Statistics and Probability Letters 192 (2023) 109679

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Note on asymptotic behavior of spatial sign autocovariance matrices

Marko Voutilainen[a,*], Pauliina Ilmonen[b], Lauri Viitasaari[c], Niko Lietzén[d]

[a] Department of Accounting and Finance, University of Turku, Finland
[b] Department of Mathematics and Systems Analysis, Aalto University, Finland
[c] Department of Mathematics, Uppsala University, Sweden
[d] Department of Mathematics and Statistics, University of Turku, Finland

ARTICLE INFO

Article history:
Received 19 November 2021
Received in revised form 12 August 2022
Accepted 9 September 2022
Available online 19 September 2022

MSC:
62H12
60F05
60G15
60G10
62H11

Keywords:
Spatial sign
Autocovariance matrix
Gaussian subordinated process

ABSTRACT

In this paper, we consider the asymptotic properties of the spatial sign autocovariance matrix for Gaussian subordinated processes with a known location parameter.

1. Introduction

In this paper, we examine the asymptotic properties of the spatial sign autocovariance matrix estimator of a subordinated Gaussian process. The spatial sign covariance matrix is a cornerstone in modern multivariate robust statistics, see Visuri et al. (2000) for the original formulation. Previously, the asymptotic properties of the spatial sign covariance matrix have been consider, under some specific models, for classical iid data, see for example Dürre et al. (2014). The classical spatial sign covariance matrix can be straightforwardly extended to take temporal dependencies into account. This so-called spatial sign autocovariance matrix is considerably more rare in robust statistics and previous work on the topic is very applied, see, e.g., Lietzén et al. (2017). To our knowledge, the asymptotic properties of the spatial sign autocovariance matrix have not previously been considered under non-trivial temporal dependency structures. We aim to fill this gap in knowledge. We utilize recent limiting theorems and approach the problem through the well-established theory concerning Gaussian subordinated processes. Note that Gaussian subordinated processes provide a large class of processes covering, e.g., stationary processes with arbitrary marginal distributions, see Viitasaari and Ilmonen (2020).

Let $Z = (Z_t)_{t\in\mathbb{N}}$, where $Z_t = [Z_t^{(1)}, \ldots, Z_t^{(n)}]$ is an n-variate centered random vector with elements given by $Z_t^{(i)} = f_i(X_t)$, such that the functions $f_i : \mathbb{R}^d \to \mathbb{R}$ are measurable and $X = (X_t)_{t\in\mathbb{N}}$ is a d-variate stationary Gaussian process. We use

* Corresponding author.
E-mail address: marko.voutilainen@utu.fi (M. Voutilainen).

https://doi.org/10.1016/j.spl.2022.109679

the notation $r_X(t) = \mathbb{E}[(X_0 - \mu)(X_t - \mu)^\top]$ for the autocovariance matrix function of a stationary process X with a mean vector μ. Moreover

$$\gamma(\tau) = \mathbb{E}\left[\frac{Z_t}{\|Z_t\|}\frac{Z_{t+\tau}^\top}{\|Z_{t+\tau}\|}\right]$$

denotes the spatial sign autocovariance matrix of Z with lag τ. Note that "spatial" here refers to the multidimensionality of the random process, whereas the parameter set is one-dimensional with the usual time-interpretation. The element (i, j) of $\gamma(\tau)$ is denoted by $\gamma_{i,j}(\tau)$. We use the following estimators,

$$\hat{\boldsymbol{\gamma}}_T(\tau) = \frac{1}{T}\sum_{t=1}^{T-\tau}\frac{Z_t}{\|Z_t\|}\frac{Z_{t+\tau}^\top}{\|Z_{t+\tau}\|}, \qquad \hat{\gamma}_{T,i,j}(\tau) = \frac{1}{T}\sum_{t=1}^{T-\tau}\frac{Z_t^{(i)}}{\|Z_t\|}\frac{Z_{t+\tau}^{(j)}}{\|Z_{t+\tau}\|}.$$

Set

$$\tilde{X}_t = \begin{bmatrix} X_t & X_{t+\tau}\end{bmatrix}^\top, \quad g_{i,j}(\tilde{X}_t) = \frac{f_i(X_t)}{\sqrt{\sum_{k=1}^n f_k(X_t)^2}}\frac{f_j(X_{t+\tau})}{\sqrt{\sum_{k=1}^n f_k(X_{t+\tau})^2}},$$

which gives,

$$\hat{\gamma}_{T,i,j}(\tau) - \gamma_{i,j}(\tau) = \frac{1}{T-\tau}\sum_{t=1}^{T-\tau}(g_{i,j}(\tilde{X}_t) - \gamma_{i,j}(\tau)) - \frac{\tau}{(T-\tau)T}\sum_{t=1}^{T-\tau} g_{i,j}(\tilde{X}_t). \tag{1}$$

Note that, if the first term of $\hat{\gamma}_{T,i,j}(\tau) - \gamma_{i,j}(\tau)$ converges (in some sense) as $T \to \infty$, then the latter term converges to zero (in some sense).

