmatrix}. \quad (20)$$

Using Equation (6) for the asymptotic variance, we observe first that

$$T^{\gamma-1} \sum_{t=1}^T \left( (f_j(\mathbf{y}_t))^2 - \mathbb{E}[(f_j(\mathbf{y}_t))^2] \right)$$

converges to a nontrivial limit only if  $\max_{i \in I} q_{2,i}(2H_i - 2) = q_{2,j}(2H_j - 2)$ . Similarly,

$$T^{\frac{\gamma}{2}-1} \sum_{t=1}^T f_j(\mathbf{y}_t)$$

converges to a nontrivial limit only if  $\max_{i \in I} q_{2,i}(2H_i - 2) = q_{1,j}(4H_j - 4)$ . Finally, if both of these conditions are satisfied, the convergence in Equation (20) follows from [3, Theorem 4].  $\square$

## Acknowledgments

The authors would like to thank Katariina Kilpinen for providing the photographs to Section 5. The authors would like to thank the two anonymous referees for their insightful comments that helped to improve this paper greatly.

## References

- [1] Adali, T., Schreier, P.J.: Optimization and estimation of complex-valued signals: Theory and applications in filtering and blind source separation. *IEEE Signal Process. Mag.* **31**(5), 112–128 (2014). [MR1824645](#). <https://doi.org/10.1109/78.847796>
- [2] Arcones, M.A.: Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22**(4), 2242–2274 (1994). [MR1331224](#)
- [3] Bai, S., Taqqu, M.: Multivariate limit theorems in the context of long-range dependence. *J. Time Ser. Anal.* **34**(6), 717–743 (2013). [MR3127215](#). <https://doi.org/10.1111/jtsa.12046>
- [4] Bai, S., Taqqu, M.: Generalized Hermite processes, discrete chaos and limit theorems. *Stoch. Process. Appl.* **124**(4), 1710–1739 (2014). [MR3163219](#). <https://doi.org/10.1016/j.spa.2013.12.011>
- [5] Beran, J.: *Statistics for Long-Memory Processes*. Chapman & Hall, New York (1994). [MR1304490](#)
- [6] Beran, J., Feng, Y., Ghosh, S., Kulik, R.: *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer (2013). [MR3075595](#). <https://doi.org/10.1007/978-3-642-35512-7>

- [7] Billingsley, P.: Probability and Measure, Anniversary Edition. Wiley, Hoboken (2012). [MR2893652](#)
- [8] Breuer, P., Major, P.: Central limit theorems for non-linear functionals of Gaussian fields. *J. Multivar. Anal.* **13**(3), 425–441 (1983). [MR0716933](#). [https://doi.org/10.1016/0047-259X\(83\)90019-2](https://doi.org/10.1016/0047-259X(83)90019-2)
- [9] Cichocki, A., Amari, S.-i.: Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, vol. 1. John Wiley & Sons (2002)
- [10] Comon, P., Jutten, C.: Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press (2010)
- [11] Dobrushin, R.L., Major, P.: Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **50**, 27–52 (1979). [MR0550122](#). <https://doi.org/10.1007/BF00535673>
- [12] Embrechts, P., Maejima, M.: Selfsimilar Processes. Princeton Series in Applied Mathematics, p. 111. Princeton University Press, Princeton, New Jersey (2002). [MR1920153](#)
- [13] Eriksson, J., Ollila, E., Koivunen, V.: Essential statistics and tools for complex random variables. *IEEE Trans. Signal Process.* **58**(10), 5400–5408 (2010). [MR2722676](#). <https://doi.org/10.1109/TSP.2010.2054085>
- [14] Fan, J., Wang, M., Yao, Q.: Modelling multivariate volatilities via conditionally uncorrelated components. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **70**(4), 679–702 (2008). [MR2523899](#). <https://doi.org/10.1111/j.1467-9868.2008.00654.x>
- [15] Goodman, N.: Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Stat.* **34**(1), 152–177 (1963). [MR0145618](#). <https://doi.org/10.1214/aoms/1177704250>
- [16] Ilmonen, P.: On asymptotic properties of the scatter matrix based estimates for complex valued independent component analysis. *Stat. Probab. Lett.* **83**(4), 1219–1226 (2013). [MR3041396](#). <https://doi.org/10.1016/j.spl.2013.01.020>
- [17] Ilmonen, P., Oja, H., Serfling, R.: On invariant coordinate system (ICS) functionals. *Int. Stat. Rev.* **80**(1), 93–110 (2012). [MR2990347](#). <https://doi.org/10.1111/j.1751-5823.2011.00163.x>
- [18] Leonenko, N.: Limit Theorems for Random Fields with Singular Spectrum, vol. 465. Springer (1999). [MR1687092](#). <https://doi.org/10.1007/978-94-011-4607-4>
- [19] Li, H., Correa, N.M., Rodriguez, P.A., Calhoun, V.D., Adali, T.: Application of independent component analysis with adaptive density model to complex-valued fMRI data. *IEEE Trans. Biomed. Eng.* **58**(10), 2794–2803 (2011)
- [20] Lietzén, N., Virta, J., Nordhausen, K., Ilmonen, P.: Minimum distance index for BSS, generalization, interpretation and asymptotics. *Aust. J. Stat.* **49**(4), 57–68 (2020)
- [21] Makogin, V., Spodarev, E.: Limit theorems for excursion sets of subordinated Gaussian random fields with long-range dependence. *Stochastics*, 1–32 (2021)
- [22] Matteson, D.S., Tsay, R.S.: Dynamic orthogonal components for multivariate time series. *J. Am. Stat. Assoc.* **106**(496), 1450–1463 (2011). [MR2896848](#). <https://doi.org/10.1198/jasa.2011.tm10616>
- [23] Miettinen, J., Nordhausen, K., Oja, H., Taskinen, S.: Statistical properties of a blind source separation estimator for stationary time series. *Stat. Probab. Lett.* **82**(11), 1865–1873 (2012). [MR2970285](#). <https://doi.org/10.1016/j.spl.2012.06.025>
- [24] Miettinen, J., Illner, K., Nordhausen, K., Oja, H., Taskinen, S., Theis, F.J.: Separation of uncorrelated stationary time series using autocovariance matrices. *J. Time Ser. Anal.* **37**(3), 337–354 (2016). [MR3512961](#). <https://doi.org/10.1111/jtsa.12159>

