

Asymptotic Equivalence of Spectral Density Estimation and Gaussian White Noise

GEORGI K. GOLUBEV, MICHAEL NUSSBAUM¹ AND HARRISON H. ZHOU¹

Université de Provence, Cornell University and Yale University

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 f . Asymptotic equivalence, in the sense of Le Cam's deficiency Δ -distance, to two Gaussian experiments with simpler structure is established. The first one is given by independent zero mean Gaussians with variance approximately $f(\omega_i)$ where ω_i is a uniform grid of points in $(-\pi, \pi)$ (nonparametric Gaussian scale regression). This approximation is closely related to well-known asymptotic independence results for the periodogram and corresponding inference methods. The second asymptotic equivalence is to a Gaussian white noise model where the drift function is the log-spectral density. This represents the step from a Gaussian scale model to a location model, and also has a counterpart in established inference methods, i.e. log-periodogram regression. The problem of simple explicit equivalence maps (Markov kernels), allowing to directly carry over inference, appears in this context but is not solved here.

1 Introduction and main results

Estimation of the spectral density $f(\omega)$, $\omega \in [-\pi, \pi]$ of a stationary process is an important and traditional problem of mathematical statistics. We observe a sample $y^{(n)} = (y(1), \dots, y(n))'$ from a real Gaussian stationary sequence $y(t)$ with $\mathbb{E}y(t) = 0$ and autocovariance function $\gamma(h) = \mathbb{E}y(t)y(t+h)$. Consider the spectral density, defined on $[-\pi, \pi]$ by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \exp(ih\omega) \quad (1.1)$$

¹Supported in part by NSF Grant DMS-03-06497

1991 Mathematics Subject Classification. 62G07, 62G20

Key words and phrases. Stationary Gaussian process, spectral density, Sobolev classes, Le Cam distance, asymptotic equivalence, Whittle likelihood, log-periodogram regression, nonparametric Gaussian scale model, signal in Gaussian white noise.

where it is assumed that $\sum_{h=-\infty}^{\infty} \gamma^2(h) < \infty$. Let Γ_n be the $n \times n$ Toeplitz covariance matrix associated with $\gamma(\cdot)$, i.e. the matrix with entries

$$(\Gamma_n)_{j,k} = \gamma(k-j) = \int_{-\pi}^{\pi} \exp(i(k-j)\omega) f(\omega) d\omega, \quad j, k = 1, \dots, n. \quad (1.2)$$

Write $\Gamma_n(f)$ for the covariance matrix corresponding to spectral density f and note that $y^{(n)}$ has a multivariate normal distribution $N_n(0, \Gamma_n(f))$. Let Σ be a nonparametric set of spectral densities to be described below. We are interested in the approximation of the statistical experiment

$$\mathcal{E}_n = (N_n(0, \Gamma_n(f)), f \in \Sigma) \quad (1.3)$$

in the sense of Le Cam's deficiency pseudodistance $\Delta(\cdot, \cdot)$; see the end of this section for a precise definition. The statistical interpretation of the Le Cam distance is as follows. For two experiments \mathcal{E} and \mathcal{F} having the same parameter space, $\Delta(\mathcal{E}, \mathcal{F}) < \varepsilon$ implies that for any decision problem with loss bounded by 1 and any statistical procedure with the experiment \mathcal{E} there is a (randomized) procedure with \mathcal{F} the risk of which evaluated in \mathcal{F} nearly matches (within ε) the risk of the original procedure evaluated in \mathcal{E} . In this statement the roles of \mathcal{E} and \mathcal{F} can also be reversed. Two sequences $\mathcal{E}_n, \mathcal{F}_n$ are said to be *asymptotically equivalent* if $\Delta(\mathcal{E}_n, \mathcal{F}_n) \rightarrow 0$.

As a guide to what can be expected, consider first the case where f_ϑ , $\vartheta \in \Theta$ is a smooth parametric family of spectral densities. Assume that Θ is a real interval; under some regularity conditions, the model is well known to fulfill the standard LAN conditions with localization rate $n^{-1/2}$ and normalized Fisher information at ϑ

$$\frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \vartheta} \log f_\vartheta(\omega) \right)^2 d\omega$$

(Davies (1973), Dzhaparidze (1985), chap. I.3, cf. also the discussion in van der Vaart (1998), Example 7.17). Consider the parametric Gaussian white noise model where the signal is the log-spectral density:

$$dZ_\omega = \log f_\vartheta(\omega) d\omega + 2\pi^{1/2} n^{-1/2} dW_\omega, \quad \omega \in [-\pi, \pi] \quad (1.4)$$

and note that in the family $(f_\vartheta, \vartheta \in \Theta)$, this model has the same asymptotic Fisher information. This is in agreement with the LAN result for the spectral density model, but it suggests that the above white noise approximation might also be true for larger (i.e. nonparametric) spectral density classes Σ .

As a second piece of evidence for the white noise approximation in the nonparametric case we take known results about the approximate spectral decomposition of the Toeplitz covariance matrix $\Gamma_n(f)$. It is a classical difficulty in time series analysis that the exact eigenvalues and eigenvectors of $\Gamma_n(f)$ cannot easily be found and used for inference about f ; in particular, the eigenvectors depend on f . However for an approximation which is a circulant matrix (denoted $\tilde{\Gamma}_n(f)$ below), the eigenvectors are independent of f and the eigenvalues are approximately $f(\omega_j)$ where ω_j are the points of an equispaced grid of size n in $[-\pi, \pi]$. If the approximation by $\tilde{\Gamma}_n(f)$ were justified, one could apply an orthogonal transformation to the data $y^{(n)}$ and obtain a Gaussian scale model

$$z_j = f^{1/2}(\omega_j) \xi_j, \quad j = 1, \dots, n \quad (1.5)$$

where ξ_j are independent standard normal. For this model, nonparametric asymptotic equivalence theory was developed in Grama and Nussbaum (1998). Results there, for certain smoothness classes $f \in \Sigma$, with f bounded away from 0, lead to the nonparametric version of the white noise model (1.4)

$$dZ_\omega = \log f(\omega)d\omega + 2\pi^{1/2}n^{-1/2}dW_\omega, \omega \in [-\pi, \pi], f \in \Sigma. \quad (1.6)$$

Our proof of asymptotic equivalence will in fact be based on the approximation of the covariance matrix $\Gamma_n(f)$ by the circulant $\tilde{\Gamma}_n(f)$, cf. Brockwell and Davis (1991), § 4.5. However we shall see that this tool does not enable a straightforward approximation of the data $y^{(n)}$ in total variation or Hellinger distance. Therefore our argument for asymptotic equivalence will be somewhat indirect, involving "bracketing" of the experiment \mathcal{E}_n by upper and lower bounds (in the sense of informativity) and also a preliminary localization of the parameter space.

To formulate our main result, define a parameter space Σ of spectral densities as follows. For $M > 0$, define a set of real valued even functions on $[-\pi, \pi]$

$$\mathcal{F}_M = \{f : M^{-1} \leq f(\omega), f(\omega) = f(-\omega), \omega \in [-\pi, \pi]\}.$$

Thus our spectral densities are assumed uniformly bounded away from 0. Let $L_2(-\pi, \pi)$ be the usual (real) L_2 -space on $[-\pi, \pi]$; for any $f \in L_2(-\pi, \pi)$, let $\gamma_f(k)$, $k \in \mathbb{Z}$ be the Fourier coefficients according to (1.1). For any $\alpha > 0$ and $M > 0$ let

$$W^\alpha(M) = \left\{ f \in L_2(-\pi, \pi) : \gamma_f^2(0) + \sum_{k=-\infty}^{\infty} |k|^{2\alpha} \gamma_f^2(k) \leq M \right\}. \quad (1.7)$$

These sets correspond to balls in the periodic fractional Sobolev scale with smoothness coefficient α . Note that for $\alpha > 1/2$, by an embedding theorem (Lemma 5.6, Appendix), functions in $W^\alpha(M)$ are also uniformly bounded. Define an a priori set for given $\alpha > 0$, $M > 0$

$$\Sigma_{\alpha, M} = W^\alpha(M) \cap \mathcal{F}_M.$$

Consider also a Gaussian scale model (1.5) where the values $f(\omega_j)$ are replaced by local averages

$$J_{j,n}(f) = n \int_{(j-1)/n}^{j/n} f(2\pi x - \pi) dx, \quad j = 1, \dots, n$$

Theorem 1.1 *Let Σ be a set of spectral densities contained in $\Sigma_{\alpha, M}$ for some $M > 0$ and $\alpha > 1/2$. Then the experiments given by observations*

$$y(1), \dots, y(n), \text{ a stationary centered Gaussian sequence with spectral density } f \\ z_1, \dots, z_n, \text{ where } z_j \text{ are independent } N(0, J_{j,n}(f))$$

with $f \in \Sigma$ are asymptotically equivalent.

Let $\|\cdot\|_{B_{p,q}^\alpha}$ be the Besov norm on the interval $[-\pi, \pi]$ with smoothness index α (see Appendix, Section 5.3). For the second main result we impose a smoothness condition involving this norm for the $\alpha > 1/2$ from above and $p = q = 6$.

Theorem 1.2 *Let Σ be a set of spectral densities as in Theorem (1.1), fulfilling additionally $\|f\|_{B_{6,6}^\alpha} \leq M$ for all $f \in \Sigma$. Then the experiments given respectively by observations*

$$z_1, \dots, z_n, \text{ where } z_j \text{ are independent } N(0, J_{j,n}(f))$$

$$dZ_\omega = \log f(\omega)d\omega + 2\pi^{1/2}n^{-1/2}dW_\omega, \omega \in [-\pi, \pi]$$

with $f \in \Sigma$ are asymptotically equivalent.

The proof of this result is in the thesis Zhou (2004). The present paper is devoted to the proof of Theorem 1.1.

In nonparametric asymptotic equivalence theory, some constructive results have recently been obtained, i.e. explicit equivalence maps have been exhibited which allow to carry over optimal decision function from one sequence of experiments to the other. Brown and Low (1996) and Brown, Low and Zhang (2002) obtained constructive results for white noise with drift and Gaussian regression with nonrandom and random design. Brown, Carter, Low and Zhang (2004) found such equivalence maps (Markov kernels) for the i.i.d. model on the unit interval (density estimation) and the model of Gaussian white noise with drift; cf. also Carter (2002). The theoretical (nonconstructive) variant of this result had earlier been established in Nussbaum (1996), in the sense of an existence proof for pertaining Markov kernels. This indirect approach relied on the well known connection to likelihood processes of experiments, cf. Le Cam and Yang (2000). In the present paper, the result of Theorem 1.1 are of nonconstructive type, using a variety of methods for bounding the Δ -distance between the time series experiment and the model of independent zero mean Gaussians. Similarly, the proof of Theorem 1.2 in Zhou (2004) is nonconstructive, but it appears likely in that a second step, relatively simple "workable" equivalence maps can be found, at least for the case of Theorem 1.1 related to the classical result about asymptotic independence of discrete Fourier transforms.

To further discuss the context of the main results, we note the following points.

1. Asymptotic independence of discrete Fourier transforms. Let

$$d_n(\omega) = \sum_{k=1}^n \exp(-\mathbf{i}k\omega) y(k), \omega \in (-\pi, \pi)$$

be the discrete Fourier transform of the time series $y(1), \dots, y(n)$. Assume n is uneven and let η_j be complex standard normal variables. It is well known that for the Fourier frequencies $\omega_j = 2\pi j/n$, $j = 1, \dots, (n-1)/2$ in $(0, \pi)$, there is an asymptotic distribution

$$(\pi n)^{-1/2} d_n(\omega_j) \approx \exp(\mathbf{i}\omega_j) f^{1/2}(\omega_j) \eta_j$$

and the values are asymptotically uncorrelated for distinct ω_j, ω_k . For a precise formulation cf. relation (2.12) below or Brockwell and Davis (1991), Proposition 4.5.2. This fact is the basis for many inference methods (e.g. Dahlhaus and Janas (1996)); see Lahiri (2003) for an extended discussion of the asymptotic independence. A linear transformation to $n-1$ independent real normals and adding a real normal according to $(2\pi n)^{-1/2} d_n(0) \approx N(0, f(0))$ suggests the Gaussian scale model (1.5).

2. Log-periodogram regression. Consider also the periodogram

$$I_n(\omega) = \frac{1}{2\pi n} |d_n(\omega)|^2.$$

Note the equality in distribution $|\eta_j|^2 \sim \chi_2^2 \sim 2e_j$, where e_j is standard exponential. As a consequence of the above result about $d_n(\omega_j)$, we have for $j = 1, \dots, (n-1)/2$

$$I_n(\omega_j) \approx f(\omega_j)e_j \tag{1.8}$$

with asymptotic independence. Assuming this model exact and taking a logarithm gives rise to the inference method of log-periodogram regression (for an account cf. Fan and Gijbels (1996), sec. 6.4)

3. The Whittle approximation. This is an approximation to $-n^{-1}$ times the log-likelihood of the time series $y(1), \dots, y(n)$. In a parametric model f_ϑ , $\vartheta \in \Theta$, with multivariate normal law $N_n(0, \Gamma_n(f_\vartheta))$, computation of the MLE involves inverting the covariance matrix $\Gamma_n(f_\vartheta)$, which is difficult since both eigenvectors and eigenvalues depend on ϑ in general. Replacing $\Gamma_n^{-1}(f_\vartheta)$ by $\Gamma_n(1/4\pi^2 f_\vartheta)$ and using an approximation to $n^{-1} \log \Gamma_n(f_\vartheta)$ leads to an expression $L^W(f) + \log 2\pi$ where

$$L^W(f) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\log f_\vartheta(\omega) + \frac{I_n(\omega)}{f_\vartheta(\omega)} \right) d\omega \tag{1.9}$$

is the Whittle likelihood (cf. Dahlhaus (1988) for a brief exposition and references). A closely related expression is obtained by assuming the model (1.8) exact: then $-n^{-1}$ times the log-likelihood is

$$L_n^W(f) = n^{-1} \sum_{j=1}^{(n-1)/2} \left(\log f_\vartheta(\omega_j) + \frac{I_n(\omega_j)}{f_\vartheta(\omega_j)} \right)$$

i. e. a discrete approximation to (1.9). For applications of the Whittle likelihood to non-parametric inference cf. Dahlhaus and Polonik (2002).

4. Asymptotics for $L^W(f)$. The accuracy of the Whittle approximation has been described as follows (Coursol and Dacunha-Castelle (1982), Dzhaparidze (1986), Theorem 1, p. 52) . Let $L_n(f)$ be the log-likelihood in the experiment (1.3); then

$$L_n(f) = -nL^W(f) - n \log 2\pi + O_P(1) \tag{1.10}$$

uniformly over $f \in \Sigma_{1/2, M}$. This justifies use of $L^W(f)$ as a contrast function, e.g. it yields asymptotic efficiency of the Whittle MLE in parametric models (Dzhaparidze (1986), Chap. II), but falls short of providing asymptotic equivalence in the Le Cam sense. Indeed if (1.10) were true with $o_P(1)$ in place of $O_P(1)$ and with $L^W(f)$ replaced by $L_n^W(f)$ then this would already imply total variation equivalence, up to an orthogonal transform, of the exact model (1.8) with $f \in \Sigma_{1/2, M}$ (via the Scheffe lemma argument of Delattre and Hoffmann (2002)). In section 2 below (cf. relation (2.18)) we note a corresponding negative result, essentially that this total variation approximation over $f \in \Sigma_{1/2, M}$ does not take place.

5. Conditions for Theorem 1.2. For a narrower parameter space, i. e. a Hölder ball with smoothness index $\alpha > 1/2$, the result of Theorem 1.2 has been proved by Grama

and Nussbaum (1998). Note that the Sobolev balls $W^\alpha(M)$ figuring in Theorem 1.1 are natural parameter sets of spectral densities since the smoothness condition is directly stated in terms of the autocovariance function $\gamma_f(\cdot)$. The Besov balls $B_{p,p}^\alpha(M)$ given in terms of the norm $\|\cdot\|_{B_{p,p}^\alpha}$ are intermediate between L_2 -Sobolev and Hölder balls. For the white noise approximation of the i.i.d. (density estimation) model, Brown, Carter, Low and Zhang (2004) succeeded in weakening the Hölder ball condition in Nussbaum (1996) to a condition that Σ is compact both in the Besov spaces $B_{2,2}^{1/2}$ and $B_{4,4}^{1/2}$ on the unit interval. This is immediately implied by $\Sigma \subset B_{4,4}^\alpha(M)$ for some $\alpha > 1/2$. Our condition for Theorem 1.2 is slightly stronger, i.e. $\Sigma \subset B_{6,6}^\alpha(M)$ for some $\alpha > 1/2$. In Remark 5.8 (Appendix) we note a sufficient condition in terms of the autocovariance function $\gamma_f(\cdot)$, i.e. give a description of the periodic version of the Besov ball.

