Asymptotic equivalence of spectral density and regression estimation

Georgii Golubev and Michael Nussbaum

Humboldt University
Institute for Statistics and Econometrics
Spandauer Straße 1
D–10178 Berlin
Germany
eMail: golubev@wiwi.hu-berlin.de
and
Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstraße 39
D – 10117 Berlin
Germany
eMail: nussbaum@wias-berlin.de

July 1998

1991 *Mathematics Subject Classification.* 62G07, 62G20
Key words and phrases. Stationary Gaussian process, spectral density, Le Cam's distance, asymptotic equivalence, local limit theorem, signal in Gaussian white noise.

Abstract

We consider the statistical experiment given by a sample $y(1),\dots,y(n)$ of a stationary Gaussian process with an unknown smooth spectral density. Asymptotic equivalence with a nonparametric regression in discrete Gaussian white noise is established. The key is a local limit theorem for an increasing number of empirical covariance coefficients.

1 Introduction and main results

Estimation of the spectral density $f(\lambda)$, $\lambda \in [-\pi,\pi]$ of a stationary process is an important and traditional problem of mathematical statistics. We are interested in the function

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} B[k]e^{\mathrm{i}\,\lambda k}$$

and we observe a sample $\mathbf{y}^n = y(1),\dots,y(n)$ from a stationary complex-valued random process $y(n)$ with $\mathbf{E}y(t) = 0$, $\mathbf{E}|y(t)|^2 < \infty$ and covariance function $B[k] = \mathbf{E}y(t)y^*(t+k)$. The practical importance of spectral density estimation is in particular due to the fact that $f(\cdot)$ reflects the energy distribution of the process $y(t)$ in the frequency domain. More precisely, it is well-known (see eg. Gikhman and Skorohod (1969)) that for any stationary process $y(t)$ there exists a stochastic orthogonal measure $Z(\lambda)$ such that

$$y(t) = \int_{-\pi}^{\pi} e^{\mathrm{i}\lambda t} Z(d\lambda)$$

and for any measurable set $A \in [-\pi,\pi]$

$$\mathbf{E}\left|\int_A Z(d\lambda)\right|^2 = \int_A f(\lambda)d\lambda.$$

In particular if $y(t)$ is a Gaussian process it is easy to check that $y(t)$ can be represented as

$$y(t) = \int_{-\pi}^{\pi} e^{\mathrm{i}\lambda t}\sqrt{f(\lambda)}dw(\lambda), \qquad (1)$$

where $w(\lambda)$, $\lambda \in [-\pi,\pi]$ is a complex-valued Brownian motion.

When dealing with a statistical problem we usually have in mind at least two questions: how can computationally reasonable estimators for the object of interest be constructed, and how can their performance be assessed? The goal of the present paper is to propose solutions in the framework of Le Cam's theory of asymptotic equivalence. The main idea of this theory is to approximate the statistical experiment by a simpler one for which the abovementioned problems can be solved more easily.

For simplicity we will assume from now on that the stationary process $y(t)$ is Gaussian and takes values in $\mathbf{R}^1$. Statistical inference about the spectral density $f(\lambda)$ is commonly based on the cumulative density of the observation $\mathbf{y}^n$. It is well known that $\mathbf{y}^n$

has the density

$$p_f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det B_f)^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top B_f^{-1}\mathbf{x}\right), \ \mathbf{x} \in \mathbf{R}^n, \tag{2}$$

where B_f is the covariance matrix with the entries

$$B_{f\,ik} = B[k-i] = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}\,(k-i)\lambda} f(\lambda)\,\mathrm{d}\lambda, \qquad i,k = 1,\dots,n.$$

Evidently this formula is not very useful from a computational point of view when the sample size n is large. Only in the case of the first order auto regression model

$$y(t) = \alpha y(t-1) + \sigma\xi(t),$$

where $\xi(t) \in \mathbf{R}^1$ is standard white Gaussian noise, formula (2) can be substantially simplified. Simple algebra easily reveals that

$$p_f(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2 - \frac{x_1^2}{2\sigma^2}(1-\alpha^2) + \frac{1}{2}\log(1-\alpha^2)\right].$$

Unfortunately, even for autoregression models of order greater that 1 exact expressions for $p_f(\mathbf{x})$ are far more intricate.

There is a simple heuristic idea to overcome this difficulty. The motivation resembles the one proposed by Mann and Wald (1943) and Whittle (1952). Let us replace (1) by its discrete counterpart. Namely, consider the periodic Gaussian process

$$\tilde{y}(t) = \sqrt{\frac{2\pi}{n}} \sum_s \sqrt{f(\lambda_s)}\mathrm{e}^{\mathrm{i}\,t\lambda_s}\xi_s, \tag{3}$$

where $\xi_s \in \mathbf{C}$ is symmetric white Gaussian noise such that $\xi_s = \xi_{-s}^*$, $\mathbf{E}|\xi_s|^2 = 1$, and the grid points λ_s are chosen in such a way that $\lambda_{s+1} - \lambda_s = 2\pi/n$ and $\lambda_{-s} = -\lambda_s$. More precisely, if n is even then the index s takes values $\{-n/2,\dots,-1,1,\dots,n/2\}$ and

$$\xi_s = \frac{\xi_s' - \mathrm{i}\,\xi_s''}{\sqrt{2}}, \quad \lambda_s = \frac{2\pi}{n}s - \frac{\pi}{n}, \quad s > 0,$$

where ξ_s' and ξ_s'' are independent $\mathcal{N}(0,1)$. Otherwise, if n is odd then s takes values $\{-n/2,\dots,-1,0,1,\dots,n/2\}$ and

$$\xi_0 = \xi_0', \quad \xi_s = \frac{\xi_s' - \mathrm{i}\,\xi_s''}{\sqrt{2}},\ s > 0; \quad \lambda_s = \frac{2\pi}{n}s, \quad s \geq 0.$$

Noting that $n^{-1/2}\exp(\mathrm{i}t\lambda_s)$ is an orthonormal system on the discrete grid one easily obtains that the probability density $\tilde{p}_f(\cdot)$ of the Gaussian vector $\tilde{\mathbf{y}}^n = \tilde{y}(1),\dots,\tilde{y}(n)$ is given by

$$\tilde{p}_f(\mathbf{x}) = \exp\left[-\frac{1}{2}\sum_s\left(\frac{I(\lambda_s,\mathbf{x})}{f(\lambda_s)} + \log(2\pi f(\lambda_s))\right)\right], \tag{4}$$

where

$$I(\lambda, \mathbf{x}) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} \mathrm{e}^{\mathrm{i}\lambda t} x_t\right|^2$$

is the periodogram. From (3) we see that the covariance matrix of the periodic process $\tilde{y}(t)$ admits the representation

$$\tilde{B}_{f\,kj} = \mathbf{E}\tilde{y}(t+k)\tilde{y}(t+j) = \frac{2\pi}{n}\sum_{s} \mathrm{e}^{\mathrm{i}\,\lambda_s(k-j)} f(\lambda_s) = \sum_{p=-\infty}^{\infty} B[k-j+np]. \tag{5}$$

Mann, Wald and Whittle proposed to base statistical inference on (4) instead of (2). Formula (4) appears computationally feasible since the periodogram $I(\lambda, \tilde{\mathbf{y}}^n)$ can be computed in $n \log n$ times. Under some regularity conditions it is well known that this device works for a large variety of finite dimensional estimation problems. We refer the reader to Dzhaparidze (1986) where various mathematical justifications of the Whittle idea are discussed.

The first goal of the current paper is to compare the statistical models (1) and (3) within the framework of Le Cam's theory of asymptotic equivalence. We shall see that under common regularity conditions these statistical experiments are asymptotically equivalent.

The cornerstone of this theory is the notion of deficiency distance between two statistical experiments $\mathcal{E} = \{P_f : f \in \Sigma\}$ and $\mathcal{F} = \{G_f : f \in \Sigma\}$ having the same parameter space. For the convenience of the reader we reproduce here the definition of this distance following Le Cam and Yang (1990).

Let $R(\mathcal{E}, W)$ be the set of functions on Σ defined in the following way: $r(f) \in R(\mathcal{E}, W)$ if there is an estimator $\hat{f}$ in $\mathcal{E}$ such that $\mathbf{E}_f W(\hat{f}, f) \le r(f)$.

Definition 1 The deficiency $\delta(\mathcal{E}, \mathcal{F})$ of $\mathcal{E}$ with respect to $\mathcal{F}$ is the smallest number $\epsilon \in [0,1]$ such that for every loss function W, $0 \le W(\cdot,\cdot) \le 1$ and every $r_2 \in R(\mathcal{F}, W)$ there is $r_1 \in R(\mathcal{E}, W)$ such that $r_1(f) \le r_2(f) + \epsilon$ for all $f \in \Sigma$.

Definition 2 The distance $\Delta(\mathcal{E}, \mathcal{F})$ between two experiments $\mathcal{E}$ and $\mathcal{F}$ is the maximum of $\delta(\mathcal{E}, \mathcal{F})$ and $\delta(\mathcal{F}, \mathcal{E})$.

Upper bounds for the Δ-distance between experiments can be obtained from the following general principle, proposed in Le Cam and Yang (1991). Assume that for different f, measures P_f and G_f are absolutely continuous. Consider the likelihood processes $\Lambda_1(f) = \mathrm{d}P_f/\mathrm{d}P_{f_0}$, $\Lambda_2(f) = \mathrm{d}G_f/\mathrm{d}G_{f_0}$ corresponding to the experiment $\mathcal{E}$ and $\mathcal{F}$ respectively. Assume that there are versions $\Lambda_i^*(f)$ of $\Lambda_i(f)$ defined on a *common probability space*. Then

$$\Delta(\mathcal{E}, \mathcal{F}) \le \frac{1}{2}\sup_{f \in \Sigma} \mathbf{E}\left|\Lambda_1^*(f) - \Lambda_2^*(f)\right|. \tag{6}$$

The main difficulty here is the construction of a common probability space (a coupling), with versions of $\Lambda_i(f)$ which are close to each other. Sometimes this probability space is straightforward, as shown by Brown and Low (1996). But in many cases more involved couplings are needed. Using functional versions of the Hungarian construction, Nussbaum (1996) established asymptotic equivalence of the i. i. d. experiment on an interval (with a density of Hölder smoothness exceeding 1/2) and a Gaussian white noise model. Another variant of the Hungarian construction was used by Grama and Nussbaum (1997) for proving asymptotic equivalence of non-Gaussian and Gaussian regression.

