Spectral-norm risk rates for multi-taper estimation of Gaussian processes

José Luis Romero a,b and Michael Speckbacher a,b

aFaculty of Mathematics, University of Vienna, Vienna, Austria; bAcoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

ABSTRACT
We consider the estimation of the covariance of a stationary Gaussian process on a multi-dimensional grid from observations taken on a general acquisition domain. We derive spectral-norm risk rates for multi-taper estimators. When applied to one-dimensional acquisition intervals, these show that Thomson’s classical multi-taper has optimal risk rates, as they match known benchmarks. We also extend existing lower risk bounds to multi-dimensional grids and conclude that multi-taper estimators associated with certain two-dimensional acquisition domains also have almost optimal risk rates.

1. Introduction
Let $X = \{X_k : k \in \mathbb{Z}^d\}$ be a stationary, real, zero mean, ergodic Gaussian process on the infinite $d$-dimensional grid $\mathbb{Z}^d$. The stochastics of $X$ are encoded in its covariance matrix

$$
\Sigma_{n,m} = \mathbb{E}\{X_n \cdot X_m\} = \sigma_{n-m}, \quad n, m \in \mathbb{Z}^d,
$$

or equivalently, in its spectral density

$$
S(\xi) = \sum_{n \in \mathbb{Z}^d} \sigma_n e^{2\pi i \langle \xi, n \rangle}, \quad \xi \in \mathbb{R}^d.
$$

Single-shot spectral estimation is the task of approximating the covariance matrix $\Sigma$ given one realisation of $X$ observed on an acquisition domain $\Omega \subset \mathbb{Z}^d$.

To compare the performance of different estimators, one must specify a suitable class of covariance matrices $\Sigma$ and an error metric for their approximations. In this article, we consider processes $X$ whose spectral density $S$ is a twice continuously differentiable function on $\mathbb{R}^d$ and often fix the normalisation

CONTACT José Luis Romero jose.luis.romero@univie.ac Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, Acoustics Research Institute, Austrian Academy of Sciences, Wohllibengasse 12-14, Vienna 1040, Austria

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
The performance of an estimator $\hat{S}$ will be measured with respect to the uniform norm

$$\| S - \hat{S} \|_\infty = \max_{\xi \in \mathbb{R}^d} | S(\xi) - \hat{S}(\xi) |,$$

as this measures the error incurred in approximating the covariance matrix $\Sigma$ in spectral norm:

$$\| \Sigma - \hat{\Sigma} \|_s := \max_{\| a \|_2 \leq 1} \| \Sigma a - \hat{\Sigma} a \|_2 = \| S - \hat{S} \|_\infty,$$

where $\hat{\Sigma}_{n,m} = \hat{\sigma}_{n-m}$ and $\hat{S}(\xi) = \sum_{n \in \mathbb{Z}^d} \hat{\sigma}_n e^{2\pi i \langle \xi, n \rangle}$, and $\| a \|_2 = (\sum_{n \in \mathbb{Z}^d} |a_n|^2)^{1/2}$. 1

The most classical setting for single-shot spectral estimation concerns dimension $d = 1$ and observations taken on a finite interval $\Omega = \{1, \ldots, N\}$. As shown in Cai, Ren, and Zhou (2013, Theorem 1), in this classical setup, the minimax risk satisfies

$$\inf_{\hat{S}} \sup_{S} \mathbb{E} \left\{ \| S - \hat{S} \|_\infty^2 \right\} \asymp \left( \frac{\log N}{N} \right)^{4/5},$$

where the supremum is taken over all stationary, real, zero mean, Gaussian processes $X$ with spectral density satisfying (2) (which are necessarily ergodic) and the infimum runs over all estimators $\hat{S}$ based on $X_1, \ldots, X_N$. Moreover, asymptotically optimal estimators, approximately achieving the infimum in (4), are known explicitly (Cai et al. 2013, Theorem 2).

A first goal of this article is to show that Thomson’s multitaper estimator (1982)

$$\hat{S}^{mt}(\xi) := \frac{1}{K} \sum_{k=0}^{K-1} \left| \sum_{n \in \Omega} X_n \cdot v_n^{(k)}(\Omega, W) \cdot e^{-2\pi i \langle \xi, n \rangle} \right|^2,$$

where $v_n^{(k)}(\Omega, W)$ denote the Slepian sequences (see Slepian 1978 and below) and are asymptotically optimal for the minimax risk (4) when the corresponding parameters are chosen adequately. More generally, we investigate multi-taper estimators on general acquisition domains and prove upper and lower minimax risk estimates.

2. Results

2.1. Multi-taper estimators

While originally introduced to study time series (Thomson 1982), we consider the multitaper estimator for general dimension $d$ and a general acquisition domain $\Omega$ with cardinality $N_\Omega$ (Bronz 1988; Hannsen 1997; Simons and Wang 2011). Given two parameters, $0 < W \leq 1$ (bandwidth) and $K \in \mathbb{N}$ (number of tapers), we first define the Slepian tapers recursively, as the set of sequences $v^{(0)}(\Omega, W), \ldots, v^{(N_\Omega-1)}(\Omega, W)$ on $\mathbb{Z}^d$ that solve the
following spectral concentration problem (Slepian 1978):

$$\text{Maximise } \int_{[-W/2,W/2]^d} \left| \sum_{n\in\Omega} v_n^{(k)}(\Omega, W) e^{2\pi i (n, \xi)} \right|^2 d\xi,$$

subject to: (i) $\sum_{n\in\Omega} |v_n^{(k)}(\Omega, W)|^2 = 1,$

(ii) $\text{supp}(v_n^{(k)}(\Omega, W)) \subseteq \Omega,$

(iii) $v_n^{(k)}(\Omega, W) \perp \{v_n^{(0)}(\Omega, W), v_n^{(1)}(\Omega, W), \ldots, v_n^{(k-1)}(\Omega, W)\}.$

Alternatively, the Slepian tapers are the normalised eigenvectors of the truncated $d$-Toeplitz matrix (Hogan and Lakey 2012, Chapters 1 and 3)

$$W^d \prod_{k=1}^d \sin(\pi W(n_k - m_k)) \quad \text{if } n, m \in \Omega,$$

$$0 \quad \text{otherwise}.$$

The multitaper estimator uses the first $K$ Slepian tapers as masks to build the averaged periodogram given in (5). It is standard to let $K$, $W$ and $N_{\Omega}$ be related by

$$K = \left\lceil N_{\Omega} \cdot W^d \right\rceil \quad (6)$$

(smallest integer $\geq N_{\Omega} \cdot W^d$) though other choices with $K < \left\lceil N_{\Omega} \cdot W^d \right\rceil$ are of practical interest (Percival and Walden 1993, Figures 203, 341; Karnik, Romberg, and Davenport 2022, Theorem 5) (see also Section 3). Corollary 2.2 gives an appropriate choice of $K$ as a function of $N_{\Omega}$, with $W$ being implicitly determined by (6).