Throughout this paper we adopt the notation that C represents a constant independent of n and the parameter (spectral density) $f \in \Sigma$, and the value of which may change at each occurrence, even on the same line.

Relations between experiments. All measurable sample spaces are assumed to be Polish (complete separable) metric spaces equipped with their Borel sigma algebra. For measures P, Q on the same sample space, let $\|P - Q\|_{TV}$ be the total variation distance. For the general case where P, Q are not necessarily on the same sample space, suppose K is a Markov kernel such that KP is a measure on the same sample space as Q . In that case, $\|Q - KP\|_{TV}$ is defined and will be used as generic notation for a Markov kernel K .

Consider now experiments (families of measures) $\mathcal{F} = (Q_f, f \in \Sigma)$ and $\mathcal{E} = (P_f, f \in \Sigma)$, with the same parameter space Σ . All experiments here are assumed dominated by a sigma-finite measure on their respective sample space. If \mathcal{E} and \mathcal{F} are on the same sample space, define their total variation distance

$$\Delta_0(\mathcal{E}, \mathcal{F}) = \sup_{f \in \Sigma} \|Q_f - P_f\|_{TV}.$$

In the general case, the deficiency of \mathcal{E} with respect to \mathcal{F} is defined as

$$\delta(\mathcal{E}, \mathcal{F}) = \inf_K \sup_{f \in \Sigma} \|Q_f - KP_f\|_{TV}$$

where \inf extends over all appropriate Markov kernels. Le Cam's pseudodistance $\Delta(\cdot, \cdot)$ between \mathcal{E} and \mathcal{F} then is

$$\Delta(\mathcal{E}, \mathcal{F}) = \max(\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, \mathcal{E})).$$

Furthermore, we will use the following notation involving experiments \mathcal{E}, \mathcal{F} or sequences of such $\mathcal{E}_n = (P_{n,f}, f \in \Sigma)$ and $\mathcal{F}_n = (Q_{n,f}, f \in \Sigma)$.

Notation.

\mathcal{E}	\preceq	\mathcal{F}	(\mathcal{F} more informative than \mathcal{E}):	$\delta(\mathcal{F}, \mathcal{E}) = 0$
\mathcal{E}	\sim	\mathcal{F}	(equivalent):	$\Delta(\mathcal{E}, \mathcal{F}) = 0$
\mathcal{E}_n	\simeq	\mathcal{F}_n	(asymptotically total variation equivalent):	$\Delta_0(\mathcal{F}_n, \mathcal{E}_n) \rightarrow 0$
\mathcal{E}_n	\lesssim	\mathcal{F}_n	(\mathcal{F}_n asymptotically more informative than \mathcal{E}_n):	$\delta(\mathcal{F}_n, \mathcal{E}_n) \rightarrow 0$
\mathcal{E}_n	\approx	\mathcal{F}_n	(asymptotically equivalent):	$\Delta(\mathcal{F}_n, \mathcal{E}_n) \rightarrow 0$

Note that "more informative" above is used in the sense of a semi-ordering, i.e. its actual meaning is "at least as informative". We shall also write the relation \simeq in a less formal way between data vectors such as $x^{(n)} \simeq y^{(n)}$, if it is clear from the context which experiments the data vectors represent.

2 The periodic Gaussian experiment

From now on we shall assume that n is uneven. Our argument for asymptotic equivalence is such that it easily allows extension to the case of general sequences $n \rightarrow \infty$ (cf. Remark 4.10 for details).

Recall that the covariance matrix $\Gamma_n = \Gamma_n(f)$ has the Toeplitz form $(\Gamma_n)_{j,k} = \gamma(k - j)$, $j, k = 1, \dots, n$, i.e.

$$\Gamma_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-2) & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \dots & \gamma(n-2) \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(n-2) & \dots & \dots & \gamma(0) & \gamma(1) \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(1) & \gamma(0) \end{pmatrix}.$$

Following Brockwell and Davis (1991), § 4.5 we shall define a circulant matrix approximation by

$$\tilde{\Gamma}_n = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(2) & \gamma(1) \\ \gamma(1) & \gamma(0) & \dots & \dots & \gamma(2) \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(2) & \dots & \dots & \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(2) & \dots & \gamma(1) & \gamma(0) \end{pmatrix}$$

where in the first row, the central element and the one following it coincide with $\gamma((n-1)/2)$. More precisely, for given uneven n define a function on integers h with $|h| < n$

$$\tilde{\gamma}_{(n),f}(h) = \begin{cases} \gamma_f(h), & |h| \leq (n-1)/2 \\ \gamma_f(n - |h|), & (n+1)/2 \leq |h| \leq n-1 \end{cases}$$

and set

$$(\tilde{\Gamma}_n)_{j,k}(f) = \tilde{\gamma}_{(n),f}(k - j), \quad j, k = 1, \dots, n. \quad (2.1)$$

We shall also write $\tilde{\Gamma}_n(f)$ for the corresponding $n \times n$ matrix, or simply $\tilde{\Gamma}_n$ and $\tilde{\gamma}_{(n)}(h)$ if the dependence on f is understood. Define

$$\omega_j = \frac{2\pi j}{n}, \quad |j| \leq (n-1)/2. \quad (2.2)$$

It is well known (see Brockwell and Davis (1991), relation 4.5.5) that the spectral decomposition of $\tilde{\Gamma}_n$ can be described as follows. We have

$$\tilde{\Gamma}_n = \sum_{|j| \leq (n-1)/2} \lambda_j \mathbf{u}_j \mathbf{u}'_j \quad (2.3)$$

where λ_j are real eigenvalues and \mathbf{u}_j are real orthonormal eigenvectors. The eigenvalues are

$$\lambda_j = \sum_{|k| \leq (n-1)/2} \gamma(k) \exp(-i\omega_j k), \quad |j| \leq (n-1)/2.$$

Note that $\lambda_j = \lambda_{-j}$, $j \neq 0$ and that the λ_j are approximate values of $2\pi f$ in the points ω_j . Indeed define

$$\tilde{f}_n(\omega) = \frac{1}{2\pi} \sum_{|k| \leq (n-1)/2} \gamma(k) \exp(ik\omega), \quad \omega \in [-\pi, \pi] \quad (2.4)$$

a truncated Fourier series approximation to f ; then \tilde{f}_n is an even function on $[-\pi, \pi]$ and

$$\lambda_j = 2\pi \tilde{f}_n(\omega_j), \quad |j| \leq (n-1)/2. \quad (2.5)$$

The eigenvectors are

$$\mathbf{u}'_0 = n^{-1/2} (1, \dots, 1), \quad (2.6)$$

$$\mathbf{u}'_j = (2/n)^{1/2} (1, \cos(\omega_j), \cos(2\omega_j) \dots, \cos((n-1)\omega_j)), \quad (2.7)$$

$$\mathbf{u}'_{-j} = (2/n)^{1/2} (0, \sin(\omega_j), \sin(2\omega_j) \dots, \sin((n-1)\omega_j)), \quad j = 1, \dots, (n-1)/2. \quad (2.8)$$

In our setting, the circulant matrix $\tilde{\Gamma}_n$ is positive definite for n large enough. Indeed, Lemma 5.6 Appendix implies that $\tilde{f}_n \geq M^{-1}/2$ uniformly over $f \in \Sigma$, for n large enough, so that $\tilde{\Gamma}_n(f)$ is a covariance matrix. Define the experiment, in analogy to (1.3),

$$\tilde{\mathcal{E}}_n = \left(N_n(0, \tilde{\Gamma}_n(f)), f \in \Sigma \right) \quad (2.9)$$

with data $\tilde{y}^{(n)}$, say. The sequence $\tilde{y}^{(n)}$ may be called a "periodic process" since it can be represented in terms of independent standard Gaussians ξ_j , as a finite sum

$$\tilde{y}^{(n)} = \sum_{|j| \leq (n-1)/2} \lambda_j^{1/2} \mathbf{u}_j \xi_j \quad (2.10)$$

where the vector \mathbf{u}_j describes a deterministic oscillation (cp. (2.6)-(2.8)). Accordingly $\tilde{\mathcal{E}}_n$ will be called a *periodic Gaussian experiment*.

The periodic process $\tilde{y}^{(n)}$ is known to approximate the original time series $y^{(n)}$ in the following sense. Define the $n \times n$ -matrix

$$U_n = (\mathbf{u}_{-(n-1)/2}, \dots, \mathbf{u}_{(n-1)/2}) \quad (2.11)$$

and consider the transforms

$$z^{(n)} = (2\pi)^{-1/2} U_n' y^{(n)}, \quad \tilde{z}^{(n)} = (2\pi)^{-1/2} U_n' \tilde{y}^{(n)}.$$

Denote $\text{Cov}(z^{(n)})$ the covariance matrix of the random vector $z^{(n)}$. Then we have (Brockwell and Davis (1991), Proposition 4.5.2), for given $f \in \Sigma$

$$\sup_{1 \leq i, j \leq n} \left| \text{Cov}(z^{(n)})_{i,j} - \text{Cov}(\tilde{z}^{(n)})_{i,j} \right| \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (2.12)$$

Since $\text{Cov}(\tilde{z}^{(n)})$ is diagonal with diagonal elements $\lambda_j/2\pi$, this means that the elements of $z^{(n)}$ are approximately uncorrelated for large n .

Note that $\tilde{z}^{(n)}$ can also be written, in accordance with (2.10) and (2.5)

$$\tilde{z}^{(n)} = \left(\tilde{f}_n^{1/2}(\omega_j) \xi_j \right)_{|j| \leq (n-1)/2} \quad (2.13)$$

which is nearly identical with the Gaussian scale model (1.5). Thus the question appears whether the approximation (2.12) can be strengthened to a total variation approximation of the respective laws $\mathcal{L}(z^{(n)}|f)$ and $\mathcal{L}(\tilde{z}^{(n)}|f)$.

The answer to that is negative; let us introduce some notation. For $n \times n$ matrices $A = (a_{jk})$ define the Euclidean norm $\|A\|$ by

$$\|A\|^2 := \text{tr}[A'A] = \sum_{j=1}^n \sum_{k=1}^n a_{jk}^2.$$

If A is symmetric, we denote the largest and smallest eigenvalues by $\lambda_{\max}(A)$, $\lambda_{\min}(A)$. For later use, we also define the operator norm of (not necessarily symmetric) A by

$$|A| := (\lambda_{\max}(A'A))^{1/2}.$$

If A is symmetric nonnegative definite then $|A| = \lambda_{\max}(A)$. The following lemma shows that the Hellinger distance between the laws of $y^{(n)}$ and $\tilde{y}^{(n)}$ depends crucially on the total Euclidean distance $\|\Gamma_n(f) - \tilde{\Gamma}_n(f)\|$ between the covariance matrices, so that an elementwise convergence as in (2.12) is not enough.

Lemma 2.1 *Let A, B be $n \times n$ covariance matrices and suppose that for some $M > 1$*

$$0 < M^{-1} \leq \lambda_{\min}(A) \text{ and } \lambda_{\max}(A) \leq M.$$

Then there exist $\epsilon = \epsilon_M > 0$ and $K = K_M > 1$ not depending on A, B and n such that $\|A - B\| \leq \epsilon$ implies

$$K^{-1} \|A - B\|^2 \leq H^2(N_n(0, A), N_n(0, B)) \leq K \|A - B\|^2.$$

where $H(\cdot, \cdot)$ is the Hellinger distance.

The proof is in section 5. To apply this lemma, set $A = \Gamma_n(f)$, $B = \tilde{\Gamma}_n(f)$ and note that, since $f \in \Sigma$ is bounded and bounded away from 0 (both uniformly over $f \in \Sigma$), the condition on the eigenvalues of $\Gamma_n(f)$ is fulfilled, also uniformly over $f \in \Sigma$ (Brockwell and Davis (1991), Proposition 4.5.3). We shall see that the expression $\|\Gamma_n(f) - \tilde{\Gamma}_n(f)\|^2$ is closely related to a Sobolev type seminorm for smoothness index 1/2. For any $f \in L_2(-\pi, \pi)$ given by (1.1) set

$$|f|_{2,\alpha}^2 := \sum_{k=-\infty}^{\infty} |k|^{2\alpha} \gamma_f^2(k), \quad \|f\|_{2,\alpha}^2 := \gamma_f^2(0) + |f|_{2,\alpha}^2 \quad (2.14)$$

provided the right side is finite; the Sobolev ball $W^\alpha(M)$ given by (1.7) is then described by $\|f\|_{2,\alpha}^2 \leq M$. Also, for any natural m define a finite dimensional linear subspace of $L_2(-\pi, \pi)$

$$L_m = \left\{ f \in L_2(-\pi, \pi) : \int f(\omega) \exp(ik\omega) d\omega = 0, |k| > m \right\}.$$

Lemma 2.2 (i) For any $f \in \Sigma$ we have

$$\left\| \Gamma_n(f) - \tilde{\Gamma}_n(f) \right\|^2 \leq 2|f|_{2,1/2}^2 \quad (2.15)$$

and for $f \in \Sigma \cap L_{(n-1)/2}$

$$|f|_{2,1/2}^2 = \left\| \Gamma_n(f) - \tilde{\Gamma}_n(f) \right\|^2.$$

(ii) For any $f, f_0 \in \Sigma$ we have

$$\left\| \Gamma_n(f) - \Gamma_n(f_0) - \left(\tilde{\Gamma}_n(f) - \tilde{\Gamma}_n(f_0) \right) \right\|^2 \leq 2|f - f_0|_{2,1/2}^2. \quad (2.16)$$

Proof. (i) From the definition of $\Gamma_n(f)$ and $\tilde{\Gamma}_n(f)$ in terms of $\gamma(\cdot)$, $\tilde{\gamma}_{(n)}(\cdot)$ we immediately obtain

$$\begin{aligned} \left\| \Gamma_n(f) - \tilde{\Gamma}_n(f) \right\|^2 &= \sum_{|k| \leq n-1} (n - |k|) \left(\gamma(k) - \tilde{\gamma}_{(n)}(k) \right)^2 \\ &= \sum_{|k|=(n+1)/2}^{n-1} (n - |k|) (\gamma(k) - \gamma(n - |k|))^2 = 2 \sum_{k=1}^{(n-1)/2} k (\gamma(k) - \gamma(n - k))^2 \quad (2.17) \\ &\leq 2 \sum_{k=1}^{(n-1)/2} 2k (\gamma^2(k) + \gamma^2(n - k)) \leq 4 \sum_{k=1}^{n-1} k \gamma^2(k) \leq 2|f|_{2,1/2}^2. \end{aligned}$$

The first inequality is proved. The second one follows immediately from (2.17).

(ii) Note that for any n , the mapping $f \rightarrow \Gamma_n(f)$ if it is defined by (1.2) for any $f \in L_2(-\pi, \pi)$ is linear, and the same is true for $f \rightarrow \tilde{\Gamma}_n(f)$ defined by (2.1). Hence

$$\Gamma_n(f) - \Gamma_n(f_0) = \Gamma_n(f - f_0), \quad \tilde{\Gamma}_n(f) - \tilde{\Gamma}_n(f_0) = \tilde{\Gamma}_n(f - f_0).$$

Now the argument is completely analogous to (i) if $\gamma(k) = \gamma_f(k)$ is replaced by $\gamma_{f-f_0}(k)$. ■

Our assumption $f \in \Sigma$, i.e. $\|f\|_{2,\alpha}^2 \leq M$ for some $\alpha > 1/2$ provides an upper bound M for $|f|_{2,1/2}^2$ but does not guarantee that this term is uniformly small. Thus we are not able to utilize Lemma 2.1 to approximate \mathcal{E}_n by $\tilde{\mathcal{E}}_n$ in Hellinger distance. In fact this Hellinger distance approximation does not take place: take a fixed m , select $f \in \Sigma \cap L_m$ such that $\|f\|_{2,1/2}^2 < \epsilon$ with ϵ from Lemma 2.1 and use the lower bound in this lemma to show that

$$H^2 \left(N_n(0, \Gamma_n(f)), N_n(0, \tilde{\Gamma}_n(f)) \right) \geq K^{-1} \epsilon^2 \quad (2.18)$$

for all sufficiently large n . Thus the direct approximation of the time series data $y^{(n)}$ by the periodic process $\tilde{y}^{(n)}$ in total variation distance fails.