It is well known that asymptotic equivalence of nonparametric experiments depends on the size of the underlying functional class. It is assumed from now on that the spectral density $f(\lambda)$ belongs to the functional class Σ_β consisting of all functions satisfying the following conditions:

- $0 < m \le f(\lambda) \le M$ for all $\lambda \in [-\pi, \pi]$
- $\sum_{k=1}^{\infty} k^{2\beta} B^2[k] \le Q < \infty$.

We consider a sample $\mathbf{y}^n$ from a stationary Gaussian process $y(t)$ with unknown $f \in \Sigma \subset \Sigma_\beta$. The corresponding statistical experiment will be denoted as

$$\mathcal{E}^n(\Sigma) = \left(\mathbf{R}^n, \mathcal{B}^n, \left(\mathbf{P}_f^n,\ f \in \Sigma\right)\right),$$

where $\mathbf{P}_f^n$ is the Gaussian measure in $\mathbf{R}^n$ with probability density (2) and $\mathcal{B}^n$ is the Borel σ-algebra. Along with the experiment $\mathcal{E}^n(\Sigma)$ we consider the experiment given by observations of the periodic Gaussian process $\tilde{y}(t)$

$$\tilde{\mathcal{E}}^n(\Sigma) = \left(\mathbf{R}^n, \mathcal{B}^n, \left(\tilde{\mathbf{P}}_f^n,\ f \in \Sigma\right)\right),$$

where $\tilde{\mathbf{P}}_f^n$ is the Gaussian measure with the density (4). Our first result is a local version of asymptotic equivalence of the experiments $\mathcal{E}^n(\Sigma)$ and $\tilde{\mathcal{E}}^n(\Sigma)$.

Theorem 1 *Let $\Sigma^n \subset \Sigma_\beta$ with $\beta > 1/2$ be a sequence of sets such that*

$$\lim_{n\to\infty} \sup_{f, f_0 \in \Sigma^n} (\|f^{(1/2)} - f_0^{(1/2)}\|^2 + \max_\lambda |f(\lambda) - f_0(\lambda)|^2) = 0,$$

where $f^{(1/2)}(\cdot)$ is the derivative of the order 1/2 of $f(\cdot)$. Then

$$\lim_{n\to\infty} \Delta(\mathcal{E}^n(\Sigma), \tilde{\mathcal{E}}^n(\Sigma)) = 0.$$

Remark 1 Note that in both experiments there exist estimators $\hat{f}_n$ such that for $\beta > 1/2$

$$\lim_{n\to\infty} \sup_{f\in\Sigma_\beta} \left[\mathbf{E}_f \|\hat{f}_n^{(1/2)} - f^{(1/2)}\|^2 + \mathbf{E}_f \max_\lambda |\hat{f}_n(\lambda) - f(\lambda)|^2 \right] n^{(2\beta-1)/2\beta} < \infty.$$

Thus the globalization arguments developed in Nussbaum (1996) and the above theorem imply that if $\Sigma \subset \Sigma_\beta$, $\beta > 1/2$ then

$$\lim_{n\to\infty} \Delta(\mathcal{E}^n(\Sigma), \tilde{\mathcal{E}}^n(\Sigma)) = 0.$$

In Theorem 1 we used a trivial construction of the common probability space:

$$\mathbf{y}^n = B^{1/2}\xi, \qquad \tilde{\mathbf{y}}^n = \tilde{B}^{1/2}\xi,$$

where ξ is $\mathcal{N}(0, E)$. Therefore asymptotic equivalence in this theorem is constructive, as in Brown and Low (1996) (i. e. the corresponding Markov kernels can easily be written down, and serve as "recipes" for obtaining optimal procedures).

The statistical experiment $\tilde{\mathcal{E}}^n(\Sigma)$ is of course simpler than $\mathcal{E}^n(\Sigma)$ but assessing the performance of an estimator still remains a difficult problem. Taking the Fourier transform in (3) we obtain an equivalent representation of $\tilde{\mathcal{E}}^n(\Sigma)$

$$X_k = \sqrt{f(\lambda_k)}\ \xi_k, \qquad \text{where} \qquad X_k = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} y(t) e^{\mathrm{i}\lambda_k t},$$

and $\xi_k \in \mathbf{C}$ is a symmetric white Gaussian noise. So if the spectral density f depends only on a finite dimensional unknown parameter, then one could use the theory of local asymptotic normality to assess the performance of an estimator. This theory is well developed, cf. Ibragimov and Khasminskii (1981) or Bickel, Klaassen, Ritov and Wellner (1993).

In the framework of Le Cam's theory our next step is to find a simpler statistical experiment than $\mathcal{E}^n(\Sigma)$. In order to reflect accurately the statistical nature of the spectral density estimation problem, we approximate this experiment by a regression model of minimal dimensionality. Let $\lambda_s^{N_n}$ be a symmetric uniform grid with step $2\pi/N_n$ and

$$\tilde{\mathcal{R}}^n(\Sigma) = \left(\mathbf{R}^n, \mathcal{B}^n, \left(\tilde{\mathbf{G}}_f^n,\ f \in \Sigma\right)\right)$$

be the experiment associated with the following Gaussian regression model

$$Y_s = f(\lambda_s^{N_n}) + \sqrt{\frac{N_n}{n}} f(\lambda_s^{N_n})\xi_s, \qquad \lambda_s^{N_n} \in (0, \pi], \tag{7}$$

where ξ_s are i.i.d. $\mathcal{N}(0, 1)$.

Theorem 2 *Let $\Sigma \subset \Sigma_\beta$ with $\beta > 1$ and let a sequence of integers N_n be such that $N_n \le \sqrt{n/\log^{1+\varepsilon} n}$, for some $\varepsilon > 0$ and*

$$\lim_{n\to\infty} \frac{n}{N_n^{2\beta}} = 0.$$

Then the experiment $\tilde{\mathcal{E}}^n(\Sigma)$ is asymptotically equivalent to $\tilde{\mathcal{R}}^n(\Sigma)$.

A final step is to apply a variance stabilizing transform to the data (7), cp. Grama and Nussbaum (1997), which in this case amounts to taking (essentially) the logarithm of the Y_s. Let

$$\mathcal{R}^n(\Sigma) = \left(\mathbf{R}^n, \mathcal{B}^n, \left(\mathbf{G}_f^n,\ f \in \Sigma\right)\right)$$

be the experiment associated with the Gaussian regression model

$$Y_s = \log(f(\lambda_s^{N_n})) + \sqrt{\frac{N_n}{n}}\xi_s, \qquad \lambda_s^{N_n} \in [0,\pi] \tag{8}$$

Theorem 3 *Under the conditions of theorem 2, the experiment $\tilde{\mathcal{E}}^n(\Sigma)$ is asymptotically equivalent to $\mathcal{R}^n(\Sigma)$.*

Remark 2 The proof of the above results is based on a local limit theorem for the empirical covariance function

$$\bar{B}[k, \mathbf{y}^n] = \frac{1}{n}\sum_{t=1}^{n} \tilde{y}(t)\tilde{y}(t+k), \quad k = 0, \dots, N_n - 1.$$

Therefore the approximations are constructive, to the same degree as the approximation of i. i. d. zero mean Gaussian data with unknown variance by a one dimensional Gaussian shift (see Nussbaum (1998)).

For extending the above result to larger functional classes than Sobolev balls with smoothness greater than 1, one could directly use the results of Grama and Nussbaum (1997). However these are based on the Hungarian construction, and therefore we can no longer exhibit realistic recipes (Markov kernels) for the asymptotic equivalence. Let Σ'_β be a subset in Σ_β such that any function from Σ'_β is of Hölder smoothness β

$$|f(y) - f(x)| \le C|y - x|^\beta.$$

Theorem 4 *If $\Sigma^n \subset \Sigma'_\beta$ with $\beta > 1/2$ then the experiments $\mathcal{E}^n(\Sigma^n)$ and $\tilde{\mathcal{E}}^n(\Sigma^n)$ are asymptotically equivalent to $\mathcal{R}^n(\Sigma)$.*

2 The periodic Gaussian experiment

Let $\mathbf{y}_0^n$ be a sample of the length n from the stationary Gaussian process $y(t)$ with the spectral density $f_0(\lambda)$ and $\tilde{\mathbf{y}}_0$ its periodic counterpart. Consider two likelihood ratios

$$\begin{aligned}\Lambda_1(f) &= \frac{p_f(\mathbf{y}_0^n)}{p_{f_0}(\mathbf{y}_0^n)} = \left(\frac{\det B_{f_0}}{\det B_f}\right)^{1/2} \exp\left(-\frac{1}{2}\mathbf{y}_0^{nT} B_f^{-1}\mathbf{y}_0^n + \frac{1}{2}\mathbf{y}_0^{nT} B_{f_0}^{-1}\mathbf{y}_0^n\right) \\ &= \left(\frac{\det B_{f_0}}{\det B_f}\right)^{1/2} \exp\left(-\frac{1}{2}\xi^T B_f^{-1} B_{f_0}\xi + \frac{1}{2}\xi^T\xi\right)\end{aligned}$$

and

$$\begin{aligned}\Lambda_2(f) &= \frac{\tilde{p}_f(\tilde{\mathbf{y}}_0^n)}{\tilde{p}_{f_0}(\tilde{\mathbf{y}}_0^n)} = \left(\frac{\det \tilde{B}_{f_0}}{\det \tilde{B}_f}\right)^{1/2} \exp\left(-\frac{1}{2}\tilde{\mathbf{y}}_0^{nT} \tilde{B}_f^{-1}\tilde{\mathbf{y}}_0^n + \frac{1}{2}\tilde{\mathbf{y}}_0^{nT} \tilde{B}_{f_0}^{-1}\tilde{\mathbf{y}}_0^n\right) \\ &= \left(\frac{\det \tilde{B}_{f_0}}{\det \tilde{B}_f}\right)^{1/2} \exp\left(-\frac{1}{2}\xi^T \tilde{B}_f^{-1} \tilde{B}_{f_0}\xi + \frac{1}{2}\xi^T\xi\right),\end{aligned}$$

where $\xi \sim \mathcal{N}(0, E)$. In the following lemma we estimate the Hellinger distance

$$H^2(\Lambda_1, \Lambda_2) = \frac{1}{2}\mathbf{E}\left[\Lambda_1^{1/2}(f) - \Lambda_2^{1/2}(f)\right]^2$$

between the above likelihood processes.