2.2. Spectral-norm mean squared risk rates for multi-tapering

We formulate risk bounds for multi-tapering in terms of the cardinality of the acquisition domain $N_{\Omega}$, its digital perimeter

$$N_{\partial\Omega} = \sum_{k \in \mathbb{Z}^d} \sum_{j=1}^d |\chi_{\Omega}(k + e_j) - \chi_{\Omega}(k)|,$$

where $\chi_{\Omega}$ denotes the characteristic (indicator) function of a domain $\Omega$, and its diameter

$$\text{diam}(\Omega) = \max\{\|k - j\|_2 : k, j \in \Omega\}.$$

The following result is an analogue of Andén and Romero (2020, Theorem 4.3) applicable to error metric (3) instead of a pointwise error estimate.

**Theorem 2.1:** Let $X$ be a stationary, real, zero mean, Gaussian process on $\mathbb{Z}^d$ with spectral density $S$ satisfying $\|S\|_{C^2} \leq 1$. Let $\Omega \subset \mathbb{Z}^d$ be a finite acquisition domain with $N_{\Omega} \geq 3$ and
consider multi-taper estimator (5) with bandwidth \( W \) and \( K = \lceil N_\Omega \cdot W^d \rceil \) tapers. Assume further that \( N_\Omega \geq (\frac{N_\Omega}{K})^{1-1/d} \). Then

\[
\mathbb{E}\{\|S - \hat{S}^{\text{mt}}\|_\infty^2\} \lesssim \max_{p \in \{1,2\}} \left( \left( \frac{\log(\text{diam}(\Omega))}{K} \right)^p + \left( \frac{K}{N_\Omega} \right)^{\frac{4}{d}} + \frac{N_\delta_\Omega^2}{N_\Omega^{2-\frac{2}{d}}} \left[ 1 + \log\left( \frac{N_\Omega}{N_\delta_\Omega} \right) \right]^2 \right). \tag{7}
\]

We now apply Theorem 2.1 to a class of acquisition domains satisfying perimeter bounds similar to those enjoyed by rectangles and disks. This regime arises, for example, when discretising an analogue domain at increasingly finer scales.

**Corollary 2.2:** Under the assumptions of Theorem 2.1, suppose that \( C > 0 \) is a constant such that

\[
N_\delta_\Omega \leq CN_\Omega^{1-1/d}, \tag{8}
\]

and that

\[
\text{diam}(\Omega) \leq \exp(N_\Omega^{1/d}). \tag{9}
\]

Then, for \( d = 1 \), the choice \( K = \lceil (\log(\text{diam}(\Omega)) \cdot N_\Omega^4)^{1/5} \rceil \) yields the upper bound

\[
\mathbb{E}\left\{\|S - \hat{S}^{\text{mt}}\|_\infty^2\right\} \lesssim \left( \frac{\log(\text{diam}(\Omega))}{N_\Omega} \right)^{4/5}, \tag{10}
\]

while, for \( d \geq 2 \), the following bound holds for \( K = \lceil (\log(\text{diam}(\Omega))^d \cdot N_\Omega^2)^{1/3} \rceil \):

\[
\mathbb{E}\left\{\|S - \hat{S}^{\text{mt}}\|_\infty^2\right\} \lesssim \left( \frac{\log(\text{diam}(\Omega))}{N_\Omega^{1/d}} \right)^{4/3}. \tag{11}
\]

The implied constants in (10) and (11) depend on the constant \( C \) in (8).

**Remark 2.3 (Risk optimality of Thomson’s multi-taper):** Applying Corollary 2.2 to a one-dimensional interval \( \Omega = \{1, \ldots, N\} \), we see that Thomson’s classical multi-taper estimator (1982) achieves minimax error rate (4).

**Remark 2.4:** In Theorem 2.1 and Corollary 2.2, the number of tapers \( K \) and the bandwidth parameter \( W \) are linked by (6). In terms of \( W \), (7) reads

\[
\mathbb{E}\{\|S - \hat{S}^{\text{mt}}\|_\infty^2\} \lesssim \max_{p \in \{1,2\}} \left( \frac{\log(\text{diam}(\Omega))}{N_\Omega W^d} \right)^p + W^4 + \frac{N_\delta_\Omega^2}{N_\Omega^{2-\frac{2}{d}}} W^2 \left[ 1 + \log\left( \frac{N_\Omega}{N_\delta_\Omega} \right) \right]^2 \,
\]

while error rates (10) and (11) hold when

\[
W = \begin{cases} 
N_\Omega^{-1/5} [\log(\text{diam}(\Omega))]^{1/5}, & d = 1, \\
N_\Omega^{-1/3d} [\log(\text{diam}(\Omega))]^{1/3}, & d \geq 2. 
\end{cases}
\]
2.3. Lower risk bounds for general acquisition domains

We are unaware of precise minimax rates for the covariance estimation problem with general acquisition domains. As a first benchmark, we derive the following bound.

Theorem 2.5: Let $\Omega \subset \mathbb{Z}^d$ be finite with $N_{\Omega} \geq 3$. Then

$$\inf_{\hat{S}} \sup_S \mathbb{E} \left\{ \|S - \hat{S}\|_{\infty}^2 \right\} \gtrsim \left( \frac{\log N_{\Omega}}{\text{diam}(\Omega)^d} \right)^{\frac{4}{4+d}},$$

where the supremum is taken over all stationary, real, zero mean, Gaussian processes on $\mathbb{Z}^d$ with spectral density satisfying $\|S\|_{C^2} \leq 1$, and the infimum runs over all estimators $\hat{S}$ based on an observation of $X$ on $\Omega$. In particular, if $\text{diam}(\Omega) \leq CN_{\Omega}^{1/d}$, then

$$\inf_{\hat{S}} \sup_S \mathbb{E} \left\{ \|S - \hat{S}\|_{\infty}^2 \right\} \gtrsim \left( \frac{\log N_{\Omega}}{N_{\Omega}^{2/d}} \right)^{\frac{4}{4+d}},$$

where the implied constant depends on $C$.

Remark 2.6 (Almost risk optimality of multi-tapering for certain two-dimensional domains): For classes of two-dimensional acquisition domains $\Omega$ satisfying $\text{diam}(\Omega) \leq CN_{\Omega}^{1/2}$, the upper bound for multi-taper estimator (11) deviates from minimax risk bound (13) only by a factor $(\log N_{\Omega})^{2/3}$.

2.4. Proofs

The upper bounds are obtained by combining the bias bounds from Abreu and Romero (2017) and Andén and Romero (2020) with methods of concentration of measure, as used in Cai et al. (2013) and Kabanava and Rauhut (2017). More specifically, Abreu and Romero (2017) and Andén and Romero (2020) bound the mean-squared error $\mathbb{E}\{|S(\xi) - \hat{S}_{\text{mt}}(\xi)|^2\}$ corresponding to the estimation of individual spectral frequencies $\xi$. Specifically, under the assumptions of Theorem 2.1, Andén and Romero (2020, Theorem 4.3) give

$$\sup_{\xi \in \mathbb{R}^d} \mathbb{E}\{|S(\xi) - \hat{S}_{\text{mt}}(\xi)|^2\} \lesssim \frac{1}{K} + \left( \frac{K}{N_{\Omega}} \right)^{\frac{4}{d}} + \frac{N_{\Omega}^2}{N_{\Omega}^{2/d}} \left[ 1 + \log \left( \frac{N_{\Omega}}{N_{\Omega}^{2/d}} \right) \right]^2. \quad (14)$$

Those estimates are here extended to the stronger error metric $\mathbb{E}\{|S - \hat{S}_{\text{mt}}|_{\infty}^2\}$, by combining them with concentration estimates for covariance estimators. Comparing (7)–(14), we see that the error bound for the uniform norm has an additional logarithmic factor.