However that does not contradict asymptotic equivalence since the latter allows for a randomization mapping (Markov kernel) applied to $\tilde{y}^{(n)}$ and $y^{(n)}$, respectively, before total variation distance of the laws is taken. We will show the existence of appropriate Markov kernels in an indirect way, via a bracketing of the original time series experiment by upper and lower bounds in the sense of informativity.

Let now \mathcal{E}_n again be the time series experiment (1.3); we shall find an asymptotic bracketing, i.e. two sequences $\hat{\mathcal{E}}_{l,n}, \hat{\mathcal{E}}_{u,n}$ such that

$$\hat{\mathcal{E}}_{l,n} \lesssim \mathcal{E}_n \lesssim \hat{\mathcal{E}}_{u,n}$$

and such that both $\hat{\mathcal{E}}_{l,n}$ and $\hat{\mathcal{E}}_{u,n}$ are asymptotically equivalent to $\tilde{\mathcal{E}}_n$ given by (2.9), and to $\hat{\mathcal{E}}_n$ representing the independent Gaussians z_1, \dots, z_n in Theorem 1.1.

3 Upper informativity bracket

The spectral representation (2.10) of the periodic sequence $\tilde{y}^{(n)} = (\tilde{y}(1), \dots, \tilde{y}(n))'$ can be written

$$\begin{aligned} \tilde{y}(t) = & (2\pi/n)^{1/2} \tilde{f}_n^{1/2}(0) \xi_0 + 2(\pi/n)^{1/2} \sum_{j=1}^{(n-1)/2} \tilde{f}_n^{1/2}(\omega_j) \cos((t-1)\omega_j) \xi_j \\ & + 2(\pi/n)^{1/2} \sum_{j=-(n-1)/2}^{-1} \tilde{f}_n^{1/2}(\omega_j) \sin((t-1)\omega_j) \xi_j, t = 1, \dots, n. \end{aligned} \quad (3.1)$$

We saw that here $\tilde{y}^{(n)}$ is a one-to-one function $\tilde{y}^{(n)} = U\tilde{z}^{(n)}$ of the n -vector of independent Gaussians $\tilde{z}^{(n)}$ (cf. (2.13)), but the approximation of $\tilde{y}^{(n)}$ to $y^{(n)}$ is not in the total variation sense (cf. (2.18)). Now take a limit in (3.1) for $n \rightarrow \infty$ and fixed t and observe that (heuristically) this yields the spectral representation of the original stationary sequence $y(t)$

$$y(t+1) = \int_{[0,\pi]} \sqrt{2} f^{1/2}(\omega) \cos(t\omega) dB_\omega + \int_{[-\pi,0]} \sqrt{2} f^{1/2}(\omega) \sin(t\omega) dB_\omega, t = 0, 1, \dots \quad (3.2)$$

where dB_ω is standard Gaussian white noise on $[-\pi, \pi]$ (cf. Brockwell and Davis (1991), Probl. 4.31). Here for any n , the vector $y^{(n)} = (y(1), \dots, y(n))'$ is represented as a functional of the continuous time process

$$dZ_\omega^* = f^{1/2}(\omega) dB_\omega, \omega \in [-\pi, \pi].$$

Thus a completely observed process $Z_\omega^*, \omega \in [-\pi, \pi]$ would represent an upper informativity bracket for any sample size n , but this experiment is statistically trivial since the observation here identifies the parameter f .

Our approach now is to construct an intermediate series $\tilde{y}^{(m,n)}$ of size n in which the uniform size n grid of points $\omega_j, |j| \leq (n-1)/2$ is replaced by a finer uniform grid of $m > n$ points in the representation (3.1). Thus $\tilde{y}^{(n,m)}$ is a functional not of n independent Gaussians but of $m > n$ of these; call their vector $\tilde{z}^{(m)}$. The random vector $\tilde{z}^{(m)}$ now represents an upper informativity bracket which remains nontrivial (asymptotically) if $m-n \rightarrow \infty$ not too quickly. An equivalent description of that idea is as follows. Consider $m > n$ and the periodic process $\tilde{y}^{(m)}$ given by (2.10) where the original sample size n is replaced by m . Then define $\tilde{y}^{(n,m)}$ as the vector of the first n components of $\tilde{y}^{(m)}$. The law of $\tilde{y}^{(n,m)}$ is $N_n(0, \tilde{\Gamma}_{n,m}(f))$ where $\tilde{\Gamma}_{n,m}(f)$ is the upper left $n \times n$ submatrix of $\tilde{\Gamma}_m(f)$.

We now easily observe the improved approximation quality of $\tilde{y}^{(n,m)}$ for $y^{(n)}$. Assume that m is also uneven. First note that for $(m+1)/2 \geq n$ we already obtain $\tilde{\Gamma}_{n,m}(f) = \Gamma_n(f)$. This follows immediately from the definition of the circular matrix $\tilde{\Gamma}_m(f)$ via the autocovariance function $\tilde{\gamma}_{(m)}(\cdot)$. However we would like to limit the increase of sample size, i.e. require $m/n \rightarrow 1$; therefore, in what follows we assume $m < 2n - 1$.

Lemma 3.1 *Assume m is uneven, $n < m < 2n - 1$. Then for any $f \in \Sigma$ we have*

$$\left\| \Gamma_n(f) - \tilde{\Gamma}_{n,m}(f) \right\|^2 \leq 4(m-n+1)^{1-2\alpha} |f|_{2,\alpha}^2,$$

and hence if $m = m_n$ is such that $m - n \rightarrow \infty$ as $n \rightarrow \infty$ then

$$\sup_{f \in \Sigma} H^2 \left(N_n(0, \Gamma_n(f)), N_n(0, \tilde{\Gamma}_{n,m}(f)) \right) \rightarrow 0. \quad (3.3)$$

Proof. From the definition of $\Gamma_n(f)$ and $\tilde{\Gamma}_{n,m}(f)$ we immediately obtain

$$\begin{aligned} \left\| \Gamma_n(f) - \tilde{\Gamma}_{n,m}(f) \right\|^2 &= \sum_{|k| \leq n-1} (n - |k|) \left(\gamma(k) - \tilde{\gamma}_{(m)}(k) \right)^2 \\ &= 2 \sum_{k=(m+1)/2}^{n-1} (n-k) (\gamma(k) - \gamma(m-k))^2 \leq 4 \sum_{k=(m+1)/2}^{n-1} (n-k) (\gamma^2(k) + \gamma^2(m-k)). \end{aligned}$$

Now note that for $m > n$, the relation $(m+1)/2 \leq k \leq n-1$ implies $k \geq (n+1)/2$ and therefore $n-k < k$, and note also $n-k < m-k$. We obtain an upper bound

$$\begin{aligned} &\leq 4 \sum_{k=(m+1)/2}^{n-1} k\gamma^2(k) + 4 \sum_{k=(m+1)/2}^{n-1} (m-k)\gamma^2(m-k) \\ &= 4 \sum_{k=(m+1)/2}^{n-1} k\gamma^2(k) + 4 \sum_{k=m-n+1}^{(m-1)/2} k\gamma^2(k) = 4 \sum_{k=m-n+1}^{n-1} k\gamma^2(k) \\ &\leq 4(m-n+1)^{1-2\alpha} \sum_{k=m-n+1}^{n-1} k^{2\alpha}\gamma^2(k) \leq 4(m-n+1)^{1-2\alpha} |f|_{2,\alpha}^2 \end{aligned}$$

where $\alpha > 1/2$. This proves the first relation. For the second, recall that $|f|_{2,\alpha}^2 \leq M$ for $f \in \Sigma$ and invoke Lemma 2.1 together with the subsequent remark on the eigenvalues of $\Gamma_n(f)$. ■

Define the experiment

$$\tilde{\mathcal{E}}_{n,m} = \left(N_n(0, \tilde{\Gamma}_{n,m}(f)), f \in \Sigma \right)$$

then (3.3) implies $\mathcal{E}_n \simeq \tilde{\mathcal{E}}_{n,m}$ if $m - n \rightarrow \infty$. Moreover, we have $\tilde{\mathcal{E}}_{n,m} \preceq \tilde{\mathcal{E}}_m$ by definition, thus

$$\mathcal{E}_n \lesssim \tilde{\mathcal{E}}_m$$

in case $m - n \rightarrow \infty$. We know that $\tilde{\mathcal{E}}_m$ is equivalent (via the linear transformation $(2\pi)^{-1/2}U'$) to observing data $\tilde{z}^{(n)}$ given by (2.13). Define $\hat{\mathcal{E}}_n$ by

$$\hat{\mathcal{E}}_n = \left(N_n(0, \hat{\Gamma}_n(f)), f \in \Sigma \right) \quad (3.4)$$

where

$$\hat{\Gamma}_n(f) = \text{Diag} (J_{j,n}(f))_{j=1,\dots,n}.$$

Note that the data z_1, \dots, z_n in Theorem 1.1 are represented by $\hat{\mathcal{E}}_n$. We shall also write $\tilde{z}^{(n)}$ for their vector, so that $\mathcal{L}(\tilde{z}^{(n)}|f) = N_n(0, \hat{\Gamma}_n(f))$.

Proposition 3.2 *We have $\hat{\mathcal{E}}_n \approx \tilde{\mathcal{E}}_n$, with corresponding equivalence maps (Markov kernels) as follows. Let $\tilde{y}^{(n)}$ and $\tilde{z}^{(n)}$ be data in $\tilde{\mathcal{E}}_n$ and $\hat{\mathcal{E}}_n$ respectively. Then, for the orthogonal matrix U_n given by (2.11)*

$$(2\pi)^{-1/2}U_n' \tilde{y}^{(n)} \simeq \tilde{z}^{(n)}, \text{ and } (2\pi)^{1/2}U_n \tilde{z}^{(n)} \simeq \tilde{y}^{(n)}.$$

Proof. Note that our first claim can also be written $\tilde{z}^{(n)} \simeq \tilde{z}^{(n)}$ where $\tilde{z}^{(n)}$ is from (2.13). To describe $\mathcal{L}(\tilde{z}^{(n)}|f)$, define $\delta_j = \tilde{f}_n(\omega_{j-(n+1)/2})$ for $j = 1, \dots, n$ and a $n \times n$ covariance matrix

$$\Delta_n(f) = \text{Diag} (\delta_j)_{j=1,\dots,n}.$$

Then $\mathcal{L}(\tilde{z}^{(n)}|f) = N_n(0, \Delta_n(f))$. The conditions on f (see also Lemma 5.6 Appendix) imply that uniformly over $j = 1, \dots, n$

$$J_{j,n}(f) \geq C^{-1}, J_{j,n}(f) \leq C$$

for some $C > 0$ not depending on f and n . Now apply Lemma 2.1 to obtain

$$H^2 \left(N_n(0, \hat{\Gamma}_n(f)), N_n(0, \Delta_n(f)) \right) \leq C \left\| \hat{\Gamma}_n(f) - \Delta_n(f) \right\|^2 = C \sum_{j=1}^n (J_{j,n}(f) - \delta_j)^2.$$

By Lemma 5.7 this is $o(1)$ uniformly in f . This implies the first relation \simeq . The second relation is an obvious consequence. ■

For a choice $m = n + r_n$, $r_n = 2 \lceil \log(n/2) \rceil$ we immediately obtain the following result. Define the upper bracket Gaussian scale experiment $\hat{\mathcal{E}}_{u,n}$ by

$$\hat{\mathcal{E}}_{u,n} := \hat{\mathcal{E}}_{n+r_n}. \quad (3.5)$$

Corollary 3.3 *Consider experiments \mathcal{E}_n and $\hat{\mathcal{E}}_{u,n}$ given respectively by (1.3) and (3.5), (3.4) with parameter space $\Sigma = \Sigma_{\alpha,M}$ where $M > 0$, $\alpha > 1/2$. Then as $n \rightarrow \infty$*

$$\mathcal{E}_n \lesssim \hat{\mathcal{E}}_{u,n}.$$

4 Lower informativity bracket

The upper bound (2.15) for the Hellinger distance of $y^{(n)}$ and the periodic process $\tilde{y}^{(n)}$ which does not tend to 0, can be improved in a certain sense if f is restricted to a shrinking neighborhood, $\Sigma_n(f_0)$ say, of some $f_0 \in \Sigma$. At this stage, f_0 is assumed known so the covariance matrices $\Gamma_n(f)$ and $\tilde{\Gamma}_n(f)$ can be used for a linear transformation of $y^{(n)}$ which brings it closer to the periodic process $\tilde{y}^{(n)}$. The linear transformation of $y^{(n)}$ which depends on f_0 can be construed as a Markov kernel mapping which yields asymptotic equivalence $\mathcal{E}_n(f_0) \approx \tilde{\mathcal{E}}_n(f_0)$ if these are the versions of \mathcal{E}_n and $\tilde{\mathcal{E}}_n$ with f restricted to $f \in \Sigma_n(f_0)$.

Such a local asymptotic equivalence can be globalized in a standard way (cf. Nussbaum (1996), Grama and Nussbaum (1998)) if *sample splitting* were available in both global experiments \mathcal{E}_n and $\tilde{\mathcal{E}}_n$. For the original stationary process that would mean that observing a series of size n is equivalent to observing two independent series of size approximately $n/2$. We will establish an asymptotic version of sample splitting for $y^{(n)}$ which involves omitting a fraction of the sample in the center of the series, i.e. omitting terms with index near $n/2$. The ensuing loss of information means that the globalization procedure only yields a lower asymptotic informativity bracket for \mathcal{E}_n , i.e. a sequence $\tilde{\mathcal{E}}_{3,n}^\#$ such that $\tilde{\mathcal{E}}_{3,n}^\# \lesssim \mathcal{E}_n$. The experiment $\tilde{\mathcal{E}}_{3,n}^\#$ will be made up of two independent periodic processes with the same parameter f and with a sample size $m \sim (n - \log n)/2$. Each of these is equivalent to a Gaussian scale model (2.13) with n replaced by m ; further arguments show that observing these two is asymptotically equivalent to a Gaussian scale model $\hat{\mathcal{E}}_{l,n} := \hat{\mathcal{E}}_{2m}$ with grid size $2m \sim n - \log n$.

A crucial step now consists in showing that in the Gaussian scale models $\hat{\mathcal{E}}_n$, the grid size n can be replaced by $n - \log n$ or $n + \log n$. This step is an analog, for the special regression model, of the well known reasoning in the i.i.d. case that additional observations may be asymptotically negligible (cf. Mammen (1986) for parametric i.i.d. models, Low and Zhou (2004) for the nonparametric case). Thus it follows that the lower and upper bracketing experiments $\hat{\mathcal{E}}_{l,n}$, $\hat{\mathcal{E}}_{u,n}$ are both asymptotically equivalent to $\hat{\mathcal{E}}_n$, and the relations

$$\hat{\mathcal{E}}_{l,n} \lesssim \mathcal{E}_n \lesssim \hat{\mathcal{E}}_{u,n}$$

then imply $\mathcal{E}_n \approx \hat{\mathcal{E}}_n$, i.e. Theorem 1.1.