We use $\mathbb{N}_0^d$ to denote d-vectors that consist of non-negative integers. We call $\mathbf{j} \in \mathbb{N}_0^d$ a multi-index of degree $|\mathbf{j}| = \sum_{i=1}^d j_i$ and we use $\|\cdot\|$ to denote the L^2 vector norm and the corresponding induced matrix norm. Next, we define multivariate Hermite polynomials according to Rahman (2017).

Definition 1.1. Let X be a d-variate centered Gaussian random vector with a positive-definite covariance matrix Σ_X and let $\phi_X(x; \Sigma_X)$ be the corresponding probability density function. Let $\mathbf{j} = (j_1, \dots, j_d) \in \mathbb{N}_0^d$ be a multi-index of degree $|\mathbf{j}| = \sum_{i=1}^d j_i$. Then, the multivariate Hermite polynomial associated with $\mathbf{j}$ is given by,

$$H_{\mathbf{j}}(x; \Sigma_X) = \frac{(-1)^{|\mathbf{j}|}}{\phi_X(x; \Sigma_X)}\left(\frac{\partial}{\partial x}\right)^{\mathbf{j}}\phi_X(x; \Sigma_X),$$

where $\left(\frac{\partial}{\partial x}\right)^{\mathbf{j}} = \frac{\partial^{|\mathbf{j}|}}{\partial x_1^{j_1}\cdots\partial x_d^{j_d}}$.

The corresponding standardized Hermite polynomials are obtained by a normalization under the given Gaussian measure. Under Definition 1.1, the standardized Hermite polynomial associated with the multi-index $\mathbf{j}$ is given by

$$\Psi_{\mathbf{j}}(x; \Sigma_X) = \frac{H_{\mathbf{j}}(x; \Sigma_X)}{\sqrt{\mathbb{E}(H_{\mathbf{j}}(X; \Sigma_X)^2)}}.$$

Similarly as in the univariate case, the first moments of the multivariate Hermite polynomials are zeroes under the Gaussian measure, when excluding the constant $H_{\mathbf{0}}(x; \Sigma_X) = 1$. Furthermore, the multivariate Hermite polynomials are weakly orthogonal wrt. the Gaussian measure in the following way.

Lemma 1.2. *Let X_1 and X_2 be two d-variate identically distributed centered jointly Gaussian random vectors with a common positive definite covariance matrix Σ_X. Let $\Sigma_{X_1X_2} := \mathbb{E}(X_1X_2^\top)$ and set $\bar{\Sigma} = \Sigma_X^{-1}\Sigma_{X_1X_2}\Sigma_X^{-1}$. Then,*

$$\mathbb{E}(H_{\mathbf{j}}(X_1; \Sigma_X)H_{\mathbf{k}}(X_2; \Sigma_X)) = \sum_{\substack{\Theta\in\mathbb{N}_0^{d\times d}\\ r(\Theta)=\mathbf{j}, c(\Theta)=\mathbf{k}}} \frac{\mathbf{j}!\mathbf{k}!(\bar{\Sigma})^{\Theta}}{\Theta!}, \qquad \text{for} \quad |\mathbf{j}| = |\mathbf{k}|, \tag{2}$$

and zero otherwise. Above the summation is over all $d \times d$ index matrices Θ with elements in $\mathbb{N}_0$ such that the row sum vector $r(\Theta)$ equals to $\mathbf{j}$ and the column sum vector $c(\Theta)$ equals to $\mathbf{k}$ with $|\mathbf{j}| = |\mathbf{k}|$. We use the notation $\mathbf{j}! = \Pi_{i=1}^d j_i!$. Furthermore, $\Theta! = \Pi_{p,q=1}^d \theta_{pq}!$, where θ_{pq} denotes the (p, q) element of Θ and

$$(\bar{\Sigma})^{\Theta} = \Pi_{p,q=1}^d \bar{\sigma}_{pq}^{\theta_{pq}},$$

where $\bar{\sigma}_{pq}$ denotes the (p, q) element of $\bar{\Sigma}$.

The proof for Lemma 1.2 is given in the online supplementary material. Lemma 1.2 can be seen as an extension of the similar result presented in Rahman (2017, Proposition 8). Note that after normalizing, the second moments of the standardized Hermite polynomials can be obtained directly using Eq. (2).

Remark 1.3. If $X_1 = X_2$, then $\bar{\Sigma} = \Sigma_X^{-1}$ in Lemma 1.2. If the components of a Gaussian vector X are independent, then the Hermite polynomials are strongly orthogonal, that is, $\mathbb{E}(H_{\mathbf{j}}(X; \Sigma_X)H_{\mathbf{k}}(X; \Sigma_X)) = 0$, for $\mathbf{j} \neq \mathbf{k}$.