The proofs of the risk bounds for the multi-taper estimator require reinspecting and adapting a portion of the arguments in Cai et al. (2013) and Kabanava and Rauhut (2017). Following this path, it would also be possible to derive deviation bounds as in Kabanava and Rauhut (2017, Theorem 3) (which are stronger than MSE bounds) (see also Karnik et al. 2022, Theorem 5). The lower (minimax risk) bounds are obtained by direct adaptation of Cai et al. (2013). Detailed proofs are provided below.
3. Conclusions

Corollary 2.2 shows that Thomson’s multi-taper has optimal risk rates for covariance estimation of stationary Gaussian processes with $C^2$ spectral densities when the bandwidth parameter $W$, the number of samples $N$, and the number of tapers $K$ are linked by $K \asymp N^{4/5} \log(N)^{1/5}$ and $K = \lceil N_W \rceil$. (Note that in many references the bandwidth interval is $[-W, W]$, so that the previous equation reads $K = \lceil 2N_W \rceil$.) However, in certain situations, notably when the spectral density exhibits a high dynamic range, practitioners choose $K < N_W$. Comments to such effect can be found in Percival and Walden (1993) and are illustrated in Percival and Walden (1993, Figures 203, 341). The recent article Karnik et al. (2022) provides analytic and numerical evidence for the superiority of certain choices $K < N_W$ in various setups, including non-smooth spectral densities, or moderate values of $N_W$, when the implied constants in (3) cannot be neglected (see Karnik et al. 2022, Theorems 2 and 5). We thus contribute to the discussion on the optimal number of tapers in Thomson’s method, by showing that the choice $K = \lceil N_W \rceil$ meets the benchmark derived in Cai et al. (2013), and hope to motivate the investigation of more adequate formal benchmarks to reflect the possible practical advantages of other choices of $K$.

Multi-taper estimators for general acquisition domains are relevant in many areas of application including geophysics (see, e.g. Simons, van der Hilst, and Zuber 2003; Simons and Wang 2011; Harig and Simons 2012) and are comparatively less explored than Thomson’s original estimator. For general acquisition geometries, we have shown that the estimates for single frequencies in Abreu and Romero (2017) and Andén and Romero (2020) extend to spectral norm estimates with a logarithmic gain in the corresponding bounds. In dimension $d = 1, 2$, Theorem 2.5 shows that such bounds are almost optimal for classes of domains whose diameter is suitably dominated. We do not know the precise mini-max rate for covariance estimation in the spectral norm outside those regimes.

4. Detailed proofs

The proofs in this section follow closely (Cai et al. 2013; Kabanava and Rauhut 2017) and, in some cases, provide more details and simplified or optimised arguments.

4.1. Proof of upper bounds

Consider now a stationary Gaussian process on $X$ on $\mathbb{Z}^d$. Multi-taper estimator (5) can be rewritten as

$$\hat{S}_{mt}(\xi) = \sum_{\ell \in (\Omega - \Omega)} e^{-2\pi i \langle \xi, \ell \rangle} \left( \sum_{n,m \in \Omega} X_n X_m \frac{1}{K} \sum_{k=0}^{K-1} \nu_n^{(k)}(\Omega, W) \nu_m^{(k)}(\Omega, W) \right),$$  \hspace{1cm} (15)

where $A - B = \{a - b : a \in A, b \in B\}$. This shows that $\hat{S}_{mt}$ (and consequently also $\mathbb{E}(\hat{S}_{mt})$) is a multivariate trigonometric polynomial of maximum component degree $\omega \equiv \lceil \text{diam}(\Omega) \rceil$. We will use the following sampling inequality: for a trigonometric polynomial
\[ p(\xi) = \sum_{\ell \in \{-n, \ldots, n\}^d} c_\ell e^{-2\pi i (\xi, \ell)}, \]
we have \[ \sup_{\xi \in [0,1]^d} |p(\xi)| \leq C_d \max_{\ell \in \{0, \ldots, 4n-1\}^d} |p(\frac{\ell}{4n})|, \]
where \( C_d \) is a constant that depends on \( d \). The one-dimensional version of the sampling inequality is classical (see, e.g. Grenander and Szegö 1958, Chapter 5.2) while the general case follows by applying the one-dimensional result to each variable iteratively; alternatively, the multi-dimensional result can be deduced from Pfister and Bresler (2018, Theorem 1).

For \( \ell \in \{1, \ldots, 4\omega\}^d \), we set \( (\xi_\ell)_k := \frac{\ell_k - 1}{4\omega} \). Applied to (15) the sampling inequality yields
\[
\| \hat{S}^{mt} - \mathbb{E}(\hat{S}^{mt}) \|_\infty = \sup_{\xi \in [0,1]^d} |\hat{S}^{mt}(\xi) - \mathbb{E}(\hat{S}^{mt})(\xi)| \\
\leq C_d \max_{\ell \in \{1, \ldots, 4\omega\}^d} |\hat{S}^{mt}(\xi_\ell) - \mathbb{E}(\hat{S}^{mt})(\xi_\ell)|,
\]
where the first identity follows from the \( \mathbb{Z}^d \)-periodicity of \( \hat{S}^{mt} \).

We now express the multi-taper as done in Lii and Rosenblatt (2008). In terms of the matrix-valued function \( V_K : [0,1]^d \to \mathbb{C}^{\Omega \times \Omega} \),
\[
(V_K(\xi))_{n,m} := e^{-2\pi i (\xi, n-m)} \frac{1}{K} \sum_{k=0}^{K-1} v_n^{(k)}(\Omega, W) \cdot v_m^{(k)}(\Omega, W), \quad n, m \in \Omega,
\]
the multi-taper estimator can be written as a quadratic operation on the restricted process \( X|_\Omega \):
\[
\hat{S}^{mt}(\xi) = \langle V_K(\xi)X|_\Omega, X|_\Omega \rangle.
\]
(We note that this formula does not correspond to ‘tapering’ as considered in Kabanava and Rauhut 2017.) At grid points, the deviation of the multi-taper estimator from its mean is, therefore,
\[
Z^K_\ell := \hat{S}^{mt}(\xi_\ell) - \mathbb{E}(\hat{S}^{mt})(\xi_\ell) = \langle V_K(\xi_\ell)X|_\Omega, X|_\Omega \rangle - \mathbb{E}(V_K(\xi_\ell)X|_\Omega, X|_\Omega).
\]

As in Kabanava and Rauhut (2017), the main tool to analyse (18) is the following generalisation of the Hanson-Wright inequality (see Adamczak 2015; Rudelson and Vershynin 2013).

**Theorem 4.1:** Let \( Y \sim \mathcal{N}(0, \Sigma) \), then for every \( A \in \mathbb{R}^{N \times N} \) and every \( t > 0 \):
\[
\mathbb{P}(\| \langle AY, Y \rangle - \mathbb{E}(\langle AY, Y \rangle) \| \geq t) \leq 2 \exp \left[ -\frac{1}{C} \min \left( \frac{t^2}{\| \Sigma \|_2^2 \| A \|_F^2}, \frac{t}{\| \Sigma \|_s \| A \|_s} \right) \right]
\]
for some universal constant \( C > 0 \).