Lemma 1 *Uniformly in $f,\ f_0 \in \Sigma_\beta$ with $\beta > 1/2$ as $n \to \infty$*

$$H^2(\Lambda_1, \Lambda_2) \le C(m, M)\left[\max_\lambda |f(\lambda) - f_0(\lambda)|^2 \|f_0^{(1/2)}\|^2 + \|f^{(1/2)} - f_0^{(1/2)}\|^2\right],$$

where $C(m, M)$ is a constant which does not depend on n, and where $f^{(1/2)}(\cdot)$ denotes the derivative of order $1/2$ of $f(\cdot)$.

Proof. By simple algebra one easily obtains

$$\begin{aligned}H^2(\Lambda_1, \Lambda_2) &= 1 - \left(\frac{\det B_f \det \tilde{B}_f}{\det B_{f_0} \det \tilde{B}_{f_0}} \det^2\left(\frac{1}{2}B_f^{-1}B_{f_0} + \frac{1}{2}\tilde{B}_f^{-1}\tilde{B}_{f_0}\right)\right)^{-1/4} \qquad (9) \\ &= 1 - \det{}^{-1/2}\left(\frac{1}{2}(B_f^{-1}B_{f_0})^{1/2}(\tilde{B}_f^{-1}\tilde{B}_{f_0})^{-1/2} + \frac{1}{2}(B_f^{-1}B_{f_0})^{-1/2}(\tilde{B}_f^{-1}\tilde{B}_{f_0})^{1/2}\right).\end{aligned}$$

Denote for brevity

$$\begin{aligned}A_+ &= \frac{1}{2}(B_f^{-1}B_{f_0})^{1/2}(\tilde{B}_f^{-1}\tilde{B}_{f_0})^{-1/2} + \frac{1}{2}(B_f^{-1}B_{f_0})^{-1/2}(\tilde{B}_f^{-1}\tilde{B}_{f_0})^{1/2}, \\ A_- &= \frac{1}{2}(B_f^{-1}B_{f_0})^{1/2}(\tilde{B}_f^{-1}\tilde{B}_{f_0})^{-1/2} - \frac{1}{2}(B_f^{-1}B_{f_0})^{-1/2}(\tilde{B}_f^{-1}\tilde{B}_{f_0})^{1/2}.\end{aligned}$$

Let $s_k[A]$ be s-numbers of the matrix A. It is well-known that if the spectral density $f(\lambda)$ is strictly bounded from below and from above then $s_1[B_f]$, $s_1[\tilde{B}_f]$, $s_1[B_f^{-1}]$, $s_1[\tilde{B}_f^{-1}]$ are bounded from above (see eg. Dzhaparidze (1986) or Davies (1973)). Therefore denoting $\Delta B = B_f - B_{f_0}$, $\Delta\tilde{B} = \tilde{B}_f - \tilde{B}_{f_0}$ and using elementary properties of s–numbers we obtain from (9)

$$\begin{aligned} H^2(\Lambda,\tilde{\Lambda}) &\le 1 - \exp\left(-\frac{1}{4}\sum_{k=1}^n \log s_k^2[A_+]\right) \le \frac{1}{4}\sum_{k=1}^n \log s_k^2[A_-] \qquad (10)\\ &\le \frac{1}{4}\sum_{k=1}^n \left(s_k^2\left[A_+\right]-1\right) = \frac{1}{4}\sum_{k=1}^n s_k^2\left[A_-\right] \le C(m,M)\sum_{k=1}^n s_k^2[B_f^{-1}B_{f_0}\tilde{B}_{f_0}^{-1}\tilde{B}_f - E]\\ &\le C(m,M)\sum_{k=1}^n s_k^2[B_{f_0}^{-1}B_f - \tilde{B}_{f_0}^{-1}\tilde{B}_f] \le C(m,M)\sum_{k=1}^n s_k^2[B_{f_0}^{-1}\Delta B - \tilde{B}_{f_0}^{-1}\Delta\tilde{B}]\\ &\le C(m,M)\sum_{k=1}^n s_k^2[\tilde{B}_{f_0}^{-1}(\Delta B - \Delta\tilde{B})] + C(m,M)\sum_{k=1}^n s_k^2[(B_{f_0}^{-1} - \tilde{B}_{f_0}^{-1})\Delta B]\\ &\le C(m,M)\sum_{k=1}^n s_k^2[\Delta B - \Delta\tilde{B}] + C(m,M)s_1^2[\Delta B]\sum_{k=1}^n s_k^2[B_{f_0} - \tilde{B}_{f_0}]. \end{aligned}$$

Next, note that

$$\begin{aligned} \sum_{k=1}^n s_k^2[B_{f_0} - \tilde{B}_{f_0}] &= \sum_{k,l=1}^n \left(\sum_{p\neq 0} B_{f_0}[k-l+np]\right)^2 \qquad (11)\\ &\le 2\sum_{k,l=1}^n \left(\sum_{|p|>1} B_{f_0}[k-l+np]\right)^2 + 2\sum_{k,l=1}^n \left(B_{f_0}[k-l+n] + B_{f_0}[k-l-n]\right)^2. \end{aligned}$$

It is easy to see that

$$\sum_{k,l=1}^n B_{f_0}^2[k-l\pm n] \le 2\sum_k |k|B_{f_0}^2[k] = \int_{-\pi}^{\pi} |f_0^{(1/2)}(\lambda)|^2\,\mathrm{d}\lambda. \qquad (12)$$

On the other hand, the Cauchy–Schwartz inequality yields

$$\begin{aligned} \sum_{k,l=1}^n \left(\sum_{|p|>1} B_{f_0}[k-l+np]\right)^2 &\le n\sum_{k=-n}^n \left(\sum_{|p|>1} B_{f_0}[k+np]\right)^2\\ &\le n\sum_{k=-n}^n \sum_{|p|>1} |k+np|^{2\beta} B_{f_0}^2[k+np] \sum_{|q|>1} |k+nq|^{-2\beta}\\ &\le n^{1-2\beta}\sum_{k=-n}^n \sum_{|p|>1} |k+np|^{2\beta} B_{f_0}^2[k+np] \le n^{1-2\beta}\sum_{|k|\ge n} |k|^{2\beta} B_{f_0}^2[k] \le \|f_0^{1/2}\|^2. \end{aligned}$$

Thus by (11) and (12)

$$\sum_{k=1}^n s_k^2[B_{f_0} - \tilde{B}_{f_0}] = C\|f_0^{(1/2)}\|^2. \qquad (13)$$

With similar arguments we get

$$\sum_{k=1}^{n} s_k^2[\Delta B - \Delta\tilde{B}] \leq C\|f^{(1/2)} - f_0^{(1/2)}\|^2.$$

Thus the assertion of the lemma follows immediately from the above inequality and (10), (13). ■

Proof of Theorem 1. This follows now directly from the above lemma, (6) and the well-known inequality $\mathbf{E}|\Lambda_1(f) - \Lambda_2(f)| \leq H(\Lambda_1, \Lambda_2)$. ■

In the sequel we will need some simple results about the Hellinger distance between Gaussian distributions. Denote by $H(\tilde{p}_f, \tilde{p}_g)$ the Hellinger distance between the densities $\tilde{p}_f(\mathbf{x})$ and $\tilde{p}_g(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^n$ defined by (4).

Lemma 2

$$H^2(\tilde{p}_f, \tilde{p}_g) \leq \frac{1}{4}\sum_s \left(\sqrt[4]{\frac{f(\lambda_s)}{g(\lambda_s)}} - \sqrt[4]{\frac{g(\lambda_s)}{f(\lambda_s)}}\right)^2.$$

Proof. By simple algebra one obtains

$$\begin{aligned} H^2(\tilde{p}_f, \tilde{p}_g) &= 1 - \left[\det \tilde{B}_f \det \tilde{B}_g \det^2\left(\frac{1}{2}\tilde{B}_f^{-1} + \frac{1}{2}\tilde{B}_g^{-1}\right)\right]^{-1/4} \\ &= 1 - \det^{-1/2}\left(\frac{1}{2}\tilde{B}_f^{-1/2}\tilde{B}_g^{1/2} + \frac{1}{2}\tilde{B}_f^{1/2}\tilde{B}_g^{-1/2}\right). \end{aligned}$$

The eigenvalues of the matrix $\tilde{B}_f^{-1/2}\tilde{B}_g^{1/2} + \tilde{B}_f^{1/2}\tilde{B}_g^{-1/2}$ take the values $f^{-1/2}(\lambda_s)g^{1/2}(\lambda_s) + f^{1/2}(\lambda_s)g^{-1/2}(\lambda_s)$. Thus we get from the above equation

$$\begin{aligned} H^2(\tilde{p}_f, \tilde{p}_g) &= 1 - \exp\left[-\frac{1}{2}\sum_s \log\left(\frac{1}{2}\sqrt{\frac{f(\lambda_s)}{g(\lambda_s)}} + \frac{1}{2}\sqrt{\frac{g(\lambda_s)}{f(\lambda_s)}}\right)\right] \\ &\leq 1 - \exp\left[-\frac{1}{2}\sum_s \left(\frac{1}{2}\sqrt{\frac{f(\lambda_s)}{g(\lambda_s)}} + \frac{1}{2}\sqrt{\frac{g(\lambda_s)}{f(\lambda_s)}} - -1\right)\right] \\ &= 1 - \exp\left[-\frac{1}{4}\sum_s \left(\sqrt[4]{\frac{f(\lambda_s)}{g(\lambda_s)}} - \sqrt[4]{\frac{g(\lambda_s)}{f(\lambda_s)}}\right)^2\right] \\ &\leq \frac{1}{4}\sum_s \left(\sqrt[4]{\frac{f(\lambda_s)}{g(\lambda_s)}} - \sqrt[4]{\frac{g(\lambda_s)}{f(\lambda_s)}}\right)^2. \end{aligned}$$