We next arrive to the important concept of Hermite ranks.

Definition 1.4. Let X be a d-variate Gaussian vector and let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function with $f(X) \in L^2$. Denote the set of polynomials $p : \mathbb{R}^d \to \mathbb{R}$ of degree m with P_m. The Hermite rank q of f with respect to X is defined as

$$q = \inf\{m : \exists\, p \in P_m \ \text{ with } \ \mathbb{E}[(f(X) - \mathbb{E}f(X))p(X)] \neq 0\}. \tag{3}$$

Hermite ranks are central when we consider Hermite polynomial expansions in the L^2 space, as can be seen from the following lemma.

Lemma 1.5. *Let X be a d-variate centered Gaussian vector with a positive definite covariance matrix. The Hermite rank of a function $f(X) \in L^2$ is the smallest degree $q \geq 1$ of Hermite polynomials present in the expansion*

$$f(X) = \sum_{\mathbf{j} \in \mathbb{N}_0^d} C_{\mathbf{j}} \Psi_{\mathbf{j}}(X; \Sigma_X). \tag{4}$$

The proof of Lemma 1.5 is again presented in the online supplementary material. For further details regarding multivariate Hermite polynomials, see Rahman (2017).

2. Weak and L^p consistency

We utilize the next two lemmas in order to show that the spatial sign autocovariance matrix estimator is consistent in L^2.

Lemma 2.1. *Let $(Y_t)_{t\in\mathbb{N}}$ be a sequence of random variables with a mutual expectation. In addition, assume that* $\operatorname{var}(Y_j) \leq C$ *and* $\left|\operatorname{cov}(Y_j, Y_k)\right| \leq g(|k - j|)$, *where $g(i) \to 0$ as $i \to \infty$. Then we have the following convergence in L^2,*

$$\frac{1}{T} \sum_{t=1}^{T} Y_t \to \mathbb{E}(Y_1).$$

For a proof of Lemma 2.1, see Voutilainen (2020, Theorem 3.10.). We use $f(x) \sim g(x)$ to denote that the two functions are asymptotically of the same order, i.e., $\lim_{x\to\infty} \frac{|f(x)|}{|g(x)|} \to C$ for some finite constant $C > 0$.

Proposition 2.2. *Let X be a d-variate stationary Gaussian process with a mean vector μ and $r_X(t) = \mathbb{E}[(X_0 - \mu)(X_t - \mu)^\top] \to 0$ as $t \to \infty$. Let $h : \mathbb{R}^d \to \mathbb{R}$, $h(X_0) \in L^2$, have Hermite rank q with respect to X. Then,*

$$\operatorname{cov}(h(X_0), h(X_t)) = \mathcal{O}\left(\max_{i,j} \left| r_X^{i,j}(t)\right|^q\right).$$

Proof of Proposition 2.2. Since the univariate case is well-known (see e.g. Beran et al., 2013, p. 223), we concentrate on the case $d \geq 2$. By Lemma A.1 of the supplementary material, we may assume that $X_t \sim \mathcal{N}(0, I)$ and $\mathbb{E}(h(X)) = 0$. Hence, we obtain a strongly orthonormal Hermite polynomial expansion without the constant zero degree term. We also use the fact that the Hermite polynomials of degree $|\mathbf{j}| = l$ form a basis for the space of orthogonal polynomials of degree l. The dimension of this vector space is $K_{d,l} = \binom{d+l-1}{l}$, see Dunkl and Xu (2014). Thus, we set $C_l = [C_{l,1}, \ldots, C_{l,K_{d,l}}]^\top$ and $\Psi_l = [\Psi_{l,1}, \ldots, \Psi_{l,K_{d,l}}]^\top$. Now, we write

$$h(X_t) = \sum_{\mathbf{j}\in\mathbb{N}_0^d} C_{\mathbf{j}} \Psi_{\mathbf{j}}(X_t; I) = \sum_{l=q}^{\infty} \sum_{i=1}^{K_{d,l}} C_{l,i} \Psi_{l,i}(X_t; I),$$

Furthermore,

$$\mathbb{E}(h(X_t)^2) = \sum_{l=q}^{\infty} \sum_{i,j=1}^{K_{d,l}} C_{l,i} C_{l,j} \mathbb{E}(\Psi_{l,i}(X_t; I)\Psi_{l,j}(X_t; I)) = \sum_{l=q}^{\infty} \|C_l\|^2 < \infty$$

by the strong orthonormality of the Hermite polynomials mentioned in Remark 1.3. Similarly,

$$\mathbb{E}(h(X_0)h(X_t)) = \sum_{l=q}^{\infty} \sum_{i,j=1}^{K_{d,l}} C_{l,i} C_{l,j} \mathbb{E}(\Psi_{l,i}(X_0; I)\Psi_{l,j}(X_t; I))$$

$$= \sum_{l=q}^{\infty} C_l^\top \mathbb{E}(\Psi_l(X_0; I)\Psi_l^\top(X_t; I)) C_l =: \sum_{l=q}^{\infty} C_l^\top A_l(t) C_l.$$