(Here, \( \| A \|_F \) is the Hilbert–Schmidt (Frobenius) norm of \( A \), while \( \| A \|_2 = \sup_{\| x \|_2 \leq 1} \| Ax \|_2 \) is the spectral norm.) To apply (18), we first carry out the following calculation.

**Lemma 4.2:** The Frobenius norm of \( V_K(\xi_\ell) \) equals \( 1/\sqrt{K} \) and its spectral norm is bounded by \( 1/K \).
Proof: Direct calculations yield

\[
\|V_K(\xi)\|_F^2 = \sum_{n \in \Omega} \sum_{m \in \Omega} e^{-2\pi i \langle \xi, n-m \rangle} \frac{1}{K} \sum_{k=0}^{K-1} v_n^{(k)}(\Omega, W) v_m^{(k)}(\Omega, W)^2
\]

\[
= \frac{1}{K^2} \sum_{n \in \Omega} \sum_{m \in \Omega} \sum_{k=0}^{K-1} \sum_{s=0}^{K-1} v_n^{(k)}(\Omega, W) v_m^{(s)}(\Omega, W) v_n^{(s)}(\Omega, W)^2
\]

\[
= \frac{1}{K^2} \sum_{k=0}^{K-1} \sum_{s=0}^{K-1} \left( \sum_{n \in \Omega} v_n^{(k)}(\Omega, W) v_n^{(s)}(\Omega, W) \right) \left( \sum_{m \in \Omega} v_m^{(k)}(\Omega, W) v_m^{(s)}(\Omega, W) \right)
\]

\[
= \frac{1}{K^2} \sum_{k=0}^{K-1} \sum_{s=0}^{K-1} \delta_{s,k} = \frac{1}{K}.
\]

To estimate the spectral norm, we first note that \(V_K(\xi)\) is a positive semi-definite matrix. Therefore, using the notation \((M_{\xi,a})_n = e^{2\pi i \langle \xi, n \rangle} a_n, n \in \Omega,\)

\[
\|V_K(\xi)\|_S = \sup_{\|Y\|_2=1} \langle V_K(\xi) Y, Y \rangle = \sup_{\|Y\|_2=1} \frac{1}{K} \sum_{k=0}^{K-1} |\langle Y, M_{\xi,\nu^{(k)}}(\Omega, W) \rangle|^2
\]

\[
\leq \sup_{\|Y\|_2=1} \frac{1}{K} \sum_{k=0}^{K-1} \|M_{\xi \nu^{(k)}}(\Omega, W)\|_2^2 = \frac{1}{K},
\]

as the Slepian tapers form an orthonormal basis. ■

We can now prove the main result.

Proof of Theorem 2.1: Step 1. Bias estimates. The following estimate was shown in Andén and Romero (2020, Proof of Theorem 4.3),

\[
|\text{Bias}(\hat{S}_\text{mt})(\xi)| = |\mathbb{E}[\hat{S}_\text{mt}](\xi) - S(\xi)|
\]

\[
\lesssim \|S\|_{C^2} \left( \frac{K}{N_{\Omega}} \right)^{2/d} + \frac{N_{\Omega}}{N_{\Omega}^{1-1/d} K^{1/d}} \left[ 1 + \log \left( \frac{N_{\Omega}}{N_{\Omega}} \right) \right].
\]

(Here it is essential that \(K = N_{\Omega} W^d + O(1).\))

Step 2. Concentration of the estimator. We show that

\[
\mathbb{E}[\|\hat{S}_\text{mt} - \mathbb{E}[\hat{S}_\text{mt}]\|_\infty^2] \lesssim \|S\|_{C^\infty}^2 \cdot \left( \frac{\log(\text{diam}(\Omega))}{K} + \left( \frac{\log(\text{diam}(\Omega))}{K} \right)^2 \right).
\]

(21)

Recall that \(\omega = \lceil \text{diam}(\Omega) \rceil\). As \(N_{\Omega} \geq 3,\)

\[
\omega \asymp \text{diam}(\Omega) \geq c_d > 1,
\]

for a dimensional constant \(c_d\).
For $\ell \in \{1, \ldots, 4\omega\}^d$, consider the matrices $V_K(\xi_\ell)$ defined in (17) and the random vectors $Z_K^\ell$, defined in (18). Let $\Sigma_{|\Omega}$ be the covariance of $X_{|\Omega}$. Then

$$\|\Sigma_{|\Omega}\|_s \leq \|\Sigma\|_s = \|S\|_\infty.$$  \hfill (23)

We apply Theorem 4.1 with $Y = X_{|\Omega}$ and $A = V_K(\xi_\ell)$ to obtain a tail bound on each $Z_K^\ell$ and use Lemma 4.2 to conclude

$$\mathbb{P}(|Z_K^\ell| \geq t) \leq 2 \cdot \exp \left[ -\frac{t}{C} \min \left( \frac{Kt^2}{\|\Sigma\|_s^2}, \frac{Kt}{\|\Sigma\|_s} \right) \right].$$  \hfill (24)

Using (24), the trivial bound $\mathbb{P}(A) \leq 1$, and the union bound leads to

$$\mathbb{P}\left( \max_{\ell \in \{1, \ldots, 4\omega\}^d} |Z_K^\ell| \geq t \right) \lesssim \min \left( 1, \omega^d \cdot \exp \left[ -\frac{t}{C} \min \left( \frac{Kt^2}{\|\Sigma\|_s^2}, \frac{Kt}{\|\Sigma\|_s} \right) \right] \right).$$

We now invoke the tail bound above together with sampling bound (16) to estimate the variance term. In doing so, we have to distinguish the cases (i) $Cd \log \omega \leq K$ and (ii) $Cd \log \omega > K$. In the former case,

$$\mathbb{E}\{\|\hat{S}^{mt} - \mathbb{E}\{\hat{S}^{mt}\}\|_\infty^2\} \leq \mathbb{E}\left\{ \max_{\ell \in \{1, \ldots, 4\omega\}^d} \left| \hat{S}^{mt}(\xi_\ell) - \mathbb{E}\{\hat{S}^{mt}\}(\xi_\ell) \right|^2 \right\}$$

$$= \mathbb{E}\left\{ \max_{\ell \in \{1, \ldots, 4\omega\}^d} |Z_K^\ell|^2 \right\}$$

$$= 2 \int_0^\infty t \cdot \mathbb{P}\left( \max_{\ell \in \{1, \ldots, 4\omega\}^d} |Z_K^\ell| > t \right) \, dt$$

$$\leq \int_0^\infty t \mathbb{P}\left( \max_{\ell \in \{1, \ldots, 4\omega\}^d} \frac{C}{K} \cdot \frac{\log \omega}{\sqrt{K}} K t \right) \, dt + \omega^d \int_0^\infty t \cdot \exp \left[ -\frac{1}{C} \frac{Kt^2}{\|\Sigma\|_s^2} \right] \, dt$$

$$+ \omega^d \int_0^\infty t \cdot \exp \left[ -\frac{1}{C} \frac{Kt}{\|\Sigma\|_s} \right] \, dt$$