4.1 Local experiments

Let \varkappa_n be a sequence $\varkappa_n \searrow 0$, fixed in the sequel. A specific choice of \varkappa_n will be made in section 4.4 below (see (4.12)). Let $\|\cdot\|_\infty$ be the sup-norm for real functions defined on $[-\pi, \pi]$, i.e.

$$\|f\|_\infty = \sup_{\omega \in [-\pi, \pi]} |f(\omega)|$$

and for $f_0 \in \Sigma$ define shrinking neighborhoods

$$\Sigma_n(f_0) = \left\{ f \in \Sigma : \|f - f_0\|_\infty + \|f - f_0\|_{2,1/2} \leq \varkappa_n \right\}. \quad (4.1)$$

The restricted experiments are

$$\mathcal{E}_n(f_0) = (N_n(0, \Gamma_n(f)), f \in \Sigma_n(f_0)), \quad \tilde{\mathcal{E}}_n(f_0) = \left(N_n(0, \tilde{\Gamma}_n(f)), f \in \Sigma_n(f_0) \right).$$

For shortness write $\Gamma = \Gamma_n(f)$, $\Gamma_0 = \Gamma_n(f_0)$ and similarly $\tilde{\Gamma} = \Gamma_n(f)$, $\tilde{\Gamma}_0 = \tilde{\Gamma}_n(f_0)$. Define a matrix

$$K_n = K_n(f_0) = \tilde{\Gamma}_0^{1/2} \Gamma_0^{-1/2} \quad (4.2)$$

and in experiment $\mathcal{E}_n(f_0)$ consider transformed observations

$$\check{y}^{(n)} := K_n(f_0) y^{(n)}.$$

Consider also the experiment $\mathcal{E}_n^*(f_0)$ given by the laws of $\check{y}^{(n)}$, i.e.

$$\mathcal{E}_n^*(f_0) = (N_n(0, K_n(f_0) \Gamma_n(f) K_n'(f_0)), f \in \Sigma_n(f_0)).$$

Clearly $\mathcal{E}_n(f_0) \sim \mathcal{E}_n^*(f_0)$; the next result proves that $\mathcal{E}_n^*(f_0) \simeq \tilde{\mathcal{E}}_n(f_0)$ and thus $\mathcal{E}_n(f_0) \approx \tilde{\mathcal{E}}_n(f_0)$.

Lemma 4.1 *We have*

$$\sup_{f_0 \in \Sigma} \sup_{f \in \Sigma_n(f_0)} H^2 \left(N_n(0, K_n(f_0) \Gamma_n(f) K_n'(f_0)), N_n(0, \tilde{\Gamma}_n(f)) \right) \leq C \varkappa_n.$$

Proof. In view of Lemma 2.1, it suffices to show that

$$\sup_{f \in \Sigma} \left(\lambda_{\max}(\tilde{\Gamma}_n) + \lambda_{\min}^{-1}(\tilde{\Gamma}_n) \right) \leq C \quad (4.3)$$

and that

$$\left\| K_n \Gamma_n K_n' - \tilde{\Gamma}_n \right\|^2 \leq C \varkappa_n.$$

Note that

$$\lambda_{\max}(\tilde{\Gamma}) = \max_{|j| \leq (n-1)/2} \left| \tilde{f}_n(\omega_j) \right|, \quad \lambda_{\min}(\tilde{\Gamma}) = \min_{|j| \leq (n-1)/2} \left| \tilde{f}_n(\omega_j) \right|$$

and that Lemma 5.6 implies

$$\sup_{f \in \Sigma} \left\| f - \tilde{f}_n \right\|_{\infty} \rightarrow 0.$$

Hence (4.3) follows immediately from $f \in \Sigma$, more specifically the fact that values of f are uniformly bounded and bounded away from 0. According to Proposition 4.5.3 in Brockwell and Davis (1991), the assumption $f \in \Sigma$ also implies a corresponding property for Γ , i.e.

$$\sup_{f \in \Sigma} \left(\lambda_{\max}(\Gamma_n) + \lambda_{\min}^{-1}(\Gamma_n) \right) \leq C. \quad (4.4)$$

Note that eigenvalues of Γ_0 and $\tilde{\Gamma}_0$ share property (4.3) since $f_0 \in \Sigma$.

Set $G = \Gamma_0^{-1/2} \Gamma \Gamma_0^{-1/2}$ and $\tilde{G} = \tilde{\Gamma}_0^{-1/2} \tilde{\Gamma} \tilde{\Gamma}_0^{-1/2}$. Since

$$\left\| K_n \Gamma_n K_n' - \tilde{\Gamma}_n \right\| \leq \left| \tilde{\Gamma}_0 \right| \left\| G - \tilde{G} \right\|$$

it now suffices to show that

$$\left\| G - \tilde{G} \right\| \leq C \varkappa_n. \quad (4.5)$$

To establish (4.5), denote $\Delta = \Gamma - \Gamma_0$, $\tilde{\Delta} = \tilde{\Gamma} - \tilde{\Gamma}_0$ and observe

$$\begin{aligned} \|G - \tilde{G}\| &= \left\| \Gamma_0^{-1/2} \Gamma \Gamma_0^{-1/2} - \tilde{\Gamma}_0^{-1/2} \tilde{\Gamma} \tilde{\Gamma}_0^{-1/2} \right\| \\ &= \left\| \Gamma_0^{-1/2} \Delta \Gamma_0^{-1/2} - \tilde{\Gamma}_0^{-1/2} \tilde{\Delta} \tilde{\Gamma}_0^{-1/2} \right\| \\ &\leq \left\| \Gamma_0^{-1/2} (\Delta - \tilde{\Delta}) \Gamma_0^{-1/2} \right\| + \left\| \Gamma_0^{-1/2} \tilde{\Delta} \tilde{\Gamma}_0^{-1/2} - \tilde{\Gamma}_0^{-1/2} \tilde{\Delta} \tilde{\Gamma}_0^{-1/2} \right\|. \end{aligned} \quad (4.6)$$

We shall now estimate the two terms on the right side separately. By elementary properties of eigenvalues we obtain

$$\left\| \Gamma_0^{-1/2} (\Delta - \tilde{\Delta}) \Gamma_0^{-1/2} \right\| \leq |\Gamma_0^{-1}| \|\Delta - \tilde{\Delta}\|$$

where $|\Gamma_0^{-1}| \leq C$ and according to Lemma 2.2 (ii)

$$\|\Delta - \tilde{\Delta}\|^2 \leq 2|f - f_0|_{2,1/2}^2.$$

Furthermore

$$\begin{aligned} &\left\| \Gamma_0^{-1/2} \tilde{\Delta} \tilde{\Gamma}_0^{-1/2} - \tilde{\Gamma}_0^{-1/2} \tilde{\Delta} \tilde{\Gamma}_0^{-1/2} \right\| \\ &= \left\| (\Gamma_0^{-1/2} - \tilde{\Gamma}_0^{-1/2}) \tilde{\Delta} \tilde{\Gamma}_0^{-1/2} + \tilde{\Gamma}_0^{-1/2} \tilde{\Delta} (\Gamma_0^{-1/2} - \tilde{\Gamma}_0^{-1/2}) \right\| \\ &\leq 2C |\tilde{\Delta}| \left\| \Gamma_0^{-1/2} - \tilde{\Gamma}_0^{-1/2} \right\| = C |\tilde{\Delta}| \left\| \Gamma_0^{-1/2} (\tilde{\Gamma}_0^{1/2} - \Gamma_0^{1/2}) \tilde{\Gamma}_0^{-1/2} \right\| \\ &\leq C |\tilde{\Delta}| \left\| \Gamma_0^{1/2} - \tilde{\Gamma}_0^{1/2} \right\|. \end{aligned}$$

Applying Lemma 5.1 and Lemma 2.2 (i) we obtain

$$\left\| \Gamma_0^{1/2} - \tilde{\Gamma}_0^{1/2} \right\|^2 \leq C \left\| \Gamma_0 - \tilde{\Gamma}_0 \right\|^2 \leq C |f_0|_{2,1/2}^2.$$

Here $|f_0|_{2,1/2}^2 \leq |f_0|_{2,\alpha}^2 \leq M$. Collecting these estimates yields

$$\|G - \tilde{G}\|^2 \leq C \left(|f - f_0|_{2,1/2}^2 + |\tilde{\Delta}|^2 \right).$$

To complete the proof, it suffices to note that, since $\tilde{\Gamma}$ and $\tilde{\Gamma}_0$ have the same set of eigenvectors (cf. (2.3) and (2.6)-(2.8))

$$\begin{aligned} |\tilde{\Delta}|^2 &= \lambda_{\max}(\tilde{\Gamma} - \tilde{\Gamma}_0)^2 = (2\pi)^2 \max_{|j| \leq (n-1)/2} \left(|\tilde{f}_n(\omega_j) - \tilde{f}_{0,n}(\omega_j)|^2 \right) \\ &\leq C \left\| \tilde{f}_n - \tilde{f}_{0,n} \right\|_{\infty}^2 \leq C \|f - f_0\|_{\infty}^2 + C n^{1-2\alpha} \log n \|f - f_0\|_{2,\alpha}^2 \end{aligned}$$

where the last inequality is a consequence of Lemma 5.6. Hence $|\tilde{\Delta}| \leq C \varkappa_n$, which establishes (4.5). ■

4.2 Sample splitting

Consider sample splitting for a stationary process: Take the observed $y^{(n)} = (y(1), \dots, y(n))$ and omit r observations in the center of the series. Recall that n was assumed uneven; assume now also r to be uneven and set $m = (n - r)/2$, then the result is the series $y(1), \dots, y(m), y(n - m + 1), \dots, y(n)$. The total covariance matrix for these reduced data is

$$\Gamma_{n,0}^{(m)}(f) := \begin{pmatrix} \Gamma_m(f) & A_{n,m} \\ A_{n,m}' & \Gamma_m(f) \end{pmatrix}$$

where the $m \times m$ matrix $A_{n,m} = A_{n,m}(f)$ contains only covariances $\gamma_f(r+1), \gamma_f(r+2)$ and of higher order. In fact A is the upper right $m \times m$ submatrix of $\Gamma_n(f)$, i.e.

$$A_{n,m} = \begin{pmatrix} \dots & \gamma(n-2) & \gamma(n-1) \\ \gamma(r+2) & \dots & \gamma(n-2) \\ \gamma(r+1) & \gamma(r+2) & \dots \end{pmatrix}.$$

In the sequel we set $r_n = 2[\log n/2] + 1$ and thus $r_n \sim \log n$, $m = (n - r_n)/2$. The corresponding experiment we denote

$$\mathcal{E}_{0,n}^\# = \left(N_{2m}(0, \Gamma_{n,0}^{(m)}(f)), f \in \Sigma \right)$$

Consider also the experiment where two independent stationary series of length m are observed, $y_1^{(m)}$ and $y_2^{(m)}$, say. The corresponding experiment is

$$\mathcal{E}_{1,n}^\# := \left(N_{2m}(0, \Gamma_{n,1}^{(m)}(f)), f \in \Sigma \right) \quad (4.7)$$

where

$$\Gamma_{n,1}^{(m)}(f) := \begin{pmatrix} \Gamma_m(f) & 0_{m \times m} \\ 0_{m \times m} & \Gamma_m(f) \end{pmatrix}.$$

Clearly we have $\mathcal{E}_{0,n}^\# \preceq \mathcal{E}_n$.

Proposition 4.2 $\mathcal{E}_{0,n}^\# \simeq \mathcal{E}_{1,n}^\#$.

Proof. Use Lemma 2.1 to compute the Hellinger distance. Take $A = \Gamma_{n,1}^{(m)}$; then the eigenvalues of A are those of $\Gamma_m(f)$, so that (4.4) can be invoked. The squared distance of the covariance matrices $\Gamma_{n,0}^{(m)}$ and $\Gamma_{n,1}^{(m)}$ is

$$\begin{aligned} \left\| \Gamma_{n,0}^{(m)} - \Gamma_{n,1}^{(m)} \right\|^2 &= 2 \|A_{n,m}\|^2 \leq 2 \sum_{k=r+1}^{n-1} (k-r)\gamma^2(k) \\ &\leq 2 \sum_{k=r+1}^{n-1} k\gamma^2(k) \leq (r+1)^{1-2\alpha} |f|_{2,\alpha}^2. \end{aligned}$$

Since $r_n \rightarrow \infty$, the result follows. ■

We have shown that two independent stationary sequences of length $m = (n - r_n)/2$ are asymptotically less informative than one sequence of length n . Having obtained a method of sample splitting for stationary sequences (with some loss of information), we can now use a localization argument to complete the proof of the lower bound.

4.3 Preliminary estimators

For the globalization procedure, we need existence of an estimator \hat{f}_n , in both of the global experiments \mathcal{E}_n and $\tilde{\mathcal{E}}_n$ (or $\hat{\mathcal{E}}_n$), such that \hat{f}_n takes values in Σ and

$$\left\| \hat{f}_n - f \right\|_{\infty} + \left\| \hat{f}_n - f \right\|_{2,1/2} = o_p(1)$$

uniformly over $f \in \Sigma$. More specifically, a rate $o_p(\kappa_n)$ with κ_n from (4.1) is needed in the above result, but κ_n has not been selected so far, and will be determined based on the results of this section (cf. (4.12) below). Select $\beta \in (1/2, \alpha)$ and consider the norm $\|f\|_{2,\beta}$ according to (2.14). Note that $\|f\|_{2,1/2} \leq C \|f\|_{2,\beta}$ and that according to Lemma 5.6 $\|f\|_{\infty} \leq C \|f\|_{2,\beta}$; therefore it suffices to show

$$\left\| \hat{f}_n - f \right\|_{2,\beta}^2 = o_p(1). \quad (4.8)$$

For this, we shall use a standard truncated orthogonal series estimator and then modify it to take values in Σ . The empirical autocovariance function is

$$\hat{\gamma}_n(k) = \frac{1}{n-k} \sum_{j=1}^{n-k} y(j)y(k+j), \quad k = 0, \dots, n-1.$$

We have unbiasedness: $E\hat{\gamma}_n(k) = \gamma_f(k)$; for the variance of $\hat{\gamma}_n(k)$ we have the following result.

Lemma 4.3 *For any spectral density $f \in L_2(-\pi, \pi)$, and any $k = 0, \dots, n-1$*

$$\text{Var}\hat{\gamma}_n(k) \leq \frac{5}{n-k} \sum_{j=0}^{n-1} \gamma_f^2(j).$$

Proof. For given k , set $m = n - k$ and $z(j) = y(j)y(j+k) - \gamma_f(k)$, $j = 1, \dots, m$. The $z(j)$ form a zero mean stationary series, with autocovariance function $\rho(j)$, say. We have

$$\begin{aligned} E \left(\frac{1}{m} \sum_{k=1}^m z(k) \right)^2 &= \frac{1}{m^2} \sum_{1 \leq j, k \leq m} \rho(k-j) = \frac{1}{m} \rho(0) + 2 \frac{1}{m^2} \sum_{k=1}^{m-1} (m-k) \rho(k) \\ &\leq \frac{2}{m} \sum_{k=0}^{m-1} \rho(k). \end{aligned} \quad (4.9)$$

The computation in Shiryaev (1996), (VI.4.5-6) gives

$$\rho(j) = \gamma^2(j) + \gamma(j-k)\gamma(j+k).$$

The inequality

$$2|\gamma(j-k)\gamma(j+k)| \leq \gamma^2(j-k) + \gamma^2(j+k)$$

now implies

$$\sum_{k=0}^{m-1} \rho(k) \leq \frac{5}{2} \sum_{k=0}^{n-1} \gamma^2(k)$$

(we bound the sum involving $\gamma^2(j-k)$ by $2\sum_{j=0}^{n-1}\gamma^2(j)$). In conjunction with (4.9) this proves the lemma. ■

For the orthogonal series estimator, define a truncation index $\tilde{n} = \lceil n^{1/(2\alpha+1)} \rceil$ and set

$$\hat{f}_n(\omega) = \sum_{|k| \leq \tilde{n}} \hat{\gamma}_n(k) \exp(ik\omega), \quad \omega \in [-\pi, \pi]. \quad (4.10)$$

Lemma 4.4 *In the experiment \mathcal{E}_n the estimator \hat{f}_n fulfills for any $\beta \in (1/2, \alpha)$ and any $\gamma \in (0, \frac{\alpha-\beta}{2\alpha+1})$*

$$\sup_{f \in \Sigma} P \left(\left\| \hat{f}_n - f \right\|_{2,\beta}^2 > n^{-\gamma} \right) \rightarrow 0. \quad (4.11)$$

Proof. By the Markov inequality, it suffices to prove

$$\sup_{f \in \Sigma} E_f \left\| \hat{f}_n - f \right\|_{2,\beta}^2 = o(n^{-\gamma}).$$