■

Consider two Gaussian processes

$$\hat{y}_f(t) = \sqrt{\frac{2\pi}{n}} \sum_s \mathrm{e}^{\mathrm{i}\,\lambda_s t} f(\lambda_s)\xi_s, \qquad \hat{y}_g(t) = \sqrt{\frac{2\pi}{n}} \sum_s \mathrm{e}^{\mathrm{i}\,\lambda_s t} g(\lambda_s)\xi_s,$$

where $\xi_s \in \mathbf{R}^1$ is a symmetric white Gaussian noise. Let $\hat{p}_f(\mathbf{x})$ and $\hat{p}_g(\mathbf{x})$, $x \in \mathbf{R}^n$ be the joint probability densities of $\hat{y}_f(0), \ldots, \hat{y}_f(N-1)$ and $\hat{y}_g(0), \ldots, \hat{y}_g(N-1)$ respectively.

Lemma 3 *Assume that $f(\lambda)$, $g(\lambda) \geq m > 0$ and $f(\lambda)$, $g(\lambda) \leq M < \infty$. Then*

$$H^2(\hat{p}_f, \hat{p}_g) \leq \frac{C(m, M)N}{n} \sum_s (f(\lambda_s) - g(\lambda_s))^2.$$

Proof. It is similar to the proof of Lemma 1 and is omitted.

3 The regression model

In this section we compare the experiment given by (3) with the regression model (7). Note that the likelihood process (see (4)) can be rewritten in the following equivalent form

$$\tilde{p}_f(\tilde{\mathbf{y}}^n) = \exp\left\{ -\frac{1}{4\pi} \sum_p \tilde{B}[p, \tilde{\mathbf{y}}^n] \sum_s \frac{\mathrm{e}^{\mathrm{i}\,\lambda_s \mathrm{p}}}{f(\lambda_s)} - \frac{1}{2} \sum_s \log\left(2\pi f\left(\lambda_s\right)\right) \right\},$$

where the empirical covariance function

$$\tilde{B}[p, \tilde{\mathbf{y}}^n] = \frac{1}{n} \sum_{t=1}^{n} \tilde{y}(t)\, \tilde{y}(t+p) = \frac{2\pi}{n} \sum_s \mathrm{e}^{\mathrm{i}\,\lambda_s p} I_n\left(\lambda_s, \tilde{\mathbf{y}}^n\right). \tag{14}$$

is a sufficient statistic. Similar to the scheme for proving asymptotic equivalence proposed in Nussbaum (1998), our first step is to study the distribution of these values as $n \to \infty$. Using (3) and (14) we obtain that

$$\kappa_p = \sqrt{n}\left(\tilde{B}[p, \tilde{\mathbf{y}}^n] - \tilde{B}_{f\,0p}\right) = \frac{2\pi}{\sqrt{n}} \sum_s \mathrm{e}^{\mathrm{i}\,\lambda_s p} f\left(\lambda_s\right) \left(|\xi_s|^2 - 1\right), \tag{15}$$

where $\tilde{B}_{f\,kj}$ is defined in (5).

Consider the accompanying Gaussian vector

$$\eta_k = \frac{2\pi}{\sqrt{n}} \sum_s f(\lambda_s) \mathrm{e}^{\mathrm{i}\,\lambda_s k} \xi'_s,$$

where $\xi'_s \in \mathbf{R}^1$ is symmetric white Gaussian noise. Denote by $p^N_\kappa(\mathbf{x})$ and $p^N_\eta(\mathbf{x})$, $\mathbf{x} \in \mathbf{R}^N$ the probability densities corresponding to random variables κ_p and η_p, for $p = 0, \ldots, N-1$. Further considerations are crucially based on the following fact.

Lemma 4 *Let $N^2 \le n \log^{-1-\varepsilon} n$ for some $\varepsilon > 0$. Then uniformly in $f(\lambda)$ such that $0 < m \le f(\lambda) \le M < \infty$*

$$\int_{\mathbf{R}^N} \left| p_\kappa^N(\mathbf{x}) - p_\eta^N(\mathbf{x}) \right| d\mathbf{x} \le \frac{CN}{\sqrt{n}} \left(1 + \sqrt{N d_N(f)} \right), \tag{16}$$

where

$$d_N(f) = \min_{c_k} \sup_\lambda \left| f(\lambda) - \sum_{|k| \le N} c_k \cos(\lambda k) \right|^2$$

and the minimum is taken over c_k fulfilling

$$\sum_{|k| \le N} c_k \cos(\lambda k) \ge m/2 > 0.$$

The proof will be given in an appendix.

Proof of Theorem 2. Since $f \in \Sigma_\beta$ with $\beta > 1$, we can find a polynomial $p(\lambda)$ of order N_n such that

$$\sum_s \left(f^{-1}(\lambda_s) - p(\lambda_s) \right)^2 \le \frac{C(m, M)n}{N_n^{2\beta}}.$$

Therefore by Lemma 2 we can assume without loss of generality that $f(\lambda) = 1/p(\lambda)$, where $p(\lambda)$ is a polynomial of the order N_n. On the other hand Lemma 4 implies that our model is equivalent to estimation of $f(\cdot)$ based on observations

$$r_k = \tilde{B}_{f0k} + \frac{2\pi}{n} \sum_s \mathrm{e}^{\mathrm{i}\,\lambda_s k} f(\lambda_s) \xi'_s, \quad |k| \le N_n,$$

where ξ'_s is symmetric white Gaussian noise and $\tilde{B}_{f0k}$ are the linear functionals of f defined in (5). Using Lemma 3 we see that this model is asymptotically equivalent to the following one:

$$r'_k = \tilde{B}_{f0k} + \frac{2\pi}{n} \sum_s \mathrm{e}^{\mathrm{i}\,\lambda_s k} \mathbf{\Pi}^{N_n/2} f(\lambda_s) \xi'_s, \quad |k| \le N_n,$$

where $\mathbf{\Pi}^M$ is the projection operator onto the space of trigonometric polynomials of the order M. Taking the discrete Fourier transform of the above equation we arrive at the equivalent model

$$Y'_p = \mathbf{\Pi}^N f(\lambda_p) + \frac{1}{\sqrt{n}} \zeta_p,$$

where

$$\zeta_p = \frac{1}{\sqrt{n}} \sum_{|k| < N_n} \sum_s \mathrm{e}^{\mathrm{i}\,(\lambda_s - \lambda_p^{N_n})k} \mathbf{\Pi}^{N_n/2} f(\lambda_s) \xi'_s.$$

Noting that for any $|k| < N_n$

$$\frac{1}{n}\sum_{|s|\le n/2}\left(\mathbf{\Pi}^{N_n/2}f(\lambda_s)\right)^2 \mathrm{e}^{\mathrm{i}\,\lambda_s k} = \frac{1}{N_n}\sum_{|s|\le N_n/2}\left(\mathbf{\Pi}^{N_n/2}f(\lambda_s^{N_n})\right)^2 \mathrm{e}^{\mathrm{i}\,\lambda_s^N k},$$

we have $\mathbf{E}\zeta_p\zeta_{p+u} = N\delta_u\left(\mathbf{\Pi}^{N_n/2}f(\lambda_p^{N_n})\right)^2$, for $p,u \ge 0$. Hence we can represent the observations Y_p' in the form

$$Y_p' = \mathbf{\Pi}^{N_n}f(\lambda_p^{N_n}) + \sqrt{\frac{N_n}{n}}\mathbf{\Pi}^{N_n/2}f(\lambda_p^{N_n})\xi_p,$$

where $\xi_s \in \mathbf{R}^1$ is a symmetric white Gaussian noise. The above formula proves the theorem since the square of the Hellinger distance between the probability densities of Y_p from (7) and Y_p' is estimated from above as

$$\frac{n}{N}\sum_p\left[\mathbf{\Pi}^{N_n/2}f(\lambda_p^{N_n})\right]^{-1}\left[f(\lambda_p^{N_n}) - \mathbf{\Pi}^{N_n/2}f(\lambda_p^{N_n})\right]^2$$
$$+\sum_p\left[\left(\frac{f(\lambda_p^{N_n})}{\mathbf{\Pi}^{N_n/2}f(\lambda_p^{N_n})}\right)^{1/4} - \left(\frac{\mathbf{\Pi}^{N_n/2}f(\lambda_p^{N_n})}{f(\lambda_p^{N_n})}\right)^{1/4}\right]^2 = o(1).$$

■

4 The variance stabilizing transform

In this section we compare the heteroscedastic regression model given by (7) with the Gaussian shift model (8). Let ξ be $\mathcal{N}(0,1)$. Consider the random variables

$$\xi(\varepsilon) = \xi\mathbf{1}\left\{\xi > -1/\varepsilon\right\}, \qquad \xi'(\varepsilon) = \log(1+\varepsilon\xi(\varepsilon))/\varepsilon.$$

Lemma 5 *As $\varepsilon \to 0$*

$$H(\mathcal{L}[\xi],\mathcal{L}[\xi(\varepsilon)]) = O(\varepsilon), \tag{17}$$
$$H(\mathcal{L}[\xi'(\varepsilon)],\mathcal{L}[\xi(\varepsilon)]) = O(\varepsilon). \tag{18}$$