The elements of $A_l(t)$ are of the form

$$\mathbb{E}(\Psi_{\mathbf{j}}(X_0;I)\Psi_{\mathbf{k}}(X_t;I)) = \frac{\mathbb{E}(H_{\mathbf{j}}(X_0;I)H_{\mathbf{k}}(X_t;I))}{\sqrt{\mathbb{E}(H_{\mathbf{j}}(X_0;I)^2)}\sqrt{\mathbb{E}(H_{\mathbf{k}}(X_0;I)^2)}}, \tag{5}$$

where $|\mathbf{j}| = |\mathbf{k}| = l$. Recall that

$$\mathbb{E}(H_{\mathbf{j}}(X_0;I)H_{\mathbf{k}}(X_t;I)) = \mathbf{j}!\mathbf{k}! \sum_{\substack{\Theta\in\mathbb{N}_0^{d\times d}\\ r(\Theta)=\mathbf{j},c(\Theta)=\mathbf{k}\\ |\mathbf{j}|=|\mathbf{k}|=l}} \frac{(\bar{\Sigma})^{\Theta}}{\Theta!}, \tag{6}$$

where now $\bar{\Sigma} = r_X(0)^{-1}r_X(t)r_X(0)^{-1} = r_X(t)$. Suppose $\max_{i,j}|r_X^{i,j}(t)| = 0$. Then also $(\bar{\Sigma})^{\Theta} = 0$ in Eq. (6) whenever $l \geq 1$. Consequently, $A_l(t) = 0$ and $\mathbb{E}(h(X_0), h(X_t)) = 0$. Now we may assume that $\max_{i,j}|r_X^{i,j}(t)| > 0$. Then

$$\begin{aligned}\frac{\mathbb{E}(h(X_0)h(X_t))}{\max_{i,j}|r_X^{i,j}(t)|^q} &= \frac{C_q^\top A_q(t)C_q}{\max_{i,j}|r_X^{i,j}(t)|^q} + \sum_{l=q+1}^{\infty}\frac{C_l^\top A_l(t)C_l}{\max_{i,j}|r_X^{i,j}(t)|^q}\\ &=: \frac{C_q^\top A_q(t)C_q}{\max_{i,j}|r_X^{i,j}(t)|^q} + \sum_{l=q+1}^{\infty} a_l(t)\end{aligned} \tag{7}$$

Next, we apply the dominated convergence theorem to show the sum above converges to 0 as $t \to \infty$. More precisely, we show that $|a_l(t)| \leq \|C_l\|^2$ when l and t are sufficiently large, and $a_l(t) \to 0$ for every $l \geq q+1$. We have that

$$|a_l(t)| \leq \frac{\|C_l\|^2\|A_l(t)\|}{\max_{i,j}|r_X^{i,j}(t)|^q}.$$

Thus, for the bound $|a_l(t)| \leq \|C_l\|^2$, it suffices to show that

$$\|A_l(t)\|_F < \max_{i,j}|r_X^{i,j}(t)|^q,$$

for large l and t, where $\|\cdot\|_F$ is the Frobenius norm. First, we note that

$$\|A_l(t)\|_F \leq K_{d,l}\max_{i,j}|(A_l(t))_{i,j}| = \binom{d+l-1}{l}\max_{i,j}|(A_l(t))_{i,j}|,$$

where for the binomial coefficient it holds by Stirling's approximation that

$$\begin{aligned}\binom{d+l-1}{l} = \frac{(d+l-1)!}{l!(d-1)!} &\sim \frac{\sqrt{d+l-1}\left(\frac{d+l-1}{e}\right)^{d+l-1}}{\sqrt{l}\left(\frac{l}{e}\right)^{l}} \sim \frac{(d+l-1)^{d+l-\frac{1}{2}}}{l^{l+\frac{1}{2}}}\\ &= \mathcal{O}\left(\frac{l^{d+l-\frac{1}{2}}}{l^{l+\frac{1}{2}}}\right) = \mathcal{O}(l^{d-1}),\end{aligned}$$

as $l \to \infty$. Thus

$$\|A_l(t)\|_F = \mathcal{O}_l\left(l^{d-1}\max_{i,j}|(A_l(t))_{i,j}|\right). \tag{8}$$

Next, we roughly evaluate the magnitude of the elements in matrix $A_l(t)$ by using Eqs. (5) and (6). Recall that $\bar{\Sigma} = r_X(t)$ in (6). Hence,

$$\frac{(\bar{\Sigma})^{\Theta}}{\max_{i,j}|r_X^{i,j}(t)|^q} \leq \max_{i,j}|r_X^{i,j}(t)|^{l-q} \tag{9}$$

In order to obtain an upper bound for the number of the elements in the sum (6), note that every such matrix Θ can be constructed by allocating l units between d^2 elements. Hence, the number of the summands is bounded by $(d^2)^l$.