A bias-variance decomposition and Lemma 4.3 yield

$$\begin{aligned} E_f \left\| \hat{f}_n - f \right\|_{2,\beta}^2 &= \sum_{|k| \leq \tilde{n}} \max(1, |k|^{2\beta}) \text{Var} \hat{\gamma}_n(k) + \sum_{|k| > \tilde{n}} |k|^{2\beta} \gamma^2(k) \\ &\leq \sum_{|k| \leq \tilde{n}} \max(1, |k|^{2\beta}) \frac{5}{n-k} \left(\sum_{j=0}^{n-1} \gamma_f^2(j) \right) + \tilde{n}^{2\beta-2\alpha} \sum_{|k| > \tilde{n}} |k|^{2\alpha} \gamma^2(k) \\ &\leq \frac{C}{n} \|f\|_2^2 \sum_{|k| \leq \tilde{n}} \max(1, |k|^{2\beta}) + \tilde{n}^{2\beta-2\alpha} |f|_{2,\alpha}^2 \\ &\leq C \|f\|_2^2 n^{-1} \tilde{n}^{2\beta+1} + C |f|_{2,\alpha}^2 \tilde{n}^{2(\beta-\alpha)} \\ &\leq C \left(\|f\|_2^2 + |f|_{2,\alpha}^2 \right) n^{2(\beta-\alpha)/(2\alpha+1)}. \end{aligned}$$

Since $\|f\|_2^2 \leq C \|f\|_{2,\alpha}^2$ and $|f|_{2,\alpha}^2 \leq \|f\|_{2,\alpha}^2$, the result follows. ■

We now turn to preliminary estimation in the periodic experiment $\tilde{\mathcal{E}}_n$ with data vector $\tilde{y}^{(n)}$. Note that this data vector can be construed as coming from a stationary sequence with autocovariance function $\tilde{\gamma}_{(n)}(\cdot)$ given by (2.1) for $|k| \leq n-1$ and $\tilde{\gamma}_{(n)}(k) = 0$ for $|k| > n-1$, i.e. the stationary sequence having spectral density \tilde{f}_n . Thus if $\hat{\gamma}_n(k)$ again denotes the empirical autocovariance function in this series then we can apply Lemma 4.3 to obtain

$$\text{Var} \hat{\gamma}_n(k) \leq \frac{5}{n-k} \sum_{j=0}^{n-1} \tilde{\gamma}_{(n),f}^2(j), \quad k = 0, \dots, n-1.$$

Obviously

$$\sum_{k=0}^{n-1} \tilde{\gamma}_{(n),f}^2(k) = \sum_{k=0}^{(n-1)/2} \gamma_f^2(k) + \sum_{k=1}^{(n-1)/2} \gamma_f^2(k) \leq 2 \|f\|_2^2.$$

Now use the estimator (4.10) with \tilde{n} as above; since $\tilde{n} = o((n-1)/2)$, we have the unbiasedness

$$E\hat{\gamma}_n(k) = \gamma_f(k), \quad k = 0, \dots, \tilde{n}.$$

Thus the proof of the following result is entirely analogous to Lemma 4.4; the estimator \hat{f}_n is also formally the same function of the data.

Lemma 4.5 *In the experiment $\tilde{\mathcal{E}}_n$ the estimator \hat{f}_n fulfills (4.11) for any $\beta \in (1/2, \alpha)$ and any $\gamma \in \left(0, \frac{\alpha-\beta}{2\alpha+1}\right)$.*

Finally consider modifications such that the estimator takes values in $\Sigma_{\alpha, M}$. Consider the space $W^\beta = \left\{f \in L_2(-\pi, \pi) : \|f\|_{2, \beta}^2 < \infty\right\}$; this is a periodic fractional Sobolev space which is Hilbert under the norm $\|f\|_{2, \beta}$. There the set $\Sigma_{\alpha, M}$ is compact and convex; hence there exists a ($\|\cdot\|_{2, \beta}$ -continuous) projection operator Π onto $\Sigma_{\alpha, M}$ in W^β (cf. Balakrishnan (1976), Definition 1.4.1). Then

$$\left\|\Pi\left(\hat{f}_n\right) - f\right\|_{2, \beta} \leq \left\|\hat{f}_n - f\right\|_{2, \beta}.$$

The modified estimators $\Pi\left(\hat{f}_n\right)$ thus again fulfill (4.11). A summary of results in this section is the following.

Proposition 4.6 *In both experiments \mathcal{E}_n and $\tilde{\mathcal{E}}_n$ there are estimators \hat{f}_n taking values in Σ and fulfilling for any $\gamma \in \left(0, \frac{\alpha-1/2}{2\alpha+1}\right)$*

$$\sup_{f \in \Sigma} P\left(\left\|\hat{f}_n - f\right\|_\infty + \left\|\hat{f}_n - f\right\|_{2, 1/2} > n^{-\gamma}\right) \rightarrow 0.$$

4.4 Globalization

In this section we denote

$$P_{f, n} := \mathcal{L}(y^{(n)}|f) = N_n(0, \Gamma_n(f)), \quad \tilde{P}_{f, n} := \mathcal{L}(\tilde{y}^{(n)}|f) = N_n\left(0, \tilde{\Gamma}_n(f)\right)$$

Consider again the experiment $\mathcal{E}_{1, n}^\#$ of (4.7) where two independent stationary series $y_1^{(m)}$ and $y_2^{(m)}$ of length $m = (n - r_n)/2$ are observed. In modified notation we now write

$$\mathcal{E}_{1, n}^\# = \mathcal{E}_m \otimes \mathcal{E}_m = (P_{f, m} \otimes P_{f, m}, f \in \Sigma).$$

We shall compare this with the experiments

$$\begin{aligned} \mathcal{E}_{2, n}^\# &: = \mathcal{E}_m \otimes \tilde{\mathcal{E}}_m = \left(P_{f, m} \otimes \tilde{P}_{f, m}, f \in \Sigma\right), \\ \mathcal{E}_{3, n}^\# &: = \tilde{\mathcal{E}}_m \otimes \tilde{\mathcal{E}}_m = \left(\tilde{P}_{f, m} \otimes \tilde{P}_{f, m}, f \in \Sigma\right). \end{aligned}$$

At this point select the shrinking rate κ_n of the neighborhoods $\Sigma_n(f_0)$ (cp. (4.1)) as

$$\kappa_n = n^{-\gamma}, \quad \gamma = \frac{\alpha - 1/2}{2(2\alpha + 1)} \tag{4.12}$$

Proposition 4.7 We have $\mathcal{E}_{2,n}^\# \lesssim \mathcal{E}_{1,n}^\#$.

Proof. We shall construct a sequence of Markov kernels M_n such that

$$\sup_{f \in \Sigma} H^2 \left(P_{f,m} \otimes \tilde{P}_{f,m}, M_n (P_{f,m} \otimes P_{f,m}) \right) \rightarrow 0.$$

Define M_n as follows: given $y_1^{(m)}$ and $y_2^{(m)}$, and A , a measurable subset of \mathbb{R}^{2m} , set

$$M_n \left(A, y_1^{(m)}, y_2^{(m)} \right) = \mathbf{1}_A \left(y_1^{(m)}, K_m(\hat{f}_m(y_1^{(m)}))y_2^{(m)} \right)$$

where $K_m(f)$ is the matrix defined by (4.2), i.e. for $f \in \Sigma$ by

$$K_m(f) = \tilde{\Gamma}_m^{1/2}(f)\Gamma_m^{-1/2}(f)$$

and \hat{f}_m is the estimator in \mathcal{E}_m of Proposition 4.6 applied to data $y_1^{(m)}$. Thus the Markov kernel M_n is in fact a deterministic map, i.e. given $y_1^{(m)}, y_2^{(m)}$, it defines a one point measure on \mathbb{R}^{2m} concentrated in $\left(y_1^{(m)}, K_m(\hat{f}_m(y_1^{(m)}))y_2^{(m)} \right)$. Thus the law $M_n (P_{f,m} \otimes P_{f,m})$ is the joint law of $y_1^{(m)}$ and $K_m(\hat{f}_m(y_1^{(m)}))y_2^{(m)}$ under f . The latter we split up into the marginal law of $y_1^{(m)}$, i.e. $P_{f,m}$ and the conditional law of $K_m(\hat{f}_m(y_1^{(m)}))y_2^{(m)}$ given $y_1^{(m)}$; write $P_{f,m}^K|y_1^{(m)}$ for the latter. We have

$$P_{f,m}^K|y_1^{(m)} = N_n \left(0, K\Gamma_m(f)K' \right) \text{ for } K = K_m(\hat{f}_m(y_1^{(m)})).$$

Now clearly

$$H^2 \left(P_{f,m} \otimes \tilde{P}_{f,m}, M_n (P_{f,m} \otimes P_{f,m}) \right) = E_f H^2 \left(\tilde{P}_{f,m}, P_{f,m}^K|y_1^{(m)} \right) \quad (4.13)$$

where E_f is taken wrt $y_1^{(m)}$ under $P_{f,m}$. Define

$$B_{f,m} := \left\{ y \in \mathbb{R}^m : \left\| \hat{f}_m(y) - f \right\|_\infty + \left\| \hat{f}_m(y) - f \right\|_{2,1/2} \leq \kappa_m \right\}.$$

By definition of $\Sigma_m(f_0)$ (cf. (4.1)) we have $f \in \Sigma_m(\hat{f}_m(y))$ if $y \in B_{f,m}$. Thus Lemma 4.1 implies

$$\sup_{y \in B_{f,m}, f \in \Sigma} H^2 \left(\tilde{P}_{f,m}, P_{f,m}^K|y \right) = o(1).$$

Moreover by Proposition 4.6

$$P_{f,m} (B_{f,m}^c) = o(1) \text{ uniformly over } f \in \Sigma. \quad (4.14)$$

Hence

$$\begin{aligned} E_f H^2 \left(\tilde{P}_{f,m}, P_{f,m}^K|y_1^{(m)} \right) &= \int_{B_{f,m}} H^2 \left(\tilde{P}_{f,m}, P_{f,m}^K|y \right) dP_{f,m}(y) + o(1) \\ &= o(1)P_{f,m}(B_{f,m}) + o(1) = o(1) \end{aligned} \quad (4.15)$$

uniformly over $f \in \Sigma$. In conjunction with (4.13) the last relation proves the claim. ■

The next result is entirely analogous if we replace the estimator \hat{f}_m based on data $y^{(m)}$ by the one based on data $\tilde{y}^{(m)}$ and formally reverse the order in the product $P_{f,m} \otimes \tilde{P}_{f,m}$.

Proposition 4.8 We have $\mathcal{E}_{3,n}^\# \lesssim \mathcal{E}_{2,n}^\#$.

Proof. We construct a sequence of Markov kernels \tilde{M}_n such that

$$\sup_{f \in \Sigma} H^2 \left(\tilde{P}_{f,m} \otimes \tilde{P}_{f,m}, \tilde{M}_n \left(P_{f,m} \otimes \tilde{P}_{f,m} \right) \right) \rightarrow 0.$$

Define \tilde{M}_n as follows: given $y_1^{(m)}$ and $\tilde{y}_2^{(m)}$, and A , a measurable subset of \mathbb{R}^{2m} , set

$$\tilde{M}_n \left(A, y_1^{(m)}, \tilde{y}_2^{(m)} \right) = \mathbf{1}_A \left(K_m(\hat{f}_m(\tilde{y}_2^{(m)})) y_1^{(m)}, \tilde{y}_2^{(m)} \right)$$

where \hat{f}_m is the estimator defined in the previous subsection, applied to data $\tilde{y}_2^{(m)}$. Analogously to (4.14) we have

$$\tilde{P}_{f,m} (B_{f,m}^c) = o(1) \text{ uniformly over } f \in \Sigma.$$

A reasoning as in (4.15) completes the proof. ■

For the experiment $\mathcal{E}_{3,n}^\#$ which consists of product measures $\tilde{P}_{f,m} \otimes \tilde{P}_{f,m}$, we can invoke Proposition 3.2, applying the equivalence map given there componentwise (i.e. to independent components $(\tilde{y}_1^{(m)}, \tilde{y}_2^{(m)})$ in $\mathcal{E}_{3,n}^\#$). A summary of the lower informativity bound results so far can thus be given as follows. For $r_n = 2 \lfloor \log(n/2) \rfloor$ define the lower bracket Gaussian scale experiment $\hat{\mathcal{E}}_{l,n}$ by

$$\hat{\mathcal{E}}_{l,n} := \hat{\mathcal{E}}_{(n-r_n)/2} \otimes \hat{\mathcal{E}}_{(n-r_n)/2}. \quad (4.16)$$

Corollary 4.9 Consider experiments \mathcal{E}_n and $\hat{\mathcal{E}}_{l,n}$ given respectively by (1.3) and (4.16), (3.4) with parameter space $\Sigma = \Sigma_{\alpha,M}$ where $M > 0$, $\alpha > 1/2$. Then as $n \rightarrow \infty$

$$\hat{\mathcal{E}}_{l,n} \lesssim \mathcal{E}_n.$$

4.5 Bracketing the Gaussian scale model

The proof of Theorem 1.1 is complete if the lower and upper informativity bounds $\hat{\mathcal{E}}_{l,n}$ and $\hat{\mathcal{E}}_{u,n}$ coincide in an asymptotic sense. Since we already established the relation $\hat{\mathcal{E}}_{l,n} \lesssim \mathcal{E}_n \lesssim \hat{\mathcal{E}}_{u,n}$ (Corollaries 3.3, 4.9), it now suffices to show that $\hat{\mathcal{E}}_{u,n} \lesssim \hat{\mathcal{E}}_{l,n}$. This essentially means that in the special nonparametric regression model $\hat{\mathcal{E}}_n$ of Gaussian scale type, having r_n additional observations does not matter asymptotically. "Additional observations" here refers to an equidistant design of higher grid size. The problem of additional observations for i. i. d. models has been discussed by Le Cam (1974) and Mammen (1986) under parametric assumptions. For nonparametric i. i. d. models, one can use the approximation by Gaussian white noise or Poisson models to bound the influence of additional observations. For simplicity, consider a Gaussian white noise model on $[0, 1]$

$$dZ_t = f(t)dt + n^{-1/2}dW_t, t \in [0, 1], f \in \Sigma$$

with parameter space Σ . Consider this experiment \mathcal{F}_n , say and also \mathcal{F}_{n+r_n} . Multiplying the data by $n^{1/2}$ gives an equivalent experiment

$$dZ_t^* = n^{1/2} f(t) dt + dW_t, \quad t \in [0, 1], \quad f \in \Sigma$$

and the corresponding one for $(n+r_n)^{1/2}$. Now, for given f , the squared Hellinger distance of the two respective measures is bounded by

$$\begin{aligned} & C \left((n+r_n)^{1/2} - n^{1/2} \right)^2 \|f\|_2^2 \\ &= C \frac{r_n^2}{n} (1 + o(1)) \|f\|_2^2 \end{aligned}$$

if $r_n = o(n)$. Thus if $r_n = o(n^{1/2})$ and $\sup_{f \in \Sigma} \|f\|_2 \leq C$ then we have $\mathcal{F}_n \approx \mathcal{F}_{n+r_n}$.

Comparable results can be obtained for nonparametric i. i. d. and regression models if these can be approximated by \mathcal{F}_n . In the present case, conversely, for the nonparametric Gaussian scale regression $\hat{\mathcal{E}}_n$, a result of type $\hat{\mathcal{E}}_n \approx \hat{\mathcal{E}}_{n+r_n}$ is a prerequisite for the Gaussian location (white noise) approximation. Note that for a narrower parameter space, given by a Lipschitz class, the white noise approximation of $\hat{\mathcal{E}}_n$ has been established (cf. Grama and Nussbaum, 1998).

Remark 4.10 The relation

$$\hat{\mathcal{E}}_{l,n} \lesssim \mathcal{E}_n \lesssim \hat{\mathcal{E}}_{u,n} \tag{4.17}$$

has been proved under the technical assumption that n is uneven. If n is even, note first that $\mathcal{E}_{n-1} \lesssim \mathcal{E}_n \lesssim \mathcal{E}_{n+1}$ (omitting one observation from \mathcal{E}_{n+1} and \mathcal{E}_n) and apply (4.17) to obtain

$$\hat{\mathcal{E}}_{l,n-1} \lesssim \mathcal{E}_n \lesssim \hat{\mathcal{E}}_{u,n+1}$$

The relation $\mathcal{E}_{u,n} \lesssim \mathcal{E}_{l,n}$ which will be proved for uneven n in the remainder of this section is easily seen to extend to $\mathcal{E}_{u,n+2} \lesssim \mathcal{E}_{l,n}$. This suffices to establish the main result Theorem 1.1 for general sample size $n \rightarrow \infty$.