Proof. We have for the densities $p_{\xi(\varepsilon)}(x)$, $p_{\xi'(\varepsilon)}(x)$

$$p_{\xi(\varepsilon)}(x) = \frac{\mathbf{1}\{x > -1/\varepsilon\}}{\sqrt{2\pi}\mathbf{P}\{\xi > -1/\varepsilon\}}\exp\left(-\frac{x^2}{2}\right),$$
$$p_{\xi'(\varepsilon)}(x) = \frac{1}{\sqrt{2\pi}\mathbf{P}\{\xi > -1/\varepsilon\}}\exp\left(-\frac{x^2}{2}\right)\exp\left(\varepsilon x - \frac{x^2}{2}g(\varepsilon x)\right),$$

where $g(t) = t^{-2}(\exp(t) - 1)^2 - 1$. By simple algebra

$$\begin{aligned} H^2(\mathcal{L}[\xi], \mathcal{L}[\xi(\varepsilon)]) &= 1 - \int_{-\infty}^{\infty} \left(p_\xi(x) p_{\xi(\varepsilon)}(x)\right)^{1/2} \, \mathrm{d}x \\ &= 1 - \frac{1}{\mathbf{P}^{1/2}\{\xi > -1/\varepsilon\}} \frac{1}{\sqrt{2\pi}} \int_{-1/\varepsilon}^{\infty} \exp\left(-\frac{x^2}{2}\right) \, \mathrm{d}x \\ &= 1 - \mathbf{P}^{1/2}\{\xi > -1/\varepsilon\} = O(\varepsilon^2). \end{aligned}$$

Let $h_\varepsilon = (2 \log \varepsilon^{-2})^{1/2}$ then we have by the Cauchy-Schwartz inequality

$$H^2(\mathcal{L}[\xi'(\varepsilon)], \mathcal{L}[\xi(\varepsilon)]) = 1 - \int_{-\infty}^{\infty} \left(p_{\xi'(\varepsilon)}(x) p_{\xi(\varepsilon)}(x)\right)^{1/2} \, \mathrm{d}x$$
$$\leq 1 - \int_{|x| \leq h_\varepsilon} \left(p_{\xi'(\varepsilon)}(x) p_{\xi(\varepsilon)}(x)\right)^{1/2} \, \mathrm{d}x + \mathbf{P}^{1/2}\{|\xi(\varepsilon)| > h_\varepsilon\} \mathbf{P}^{1/2}\{|\xi'(\varepsilon)| > h_\varepsilon\}.$$

According to the definition of h_ε

$$\begin{aligned} \mathbf{P}\{|\xi(\varepsilon)| > h_\varepsilon\} &= O(\varepsilon^2), \\ \mathbf{P}\{|\xi'(\varepsilon)| > h_\varepsilon\} &= O(\varepsilon^2). \end{aligned}$$

Also, noting that $g(t) = t + t^2/4 + O(t^3),\ t \to 0$ and using Taylor expansion one obtains

$$\begin{aligned} &\int_{|x| \leq h_\varepsilon} \left(p_{\xi'(\varepsilon)}(x) p_{\xi(\varepsilon)}(x)\right)^{1/2} \, \mathrm{d}x \\ &= \frac{1}{\mathbf{P}^{1/2}\{\xi > -1/\varepsilon\}} \frac{1}{\sqrt{2\pi}} \int_{-h_\varepsilon}^{h_\varepsilon} \exp\left(-\frac{x^2}{2} + \varepsilon x - \frac{x^2}{2} g(\varepsilon x)\right) \, \mathrm{d}x \\ &= \frac{1}{\mathbf{P}^{1/2}\{\xi > -1/\varepsilon\}} \frac{1}{\sqrt{2\pi}} \int_{-h_\varepsilon}^{h_\varepsilon} \exp\left(-\frac{x^2}{2}\right) \left(1 + \varepsilon x - \frac{\varepsilon x^3}{2} + O(\varepsilon^2 x^6)\right) \, \mathrm{d}x \\ &= 1 + O(\varepsilon^2). \end{aligned}$$

■

Proof of Theorem 3. Let us denote $\mathcal{S}_n = \{s : \lambda_s^{N_n} \in (0, \pi]\}$ and $\varepsilon_n = (N_n/n)^{1/2}$; we use $\lambda_s = \lambda_s^{N_n}$ for the grid points. The experiment generated by the observations

$$\log \max\{\tilde{Y}_s, 0\}, \quad s \in \mathcal{S}_n$$

where

$$\tilde{Y}_s = f(\lambda_s) + \varepsilon_n f(\lambda_s) \xi_s$$

is equivalent to the experiment generated by the observations

$$Z_s = \log Y_s^{'}, \quad s \in \mathcal{S}_n$$

where

$$Y_s^{'} = f(\lambda_s) + \varepsilon_n f(\lambda_s) \xi_s(\varepsilon_n),$$

$$\xi_s(\varepsilon_n) = \begin{cases} \xi_s, & \text{when } \xi_i \geq -1/\varepsilon_n \\ 0, & \text{otherwise,} \end{cases}$$

where $\xi_s \sim N(0,1)$ are i. i. d. Let us prove that this experiment is in turn asymptotically equivalent to the experiment generated by observations

$$Z_s' = \log f(\lambda_s) + \varepsilon_n \xi_s(\varepsilon_n).$$

Using that $H^2(\mathcal{L}[\zeta_1], \mathcal{L}[\zeta_2]) = H^2(\mathcal{L}[a\zeta_1 + b], \mathcal{L}[a\zeta_2 + b])$, $a > 0$ we have by (18)

$$\begin{aligned} H^2\left(\mathcal{L}[(Z_s)_{s\in\mathcal{S}_n}], \mathcal{L}[(Z_s')_{s\in\mathcal{S}_n}]\right) &\leq 2\sum_{s\in\mathcal{S}_n} H^2\left(\mathcal{L}[Z_s], \mathcal{L}[Z_s']\right) \\ &= N_n H^2(\mathcal{L}[\xi_1(\varepsilon_n)], \mathcal{L}[\varepsilon^{-1}\log(1+\varepsilon\xi_1(\varepsilon_n)]) \leq O\left(N_n^2/n\right) = o(1). \end{aligned}$$

Let Y_s be given by (8). Then with the same arguments and (17) we get

$$\begin{aligned} H^2\left(\mathcal{L}((Z_s')_{s\in\mathcal{S}_n}), \mathcal{L}((Y_s)_{s\in\mathcal{S}_n})\right) &\leq 2\sum_{s\in\mathcal{S}_n} H^2\left(\mathcal{L}(Z_s'), \mathcal{L}(Y_s)\right) \\ &= N_n H^2(\mathcal{L}[\xi_1(\varepsilon_n)], \mathcal{L}[\xi_1]) \leq O\left(N_n^2/n\right) = o(1). \end{aligned}$$

■

5 Appendix

We begin the proof of Lemma 4 with some simple results about Gaussian random variables.

Lemma 6 *Let $\xi_k \in \mathbf{R}^1$ be i.i.d. $\mathcal{N}(0,1)$. Then uniformly in m*

$$\mathbf{E}\left(\sum_{i=1}^N \xi_i^2\right)^m \leq (N+2m)^m.$$

Proof. It easily follows from the characteristic functions method. Let $\mathcal{D}^m$ be the operator that takes the derivative of order m of a function $q(t)$ at $t = 0$

$$\mathcal{D}^m q(t) = \left.\frac{\mathrm{d}^m q(t)}{\mathrm{d}t^m}\right|_{t=0}.$$

Then evidently

$$\mathbf{E}\left(\sum_{i=1}^N \xi_i^2\right)^m = \mathcal{D}^m \exp\left\{t\sum_{i=1}^N \xi_i^2\right\} = \mathcal{D}^m(1-2t)^{-N/2} = \prod_{p=1}^m (N+2p).$$

Sightly more complicated arguments lead to the following result:

Lemma 7 *Let $\xi_i \in \mathbf{R}^1$ be i.i.d. $\mathcal{N}(0,1)$. Then uniformly in m and k*

$$\mathbf{E}\left(\sum_{i=1}^{N} |\xi_i|^k\right)^m \leq \left((Cm)^{k/2-1} k^{k/2}\right)^m (N+m)^m, \tag{19}$$

$$\mathbf{E}\left(\sum_{i=1}^{N} \xi_i^{2k+1}\right)^m \leq \left((Cm)^{2k-1} (2k+1)^{2k+1}\right)^{m/2} (N+m)^{m/2}, \tag{20}$$

where C is a generic constant.

Proof. We have

$$\begin{aligned}
\mathbf{E}\left(\sum_{i=1}^{N} |\xi_i|^k\right)^m &= \lim_{A\to\infty} \mathbf{E}\left(\sum_{i=1}^{N} |\xi_i|^k \mathbf{1}\{|\xi_i| < A\}\right)^m \\
&= \lim_{A\to\infty} \mathcal{D}^m \mathbf{E} \exp\left(t \sum_{i=1}^{N} |\xi_i|^k \mathbf{1}\{|\xi_i| < A\}\right) \\
&= \lim_{A\to\infty} \mathcal{D}^m \left(\sum_{p=0}^{\infty} \frac{t^p}{p\,!} \mathbf{E}|\xi_i|^{kp} \mathbf{1}\{|\xi_i| < A\}\right)^N \\
&= \lim_{A\to\infty} \mathcal{D}^m \left(\sum_{p=0}^{m} \frac{t^p}{p\,!} \mathbf{E}|\xi_i|^{kp} \mathbf{1}\{|\xi_i| < A\}\right)^N = \mathcal{D}^m \left(\sum_{p=0}^{m} \frac{t^p}{p\,!} \mathbf{E}|\xi_i|^{kp}\right)^N.
\end{aligned}$$

On the other hand

$$\begin{aligned}
\mathcal{D}^m \left(\sum_{p=0}^{m} \frac{t^p}{p\,!} \mathbf{E}|\xi_i|^{kp}\right)^N &\leq \mathcal{D}^m \left(\sum_{p=0}^{m} \frac{t^p}{p\,!} (kp)^{kp/2}\right)^N \\
&\leq \mathcal{D}^m \left(\sum_{p=0}^{m} t^p (kp)^{kp/2} (Cp)^{-p}\right)^N \leq \mathcal{D}^m \left(\sum_{p=0}^{m} t^p k^{kp/2} (Cm)^{(k/2-1)p}\right)^N \\
&= \mathcal{D}^m \left(\sum_{p=0}^{\infty} t^p k^{kp/2} (Cm)^{(k/2-1)p}\right)^N = \mathcal{D}^m \left(1 - t k^{k/2} (Cm)^{(k/2-1)}\right)^{-N} \\
&\leq (N+m)^m k^{km/2} (Cm)^{m(k/2-1)}
\end{aligned}$$

proving (19). The inequality (20) is proved in the same way. ∎

Later on we will use the following result, which is an immediate consequence of the convexity of the function $|x|^m$, $m \geq 1$.