$$\leq \frac{C\|\Sigma\|_s^2 d \log \omega}{2K} + \frac{C\|\Sigma\|_s^2 \omega^d}{2K} \int_0^\infty e^{-t} \, dt + \frac{(C\|\Sigma\|_s^2 \omega^d)}{K^2} \int_0^\infty t \cdot e^{-t} \, dt$$

$$= \frac{C\|\Sigma\|_s^2 d \log \omega}{2K} + \frac{C\|\Sigma\|_s^2 \omega^d}{2K} + \frac{(C\|\Sigma\|_s^2 \omega^d)}{K^2} \cdot \Gamma\left( 2, \frac{K}{C} \right),$$

where $\Gamma(s, x)$ denotes the upper incomplete gamma function. As $\Gamma(s + 1, x) = s\Gamma(s, x) + x^s e^{-x}$, we find

$$\mathbb{E}\{\|\hat{S}^{mt} - \mathbb{E}\{\hat{S}^{mt}\}\|_\infty^2\} \leq \frac{\|\Sigma\|_s^2}{K} \left( \log \omega + \omega^d e^{-K/C} \right) \leq \frac{\|\Sigma\|_s^2 \log \omega}{K},$$

\hfill (25)

where the last inequality follows from the assumption $Cd \log \omega \leq K$. 


If \( Cd \log \omega > K \), then, following the same line of arguments, we obtain

\[
\mathbb{E} \{ \| \hat{S}^{\text{mt}} - \mathbb{E} \{ \hat{S}^{\text{mt}} \} \|_\infty^2 \} \lesssim \int_0^\infty \frac{C \| \Sigma \| \log \omega}{K} t \ dt + \omega^d \int_0^\infty \frac{C \| \Sigma \| \log \omega}{K^2} t \cdot \exp \left[ -\frac{1}{C} \frac{K t}{\| \Sigma \| s} \right] dt
\]

\[
= \frac{1}{2} \left( \frac{C \| \Sigma \| \log \omega}{K} \right)^2 + \frac{(C \| \Sigma \| \log \omega)^2}{K^2} \int_0^\infty t \cdot e^{-t} \ dt
\]

\[
= \frac{1}{2} \left( \frac{C \| \Sigma \| \log \omega}{K} \right)^2 + \frac{(C \| \Sigma \| \log \omega)^2}{K^2} \cdot \Gamma(2, d \log \omega)
\]

\[
\lesssim \left( \frac{\| \Sigma \| \log \omega}{K} \right)^2.
\]  \hspace{1cm} (26)

Estimates (25) and (26), together with (23), yield (21).

**Step 3.** We decompose the mean-squared error as

\[
\mathbb{E} \left\{ \| S - \hat{S}^{\text{mt}} \|_\infty^2 \right\} \lesssim \mathbb{E} \{ \| \hat{S}^{\text{mt}} - \mathbb{E} \{ \hat{S}^{\text{mt}} \} \|_\infty^2 \} + \| \mathbb{E} \{ \hat{S}^{\text{mt}} \} - S \|_\infty^2
\]

and combine (20) and (21) to obtain (7).

**Proof of Corollary 2.2:** For every set \( \Omega \subset \mathbb{Z}^d \), one has \( N_\Omega \leq (2 \cdot \text{diam}(\Omega) + 1)^d \lesssim \text{diam}(\Omega)^d \), where the second inequality uses that \( \text{diam}(\Omega) > 1 \) (as \( N_\Omega \geq 3 \)). It then follows from (7) and (8) that

\[
\mathbb{E} \left\{ \| S - \hat{S}^{\text{mt}} \|_\infty^2 \right\} \lesssim \frac{\log(\text{diam}(\Omega))}{K} + \frac{(\log(\text{diam}(\Omega)))^2}{K^\frac{d}{2}} + \frac{\left( \frac{K}{N_\Omega} \right)^{\frac{d}{2}}}{K^\frac{d}{2}}.
\]

If \( d = 1 \), we assume for the moment that \( \log(\text{diam}(\Omega)) \leq K \) which implies

\[
\mathbb{E} \left\{ \| S - \hat{S}^{\text{mt}} \|_\infty^2 \right\} \lesssim \frac{\log(\text{diam}(\Omega))}{K} + \left( \frac{K}{N_\Omega} \right)^{\frac{d}{2}}.
\]

Optimising the right-hand side with respect to \( K \) yields the possibly non-integer value \( K^* = \left( \frac{1}{4} \log(\text{diam}(\Omega)) N_\Omega^4 \right)^{1/5} \). If we choose \( K \) slightly larger than the optimum, namely \( K = \lceil (\log(\text{diam}(\Omega)) N_\Omega^4)^{1/5} \rceil \), then

\[
\frac{\log(\text{diam}(\Omega))}{K} \leq \left( \frac{\log(\text{diam}(\Omega))}{N_\Omega} \right)^{4/5} \leq 1,
\]

by assumption (9). A direct computation yields (10).

Note that there are at most \( N_\Omega \) orthogonal tapers. It is, therefore, necessary to ensure that our choice of \( K \) satisfies \( K \leq N_\Omega \) which is also done by (9).

For \( d \geq 2 \), one has \( \log(\text{diam}(\Omega))/K \leq \log(\text{diam}(\Omega))^2/K^{2/d} \) showing that

\[
\mathbb{E} \left\{ \| S - \hat{S}^{\text{mt}} \|_\infty^2 \right\} \lesssim \frac{(\log \text{diam}(\Omega))^2}{K^\frac{d}{2}} \left( \frac{K}{N_\Omega} \right)^{\frac{d}{2}},
\]

and the previous optimisation steps can be repeated to finish the proof.
4.2. Proof of lower bounds

The Kullback–Leibler divergence of two probability distributions \( P, Q \) is defined by

\[
K(P, Q) = \int \log \frac{dP}{dQ} \, dP,
\]

provided that \( P \) is absolutely continuous with respect to \( Q \) (denoted \( P \ll Q \)). For \( d \)-dimensional normal distributions, the Kullback–Leibler divergence reads

\[
K(\mathcal{N}(0, \Sigma_1), \mathcal{N}(0, \Sigma_2)) = \frac{1}{2} \left( \text{trace}(\Sigma_2^{-1} \Sigma_1) - d - \log \left( \frac{\det \Sigma_1}{\det \Sigma_2} \right) \right).
\] (27)

We will need the following version of Fano’s lemma which can be found, for instance, in Tsybakov (2009, Theorem 2.7).

**Proposition 4.3:** Let \( S_M = \{S_0, S_1, \ldots, S_M\} \) be a family of bounded functions on \( \mathbb{R}^d \), \( M \geq 1 \), \( \{Z_0, \ldots, Z_M\} \) a set of random variables with corresponding probability measures \( \{P_0, \ldots, P_M\} \), and \( d : S_M \times S_M \rightarrow [0, \infty) \) a metric. If

(i) \( d(S_i, S_j) \geq 2\delta > 0, 0 \leq i < j \leq M \),
(ii) \( P_j \ll P_0, j = 1, \ldots, M \), and for some \( 0 < \alpha < 1/8 \)

\[
\sum_{j=1}^{M} K(P_j, P_0) \leq \alpha M \log M,
\] (28)

then we have

\[
\inf_{\hat{S}} \sup_{S \in S_M} \mathbb{E} \left\{ d(\hat{S}, S)^2 \right\} \geq C\delta^2,
\] (29)

where the infimum is taken over all estimators based on an observation of \( Z_0, \ldots, Z_M \) and \( C \) only depends on \( \alpha \).