4.5.1 First part of the bracketing argument

Denote again $m = (n - r_n)/2$ where $r_n = 2[(\log n)/2] + 1$.

Lemma 4.11 For $\hat{\mathcal{E}}_{l,n} = \hat{\mathcal{E}}_m \otimes \hat{\mathcal{E}}_m$ we have

$$\hat{\mathcal{E}}_m \otimes \hat{\mathcal{E}}_m \approx \hat{\mathcal{E}}_{2m}.$$

Proof. Note that the measures in $\hat{\mathcal{E}}_m \otimes \hat{\mathcal{E}}_m$ are product measures, which can be described, after a rearrangement of components, as

$$Q_{1,m} := \bigotimes_{j=1}^m (N(0, J_{j,m}(f)) \otimes N(0, J_{j,m}(f)))$$

whereas the measures in $\hat{\mathcal{E}}_{2m}$ are

$$Q_{2,m} := \bigotimes_{j=1}^m (N(0, J_{2j-1,2m}(f)) \otimes N(0, J_{2j,2m}(f))).$$

Now Lemma 2.1 yields

$$H^2(Q_{1,m}, Q_{2,m}) \leq C \sum_{j=1}^m \left((J_{2j-1,2m}(f) - J_{j,m}(f))^2 + (J_{2j,2m}(f) - J_{j,m}(f))^2 \right).$$

Define a partition of $(-\pi, \pi)$ into n intervals $W_{j,n}$, $j = 1, \dots, n$ of equal length and for any $f \in L_2(-\pi, \pi)$, let

$$\bar{f}_n = \sum_{j=1}^n J_{j,n}(f) \mathbf{1}_{W_{j,n}} \quad (4.18)$$

be the L_2 -projection of f onto piecewise constant functions wrt the partition. Note that we have

$$\|\bar{f}_{2m} - \bar{f}_m\|_2^2 = \frac{2\pi}{m} \sum_{j=1}^m \left((J_{2j-1,2m}(f) - J_{j,m}(f))^2 + (J_{2j,2m}(f) - J_{j,m}(f))^2 \right)$$

so that

$$H^2(Q_{1,m}, Q_{2,m}) \leq Cm \|\bar{f}_{2m} - \bar{f}_m\|_2^2 \leq Cm \left(\|f - \bar{f}_{2m}\|_2^2 + \|f - \bar{f}_m\|_2^2 \right).$$

The result now follows from

$$\sup_{f \in \Sigma} m \|f - \bar{f}_m\|_2^2 \rightarrow 0. \quad (4.19)$$

which is a consequence of Lemmas 5.3 and 5.5. ■

4.5.2 Second part of the bracketing argument

In view of $\hat{\mathcal{E}}_{2m} = \hat{\mathcal{E}}_{n-r_n}$, our next aim is to show

$$\hat{\mathcal{E}}_{n-r_n} \approx \hat{\mathcal{E}}_n$$

where r_n does not grow too quickly. Previously we defined $r_n = 2[(\log n)/2] + 1$, but we will assume more generally now that $r_n = o(n^{1/2})$.

Consider the gamma density with shape parameter $a > 0$

$$g_a(x) = \frac{1}{\Gamma(a)} x^{a-1} \exp(-x), \quad x \geq 0$$

where $\Gamma(a)$ is the gamma function, and more generally the density with additional scale parameter $s > 0$

$$g_{a,s}(x) = \frac{1}{\Gamma(a)} s^{-a} x^{a-1} \exp(-xs^{-1}), \quad x \geq 0.$$

We will call the respective law the $\Gamma(a, s)$ law. Clearly if $X \sim \Gamma(a, 1)$ then $sX \sim \Gamma(a, s)$. It is well known that $\Gamma(n/2, 2) = \chi_n^2$ and that the following result holds. Assume $X \sim \Gamma(a, s)$ and $Y \sim \Gamma(b, s)$; then $X+Y$, $X/(X+Y)$ are independent random variables, and $X+Y \sim \Gamma(a+b, s)$ while $X/(X+Y)$ has a Beta(a, b) distribution (Bickel and Doksum (2001), Theorem B.2.3, p. 489).

Furthermore, for fixed $a > 0$ consider the family of laws

$$(\Gamma(a, s), s > 0). \quad (4.20)$$

Clearly this is a one parameter exponential family; the shape of this exponential family implies that in a product family

$$(\Gamma^{\otimes n}(a, s), s > 0)$$

with n i.i.d. observations X_1, \dots, X_n , the sum $\sum_{i=1}^n X_i$ is a sufficient statistic. This sufficient statistic has law $\Gamma(na, s)$; hence for any subset $S \subset (0, \infty)$ we have the equivalence of experiments

$$(\Gamma^{\otimes n}(a, s), s \in S) \sim (\Gamma(na, s), s \in S). \quad (4.21)$$

Lemma 4.12 *For all $a > 0$ and for $s, t > 0$*

$$H^2(\Gamma(a, s), \Gamma(a, t)) = 2 \left(1 - \left(1 - \frac{(s^{1/2} - t^{1/2})^2}{s + t} \right)^a \right).$$

Proof. We have

$$H^2(\Gamma(a, s), \Gamma(a, t)) = 2 \left(1 - \int g_{a,s}^{1/2}(x) g_{a,t}^{1/2}(x) dx \right),$$

$$\int g_{a,s}^{1/2}(x) g_{a,t}^{1/2}(x) dx = \frac{1}{\Gamma(a)} \int_0^\infty x^{a-1} s^{-a/2} t^{-a/2} \exp\left(-x \left(\frac{1}{2s} + \frac{1}{2t}\right)\right) dx.$$

With a substitution $u = x \left(\frac{1}{2s} + \frac{1}{2t}\right)$ this becomes

$$\begin{aligned} & \frac{1}{\Gamma(a)} \int_0^\infty \left(\frac{2s^{1/2}t^{1/2}}{s+t}\right)^a u^{a-1} \exp(-u) du \\ &= \left(\frac{2s^{1/2}t^{1/2}}{s+t}\right)^a = \left(1 - \frac{(s^{1/2} - t^{1/2})^2}{s+t}\right)^a. \end{aligned}$$

■

Lemma 4.13 *We have, for all $s > 0$ and $a, b > 0$*

$$H^2(\Gamma(a, s), \Gamma(b, s)) = 2 \left(1 - \frac{\Gamma((a+b)/2)}{(\Gamma(a)\Gamma(b))^{1/2}} \right).$$

Proof. In this case

$$\begin{aligned} \int g_{a,s}^{1/2}(x)g_{b,s}^{1/2}(x)dx &= \frac{1}{\Gamma^{1/2}(a)\Gamma^{1/2}(b)} \int_0^\infty x^{(a+b)/2-1} s^{-(a+b)/2} \exp(-xs^{-1}) dx \\ &= \frac{\Gamma((a+b)/2)}{\Gamma^{1/2}(a)\Gamma^{1/2}(b)}. \end{aligned}$$

■

In $\hat{\mathcal{E}}_n$ we observe (cp. (3.4))

$$z_j = J_{j,n}^{1/2}(f)\xi_j, \quad j = 1, \dots, n$$

for independent standard normals ξ_j , which by sufficiency is equivalent to observing $z_j^2 = J_{j,n}(f)\xi_j^2$. Thus $\hat{\mathcal{E}}_n$ is equivalent to

$$\hat{\mathcal{E}}_{n,1} := \left(\bigotimes_{j=1}^n \Gamma(1/2, 2J_{j,n}(f)), f \in \Sigma \right). \quad (4.22)$$

Set again $m = n - r_n$. The above experiment in turn is equivalent, by the sufficiency argument for the scaled gamma law invoked in (4.21), to

$$\hat{\mathcal{E}}_{n,m} := \left(\bigotimes_{j=1}^n \Gamma^{\otimes m}(1/2m, 2J_{j,n}(f)), f \in \Sigma \right).$$

Analogously we have

$$\hat{\mathcal{E}}_m \sim \hat{\mathcal{E}}_{m,1} \sim \hat{\mathcal{E}}_{m,n} := \left(\bigotimes_{j=1}^m \Gamma^{\otimes n}(1/2n, 2J_{j,m}(f)), f \in \Sigma \right). \quad (4.23)$$

Introduce an intermediate experiment

$$\hat{\mathcal{E}}_{m,n}^* := \left(\bigotimes_{j=1}^m \Gamma^{\otimes n}(1/2m, 2J_{j,m}(f)), f \in \Sigma \right)$$

Lemma 4.14 *We have the total variation asymptotic equivalence*

$$\hat{\mathcal{E}}_{m,n}^* \simeq \hat{\mathcal{E}}_{n,m} \text{ as } n \rightarrow \infty.$$

Proof. Write the measures in $\hat{\mathcal{E}}_{n,m}$ as a product of mn components, i.e. as $\bigotimes_{i=1}^{mn} Q_{1,i}$ where the component measures $Q_{1,i}$ are defined as follows. For every $i = 1, \dots, mn$, let $j(1, i)$ be the unique index $j \in \{1, \dots, n\}$ such that there exists $k \in \{1, \dots, m\}$ for which $i = (j-1)m + k$. Then

$$Q_{1,i} := \Gamma(1/2m, 2J_{j(1,i),n}(f)), \quad i = 1, \dots, mn.$$

Analogously, let $j(2, i)$ be the unique index $j \in \{1, \dots, m\}$ such that there exists $k \in \{1, \dots, n\}$ for which $i = (j-1)n + k$. Then the measures in $\hat{\mathcal{E}}_{m,n}^*$ are a product of mn components, i. e. are $\bigotimes_{i=1}^{mn} Q_{2,i}$ where

$$Q_{2,i} = \Gamma(1/2m, 2J_{j(2,i),m}(f)), \quad i = 1, \dots, mn.$$

Then the Hellinger distance between measures in $\hat{\mathcal{E}}_{n,m}$ and $\hat{\mathcal{E}}_{m,n}^*$ is, using Lemma 2.19 in Strasser (1985) and then Lemma 4.12

$$\begin{aligned} H^2 \left(\bigotimes_{i=1}^{mn} Q_{1,i}, \bigotimes_{i=1}^{mn} Q_{2,i} \right) &\leq 2 \sum_{i=1}^{mn} H^2(Q_{1,i}, Q_{2,i}) \\ &= 4 \sum_{i=1}^{mn} \left(1 - \left(1 - \frac{\left(J_{j(1,i),n}^{1/2}(f) - J_{j(2,i),m}^{1/2}(f) \right)^2}{J_{j(1,i),n}(f) + J_{j(2,i),m}(f)} \right)^{1/2m} \right). \end{aligned} \quad (4.24)$$

By using the inequality

$$\frac{(s^{1/2} - t^{1/2})^2}{s + t} = \frac{(s - t)^2}{(s + t)(s^{1/2} + t^{1/2})^2} \leq \frac{(s - t)^2}{s^2}$$

and observing that for $f \in \Sigma$, we have $J_{j,n}(f) \geq M^{-1}$, we obtain an upper bound for (4.24)

$$4 \sum_{i=1}^{mn} \left(1 - \left(1 - M^2 (J_{j(1,i),n}(f) - J_{j(2,i),m}(f))^2 \right)^{1/2m} \right). \quad (4.25)$$

The expression $J_{j(1,i),n}(f) - J_{j(2,i),m}(f)$ can be described as follows. For any $x \in (\frac{i-1}{mn}, \frac{i}{mn})$, $i = 1, \dots, mn$ we have

$$J_{j(1,i),n}(f) - J_{j(2,i),m}(f) = \bar{f}_n(x) - \bar{f}_m(x). \quad (4.26)$$

where \bar{f}_n is defined by (4.18). Now as a consequence of Lemmas 5.4 and 5.5

$$\sup_{f \in \Sigma} \|\bar{f}_n - \bar{f}_m\|_\infty \leq \sup_{f \in \Sigma} \|f - \bar{f}_n\|_\infty + \sup_{f \in \Sigma} \|f - \bar{f}_m\|_\infty = o(1). \quad (4.27)$$

Note that for $m \rightarrow \infty$ and $z \rightarrow 0$ we have

$$\begin{aligned} (1 - Cz^2)^{1/2m} &= \exp \left(\frac{1}{2m} \log(1 - Cz^2) \right) \\ &= \exp \left(-\frac{1}{2m} (Cz^2 + O(z^4)) \right) = 1 - \frac{1}{2m} (Cz^2 + O(z^4)) + o \left(\frac{z^2}{m} \right). \end{aligned}$$

Thus from (4.25) we obtain in view of (4.27)

$$H^2 \left(\bigotimes_{i=1}^{mn} Q_{1,i}, \bigotimes_{i=1}^{mn} Q_{2,i} \right) \leq C \sum_{i=1}^{mn} \frac{1}{m} (J_{j(1,i),n}(f) - J_{j(2,i),m}(f))^2 (1 + o(1)).$$

As a consequence of (4.26) we obtain

$$\|\bar{f}_n - \bar{f}_m\|_2^2 = \sum_{i=1}^{mn} \frac{1}{mn} (J_{j(1,i),n}(f) - J_{j(2,i),m}(f))^2$$

which implies

$$H^2 \left(\bigotimes_{i=1}^{mn} Q_{1,i}, \bigotimes_{i=1}^{mn} Q_{2,i} \right) \leq Cn \|\bar{f}_n - \bar{f}_m\|_2^2 \leq Cn \|f - \bar{f}_m\|_2^2 + Cn \|f - \bar{f}_n\|_2^2.$$

Now as in (4.19) this upper bound is $o(1)$ uniformly over $f \in \Sigma$. ■

Lemma 4.15 *We have the asymptotic equivalence*

$$\hat{\mathcal{E}}_{m,n}^* \simeq \hat{\mathcal{E}}_{m,n} \text{ as } n \rightarrow \infty.$$

Proof. We know (cf. (4.23), (4.22)) that $\hat{\mathcal{E}}_{m,n} \sim \hat{\mathcal{E}}_{m,1}$ where

$$\hat{\mathcal{E}}_{m,1} = \left(\bigotimes_{j=1}^m \Gamma(1/2, 2J_{j,m}(f)), f \in \Sigma \right).$$

Analogously, using (4.21) again, we obtain

$$\hat{\mathcal{E}}_{m,n}^* \sim \bar{\mathcal{E}}_{m,1}^* := \left(\bigotimes_{j=1}^m \Gamma(n/2m, 2J_{j,m}(f)), f \in \Sigma \right)$$

For given $f \in \Sigma$, the Hellinger distance between the two respective product measures is bounded by (using Lemma 2.19 in Strasser (1985) and then Lemma 4.13)

$$2 \sum_{j=1}^m H^2(\Gamma(1/2, 2J_{j,m}(f)), \Gamma(n/2m, 2J_{j,m}(f))) = 4 \sum_{j=1}^m \left(1 - \frac{\Gamma(1/4 + n/4m)}{(\Gamma(1/2)\Gamma(n/2m))^{1/2}} \right).$$

Note that this bound does not depend on $f \in \Sigma$. Write $n/m = 1 + \delta$ where $\delta = r_n/m$; the above is

$$4 \sum_{j=1}^m \frac{(\Gamma(1/2)\Gamma(1/2 + \delta/2))^{1/2} - \Gamma(1/2 + \delta/4)}{(\Gamma(1/2)\Gamma(1/2 + \delta/2))^{1/2}}. \quad (4.28)$$

The Gamma function is infinitely differentiable on $(0, \infty)$; by a Taylor expansion we obtain

$$\begin{aligned} \Gamma(1/2 + \delta/4) &= \Gamma(1/2) + \Gamma'(1/2)\frac{\delta}{4} + O(\delta^2), \\ \Gamma^{1/2}(1/2 + \delta/2) &= \Gamma^{1/2}(1/2) + \frac{1}{2}\Gamma^{-1/2}(1/2)\Gamma'(1/2)\frac{\delta}{2} + O(\delta^2). \end{aligned}$$

Consequently

$$(\Gamma(1/2)\Gamma(1/2 + \delta/2))^{1/2} - \Gamma(1/2 + \delta/4) = O(\delta^2)$$

so that (4.28) becomes

$$\sum_{j=1}^m \frac{O(\delta^2)}{\Gamma(1/2)(1 + o(1))} = m\delta^2 O(1) \leq \frac{r_n^2}{m} O(1).$$