Lemma 8 *Let $\zeta_0, \ldots, \zeta_{N-1}$ be zero mean Gaussian random variables in $\mathbf{C}$ with the covariance matrix B. Then*

$$\mathbf{E}\left(\sum_s \left|\sum_{p=0}^{N-1} \exp(\mathrm{i}\,\lambda_s p) \zeta_p\right|^k\right)^m \leq (s_1[B]N)^{km/2} \left(\frac{n}{N}\right)^m \mathbf{E}\left(\sum_{l=0}^{N-1} |\xi_l|^k\right)^m,$$

where $s_1[B]$ is the first eigenvalue of B and $\xi_k \in \mathbf{C}$ are independent $\mathcal{N}(0,1)$.

Proof. Since the matrix $s_1[B]E - B$ is nonnegative definite, there exist Gaussian random variables $\zeta'_0, \dots, \zeta'_{N-1}$ which do not depend on $\zeta_0, \dots, \zeta_{N-1}$ and which have the covariance matrix $s_1[B]E - B$. Therefore $\zeta_0 + \zeta'_0, \dots, \zeta_{N-1} + \zeta'_{N-1}$ are independent with the variance $s_1[B]$. Thus by Anderson's lemma one obtains

$$\begin{aligned}
\mathbf{E}\left(\sum_s \left|\sum_{p=0}^{N-1} \exp(\mathrm{i}\,\lambda_s p)\zeta_p\right|^k\right)^m &\leq \mathbf{E}\left(\sum_s \left|\sum_{p=0}^{N-1} \exp(\mathrm{i}\,\lambda_s p)(\zeta_p + \zeta'_p)\right|^k\right)^m \qquad (21)\\
&\leq (s_1[B])^{km/2}\mathbf{E}\left(\sum_s \left|\sum_{p=0}^{N-1} \exp(\mathrm{i}\,\lambda_s p)\xi'_p\right|^k\right)^m,
\end{aligned}$$

where ξ'_k are independent $\mathcal{N}(0,1)$. Next, by convexity of $|x|^m$ we get

$$\begin{aligned}
&\mathbf{E}\left(\sum_s \left|\sum_{p=0}^{N-1} \exp(\mathrm{i}\,\lambda_s p)\xi'_p\right|^k\right)^m \qquad (22)\\
&\leq \mathbf{E}\left(\sum_{j=0}^{n/N}\sum_{k=0}^{N-1}\left|\sum_{p=0}^{N-1} \exp\left[\mathrm{i}\,p\left(\frac{2\pi k}{N} + \frac{2\pi j}{n}\right)\right]\xi'_p\right|^k\right)^m\\
&\leq \left(\frac{n}{N}\right)^m \max_j \mathbf{E}\left(\sum_{k=0}^{N-1}\left|\sum_{p=0}^{N-1} \exp\left[\mathrm{i}\,p\left(\frac{2\pi k}{N} + \frac{2\pi j}{n}\right)\right]\xi'_p\right|^k\right)^m.
\end{aligned}$$

Finally noting that the following Gaussian random variables

$$\xi_k = \sum_{p=0}^{N-1} \exp\left[\mathrm{i}\,p\left(\frac{2\pi k}{N} + \frac{2\pi j}{n}\right)\right]\xi'_p$$

are independent $\mathcal{N}(0, \sqrt{N})$ and using (21), (22) we arrive at the assertion of the lemma. ■

With (15) it is not difficult to compute the characteristic function of the random variables $\kappa_p,\ p \in [0, N-1]$:

$$\begin{aligned}
\varphi_\kappa^N(\mathbf{t}) &= \mathbf{E}\exp\left(\mathrm{i}\sum_{p=0}^{N-1} t_p\kappa_p\right) \qquad (23)\\
&= \exp\left\{-\frac{\mathrm{i}}{\sqrt{n}}\sum_{p=0}^{N-1} t_p\tilde{B}_{f\,0p} - \frac{1}{2}\sum_s \log\left(1 - \frac{4\pi\mathrm{i}}{\sqrt{n}} f(\lambda_s)\,Q(\lambda_s, \mathbf{t})\right)\right\},
\end{aligned}$$

where

$$Q(\lambda, \mathbf{t}) = \sum_{p=0}^{N-1} t_p \cos(\lambda p).$$

Denote for brevity by

$$\varphi_\eta^N(\mathbf{t}) = \exp\left\{-\frac{4\pi^2}{n}\sum_s f^2(\lambda_s)Q^2(\lambda_s,\mathbf{t})\right\} \tag{24}$$

the characteristic function of the random variables η_p, $p \in [0, N-1]$.

Lemma 9 *Let $N < n/2$. Then*

$$\left|\varphi_\kappa^N(\mathbf{t})\right| \le \left(1 + \frac{32\pi^2 m^2 N}{n}\sum_{p=0}^{N-1} t_p^2\right)^{-n/(8N)}. \tag{25}$$

Proof. Since $f(\lambda) \ge m$, we have the following upper bound for the absolute value of the characteristic function:

$$\begin{aligned} \left|\varphi_\kappa^N(\mathbf{t})\right| &\le \prod_s \left(1 + \frac{16\pi^2}{n} f^2(\lambda_s)\, Q^2(\lambda_s,\mathbf{t})\right)^{-1/4} \\ &\le \exp\left(-\frac{1}{4}\sum_s \log\left(1 + \frac{16\pi^2 m^2}{n} Q^2(\lambda_s,\mathbf{t})\right)\right) \end{aligned} \tag{26}$$

On the other hand, the Cauchy–Schwartz inequality yields

$$\begin{aligned} \sum_s Q^4(\lambda_s,\mathbf{t}) &\le \max_k Q^2(\lambda_k,\mathbf{t}) \sum_s Q^2(\lambda_s,\mathbf{t}) \\ &\le \left(\sum_{p=0}^{N-1} |t_p|\right)^2 \sum_s Q^2(\lambda_s,\mathbf{t}) \le N \sum_{p=0}^{N-1} t_p^2 \ \sum_s Q^2(\lambda_s,\mathbf{t}) \\ &\le \frac{2N}{n}\left(\sum_s Q^2(\lambda_s,\mathbf{t})\right)^2. \end{aligned} \tag{27}$$

Denote for brevity

$$E(\mathbf{t}) = \frac{2}{n}\sum_s Q^2(\lambda_s,\mathbf{t}) = 2t_0^2 + \sum_{p=0}^{N-1} t_p^2$$

and

$$\mathcal{G}_{E(\mathbf{t})} = \left\{g(\lambda_s) \ge 0 : \sum_s g(\lambda_s) = nE(\mathbf{t})/2, \quad \sum_s g^2(s) \le NnE^2(\mathbf{t})/2\right\}.$$

Then according to (27) we have

$$\sum_s \log\left(1 + \frac{16\pi^2 m^2}{n} Q^2(\lambda_s,\mathbf{t})\right) \ge \min_{g\in\mathcal{G}_{E(\mathbf{t})}} \sum_s \log\left(1 + \frac{16\pi^2 m^2}{n} g(\lambda_s)\right).$$

It is easy to see from the Lagrange multiplier principle that the minimum in the right-hand side of the above equation is attained at

$$g(\lambda_s) = g^*(\lambda_s) = \begin{cases} G, & s \in [1,K], \\ 0, & s \notin [1,K]. \end{cases}$$

Since $g^*(\lambda_s)$ belongs to $\mathcal{G}_{E(\mathbf{t})}$ we get $K = n/(2N)$, $G = NE(\mathbf{t})$, and therefore

$$\begin{aligned}\sum_s \log\left(1+\frac{16\pi^2 m^2}{n}Q^2(s,\mathbf{t})\right) &\geq \sum_s \log\left(1+\frac{16\pi^2 m^2}{n}g^*(\lambda_s)\right)\\ &= \frac{n}{2N}\log\left(1+\frac{16\pi^2 m^2 NE(\mathbf{t})}{n}\right).\end{aligned}$$

Hence from (26) we arrive at (25). ■

Proof of Lemma 4. The main idea of the proof is straightforward. We compute a sufficiently good approximation for the characteristic function of the random variables κ_l, $l \in [0, N-1]$, then take its inverse Fourier transform to get an approximation for the density $p_\eta^N(\cdot)$, and finally we evaluate the L_1-distance between the Gaussian density $p_\kappa^N(\cdot)$ and the approximation.