Let \( S \) be the class of all \( C^2 \) functions on the line, with \( \|S\|_{C^2} \leq 1 \), that are the spectral densities of zero mean, stationary, stochastic Gaussian processes on \( \mathbb{Z}^d \). Each function in \( S \) is in fact the spectral density of a unique stationary, real, zero mean, Gaussian process (which is necessarily ergodic).

The partial Fourier sum of a function \( S \) is denoted by

\[
F_p(S)(\xi) = \sum_{\|k\|_{\infty} \leq p} \mathcal{F}(S)(k) e^{2\pi i \langle \xi, k \rangle}.
\]

The following lemma provides a class that allows us to apply Fano’s lemma.

**Lemma 4.4:** There exists \( \varepsilon > 0 \) such that for all \( M \in \mathbb{N} \) and \( \tau \in (0, \varepsilon) \) a class of functions \( S_M = \{S_0, S_1, \ldots, S_M\} \subseteq S \) with the following properties exists:

(i) \( S_0 \equiv 1/2 \).
(ii) \( \|S_n - S_0\|_\infty \asymp \|S_n - S_m\|_\infty \asymp \frac{\tau}{M^{2/d}}, 1 \leq n \neq m \leq M, \)

(iii) \( \|S_n - S_0\|_2 \asymp \frac{\tau^2}{M^{1+4/d}}, 1 \leq n \leq M, \)

(iv) \( \|F_p(S_n) - S_0\|_\infty \leq 1/4, 1 \leq n \leq M, p \geq 0, \)

(v) \( F_p(S_n) \geq 0, 1 \leq n \leq M, p \geq 0. \)

(The implicit constants and \( \varepsilon \) only depend on the ambient dimension \( d \).)

**Proof:** Step 1. Let us define \( A \in C^\infty(\mathbb{R}^d) \) by

\[
A(x) = \exp \left( -\frac{1}{1-4\|x\|_2^2} \right) \chi_{\|x\|_2 < 1/2}(x), \quad x \in \mathbb{R}^d.
\]

Set \( K := \lceil M^{1/d} \rceil + 1 \) and, for \( n \in \{1, \ldots, K-1\}^d, \)

\[
S_n(\xi) := S_0 + \frac{\tau^2}{K^2} \sum_{\xi \in \{-1,1\}^d} A(2K\xi - \varphi(n,\varepsilon)), \quad \xi \in [-1/2,1/2]^d,
\]

(30)

where \( \varphi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) is defined by

\[
\varphi(x,y)_j = \begin{cases} x_j, & \text{if } y_j \geq 0, \\ -x_j, & \text{if } y_j < 0. \end{cases}
\]

Now pick \( M \) functions out of \( \{S_n\}_{n \in \{1,\ldots,K-1\}^d} \), define (after relabelling the indices) \( S_M := \{S_0, S_1, \ldots, S_M\} \), choose \( \varepsilon \) such that \( \|S_n\|_{C^2} \leq 1/2 + 4\varepsilon \|A\|_{C^2} \leq 1 \), and let \( \tau \in (0, \varepsilon). \)

Step 2. Let us show that the Fourier coefficients of \( S_n \) only depend on the componentwise magnitude of their index:

\[
\mathcal{F}(S_n)(m) = \mathcal{F}(S_n)((|m_1|, \ldots, |m_d|)), \quad m \in \mathbb{Z}^d.
\]

We first note that \( \varphi \) is linear in the first argument and that \( \varphi(\varphi(x,y),y) = x \). Moreover, if \( y_j \neq 0 \), for every \( j = 1, \ldots, d \), then \( \varphi(\varphi(x,y),z) = \varphi(x, \varphi(y,z)) \). These properties together with the symmetry of \( A \) w.r.t. the origin imply that

\[
A(2K\varphi(\xi, m) - \varphi(n,\varepsilon)) = A(2K\varphi(\xi, m) - \varphi(\varphi(n,\varepsilon), m), m))
\]

\[
= A(2K\xi - \varphi(n,\varepsilon), m))
\]

\[
= A(2K\xi - \varphi(n,\varepsilon), m))
\]

\[
= A(2K\xi - \varphi(n,\varepsilon)),
\]

for \( \tilde{\varepsilon} := \varphi(e, m) \in \{-1,1\}^d \). Since \( \varphi(\cdot,y) : \{-1,1\}^d \to \{-1,1\}^d \) is bijective for every \( y \in \mathbb{R}^d \), we conclude that \( S_n(\varphi(\xi, m)) = S_n(\xi) \). Hence, writing \( |m| = (|m_1|, \ldots, |m_d|) \),

\[
\mathcal{F}(S_n)(m) = \int_{[-\frac{1}{2},\frac{1}{2}]^d} S_n(\xi) e^{2\pi i (\xi, m)} \, d\xi = \int_{[-\frac{1}{2},\frac{1}{2}]^d} S_n(\xi) e^{2\pi i (\varphi(\xi, m), |m|)} \, d\xi
\]

\[
= \int_{[-\frac{1}{2},\frac{1}{2}]^d} S_n(\varphi(\xi, m)) e^{2\pi i (\xi, |m|)} \, d\xi = \mathcal{F}(S_n)(|m|).
\]
Step 3. Note that for $e \in \{-1, 1\}^d$ and $n \in \{1, \ldots, K - 1\}^d$,
\[ \operatorname{supp}[A(2K\xi - \varphi(n, e))] \subseteq B_{1/4K}(\varphi(n, e)/2K). \]

As $B_{1/4K}(\varphi(n_1, e_1)/2K)$ and $B_{1/4K}(\varphi(n_2, e_2)/2K)$ are disjoint whenever $e_1 \neq e_2$ or $n_1 \neq n_2$, a direct inspection of (30) yields (ii) and (iii).

In addition, $\|S_0 - S_n\|_{C^2} \leq C\epsilon \leq C\epsilon$, and therefore, $\|F_p(S_0 - S_n)\|_\infty = \|F_p(S_n) - S_0\|_\infty \leq C\epsilon$, for a constant $C > 0$ (see, e.g. Schultz 1969, Theorem 4.4). Hence, (iv) holds, provided that $\epsilon$ is chosen small enough. Finally, by (31), $F_p(S_n) \in \mathbb{R}$ for every $p \geq 0$, so (iv) implies (v).

We can now prove the desired lower bounds.

**Proof of Theorem 2.5:** Set $\omega := \lceil \text{diam}(\Omega) \rceil$ and
\[ M := \left( \frac{\omega}{(\log \omega)^{1/d}} \right)^{d^2/(4+d)} \geq 2. \tag{32} \]
Since $N_\Omega \geq 3$ we have $\omega \geq 2$ and consequently
\[ M \asymp \left( \frac{\omega}{(\log \omega)^{1/d}} \right)^{d^2/(4+d)}. \tag{33} \]

Fix a small parameter $\tau > 0$ to be specified and invoke Lemma 4.4 to obtain functions $\{S_0, S_1, \ldots, S_M\}$.