Next, we turn to the denominator $\Theta!$. The sum of the elements of Θ is l. It is clear that $\Theta!$ obtains its minimum when we allocate $\left\lfloor \frac{l}{d^2}\right\rfloor$ units to every d^2 elements, and an additional unit to $l - \left\lfloor \frac{l}{d^2}\right\rfloor d^2$ elements, of Θ. In this case

$$\Theta! = \left(\left\lfloor \frac{l}{d^2}\right\rfloor !\right)^{d^2}\left(\left\lfloor \frac{l}{d^2}\right\rfloor + 1\right)^{l-\left\lfloor \frac{l}{d^2}\right\rfloor d^2},$$

where the latter of the two terms is at least one. As $l \to \infty$, Stirling's approximation gives

$$\begin{aligned}\left(\left\lfloor \frac{l}{d^2} \right\rfloor !\right)^{d^2} &\sim \left(\sqrt{2\pi \left\lfloor \frac{l}{d^2} \right\rfloor}\left(\frac{\left\lfloor \frac{l}{d^2} \right\rfloor}{e}\right)^{\left\lfloor \frac{l}{d^2} \right\rfloor}\right)^{d^2} \sim \frac{\left(\left\lfloor \frac{l}{d^2} \right\rfloor\right)^{d^2\left(\left\lfloor \frac{l}{d^2} \right\rfloor + \frac{1}{2}\right)}}{e^{d^2\left\lfloor \frac{l}{d^2} \right\rfloor}} \\ &\geq \frac{(\frac{l}{d^2} - 1)^{d^2(\frac{l}{d^2} - \frac{1}{2})}}{e^l} = \frac{(\frac{l}{d^2} - 1)^{l - \frac{1}{2}d^2}}{e^l},\end{aligned} \tag{10}$$

where $(\frac{l}{d^2} - 1)^{l - \frac{1}{2}d^2}$ is of order $\left(\frac{l}{d^2}\right)^{l - \frac{1}{2}d^2}$.

Finally, we evaluate the normalization factors $\sqrt{\mathbb{E}(H_{\mathbf{j}}(X_0; I)^2)}$ of (5). Note that now in (6) $\bar{\Sigma} = I$, which gives,

$$\mathbb{E}(H_{\mathbf{j}}(X_0; I)^2) = (\mathbf{j}!)^2 \sum_{\substack{\Theta \in \mathbb{N}_0^{d \times d} \\ r(\Theta) = c(\Theta) = \mathbf{j} \\ |\mathbf{j}| = l}} \frac{(I)^{\Theta}}{\Theta!},$$

where the only nonzero contribution to the sum comes from the diagonal matrix Θ whose diagonal vector equals to $\mathbf{j}$. In this case, $(I)^{\Theta}$ is obviously equal to one and $\Theta! = \mathbf{j}!$. On the other hand, $\mathbf{j}!$ gets maximized when we allocate all the weight l to a single element giving $\mathbf{j}! = l!$ Hence, $\Theta! \leq l!$, where again by Stirling's approximation

$$l! \sim \sqrt{l}\left(\frac{l}{e}\right)^l = \left(\frac{l^{l+\frac{1}{2}}}{e^l}\right), \tag{11}$$

as $l \to \infty$. By (8), (9), (10), (11) and since the number of summands in (6) is bounded by d^{2l} we get

$$\frac{\|A_l(t)\|_F}{\max_{i,j} |r_X^{i,j}(t)|^q} = \mathcal{O}_l\left(l^{\frac{1}{2}(d^2+2d-1)}\left(\max_{i,j} |r_X^{i,j}(t)|^{1-\frac{q}{l}} d^4\right)^l\right). \tag{12}$$

Since $r_X(t) \to 0$, there exists T such that $\max_{i,j} |r_X^{i,j}(t)| < \frac{1}{d^{4(q+1)}}$ for all $t \geq T$. In this case, for $l \geq q+1$,

$$\max_{i,j} |r_X^{i,j}(t)|^{1-\frac{q}{l}} d^4 < d^4 \left(\frac{1}{d^{4(q+1)}}\right)^{1-\frac{q}{l}} \leq d^4 \left(\frac{1}{d^{4(q+1)}}\right)^{1-\frac{q}{q+1}} = 1.$$

Since (12) is of order $o_l(1)$, the dominated convergence theorem applies for the sum in (7) as $t \to \infty$.