The condition $r_n = o(n^{1/2})$ now implies that this upper bound is $o(1)$. We thus established total variation asymptotic equivalence $\hat{\mathcal{E}}_{m,1} \simeq \bar{\mathcal{E}}_{m,1}^*$. ■

5 Appendix: auxiliary statements and analytic facts

5.1 Proof of Lemma 2.1

Consider the spectral decompositions of A and B :

$$A = C_1 \Lambda_1 C_1', \quad B = C_2 \Lambda_2 C_2'$$

where Λ_i are $n \times n$ diagonal matrices and C_i are orthogonal matrices. Recall the simultaneous diagonalization of A and B : setting $D = \Lambda_1^{-1/2} C_1'$, we obtain

$$DAD' = I_n, \quad \tilde{B} := DBD'$$

and letting $\tilde{B} = \tilde{C} \tilde{\Lambda} \tilde{C}'$ be the spectral decomposition of \tilde{B} , we obtain with $\tilde{D} := \tilde{C}' D$

$$\tilde{D}A\tilde{D}' = I_n, \quad \tilde{D}B\tilde{D}' = \tilde{\Lambda}.$$

We now claim that

$$\left\| I_n - \tilde{\Lambda} \right\|^2 \leq M^2 \|A - B\|^2. \quad (5.1)$$

Indeed we have

$$\begin{aligned} \left\| I_n - \tilde{\Lambda} \right\|^2 &= \text{tr} \left[(I_n - \tilde{\Lambda})(I_n - \tilde{\Lambda}) \right] \\ &= \text{tr} \left[\left(\tilde{D}A\tilde{D}' - \tilde{D}B\tilde{D}' \right) \left(\tilde{D}A\tilde{D}' - \tilde{D}B\tilde{D}' \right) \right] \\ &= \text{tr} \left[\tilde{D}'\tilde{D}(A - B)\tilde{D}'\tilde{D}(A - B) \right] \end{aligned}$$

Now for eigenvalues $\lambda_{\max}(\cdot)$ we have

$$\begin{aligned} \lambda_{\max}(\tilde{D}'\tilde{D}) &= \lambda_{\max}(\tilde{D}\tilde{D}') = \lambda_{\max}(\tilde{C}'DD'\tilde{C}) = \lambda_{\max}(DD') \\ &= \lambda_{\max}(C_1\Lambda_1^{-1/2}\Lambda_1^{-1/2}C_1') = \lambda_{\min}^{-1}(A) \leq M, \end{aligned} \quad (5.2)$$

hence

$$\begin{aligned} \left\| I_n - \tilde{\Lambda} \right\|^2 &\leq M \text{tr} \left[\tilde{D}'\tilde{D}(A - B)(A - B) \right] \\ &\leq M^2 \text{tr}[(A - B)(A - B)] = M^2 \|A - B\|^2 \end{aligned}$$

so that (5.1) is proved. Similarly to (5.2) we obtain a bound from below

$$\lambda_{\min}(\tilde{D}'\tilde{D}) = \lambda_{\max}^{-1}(A) \geq M^{-1}$$

which yields analogously to (5.1)

$$\left\| I_n - \tilde{\Lambda} \right\|^2 \geq M^{-2} \|A - B\|^2. \quad (5.3)$$

Consider now the Hellinger affinity $A_H(\cdot, \cdot)$ between the one dimensional normals $N(0, 1)$ and $N(0, \sigma^2)$: if φ is the standard normal density then

$$\begin{aligned} A_H(N(0, 1), N(0, \sigma^2)) &= \sigma^{-1/2} \int \varphi^{1/2}(x) \varphi^{1/2}(x\sigma^{-1}) dx \\ &= \left(\frac{2\sigma}{1 + \sigma^2} \right)^{1/2} = \left(1 - \frac{(1 - \sigma)^2}{1 + \sigma^2} \right)^{1/2} \\ &= \left(1 - \frac{(1 - \sigma^2)^2}{(1 + \sigma)^2(1 + \sigma^2)} \right)^{1/2} \end{aligned}$$

Let $h = \sigma^2 - 1$; then as $h \rightarrow 0$

$$\log A_H(N(0, 1), N(0, \sigma^2)) = -\frac{h^2}{16}(1 + o(1)). \quad (5.4)$$

The matrix \tilde{D} is nonsingular, and since the Hellinger distance is invariant under one-to-one transformations,

$$\begin{aligned} H^2(N_n(0, A), N_n(0, B)) &= H^2(N_n(0, I_n), N_n(0, \tilde{\Lambda})) = \\ &= 2 \left(1 - A_H(N_n(0, I_n), N_n(0, \tilde{\Lambda})) \right) = 2 \left(1 - \prod_{i=1}^n A_H(N(0, 1), N(0, \tilde{\lambda}_i)) \right) \\ &= 2 \left(1 - \exp \left(\sum_{i=1}^n \log \left(A_H(N(0, 1), N(0, \tilde{\lambda}_i)) \right) \right) \right) \end{aligned} \quad (5.5)$$

where $\tilde{\lambda}_i$, $i = 1, \dots, n$ are the diagonal elements of $\tilde{\Lambda}$. Let us assume that $\|A - B\| \leq \epsilon \rightarrow 0$ where the dimension n of A, B may vary arbitrarily. Since

$$\sup_{i=1, \dots, n} |1 - \tilde{\lambda}_i|^2 \leq \sum_{i=1}^n (1 - \tilde{\lambda}_i)^2 = \|I_n - \tilde{\Lambda}\|^2$$

we may write, in view of (5.1) and (5.4)

$$\log \left(A_H(N(0, 1), N(0, \tilde{\lambda}_i)) \right) = -\frac{1}{16} (1 - \tilde{\lambda}_i)^2 (1 + \rho_i)$$

where $\sup_{i=1, \dots, n} |\rho_i| \rightarrow 0$ as $\epsilon \rightarrow 0$. Since

$$\left| \sum_{i=1}^n (1 - \tilde{\lambda}_i)^2 \rho_i \right| \leq \left(\sup_{i=1, \dots, n} |\rho_i| \right) \sum_{i=1}^n (1 - \tilde{\lambda}_i)^2$$

we obtain for $\epsilon \rightarrow 0$

$$\begin{aligned} -16 \sum_{i=1}^n \log \left(A_H(N(0, 1), N(0, \tilde{\lambda}_i)) \right) &= \sum_{i=1}^n (1 - \tilde{\lambda}_i)^2 (1 + \rho_i) \\ &= \sum_{i=1}^n (1 - \tilde{\lambda}_i)^2 + o(1) \sum_{i=1}^n (1 - \tilde{\lambda}_i)^2 = \|I_n - \tilde{\Lambda}\|^2 (1 + o(1)) \end{aligned}$$

and as a consequence from (5.5)

$$\begin{aligned} H^2(N_n(0, A), N_n(0, B)) &= 2 \left(1 - \exp \left(-\frac{1}{16} \|I_n - \tilde{\Lambda}\|^2 (1 + o(1)) \right) \right) \\ &= \frac{1}{8} \|I_n - \tilde{\Lambda}\|^2 (1 + o(1)). \end{aligned}$$

In conjunction with (5.1) and (5.3), the last relation proves the lemma.

5.2 An auxiliary result for the proof of Lemma 4.1

Let A, B be two $n \times n$ covariance matrices. Recall that for every covariance matrix A there is a uniquely defined symmetric square root matrix $A^{1/2}$: if $A = D\Lambda D^\top$ is a spectral decomposition (D orthogonal, Λ diagonal) of A then $A^{1/2} = D\Lambda^{1/2}D^\top$.

Lemma 5.1 *Let A, B be two $n \times n$ covariance matrices. Then*

$$\|A^{1/2} - B^{1/2}\| \lambda_{\min}(A^{1/2} + B^{1/2}) \leq \|A - B\|.$$

Proof. Observe that

$$\begin{aligned} (A^{1/2} - B^{1/2})B^{1/2} + A^{1/2}(A^{1/2} - B^{1/2}) &= A - B \\ (A^{1/2} - B^{1/2})A^{1/2} + B^{1/2}(A^{1/2} - B^{1/2}) &= A - B \end{aligned}$$

Add up the two equations and set $S = (A^{1/2} + B^{1/2})$; $D = (A^{1/2} - B^{1/2})$; then

$$DS + SD = 2(A - B). \tag{5.6}$$

Take the squared norm $\|\cdot\|^2$ on both sides and observe

$$\begin{aligned} \|DS + SD\|^2 &= \text{tr}[(DS + SD)(DS + SD)] \\ &= 2\text{tr}[DSSD] + 2\text{tr}[DSDS]. \end{aligned}$$

Clearly we have

$$\begin{aligned} \text{tr}[DSSD] &\geq (\lambda_{\min}(S))^2 \text{tr}[DD] \\ \text{tr}[DSDS] &= \text{tr}\left[S^{1/2}DSDS^{1/2}\right] \geq \lambda_{\min}(S) \text{tr}\left[S^{1/2}DD S^{1/2}\right] \\ &\geq (\lambda_{\min}(S))^2 \text{tr}[DD]. \end{aligned}$$

The last two displays imply

$$\|DS + SD\|^2 \geq 4(\lambda_{\min}(S))^2 \|D\|^2$$

which in conjunction with (5.6) yields

$$\|D\| \lambda_{\min}(S) \leq \|A - B\|.$$

■

5.3 Besov spaces on an interval

Let f be a function defined on $I = [0, 1]$ and for $0 < h < 1$ define

$$\begin{aligned} \|\Delta_h f\|_p^p & : = \int_0^{1-h} |f(x) - f(x+h)|^p dx \text{ for } 1 \leq p < \infty, \\ \|\Delta_h f\|_\infty & = \sup_{0 \leq x \leq 1-h} |f(x) - f(x+h)|. \end{aligned}$$

For $1 \leq p \leq \infty$, the modulus of smoothness is defined as

$$\omega(f, t)_p := \sup_{0 < h \leq t} \|\Delta_h f\|_p.$$

For $0 < \alpha < 1$ and $1 \leq q \leq \infty$ define a Besov type seminorm $|f|_{B_{p,q}^\alpha}$ by

$$\begin{aligned} |f|_{B_{p,q}^\alpha} & : = \left(\int_0^\infty \left(\frac{\omega(f, t)_p}{t^\alpha} \right)^q \frac{dt}{t} \right)^{1/q} \text{ for } 1 \leq q < \infty, \\ |f|_{B_{p,q}^\alpha} & = \sup_{t>1} \omega(f, t)_p \text{ for } q = \infty \end{aligned}$$

and a norm $\|f\|_{B_{p,q}^\alpha}$ by

$$\|f\|_{B_{p,q}^\alpha} := \|f\|_p + |f|_{B_{p,q}^\alpha}.$$

The Besov space $B_{p,q}^\alpha$ (for $1 \leq p < \infty$, $0 < \alpha < 1$) is the set of f where $\|f\|_{B_{p,q}^\alpha} < \infty$, equipped with the norm $\|\cdot\|_{B_{p,q}^\alpha}$. Define also the Hölder norm

$$\|f\|_{C^\alpha} := \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \quad (5.7)$$

and the corresponding Hölder space C^α . For two different spaces, B and B' say, an *embedding theorem* (written $B \hookrightarrow B'$) is a norm inequality

$$\|f\|_{B'} \leq C \|f\|_B$$

where C depends on B', B . Thus the embedding implies the set inclusion $B \subset B'$. We cite the basic embedding theorem for our case, which is obtained by combining Theorems 18.4, 18.5–18.8 in Besov, Il'in and Nikol'skii (1979) with Theorems 3.3.1 and 2.5.7. in Triebel (1983), for the special case of a domain $[0, 1]$.

Proposition 5.2 *Let $0 < \alpha' < \alpha < 1$ and $1 \leq p, q, p', q' \leq \infty$. Then*

i) if $q < q'$ then

$$B_{p,q}^\alpha \hookrightarrow B_{p,q'}^\alpha$$

ii) if $p < p'$ and $\alpha - (1/p - 1/p') > 0$ then for $\alpha' = \alpha - (1/p - 1/p')$

$$B_{p,q}^\alpha \hookrightarrow B_{p',q}^{\alpha'}$$

iii) if $p \geq p'$ then

$$B_{p,q}^\alpha \hookrightarrow B_{p',q'}^{\alpha'}.$$

iv) we have $B_{\infty,\infty}^\alpha = C^\alpha$ in the sense of equivalence of norms:

$$B_{\infty,\infty}^\alpha \hookrightarrow C^\alpha \text{ and } C^\alpha \hookrightarrow B_{\infty,\infty}^\alpha.$$

Approximation by step functions. Consider a partition of $[0, 1]$ into n intervals $W_{j,n}$ $j = 1, \dots, n$ of equal length and for any $f \in L_2(0, 1)$, let \bar{f}_n be the L_2 -projection onto the piecewise constant functions, i.e.

$$\bar{f}_n = \sum_{j=1}^n J_{j,n}(f) \mathbf{1}_{W_{j,n}}, \text{ where } J_{j,n}(f) = n \int_{W_{j,n}} f(x) dx. \quad (5.8)$$

Lemma 5.3 For $0 < \alpha < 1$ and $f \in B_{2,2}^\alpha$ we have

$$\|f - \bar{f}_n\|_2^2 \leq 4 n^{-2\alpha} |f|_{B_{2,2}^\alpha}^2.$$

Proof. Note first

$$\|f - \bar{f}_n\|_2^2 = \sum_{j=1}^n \int_{W_{j,n}} (f(x) - J_{j,n}(f))^2 dx. \quad (5.9)$$

For any interval (a, b) and $\varepsilon = b - a$ we have

$$\begin{aligned} \int_a^b \left(f(x) - \varepsilon^{-1} \int_a^b f(u) du \right)^2 dx &= \int_a^b \left(\varepsilon^{-1} \int_a^b (f(x) - f(u)) du \right)^2 dx \\ \text{(by Jensen's inequality)} &\leq \int_a^b \varepsilon^{-1} \int_a^b (f(x) - f(u))^2 dudx \\ &= 2\varepsilon^{-1} \int_a^b \int_x^b (f(x) - f(u))^2 dudx. \end{aligned}$$

With a change of variable $h = u - x$ the above equals

$$\begin{aligned} &2\varepsilon^{-1} \int_a^b \int_0^{b-x} (f(x) - f(x+h))^2 dh dx \\ &\leq 2\varepsilon^{-1} \int_a^b \int_0^{b-x} (b-x)^{2\alpha+1} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dh dx \\ &\leq 2\varepsilon^{2\alpha} \int_a^b \int_0^{b-x} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dh dx. \end{aligned}$$

Setting now $(a, b) = W_{j,n}$, $b = j/n$ and $\varepsilon = n^{-1}$ we obtain for $j = 1, \dots, n-1$

$$\begin{aligned} \int_{W_{j,n}} (f(x) - J_{j,n}(f))^2 dx &\leq 2n^{-2\alpha} \int_{W_{j,n}} \int_0^{j/n-x} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dh dx \\ &\leq 2n^{-2\alpha} \int_{W_{j,n}} \int_0^{1/n} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dh dx \end{aligned} \quad (5.10)$$

whereas for $j = n$ we have only the bound (5.10), i.e

$$\int_{W_{n,n}} (f(x) - J_{n,n}(f))^2 dx \leq 2n^{-2\alpha} \int_{W_{n,n}} \int_0^{1-x} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dh dx.$$

Hence from (5.9) by adding the upper bounds

$$\begin{aligned} \|f - \bar{f}_n\|_2^2 &= \sum_{j=1}^{n-1} \int_{W_{j,n}} (f(x) - J_{j,n}(f))^2 dx + \int_{W_{n,n}} (f(x) - J_{n,n}(f))^2 dx, \\ \sum_{j=1}^{n-1} \int_{W_{j,n}} (f(x) - J_{j,n}(f))^2 dx &\leq 2n^{-2\alpha} \int_0^{1/n} \int_0^{1-1/n} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dx dh \\ &\leq 2n^{-2\alpha} \int_0^{1/n} \|\Delta_h f\|_2^2 h^{-(2\alpha+1)} dh \end{aligned}$$

and

$$\int_{W_{n,n}} (f(x) - J_{n,n}(f))^2 dx \leq 2n^{-2\alpha} \int_{1-1/n}^1 \int_0^{1-x} \frac{(f(x) - f(x+h))^2}{h^{2\alpha+1}} dh dx. \quad (5.11)$$