**Step 1. A more informative model with circulant covariance.** Let $n^* \in \mathbb{Z}^d$ be such that
$\Omega \subseteq n^* + \{0, \ldots, \omega\}^d =: \tilde{\Omega}$.

The samples $\{X_n : n \in \tilde{\Omega}\}$ are distributed according to the covariance matrix $\tilde{\Sigma} = \Sigma_{\tilde{\Omega}}$. Let $Y$ be the $(2\omega + 1)^d$-dimensional Gaussian random variable defined by the multidimensional circulant covariance matrix $\Sigma_Y$,
\[ (\Sigma_Y)_{n,m} := \sigma_{u(n-m)}, \quad m, n \in \{-\omega, \ldots, \omega\}^d, \]
where
\[ u(n - m) = \begin{cases} |n_j - m_j|, & \text{when } |n_j - m_j| \leq \omega, \\ 2\omega + 1 - |n_j - m_j|, & \text{when } \omega + 1 \leq |n_j - m_j| \leq 2\omega, \end{cases} \quad j = 1, \ldots, d. \]

The minimax error of estimating $S$ from samples of $X$ on $\Omega$ is larger than that based on samples on $\tilde{\Omega}$. This latter error is in turn larger than the one corresponding to the estimation of $S$ based on samples of $Y$, as the random vector $Y_{\{0,\ldots,\omega\}^d}$ is distributed exactly as $X_{\tilde{\Omega}}$. Thus,
\[ \inf_{\hat{S}_{X,\tilde{\Omega}}} \sup_{S \in \mathcal{S}} \mathbb{E}\left\{ \|S - \hat{S}_{X,\Omega}\|_\infty^2 \right\} \geq \inf_{\hat{S}_Y} \sup_{S \in \mathcal{S}} \mathbb{E}\left\{ \|S - \hat{S}_Y\|_\infty^2 \right\}, \tag{34} \]
where $\hat{S}_{X,\Omega}$ denotes an estimator based on samples on $\Omega$ of a random process $X$ with spectral density $S$ and $\hat{S}_Y$ denotes an estimator based on samples of $Y$. 


**Step 2. Diagonalisation.** Let $S \in S$ and assume that $F_\omega(S) \geq 0$. As $\Sigma_Y$ is a multidimensional circulant matrix, it is diagonalised by the multidimensional Fourier matrix

$$(U)_{n,m} := (2\omega + 1)^{-d/2} e^{2\pi i \frac{(n,m)}{2\omega + 1}}, \quad n, m \in \{-\omega, \ldots, \omega\}^d.$$ 

Explicitly,

$$USYU^* = \text{diag} \left\{ \left[ F_\omega(S) (\xi_k) \right]_{k \in \{-\omega, \ldots, \omega\}^d} \right\},$$

where

$$\xi_k := \frac{k}{2\omega + 1}.$$ 

Let $Z := UY$. Since the covariance matrix $\Sigma_Z = USYU^*$ is diagonal,

$$Z_k \sim \left[ F_\omega(S) (\xi_k) \right]^{1/2} \eta_k, \quad \|k\|_\infty \leq \omega,$$

where $\eta_k \sim \mathcal{N}(0,1)$. As the experiment of taking a sample from $Y$ contains exactly as much information as taking a sample from $Z$, we get

$$\inf \sup \mathbb{E} \left\{ \left\| S - \widehat{S}_Y \right\|_\infty^2 \right\} = \inf \sup \mathbb{E} \left\{ \left\| S - \widehat{S}_Z \right\|_\infty^2 \right\},$$

where $\widehat{S}_Z$ denotes an estimator based on samples of $Z$.

**Step 3. Estimation of Kullback–Leibler divergences.** Recall that $S_0 \equiv 1/2$ while $F_\omega(S_n)(\xi_k) \geq 0$, for $n = 1, \ldots, M$. We can, therefore, define the random vectors $Z_n$ by

$$(Z_n)_k = \left[ F_\omega(S_n)(\xi_k) \right]^{1/2} \eta_k.$$ 

Then, with the notation of (35),

$$Z_n \sim \mathcal{N}(0, \text{diag} \{ F_\omega(S_n)(\xi_k) \}_{k \in \{-\omega, \ldots, \omega\}^d}).$$

Let $\mathbb{P}_n$ denote the distribution of $Z_n$. By (27),

$$K(\mathbb{P}_n, \mathbb{P}_0)$$

$$= \frac{1}{2} \left[ \text{trace} \left( \text{diag} \left\{ \frac{F_\omega(S_n)(\xi_k)}{S_0} \right\} \right) - (2\omega + 1)^d - \log \left( \frac{\det(\text{diag}(F_\omega(S_n)(\xi_k)))}{\det(\text{diag}(S_0))} \right) \right]$$

$$= \frac{1}{2} \left[ \sum_{\|k\|_\infty \leq \omega} \frac{F_\omega(S_n)(\xi_k)}{S_0} - (2\omega + 1)^d - \log \left( \prod_{\|k\|_\infty \leq \omega} \frac{F_\omega(S_n)(\xi_k)}{S_0} \right) \right]$$

$$= \frac{1}{2} \left[ \sum_{\|k\|_\infty \leq \omega} \left[ \frac{F_\omega(S_n)(\xi_k)}{S_0} - 1 - \log \left( \frac{F_\omega(S_n)(\xi_k)}{S_0} \right) \right] \right].$$

For $\tau$ small enough, we have that $|F_\omega(S_n)(\xi_k)/S_0 - 1| \leq 1/2$. Since $a - \log(1 + a) \leq a^2$, for $|a| \leq 1/2$, it thus follows that

$$\frac{F_\omega(S_n)(\xi_k)}{S_0} - 1 - \log \left( \frac{F_\omega(S_n)(\xi_k)}{S_0} \right) \leq \frac{1}{S_0^2} \left( F_\omega(S_n)(\xi_k) - S_0 \right)^2.$$
Let \( \nu_n(m) \) denote the \( m \)th Fourier coefficients of \( S_n - S_0 \). Then, using Parseval’s identity and \( S_0 \equiv 1/2 \), we estimate

\[
\sum_{n=1}^{M} K(\mathbb{P}_n, \mathbb{P}_0) = \frac{1}{2} \sum_{n=1}^{M} \sum_{\|k\|_{\infty} \leq \omega} \left[ \frac{F_{\omega}(S_n)(\xi_k)}{S_0} - 1 - \log \left( \frac{F_{\omega}(S_n)(\xi_k)}{S_0} \right) \right]
\]

\[
\leq \frac{1}{2S_0^2} \sum_{n=1}^{M} \sum_{\|k\|_{\infty} \leq \omega} |F_{\omega}(S_n)(\xi_k) - S_0|^2
\]

\[
= 2 \sum_{n=1}^{M} \sum_{\|k\|_{\infty} \leq \omega} \left| \sum_{\|m\|_{\infty} \leq \omega} \nu_n(m) e^{2\pi i \langle \xi_k, m \rangle} \right|^2
\]

\[
= 2 \sum_{n=1}^{M} \sum_{\|k\|_{\infty} \leq \omega} \sum_{\|m\|_{\infty} \leq \omega} \sum_{\|\ell\|_{\infty} \leq \omega} \nu_n(m) \overline{\nu_n(\ell)} e^{2\pi i \langle \xi_k, m - \ell \rangle}
\]

\[
= 2(2\omega + 1)^d \sum_{n=1}^{M} \sum_{\|m\|_{\infty} \leq \omega} |\nu_n(m)|^2 \leq 2(2\omega + 1)^d \sum_{n=1}^{M} \sum_{m \in \mathbb{Z}^d} |\nu_n(m)|^2
\]

\[
= 2(2\omega + 1)^d \sum_{n=1}^{M} \|S_n - S_0\|_2^2.
\]