Moreover, $a_l(t) \to 0$ for all $l \geq q+1$, since

$$\frac{\|A_l(t)\|_F}{\max_{i,j} |r_X^{i,j}(t)|^q} \leq K_{d,l} \frac{\max_{i,j} |(A_l(t))_{i,j}|}{\max_{i,j} |r_X^{i,j}(t)|^q} \to 0$$

by (9) and the assumption. It remains to show that first term of (7) is bounded as $t \to \infty$. Now,

$$\frac{\|A_q(t)\|_F}{\max_{i,j} |r_X^{i,j}(t)|^q} \leq K_{d,q} \frac{\max_{i,j} |(A_q(t))_{i,j}|}{\max_{i,j} |r_X^{i,j}(t)|^q},$$

which is bounded since (9) is now bounded by one. Therefore, we conclude,

$$\mathbb{E}(h(X_0), h(X_t)) = \mathcal{O}\left(\max_{i,j} |r_X^{i,j}(t)|^q\right). \quad \square$$

Theorem 2.3. *Let X be a d-variate stationary Gaussian process with a mean vector μ and $r_X(t) \to 0$. Let $h : \mathbb{R}^d \to \mathbb{R}$ be such that $h(X_0) \in L^2$. Then,*

$$\frac{1}{T}\sum_{t=1}^{T} h(X_t) \to \mathbb{E}(h(X_1)) \quad \text{in } L^2.$$

If h is bounded, then the convergence takes place in L^p for all $p \geq 1$.

Proof. The first part of the claim follows directly from Lemma 2.1 and Proposition 2.2. Note that

$$Y_T =: \frac{1}{T}\sum_{t=1}^{T}(h(X_t) - \mathbb{E}(h(X_1)))$$

is a bounded sequence converging to zero in L^2. Hence, Y_T^p is uniformly integrable converging to zero in probability for all $p \geq 1$. Consequently, $Y_T^p \to 0$ in L^1 giving the second part of the claim. □

Lemma 2.4. *Let Y_T be a sequence of n-variate random vectors such that the elements of Y_T convergence to the elements of Y in L^p for $p \geq 2$. Then*

$$\|Y_T - Y\|_p \to 0,$$

where $\|\cdot\|_p = \sqrt[p]{\mathbb{E}(\|\cdot\|^p)}$ denotes the L^p norm of random vectors.

Proof. The case $p = 2$ follows directly. Hence, let $p > 2$. We use the Hölder's inequality with $r = \frac{p}{2}$ and $s = \frac{p}{p-2}$, which produces,

$$\begin{aligned}\|Y_T - Y\|^2 = \sum_{i=1}^{n}(Y_T^{(i)} - Y^{(i)})^2 &\leq \left(\sum_{i=i}^{n} |Y_T^{(i)} - Y^{(i)}|^p\right)^{\frac{2}{p}} \left(\sum_{i=1}^{n} 1\right)^{\frac{p-2}{p}} \\ &= C_{p,n}\left(\sum_{i=i}^{n} |Y_T^{(i)} - Y^{(i)}|^p\right)^{\frac{2}{p}}.\end{aligned}$$

Therefore

$$\|Y_T - Y\|_p^p \leq C_{p,n}\mathbb{E}\sum_{i=1}^{n} |Y_T^{(i)} - Y^{(i)}|^p \to 0. \quad \square$$

Theorem 2.5. *Let X be a d-variate stationary Gaussian process with a mean vector μ and $r_X(t) = \mathbb{E}((X_0 - \mu)(X_t - \mu)^\top) \to 0$. Then*

$$\|\hat{\boldsymbol{\gamma}}_T(\tau) - \boldsymbol{\gamma}(\tau)\| \to 0 \quad \textit{in probability} \tag{13}$$

and

$$\|\operatorname{vec}(\hat{\boldsymbol{\gamma}}_T(\tau) - \boldsymbol{\gamma}(\tau))\|_p \to 0 \quad \textit{for all } p \geq 1,$$

where $\|\cdot\|_p$ denotes the L^p norm of random vectors.

Proof. The process $\tilde{X}$ and the function $g_{i,j}$ satisfy the assumptions of Theorem 2.3. Hence

$$\frac{1}{T-\tau}\sum_{t=1}^{T-\tau} g_{i,j}(\tilde{X}_t) \to \mathbb{E}(g_{i,j}(\tilde{X}_1)) = \gamma_{i,j}(\tau)$$

in L^p for all $p \geq 1$. Consequently, by (1), the elements of $\boldsymbol{\gamma}_T(\tau)$ converge to the elements of $\boldsymbol{\gamma}(\tau)$ in probability and in L^p for all $p \geq 1$. This directly gives (13). The second part of the claim follows from Lemma 2.4 (and Hölder's inequality for $1 \leq p < 2$). □