- [25] Miettinen, J., Matilainen, M., Nordhausen, K., Taskinen, S.: Extracting conditionally heteroskedastic components using independent component analysis. *J. Time Ser. Anal.* **41**(2), 293–311 (2020). [MR4086188](#). <https://doi.org/10.1111/jtsa.12505>
- [26] Nordhausen, K., Oja, H., Tyler, D.E., et al.: Tools for exploring multivariate data: The package ICS. *J. Stat. Softw.* **28**(6), 1–31 (2008)
- [27] Nourdin, I., Peccati, G.: Stein’s method on Wiener chaos. *Probab. Theory Relat. Fields* **145**(1), 75–118 (2009). [MR2520122](#). <https://doi.org/10.1007/s00440-008-0162-x>
- [28] Nourdin, I., Peccati, G.: Stein’s method and exact Berry-Esseen asymptotics for functionals of Gaussian fields. *Ann. Probab.* **37**(6), 2231–2261 (2010). [MR2573557](#). <https://doi.org/10.1214/09-AOP461>
- [29] Nourdin, I., Peccati, G.: Normal Approximations Using Malliavin Calculus: From Stein’s Method to Universality. Cambridge University Press (2012). [MR2962301](#). <https://doi.org/10.1017/CBO9781139084659>
- [30] Nourdin, I., Peccati, G., Podolskij, M.: Quantitative Breuer–Major theorems. *Stoch. Process. Appl.* **121**(4), 793–812 (2011). [MR2770907](#). <https://doi.org/10.1016/j.spa.2010.12.006>
- [31] Picinbono, B.: Second-order complex random vectors and normal distributions. *IEEE Trans. Signal Process.* **44**(10), 2637–2640 (1996)
- [32] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2018). R Foundation for Statistical Computing. <https://www.R-project.org/>
- [33] Samorodnitsky, G.: Long range dependence. *Found. Trends Stoch. Syst.* **1**(3), 163–257 (2007). [MR2379935](#). <https://doi.org/10.1561/09000000004>
- [34] Taqqu, M.: Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probab. Theory Relat. Fields* **31**(4), 287–302 (1975). [MR0400329](#). <https://doi.org/10.1007/BF00532868>
- [35] Tong, L., Soon, V., Huang, Y.-F., Liu, R.: Amuse: a new blind identification algorithm. In: *IEEE International Symposium on Circuits and Systems*, pp. 1784–1787. IEEE (1990).
- [36] Tudor, C.A.: *Analysis of Variations for Self-similar Processes: A Stochastic Calculus Approach*. Springer (2013). [MR3112799](#). <https://doi.org/10.1007/978-3-319-00936-0>
- [37] Urbanek, S.: Jpeg: Read and Write JPEG Images. (2014). R package version 0.1-8. <https://CRAN.R-project.org/package=jpeg>
- [38] Van der Vaart, A.W.: *Asymptotic Statistics*, vol. 3. Cambridge University Press (2000). [MR1652247](#). <https://doi.org/10.1017/CBO9780511802256>
- [39] Viitasaari, L., Ilmonen, P.: On modeling a class of weakly stationary processes. *Front. Appl. Math. Stat.* **15** (2020).
- [40] Wendler, M., Wu, W.B.: Central limit theorems for nearly long range dependent subordinated linear processes. *J. Appl. Probab.* **57**(2), 637–656 (2020). [MR4125469](#). <https://doi.org/10.1017/jpr.2020.10>
- [41] Zarzoso, V., Comon, P.: Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *IEEE Trans. Neural Netw.* **21**(2), 248–261 (2009)

## Funding

N. Lietzén gratefully acknowledges financial support from the Emil Aaltonen Foundation (Grant 180144 N) and from the Academy of Finland (Grant 321968).