Let $\mathcal{K}_n = \{x \in \mathbf{R}^N : \ |x_k| \leq A_n\}$, where $A_n = 4\pi M\sqrt{\log n}$. From the Markov inequality with $\mu = A_n/(8\pi^2 M^2)$ and by Taylor expansion we get

$$\begin{aligned}\int_{\mathbf{x}\notin\mathcal{K}_n} p_\kappa^N(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \mathbf{P}\left\{\max_p |\kappa_p| > A_n\right\}\\ &\leq \exp(-\mu A_n)\sum_{p=0}^{N-1} \mathbf{E}^N\exp(\mu\kappa_p) + \mathbf{E}^N\exp(-\mu\kappa_p)\\ &\leq 2\exp(-\mu A_n)\sum_{p=0}^{N-1}\mathbf{E}\exp\left\{\frac{2\pi\mu}{\sqrt{n}}\sum_s \cos(\lambda_s p)f(\lambda_s)(\xi_s^2-1)\right\}\\ &= 2\exp(-\mu A_n)\sum_{p=0}^{N-1}\exp\left\{-\frac{1}{2}\sum_s\log\left(1-\frac{4\pi\mu}{\sqrt{n}}\cos(\lambda_s p)f(\lambda_s)\right)\right\}\\ &\leq 2N\exp\left[-\mu A_n + 4\pi^2M^2\mu^2(1+o(1))\right] \leq Nn^{-1}.\end{aligned}$$

Since η_k are Gaussian with zero mean and $\mathbf{E}\eta_k^2 \leq 8\pi^2 M^2$ we evidently have

$$\int_{\mathbf{x}\notin\mathcal{K}_n} p_\eta^N(\mathbf{x})\,\mathrm{d}\mathbf{x} \leq Nn^{-1}.$$

Therefore, in order to prove the lemma it remains to evaluate from above

$$D(p_\kappa, p_\eta) = \int_{\mathbf{x}\in\mathcal{K}_n}\left|p_\kappa^N(\mathbf{x}) - p_\eta^N(\mathbf{x})\right|\,\mathrm{d}\mathbf{x}.$$

Introduce the ℓ_2 and ℓ_1–norms in $\mathbf{R}^N$

$$\|\mathbf{t}\|_2 = \left(\sum_{p=0}^{N-1} t_k^2\right)^{1/2}, \qquad \|\mathbf{t}\|_1 = \sum_{p=0}^{N-1}|t_k|.$$

Assuming that $\mathbf{t} \in \mathcal{T}_n$, where $\mathcal{T}_n = \{\|\mathbf{t}\|_1 \leq AN\sqrt{\log n}\}$ and A is a sufficiently large constant, one obtains by Taylor expansion and by (23), (24)

$$\varphi_\kappa^N(\mathbf{t}) = \varphi_\eta^N(\mathbf{t}) \exp\left(R^N(\mathbf{t}) + O(1) \left(\frac{A^2 N^2 \log n}{n} \right)^{WN/2} \|\mathbf{t}\|_2^2 \right), \tag{28}$$

where

$$R^N(\mathbf{t}) = \frac{1}{2} \sum_{k=3}^{WN} \frac{(-1)^{k+1}}{k} \sum_s \left(\frac{4\pi \mathrm{i}}{\sqrt{n}} f(\lambda_s) Q(\lambda_s, \mathbf{t}) \right)^k$$

and W is a sufficiently large integer depending on n, which will be chosen later on.

Inverting the Fourier transform we get

$$\begin{aligned} D(p_\kappa, p_\eta) &= \frac{1}{(2\pi)^N} \int_{\mathbf{x} \in \mathcal{K}_n} \left| \int_{\mathbf{R}^N} \mathrm{e}^{\mathrm{i}(\mathbf{t},\mathbf{x})} \left(\varphi_\kappa^N(\mathbf{t}) - \varphi_\eta^N(\mathbf{t}) \right) \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} \\ &\leq \frac{1}{(2\pi)^N} \int_{\mathbf{x} \in \mathcal{K}_n} \left| \int_{\mathbf{t} \in \mathcal{T}_n} \mathrm{e}^{\mathrm{i}(\mathbf{t},\mathbf{x})} \left(\varphi_\kappa^N(\mathbf{t}) - \varphi_\eta^N(\mathbf{t}) \right) \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} \\ &\quad + \frac{\operatorname{mes} \mathcal{K}_n}{(2\pi)^N} \int_{\mathbf{t} \in \mathcal{T}_n} \left(|\varphi_\kappa^N(\mathbf{t})| + |\varphi_\eta^N(\mathbf{t})| \right) \mathrm{d}\mathbf{t} \,. \end{aligned} \tag{29}$$

The last term in the right-hand side of the above inequality is evaluated by (24) and by the Markov inequality

$$\begin{aligned} \frac{1}{(2\pi)^{N/2}} \int_{\mathbf{t} \in \mathcal{T}_n} |\varphi_\eta^N(\mathbf{t})| \, \mathrm{d}\mathbf{t} &\leq \frac{1}{(2\pi)^{N/2}} \int_{\mathbf{t} \in \mathcal{T}_n} \exp\left\{ -4\pi^2 m^2 \|\mathbf{t}\|_2^2 \right\} \mathrm{d}\mathbf{t} \\ &\leq C^{-N/2} \mathbf{P} \left\{ \sum_{p=0}^{N-1} |\xi_p| \geq AN\sqrt{\log n} \right\} \leq C^{-N/2} \exp\left(N - AN \log n / 2 \right). \end{aligned}$$

Therefore we obtain

$$\frac{\operatorname{mes} \mathcal{K}_n}{(2\pi)^N} \int_{\mathbf{t} \in \mathcal{T}_n} |\varphi_\eta^N(\mathbf{t})| \, \mathrm{d}\mathbf{t} \leq (C \log n)^{N/2} \, \mathrm{e}^{-AN \log n / 2} \leq n^{-1}. \tag{30}$$

On the other hand we get from Lemma 9

$$\begin{aligned} &\frac{\operatorname{mes} \mathcal{K}_n}{(2\pi)^N} \int_{\mathbf{t} \in \mathcal{T}_n} |\varphi_\kappa^N(\mathbf{t})| \, \mathrm{d}\mathbf{t} \\ &\leq \frac{\operatorname{mes} \mathcal{K}_n}{(2\pi)^N} \int_{\mathbf{t} \in \mathcal{T}_n} \left(1 + \frac{32\pi^2 m^2 N}{n} \|\mathbf{t}\|_2^2 \right)^{-n/(8N)} \mathrm{d}\mathbf{t} \\ &\leq \left(\frac{Cn \log n}{N} \right)^{N/2} \max_{\|\mathbf{t}\|_1 \geq CAN\sqrt{N \log n / n}} \left(1 + \|\mathbf{t}\|_2^2 \right)^{-n/(8N)+N} \int_{\mathbf{R}^N} \frac{\mathrm{d}\mathbf{t}}{\left(1 + \|\mathbf{t}\|_2^2 \right)^N} \\ &\leq \left(\frac{Cn \log n}{N} \right)^{N/2} \left(1 + C^2 A^2 N^2 \log n / n \right)^{-n/(8N)+N} \leq n^{-1}. \end{aligned} \tag{31}$$

In order to estimate the first term in the right-hand side of (29) note that according to the assumption of the theorem we can chose the number W such that

$$\operatorname{mes} \mathcal{K}_n \left(\frac{N^2 \log n}{n} \right)^{WN/2} \leq n^{-1}$$

and therefore by Taylor expansion and (28)–(31) one obtains

$$\begin{aligned} D(p_\kappa, p_\eta) &\leq \frac{1}{(2\pi)^N} \int_{\mathbf{x}\in\mathcal{K}_n} \left| \int_{\mathbf{t}\in\mathcal{T}_n} \mathrm{e}^{\mathrm{i}\mathbf{t}^T\mathbf{x}} \left[\varphi_\kappa^N(\mathbf{t}) - \varphi_\eta^N(\mathbf{t}) \right] \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} + O\left(n^{-1}\right) \qquad (32) \\ &\leq \frac{1}{(2\pi)^N} \sum_{m=1}^{WN} \frac{1}{m!} \int \left| \int_{\mathbf{t}\in\mathcal{T}_n} \mathrm{e}^{\mathrm{i}\mathbf{t}^T\mathbf{x}} \varphi_\eta^N(\mathbf{t}) \left[R^N(\mathbf{t}) \right]^m \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} + O\left(n^{-1}\right). \end{aligned}$$

Our next step is to estimate the leading term in the right-hand side of the above equation. It is convenient to introduce the matrix A with the entries

$$A_{kl} = \frac{8\pi^2}{n} \sum_s f^2(\lambda_s) \cos(\lambda_s l) \cos(\lambda_s k). \qquad (33)$$

Then by simple algebra one obtains

$$\begin{aligned} &\frac{1}{(2\pi)^N} \int \left| \int_{\mathbf{t}\in\mathcal{T}_n} \mathrm{e}^{\mathrm{i}\,\mathbf{t}^T\mathbf{x}} \varphi_\eta^N(\mathbf{t}) \left[R^N(\mathbf{t}) \right]^m \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} \qquad (34) \\ &= \frac{1}{(2\pi)^N} \int \left| \int_{\mathbf{t}\in\mathcal{T}_n} \mathrm{e}^{\mathrm{i}\,\mathbf{t}^T\mathbf{x} - \mathbf{t}^T A\mathbf{t}/2} \left[R^N(\mathbf{t}) \right]^m \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} \\ &= \frac{1}{(2\pi)^N} \int_{\mathbf{R}^N} \left| \int_{\mathbf{R}^N} \mathrm{e}^{-\mathbf{x}^T\mathbf{x}/2 - \mathbf{t}^T\mathbf{t}/2} \left[R^N\left(A^{-1/2}(\mathbf{t} - \mathrm{i}\mathbf{x}) \right) \right]^m \right. \\ &\quad \left. \times\ \mathbf{1}\{\mathbf{t} - \mathrm{i}\,\mathbf{x} \in \mathcal{T}_n\}\, \mathrm{d}\mathbf{t} \right| \mathrm{d}\mathbf{x} \\ &= \mathbf{E} \left| \mathbf{E} \left\{ \left[R^N(\zeta - \mathrm{i}\zeta') \right]^m \mathbf{1}\{\zeta - \mathrm{i}\,\zeta' \in \mathcal{T}_n\} | \zeta \right\} \right|, \end{aligned}$$

where ζ and ζ' are independent Gaussian vectors with zero mean and the covariance matrix A^{-1}. The main difficulty is the evaluation of the right-hand side of the above inequality, involving the first order term

$$r_3^m = \mathbf{E} \left| \sum_s \left(f(\lambda_s) \sum_{p=0}^{N-1} \cos(\lambda_s p) \zeta_p \right)^3 \right|^m.$$