Invoking Lemma 4.4(ii), we conclude

\[
\sum_{n=1}^{M} K(\mathbb{P}_n, \mathbb{P}_0) \lesssim \tau^2 M \frac{\omega^d}{M^{1+4/d}}. \tag{37}
\]

**Step 4. Application of Fano’s inequality.** By our choice of \( M \) in (32),

\[
\frac{\omega^d}{M^{1+4/d}} \leq \log \omega \lesssim \log M.
\]

Combining this with (37), we conclude that

\[
\sum_{n=1}^{M} K(\mathbb{P}_n, \mathbb{P}_0) \leq C \tau^2 M \log(M),
\]

for a constant \( C > 0 \). We choose \( \tau \) so that \( C \tau^2 < 1/8 \) and, consequently, assumption (ii) in Proposition 4.3 is satisfied. Since \( \frac{N{\Omega}}{\Omega} \geq 3 \), \( \text{diam}(\Omega) > 1 \) and \( \omega \asymp \text{diam}(\Omega) \). Note also that \( \frac{N{\Omega}}{d} \leq \text{diam}(\Omega) \). We now invoke Proposition 4.3, and combine its conclusion with
the previous observations and (33), (34) and (36) to estimate
\[
\inf_{\hat{S} \in \mathcal{S}} \sup_{S \in \mathcal{S}} \mathbb{E} \{ \| \hat{S} - S \|_{\infty}^2 \} \gtrsim M^{-4/d} \gtrsim \left( \frac{\log \omega}{\omega} \right)^{\frac{4d}{4+d}} \gtrsim \left( \frac{\log N_\Omega}{\text{diam}(\Omega)} \right)^{\frac{4d}{4+d}}
\]
as desired. \hfill \blacksquare

**Notes**

1. Indeed, by (1), the covariance matrix $\Sigma$ represents a convolution operator on $\ell^2(\mathbb{Z}^d)$, which is unitarily equivalent via a Fourier expansion to a multiplication operator on $L^2([0,1]^d)$ with symbol $S$ (see, e.g. Grenander and Szegö 1958 for more background).
2. Here, and throughout, we write $f \lesssim g$ for two functions if there exists a constant $C > 0$ such that $f(x) \leq C g(x)$ for all $x$, while $f \asymp g$ means $f \lesssim g$ and $g \lesssim f$.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

J. L. R. and M. S. gratefully acknowledge support from the Austrian Science Fund (FWF) [Y 1199 and J 4254].

**ORCID**

José Luis Romero [http://orcid.org/0000-0002-2491-8962](http://orcid.org/0000-0002-2491-8962)

Michael Speckbacher [http://orcid.org/0000-0002-5393-5163](http://orcid.org/0000-0002-5393-5163)

**References**

Abreu, L.D., and Romero, J.L. (2017), ‘MSE Estimates for Multitaper Spectral Estimation and Off-grid Compressive Sensing’, *IEEE Transactions on Information Theory*, 63(12), 7770–7776.

Adamczak, R. (2015), ‘A Note on the Hanson-Wright Inequality for Random Vectors with Dependencies’, *Electronic Communications in Probability*, 20, 1–13.

Andén, J., and Romero, J.L. (2020), ‘Multitaper Estimation on Arbitrary Domains’, *SIAM Journal on Imaging Sciences*, 13(3), 1565–1594.

Bronze, T.P. (1988), ‘Spectral Estimation of Irregularly Sampled Multidimensional Processes by Generalized Prolate Spheroidal Sequences’, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(12), 1862–1873.

Cai, T., Ren, Z., and Zhou, H. (2013), ‘Optimal Rates of Convergence for Estimating Toeplitz Covariance Matrices’, *Probability Theory and Related Fields*, 156(1–2), 101–143.

Grenander, U., and Szegö, G. (1958), *Toeplitz Forms and Their Applications*, California Monographs in Mathematical Sciences, Berkeley, Los Angeles: University of California Press.

Hanssen, A. (1997), ‘Multidimensional Multitaper Spectral Estimation’, *Signal Processing*, 58(3), 327–332.

Harig, C., and Simons, F.J. (2012), ‘Mapping Greenland’s Mass Loss in Space and Time’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(49), 19934–19937.

Hogan, J.A., and Lakey, J.D. (2012), *Duration and Bandwidth Limiting*, Applied and Numerical Harmonic Analysis, New York: Birkhäuser/Springer.

Kabanava, M., and Rauhut, H. (2017), Masked covariance estimation. arXiv:1709.09377.

Karnik, S., Romberg, J., and Davenport, M.A. (2022), ‘Thomson’s Multitaper Method Revisited’, *IEEE Transactions on Information Theory*, to appear. doi:10.1109/TIT.2022.3151415
Lii, K.S., and Rosenblatt, M. (2008), ‘Prolate Spheroidal Spectral Estimates’, *Statistics & Probability Letters*, 78(11), 1339–1348.
Percival, D.B., and Walden, A.T. (1993), *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge: Cambridge University Press.
Pfister, L., and Bresler, Y. (2018), ‘Bounding Multivariate Trigonometric Polynomials with Applications to Filter Bank Design’, *IEEE Transactions on Signal Processing*, 67(3), 700–707. doi:10.1109/TSP.2018.2883925
Rudelson, M., and Vershynin, R. (2013), ‘Hanson–Wright Inequality and Sub-Gaussian Concentration’, *Electronic Communications in Probability*, 18(82), 1–9.
Schultz, M.H. (1969), ‘$L^\infty$-Multivariate Approximation Theory’, *SIAM Journal on Numerical Analysis*, 6(2), 161–183.
Simons, F.J., van der Hilst, R.D., and Zuber, M.T. (2003), ‘Spatiospectral Localization of Isostatic Coherence Anisotropy in Australia and Its Relation to Seismic Anisotropy: Implications for Lithospheric Deformation’, *Journal of Geophysical Research: Solid Earth*, 108(B5), 291–291.
Simons, F.J., and Wang, D.V. (2011), ‘Spatiospectral Concentration in the Cartesian Plane’, *GEM – International Journal on Geomathematics*, 2(1), 1–36.
Slepian, D. (1978), ‘Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty–V: The Discrete Case’, *Bell System Technical Journal*, 57(5), 1371–1430.
Thomson, D.J. (1982), ‘Spectrum Estimation and Harmonic Analysis’, *Proceedings of the IEEE*, 70(9), 1055–1096.
Tsybakov, A.B. (2009), *Introduction to Nonparametric Estimation*. New York: Springer.