Now set

$$\begin{aligned} g(x, h) &= (f(x) - f(x+h))^2 h^{-(2\alpha+1)}, \\ A &= \{(x, h) : 0 \leq h \leq 1/n, 0 \leq x \leq 1-h\}. \end{aligned}$$

Then

$$\int_0^{1/n} \|\Delta_h f\|_2^2 h^{-(2\alpha+1)} dh = \int_A g(x, h) d(x, h)$$

and the second integral in (5.11) can be written in the same way but over a domain

$$A^* = \{(x, h) : 0 \leq h \leq 1-x, 1-1/n \leq x \leq 1\}.$$

Since $A^* \subset A$ and $g(x, h) \geq 0$, we obtain

$$\begin{aligned} \|f - \bar{f}_n\|_2^2 &\leq 2n^{-2\alpha} \int_A g(x, h) d(x, h) + 2n^{-2\alpha} \int_{A^*} g(x, h) d(x, h) \\ &\leq 4n^{-2\alpha} \int_0^{1/n} \|\Delta_h f\|_2^2 h^{-(2\alpha+1)} dh \leq 4n^{-2\alpha} \|f\|_{B_{2,2}^\alpha}^2. \end{aligned}$$

■

Lemma 5.4 For $1/2 < \alpha < 1$ and $f \in B_{2,2}^\alpha$ we have

$$\|f - \bar{f}_n\|_\infty \leq C_\alpha n^{1/2-\alpha} \|f\|_{B_{2,2}^\alpha}.$$

Proof. For $0 < \beta < 1$, consider the Hölder space C^β with norm $\|f\|_{C^\beta}$ (cf. (5.7)). For $f \in C^\beta$, the result

$$\|f - \bar{f}_n\|_\infty \leq n^{-\beta} \|f\|_{C^\beta}$$

is immediate. By Proposition 5.2 (ii),(i) and (iv), we have the embeddings

$$B_{2,2}^\alpha \hookrightarrow B_{\infty,2}^{\alpha-1/2} \hookrightarrow B_{\infty,\infty}^{\alpha-1/2} \hookrightarrow C^{a-1/2}. \quad (5.12)$$

Setting $\beta = a - 1/2$, we obtain the result. ■

Periodic spaces. For any $f \in L_2(0, 1)$ and $0 < \alpha < 1$, let $\|f\|_{2,\alpha}$ be the norm defined in terms of Fourier coefficients analogous to (2.14), i.e.

$$\|f\|_{2,\alpha}^2 := \gamma_f^2(0) + \sum_{j=-\infty}^{\infty} |j|^{2\alpha} \gamma_f^2(j), \text{ where } \gamma_f(j) = \int_0^1 \exp(2\pi i j x) f(x) dx.$$

Let W^α be the set of f where $\|f\|_{2,\alpha} < \infty$ equipped with this norm. This is the periodic version of the Besov-Sobolev space $B_{2,2}^\alpha$ (thus a standard notation for W^α would be $\tilde{B}_{2,2}^\alpha$); we will prove one part of this claim via the embedding below. For a more comprehensive treatment cf. Triebel (1983), Theorem 9.2.1.

Lemma 5.5 For $0 < \alpha < 1$ we have

$$W^\alpha \hookrightarrow B_{2,2}^\alpha.$$

Proof. We will first establish the inequality

$$\|f\|_{B_{p,q}^\alpha}^2 \leq C_\alpha \left(\sum_{n=1}^{\infty} n^{2\alpha-1} \omega^2(f, n^{-1})_2 + \|f\|_2^2 \right). \quad (5.13)$$

To this end, note that for $h \geq 1$ we have $\omega^2(f, h)_2 = \omega^2(f, 1)_2$ and therefore, by integrating over intervals $((n+1)^{-1}, n^{-1})$

$$\int_0^\infty \left(\frac{\omega(f, t)_2^2}{t^{2\alpha}} \right) \frac{dt}{t} \leq \sum_{n=1}^{\infty} (n+1)^{2\alpha+1} \omega^2(f, n^{-1})_2 \left(\frac{1}{n(n+1)} \right) + \int_1^\infty \left(\frac{\omega(f, 1)_2^2}{t^{2\alpha}} \right) \frac{dt}{t}$$

which in view of $\omega(f, 1)_2^2 \leq 4 \|f\|_2^2$ gives a bound

$$\|f\|_{B_{2,2}^\alpha}^2 \leq 2^{2\alpha} \sum_{n=1}^{\infty} n^{2\alpha-1} \omega^2(f, n^{-1})_2 + 4 \|f\|_2^2 \int_1^\infty t^{-(2\alpha+1)} dt.$$

and thus (5.13). Define a periodic version of $\|\Delta_h f\|_2^2$ by first extending the function f outside $[0, 1]$ periodically, and then setting

$$\|\tilde{\Delta}_h f\|_2^2 := \int_0^1 |f(x) - f(x+h)|^2 dx.$$

The periodic modulus of smoothness is then

$$\tilde{\omega}(f, t)_2 := \sup_{0 < h \leq t} \|\tilde{\Delta}_h f\|_2. \quad (5.14)$$

Evidently we have $\omega(f, t)_2 \leq \tilde{\omega}(f, t)_2$ for all $t > 0$. Now

$$\begin{aligned} \|\tilde{\Delta}_h f\|_2^2 &= \sum_{j=-\infty}^{\infty} \gamma_f^2(j) |1 - \exp(ijh)|^2 = 4 \sum_{j=-\infty}^{\infty} \gamma_f^2(j) \sin^2\left(\frac{jh}{2}\right) \\ &\leq \sum_{|j| \leq n} \gamma_f^2(j) j^2 h^2 + 4 \sum_{|j| > n} \gamma_f^2(j). \end{aligned}$$

Consequently

$$\begin{aligned}
& \sum_{n=1}^{\infty} n^{2\alpha-1} \omega^2(f, n^{-1})_2 \leq \sum_{n=1}^{\infty} n^{2\alpha-1} \tilde{\omega}^2(f, n^{-1})_2 \\
& \leq \sum_{n=1}^{\infty} n^{2\alpha-1} \sum_{|j| \leq n} \gamma_f^2(j) j^2 n^{-2} + 4 \sum_{n=1}^{\infty} n^{2\alpha-1} \sum_{|j| > n} \gamma_f^2(j) \\
& = \sum_{j=-\infty}^{\infty} \gamma_f^2(j) j^2 \sum_{n \geq |j|} n^{2\alpha-3} + 4 \sum_{j=-\infty}^{\infty} \gamma_f^2(j) \sum_{n < |j|} n^{2\alpha-1} \\
& \leq C_\alpha \sum_{j=-\infty}^{\infty} \gamma_f^2(j) |j|^{2\alpha} + C_\alpha \sum_{j=-\infty}^{\infty} \gamma_f^2(j) |j|^{2\alpha} \leq C_\alpha \|f\|_{2,\alpha}^2
\end{aligned}$$

which in conjunction with (5.13) proves the claim. ■

For any $f \in L_2(0, 1)$ (real-valued) and uneven n , let

$$\tilde{f}_n(x) = \sum_{|j| \leq (n-1)/2} \gamma_f(j) \exp(2\pi i j x)$$

be its truncated Fourier series. The letter C denotes generic constants depending on α but not on f .

Lemma 5.6 *For $1/2 < \alpha < 1$ we have*

$$\begin{aligned}
W^\alpha & \hookrightarrow C^{\alpha-1/2}, \\
\|f - \tilde{f}_n\|_\infty & \leq C n^{1/2-\alpha} \|f\|_{2,\alpha}.
\end{aligned}$$

Proof. The first relation follows from Lemma 5.5 and the embedding (5.12). The second then follows from the Cauchy-Schwartz inequality via

$$\|f - \tilde{f}_n\|_\infty^2 \leq \left(\sum_{|j| > (n-1)/2} j^{2\alpha} \gamma_f^2(j) \right) \left(\sum_{|j| > (n-1)/2} j^{-2\alpha} \right) \leq \|f\|_{2,\alpha}^2 C n^{1-2\alpha}.$$

■

Let $\omega_{j,n}$ be the midpoint of the interval $W_{j,n}$ $j = 1, \dots, n$ (cf. (5.8)).

Lemma 5.7 *For $1/2 < \alpha < 1$ and $f \in \tilde{B}_{2,2}^\alpha$ we have along uneven n*

$$\sum_{j=1}^n \left(\tilde{f}_n(\omega_{j,n}) - J_{j,n}(f) \right)^2 \leq C_\alpha n^{1-2\alpha} \|f\|_{2,\alpha}^2.$$

Proof. Note that

$$\sum_{j=1}^n \left(\tilde{f}_n(\omega_{j,n}) - J_{j,n}(f) \right)^2 \leq$$

$$\leq 2 \sum_{j=1}^n \left(J_{j,n}(\tilde{f}_n) - J_{j,n}(f) \right)^2 + 2 \sum_{j=1}^n \left(\tilde{f}_n(\omega_{j,n}) - J_{j,n}(\tilde{f}_n) \right)^2.$$

Here by Parseval's relation and the projection property of \bar{f}_n the first term is bounded by

$$2n \left\| f - \tilde{f}_n \right\|_2^2 \leq 2n \left(\frac{2}{n-1} \right)^{2\alpha} \|f\|_{2,\alpha}^2 \leq C_\alpha n^{1-2\alpha} \|f\|_{2,\alpha}^2.$$

Thus it remains to show that

$$\sum_{j=1}^n \left(\tilde{f}_n(\omega_{j,n}) - J_{j,n}(\tilde{f}_n) \right)^2 \leq C_\alpha n^{1-2\alpha} \|f\|_{2,\alpha}^2. \quad (5.15)$$

We have

$$\begin{aligned} \left| \tilde{f}_n(\omega_{j,n}) - J_{j,n}(\tilde{f}_n) \right| &\leq n \int_{W_{j,n}} \left| \tilde{f}_n(x) - \tilde{f}_n(\omega_{j,n}) \right| dx \leq n \int_{W_{j,n}} \left| \int_x^{\omega_{j,n}} D\tilde{f}_n(t) dt \right| dx \\ &\leq n \int_{W_{j,n}} |x - \omega_{j,n}|^{1/2} \left(\int_x^{\omega_{j,n}} (D\tilde{f}_n(t))^2 dt \right)^{1/2} dx \\ &\leq n^{-1/2} \left(\int_{W_{j,n}} (D\tilde{f}_n(t))^2 dt \right)^{1/2}. \end{aligned}$$

Consequently

$$\sum_{j=1}^n \left(\tilde{f}_n(\omega_{j,n}) - J_{j,n}(\tilde{f}_n) \right)^2 \leq n^{-1} \int_{[0,1]} (D\tilde{f}_n(t))^2 dt.$$

By termwise differentiation and Parseval's relation the right side equals

$$\begin{aligned} n^{-1} \sum_{|j| \leq (n-1)/2} j^2 \gamma_f^2(j) &= \sum_{|j| \leq (n-1)/2} \left(\frac{|j|^{2-2\alpha}}{n} \right) |j|^{2\alpha} \gamma_f^2(j) \\ &\leq n^{1-2\alpha} \|f\|_{2,\alpha}^2 \end{aligned}$$

which establishes (5.15). ■

Remark 5.8 *Periodic Besov spaces* $\tilde{B}_{p,q}^\alpha$. These can be defined for $0 < \alpha < 1$ and $1 \leq p, q \leq \infty$ analogously to the spaces $B_{p,q}^\alpha$ as above, using an periodic increment norm $\left\| \tilde{\Delta}_h f \right\|_p$ and a periodic modulus of smoothness $\tilde{\omega}(f, t)_p$ defined analogously to (5.14). Clearly then $\tilde{B}_{p,q}^\alpha \hookrightarrow B_{p,q}^\alpha$. An intrinsic characterization in terms of Fourier coefficients $\gamma_f(k)$ is as follows: for $p \geq 2$ and $1/2 < \alpha < 1$ the expression

$$\left(\sum_{j=0}^{\infty} 2^{j\alpha q} \left\| \sum_{2^{j-1}-1 < |k| < 2^j} \gamma_f(k) \exp(2\pi i k x) \right\|_p^q \right)^{1/q}$$

is an equivalent norm in $\tilde{B}_{p,q}^\alpha$; cf. Nikolskii, sec. 5.6, relation (6) (cp. also Triebel (1983), definition 2.3.1/2).

References

- [1] Bickel, P. and Doksum, K. (2001). *Mathematical Statistics, Vol 1 (2nd ed.)*. Prentice Hall, Upper Saddle River, N.J.
- [2] Balakrishnan, A.V. (1976). *Applied Functional Analysis*. Springer, New York.
- [3] Besov, O.V., Il'in, V. P. and Nikol'skii, S. M. (1979). *Integral Representation of Functions and Imbedding Theorems, Vol II*. V.H. Winston & Sons, Washington, DC.
- [4] Brockwell, P. J, and Davis, R. A.(1991). *Time Series: Theory and Methods*. Second Edition. Springer-Verlag, New York.
- [5] Brown, L.D. and Low M. (1996). Asymptotic equivalence of non-parametric regression and white noise. *Ann. Statist.* **24** 2384–2398
- [6] Brown, L. D., Low M. and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30** 688-707
- [7] Brown, L.D., Carter, A.V., Low M.G. and Zhang, C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *Ann. Statist.* **32**, 2074-2097
- [8] Carter, A. V. (2002). Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.* **30** 708-730
- [9] Coursol, J. and Dacunha-Castelle, D. (1982). Remarks on the approximation of the likelihood function of a stationary Gaussian process. *Theory Probab. Applic.* **27** (1) 162-167
- [10] Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *Ann. Statist.* **16** 808-841
- [11] Dahlhaus, R. and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *Ann. Statist.* **24** 1934–1963
- [12] Dahlhaus, R and Polonik, W. (2002). Empirical spectral processes and nonparametric maximum likelihood estimation for time series. In: *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch, M. Sørensen, Eds.), Birkhäuser, Boston, 275-298.
- [13] Davies, R.B. (1973). Asymptotic inference in stationary Gaussian time-series, *Adv. Appl. Probab.* **5**, 469–497.
- [14] Delattre, S. and Hoffmann, M. (2002). Asymptotic equivalence for a null recurrent diffusion. *Bernoulli* **8** 139-174
- [15] Dzhaparidze K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer-Verlag, New York Inc.
- [16] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London, New York.

- [17] Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Prob. Theor. Rel. Fields*, **111**, 167-214
- [18] Lahiri, S.N. (2003). A necessary and sufficient condition for asymptotic independence of discrete Fourier transforms under short- and long- range dependence. *Ann. Statist.* **31** 613-641
- [19] Le Cam, L. (1974). On the information contained in additional observations. *Ann. Statist.* **2** 630-649.
- [20] Le Cam, L. and Yang, G. (2000). *Asymptotics in Statistics*. 2nd Edition, Springer-Verlag, New-York.
- [21] Mammen, E. (1986). The statistical information contained in additional observations. *Ann. Statist.* **14** 665-678
- [22] Nikol'skii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems*. Springer-Verlag, Berlin.
- [23] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399–2430.
- [24] Shiryaev, A. N. (1996). *Probability*. 2nd ed. Springer, New York.
- [25] Strasser, H. (1985). *Mathematical Theory of Statistics*. Walter de Gruyter, Berlin.
- [26] Triebel, H. (1983). *Theory of Function Spaces*. Birkhäuser, Basel.
- [27] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- [28] Zhou, H. H. (2004). *Minimax Estimation with Thresholding and Asymptotic Equivalence for Gaussian Variance Regression*. Ph. D. Thesis, Cornell University, Ithaca, NY, <http://www.stat.yale.edu/~hz68/> (Technical Reports).

UNIVERSITÉ DE PROVENCE
 CMI
 39, RUE F.JOLIOT-CURIE
 13453 MARSEILLE, CEDEX 13, FRANCE
 E-MAIL: golubev@gyptis.univ-mrs.fr

DEPARTMENT OF MATHEMATICS
 MALOTT HALL
 CORNELL UNIVERSITY
 ITHACA NY 14853
 E-MAIL: nussbaum@math.cornell.edu

DEPARTMENT OF STATISTICS
 YALE UNIVERSITY
 P.O. BOX 208290
 NEW HAVEN CT 06520
 E-MAIL: huibin.zhou@yale.edu