Represent the function $f(\lambda)$ in the form $f^N(\lambda) + f(\lambda) - f^N(\lambda)$, where $f^N(\lambda)$ is a polynomial of order $N/2$ which is symmetric and strictly bounded from below. Since the norm of A^{-1} is strictly bounded from above (see Davies (1973)) we get from the Hölder inequality and Lemma 8

$$\begin{aligned} r_3^m &\leq C^m \mathbf{E} \left| \sum_s \left(f^N(\lambda_s) \sum_{p=0}^{N-1} \cos(\lambda_s p) \zeta_p \right)^3 \right|^m \qquad (35) \\ &\quad + C^m \max_\lambda |f^N(\lambda) - f(\lambda)|^m \left(\frac{n}{N} \right)^m \mathbf{E} \left(\sum_{l=0}^{N-1} |\xi_l|^k \right)^m, \end{aligned}$$

where ξ_l are i.i.d. $\mathcal{N}(0,1)$. Representing A as $A = \tilde{A} + A - \tilde{A}$, with

$$\tilde{A}_{lk} = \frac{8\pi^2}{n} \sum_s \left[f^N(\lambda_s)\right]^2 \cos(\lambda_s l)\cos(\lambda_s k),$$

and once again applying the Hölder inequality and Lemma 8, we get for

$$\hat{r}_3^m = \mathbf{E}\left|\sum_s \left(f^N(\lambda_s)\sum_{p=0}^{N-1}\cos(\lambda_s p)\zeta_p\right)^3\right|^m$$

the upper bound

$$\begin{aligned}\hat{r}_3^m \;\le\; & \mathbf{E}\left|\sum_s \left(f^N(\lambda_s)\sum_{p=0}^{N-1}\cos(\lambda_s p)\tilde{A}^{-1/2}\xi_p\right)^3\right|^m \\ & + C^m s_1^{m/2}\left[(A^{-1/2}-\tilde{A}^{-1/2})(A^{-1/2}-\tilde{A}^{-1/2})^T\right]\left(\frac{n}{N}\right)^m \mathbf{E}\left(\sum_{l=0}^{N-1}|\xi_l|^k\right)^m.\end{aligned} \tag{36}$$

Let $\lambda_s^N = 2\pi s/N$ be a uniform grid on $[-\pi,\pi]$. Since $\left[f^N(\lambda)\right]^2$ is a symmetric trigonometric polynomial of order N, we can represent $\tilde{A}$ in the form

$$\tilde{A}_{kl} = \frac{8\pi^2}{N}\sum_s \left[f^N(\lambda_s^N)\right]^2 \cos(\lambda_s^N l)\cos(\lambda_s^N k).$$

Since $\{\cos(\lambda_s^N p),\ p\in[0,N-1]\}$ are the eigenvectors of the matrix $\tilde{A}$, the matrix $\tilde{A}^{-1/2}$ can be represented as

$$\tilde{A}_{kl}^{-1/2} = \frac{1}{\pi N}\sum_s \frac{1}{f^N(\lambda_s^N)}\cos(\lambda_s^N l)\cos(\lambda_s^N k).$$

Hence

$$\tilde{A}^{-1/2}\xi_l = \sqrt{\frac{2}{N}}\sum_s \frac{1}{2\pi f^N(\lambda_s^N)}\cos(\lambda_s^N l)\xi_s,$$

where ξ_s are independent $\mathcal{N}(0,1)$. Equivalently,

$$\frac{\xi_s}{2\pi f^N(\lambda_s^N)} = \sqrt{\frac{2}{N}}\sum_s \cos(\lambda_s^N l)\tilde{A}^{-1/2}\xi_l.$$

Thus we get

$$\begin{aligned}& \mathbf{E}\left|\sum_s \left(f(\lambda_s)\sum_{p=0}^{N-1}\cos(\lambda_s p)\tilde{A}^{-1/2}\xi_l\right)^3\right|^m \\ & = \left(\frac{n}{N}\right)^m \mathbf{E}\left|\sum_s \left(f(\lambda_s^N)\sum_{p=0}^{N-1}\cos(\lambda_s^N p)\tilde{A}^{-1/2}\xi_p\right)^3\right|^m \le (CN)^{3m/2}\mathbf{E}\left|\sum_{p=0}^{N-1}\xi_s^3\right|^m.\end{aligned} \tag{37}$$

Noting that

$$\begin{aligned} s_1\left[(A^{-1/2}-\tilde{A}^{-1/2})(A^{-1/2}-\tilde{A}^{-1/2})^T\right] &\leq Cs_1\left[(A-\tilde{A})(A-\tilde{A})^T\right] \\ &\leq C\sup_\lambda |f(\lambda)-f^N(\lambda)|^2 \end{aligned}$$

one obtains from the above equation and from (35) – (37)

$$r_3^m \leq \left(\frac{CN}{n}\right)^{m/2}\left[m^{m/2}N^{m/2}+(d_N(f))^{m/2}m^{m/2}N^m\right]. \tag{38}$$

Now consider the remainder terms in the right-hand side of (34). Denote for brevity $N(m) = NW\log n/m$. We obtain

$$\begin{aligned} &\mathbf{E}\left|\mathbf{E}\left\{\sum_{k=4}^{WN}\frac{1}{2k}\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\mathbf{1}\{\zeta-\mathrm{i}\,\zeta'\in\mathcal{T}_n\}|\zeta'\right\}\right|^m \qquad (39)\\ &\leq [\log(WN)]^m\sum_{k=4}^{WN}\frac{1}{2k}\mathbf{E}\left|\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\mathbf{1}\{\zeta-\mathrm{i}\,\zeta'\in\mathcal{T}_n\}\right|^m\\ &\leq [\log(WN)]^m\sum_{k=4}^{N(m)}\frac{1}{2k}\mathbf{E}\left|\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\right|^m\\ &+\ [\log(WN)]^m\sum_{k=W(m)}^{WN}\frac{1}{2k}\mathbf{E}\left|\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\mathbf{1}\{\zeta-\mathrm{i}\,\zeta'\in\mathcal{T}_n\}\right|^m \end{aligned}$$

Since the norm of the matrix A^{-1} is strictly bounded from above by some constant which does not depend on N (cf. Davies (1973)), we get from Lemmas 7, 8

$$\mathbf{E}\left|\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\right|^m \leq \left(\frac{CmkN}{n}\right)^{mk/2}\left[\left(\frac{n}{m}\right)^m+\left(\frac{n}{N}\right)^m\right]$$

and therefore

$$\sum_{k=4}^{W(m)}\frac{1}{2k}\mathbf{E}\left|\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\right|^m \leq \left(\frac{CN^2}{n}\right)^m+N\left(\frac{CNm^2}{n}\right)^m. \tag{40}$$

If $k > N(m)$ we get from Lemma 6

$$\begin{aligned} &\mathbf{E}\left|\sum_s\left(\frac{4\pi\mathrm{i}}{\sqrt{n}}f(\lambda_s)Q(\lambda_s,\zeta-\mathrm{i}\zeta')\right)^k\mathbf{1}\{\zeta-\mathrm{i}\,\zeta'\in\mathcal{T}_n\}\right|^m\\ &\leq\ \frac{C^{mk}n^m}{n^{km/2}}\mathbf{E}\left[\frac{1}{n}\sum_s|Q(\lambda_s,\zeta-\mathrm{i}\zeta')|^k\right]^m\mathbf{1}\{\zeta-\mathrm{i}\,\zeta'\in\mathcal{T}_n\}\\ &\leq\ \frac{C^{mk}n^m}{n^{km/2}}\mathbf{E}\left[\frac{(AN\sqrt{\log n})^{k-2}}{n}\sum_s|Q(\lambda_s,\zeta-\mathrm{i}\zeta')|^2\right]^m\\ &\leq\ (nN)^m\left(\frac{CA^2N^2\log n}{n}\right)^{(k/2-1)m} \end{aligned}$$

and hence

$$\sum_{k=N(m)}^{WN} \frac{1}{2k}\mathbf{E}\left|\sum_s \left(\frac{4\pi\mathrm{i}}{\sqrt{n}} f(\lambda_s) Q(\lambda_s, \zeta - \mathrm{i}\zeta')\right)^k \mathbf{1}\{\zeta - \mathrm{i}\,\zeta' \in \mathcal{T}_n\}\right|^m \tag{41}$$
$$\leq \; (nN)^m \sum_{k=N(m)}^{\infty} \left(\frac{CN^2 \log n}{n}\right)^{(k/2-1)m} \leq (nN)^m \left(\frac{CN^2 \log n}{n}\right)^{WN \log n/2}.$$

The proof of the lemma follows now from (32) and (38) – (41). ■

References

[1] Bickel, P. J. Klaassen, C. A. Ritov, and Y. Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore.

[2] Brown, L.D. and Low M. (1996). Asymptotic equivalence of non-parametric regression and white noise. *Ann. Statist.* **24**, 2384–2398

[3] Davies, R.B. (1973). Asymptotic inference in stationary Gaussian time-series, *Adv. Appl. Probab.* **5**, 469–497.

[4] Dzhaparidze K. (1986) *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer-Verlag, New York Inc.

[5] Gikhman, I.I. and Skorokhod, A.V. (1969). *Introduction to the Theory of Random Processes.* Saunders Co.

[6] Grama, I. and Nussbaum, M. (1997). Asymptotic equivalence for nonparametric generalized models. Preprint No. 289, Weierstrass Institute, Berlin (To appear, *Prob. Theor. Rel. Fields*)

[7] Ibragimov, I.A. and Khasminskii R.Z. (1981) *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

[8] Le Cam, L. and Yang, G. (1990). *Asymptotics in Statistics.* Springer-Verlag, New-York.

[9] Mann, H. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations, *Econometrics,* **11**, 173-220.

[10] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399–2430.

[11] Nussbaum, M. (1998). Approximation of statistical experiments for ill–posed function estimation problems. *Manuscript.*

[12] Whittle, P. (1952). Estimation and information in time series analysis, *Skand. Aktuar.,* **35**, 48-60.