3. Strong consistency and asymptotic distribution

Lemma 3.1. *Let $(Y_t)_{t\in\mathbb{N}}$ be a weakly stationary process with the autocovariance function $r_Y(t)$ satisfying*

$$\sum_{t=0}^{\infty} |r_Y(t)| < \infty.$$

Then we have the following convergence almost surely,

$$\frac{1}{T}\sum_{t=1}^{T} Y_t \to \mathbb{E}(Y_1).$$

Proof. This follows directly from White (2014, Theorem 3.57). □

Theorem 3.2. *Let X be a d-variate stationary Gaussian process. Let the Hermite rank of $h : \mathbb{R}^d \to \mathbb{R}$, with $h(X_0) \in L^2$, be q. If*

$$\sum_{t=0}^{\infty} |r_X^{i,j}(t)|^q < \infty, \quad \text{for all} \quad i, j \in \{1, \ldots, d\},$$

then

$$\frac{1}{T}\sum_{t=1}^{T} h(X_t) \to \mathbb{E}(h(X_1)) \quad \textit{almost surely.}$$

Proof. The result follows from Proposition 2.2 together with Lemma 3.1. □

We next recall the celebrated theorem from Breuer and Major (1983).

Theorem 3.3. *Let X be a d-variate stationary Gaussian process and let $\{f_i : i = 1, \ldots, n\}$ be a set of functions $f_i : \mathbb{R}^d \to \mathbb{R}$ whose Hermite ranks with respect to X_t are at least q. If*

$$\sum_{t=0}^{\infty} |r_X^{i,j}(t)|^q < \infty, \quad \text{for all} \quad i, j \in \{1, \ldots, d\},$$

then the sequence of random vectors

$$\frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}(f_1(X_t) - \mathbb{E}f_1(X_t)) \cdots \sum_{t=1}^{T}(f_n(X_t) - \mathbb{E}f_n(X_t))\right]^\top$$

converges in distribution to $\mathcal{N}(0, \Sigma)$. The elements σ_{ij}, $i, j \in \{1, \ldots n\}$, of the limiting covariance matrix are given by

$$\sigma_{ij} = \mathrm{cov}(f_i(X_1), f_j(X_1)) + \sum_{\tau=1}^{\infty} \mathrm{cov}(f_i(X_1), f_j(X_{1+\tau})) + \mathrm{cov}(f_j(X_1), f_i(X_{1+\tau})).$$

The above Breuer–Major theorem is conventionally stated for a centered X. However, Step 1 of Lemma A.1 from the supplementary material reveals that this assumption is not necessary.

Theorem 3.4. *Let X be a d-variate stationary Gaussian process. Let the process $\tilde{X}$ and the functions $g_{i,j}$ be as in Section 1. Assume that the Hermite ranks of the functions $g_{i,j}$ with respect to $\tilde{X}$ are at least q and*

$$\sum_{t=0}^{\infty} |r_X^{i,j}(t)|^q < \infty, \quad \text{for all} \quad i, j \in \{1, \ldots, d\}. \tag{14}$$

Then

$$\hat{\boldsymbol{\gamma}}_T(\tau) \to \boldsymbol{\gamma}(\tau)$$

almost surely and

$$\sqrt{T}\,\mathrm{vec}(\hat{\boldsymbol{\gamma}}_T(\tau) - \boldsymbol{\gamma}(\tau)) \to \mathcal{N}(0, \Sigma)$$

in distribution. The elements of the limiting $n^2 \times n^2$ covariance matrix are,

$$\begin{aligned}
&\mathrm{cov}(g_{i,j}(\tilde{X}_1), g_{k,l}(\tilde{X}_1)) \\
&\quad + \sum_{m=1}^{\infty}\left(\mathrm{cov}(g_{i,j}(\tilde{X}_1), g_{k,l}(\tilde{X}_{m+1})) + \mathrm{cov}(g_{i,j}(\tilde{X}_{m+1}), g_{k,l}(\tilde{X}_1))\right) \\
&= \mathrm{cov}\left(\frac{Z_0^{(i)}Z_\tau^{(j)}}{\sqrt{\sum_{r=1}^n (Z_0^{(r)})^2 \sum_{r=1}^n (Z_\tau^{(r)})^2}}, \frac{Z_0^{(k)}Z_\tau^{(l)}}{\sqrt{\sum_{r=1}^n (Z_0^{(r)})^2 \sum_{r=1}^n (Z_\tau^{(r)})^2}}\right) \\
&\quad + \sum_{m=1}^{\infty}\mathrm{cov}\left(\frac{Z_0^{(i)}Z_\tau^{(j)}}{\sqrt{\sum_{r=1}^n (Z_0^{(r)})^2 \sum_{r=1}^n (Z_\tau^{(r)})^2}}, \frac{Z_m^{(k)}Z_{m+\tau}^{(l)}}{\sqrt{\sum_{r=1}^n (Z_m^{(r)})^2 \sum_{r=1}^n (Z_{m+\tau}^{(r)})^2}}\right) \\
&\quad + \sum_{m=1}^{\infty}\mathrm{cov}\left(\frac{Z_m^{(i)}Z_{m+\tau}^{(j)}}{\sqrt{\sum_{r=1}^n (Z_m^{(r)})^2 \sum_{r=1}^n (Z_{m+\tau}^{(r)})^2}}, \frac{Z_0^{(k)}Z_\tau^{(l)}}{\sqrt{\sum_{r=1}^n (Z_0^{(r)})^2 \sum_{r=1}^n (Z_\tau^{(r)})^2}}\right),
\end{aligned}$$

where $i, j, k, l \in \{1, \ldots, n\}$.

Proof. First we note that the summability condition (14) for $\tilde{X}$ follows directly from the assumption on X. Hence, by Theorem 3.2, the usual estimators of the mean of $g_{i,j}(\tilde{X}_t)$ converge almost surely to $\gamma_{i,j}(\tau)$, which gives the first part.

The elements of $\sqrt{T}\,\mathrm{vec}(\hat{\gamma}_T(\tau)-\gamma(\tau))$ are of the form $\sqrt{T}(\hat{\gamma}_{T,i,j}(\tau)-\gamma_{i,j}(\tau))$. The joint convergence of these elements is, in the light of the decomposition (1), dictated by the joint convergence of

$$\frac{\sqrt{T}}{T-\tau}\sum_{t=1}^{T-\tau}(g_{i,j}(\tilde{X}_t)-\mathbb{E}g_{i,j}(\tilde{X}_t)).$$

The constant shift τ in the rate function $\sqrt{\cdot}$ does not affect the limit and hence, Theorem 3.3 holds. □

If the components $r_X^{i,j}(t)$ are absolutely summable, then Theorems 3.2, 3.3 and 3.4 are valid regardless of the Hermite ranks of the involved functions. Theorem 3.4 covers also cases, where the subordinated process Z itself is long-range dependent. For example, it is well-known that in the case of a univariate X, the autocovariance $r_{Z^{(i)}}(t)$ of $Z_t^{(i)} = f_i(X_t)$ satisfies

$$r_{Z^{(i)}}(t) \sim r_X(t)^{q_i},$$

where q_i is the Hermite rank of f_i. Now, the autocovariance of $Z^{(i)}$ is not necessarily absolutely summable, if $q_i < q$. In this case, the process Z exhibits long-range dependence although the processes $g_{i,j}(\tilde{X})$ are short-range dependent as shown by Proposition 2.2. This type of situation can occur by the instability of Hermite ranks (see e.g. the related papers Bai and Taqqu, 2019 and Bai et al., 2018).

4. Future prospects

In the future, it would be interesting to consider the spatial sign autocovariance matrix estimator in more general settings that may yield non-Gaussian limiting distributions. Furthermore, the results of this paper could be extended to cover models that contain an unknown location parameter. Especially, the use of a robust measure of location, such as the spatial median, would be well-motivated.

Data availability

No data was used for the research described in the article.

Acknowledgments

The work of NL was supported by the Academy of Finland (Grant 321968).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2022.109679.

References

Bai, S., Taqqu, M.S., 2019. Sensitivity of the Hermite rank. Stochastic Process. Appl. 129 (3), 822–840.
Bai, S., Taqqu, M.S., et al., 2018. How the instability of ranks under long memory affects large-sample inference. Statist. Sci. 33 (1), 96–116.
Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-Memory Processes: Probabilistic Properties and Statistical Methods. (155), Springer-Verlag Berlin Heidelberg.
Breuer, P., Major, P., 1983. Central limit theorems for non-linear functionals of Gaussian fields. J. Multivariate Anal. 13 (3), 425–441.
Dunkl, C.F., Xu, Y., 2014. Orthogonal Polynomials of Several Variables. (155), Cambridge University Press.
Dürre, A., Vogel, D., Tyler, D.E., 2014. The spatial sign covariance matrix with unknown location. J. Multivariate Anal. 130, 107–117.
Lietzén, N., Nordhausen, K., Ilmonen, P., 2017. Complex valued robust multidimensional SOBI. In: International Conference on Latent Variable Analysis and Signal Separation. Springer, pp. 131–140.
Rahman, S., 2017. Wiener–Hermite polynomial expansion for multivariate Gaussian probability measures. J. Math. Anal. Appl. 454 (1), 303–334.
Viitasaari, L., Ilmonen, P., 2020. On modeling a class of weakly stationary processes. Front. Appl. Math. Stat. 15.
Visuri, S., Koivunen, V., Oja, H., 2000. Sign and rank covariance matrices. J. Statist. Plann. Inference 91 (2), 557–575.
Voutilainen, M., 2020. New approaches for modeling and estimation of discrete and continuous time stationary processes (Doctoral thesis). Aalto University School of Science, URL http://urn.fi/URN:ISBN:978-952-64-0183-6.
White, H., 2014. Asymptotic Theory for Econometricians. Academic Press.