



**HAL**  
open science

## Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors

Willem Kruijer, Judith Rousseau

► **To cite this version:**

Willem Kruijer, Judith Rousseau. Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors. 2012. hal-00672772

**HAL Id: hal-00672772**

**<https://hal.science/hal-00672772>**

Preprint submitted on 21 Feb 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors.

Willem Kruijer and Judith Rousseau

February 22, 2012

## Abstract

For a Gaussian time series with long-memory behavior, we use the FEXP-model for semi-parametric estimation of the long-memory parameter  $d$ . The true spectral density  $f_o$  is assumed to have long-memory parameter  $d_o$  and a FEXP-expansion of Sobolev-regularity  $\beta > 1$ . We prove that when  $k$  follows a Poisson or geometric prior, or a sieve prior increasing at rate  $n^{\frac{1}{1+2\beta}}$ ,  $d$  converges to  $d_o$  at a suboptimal rate. When the sieve prior increases at rate  $n^{\frac{1}{2\beta}}$  however, the minimax rate is almost obtained. Our results can be seen as a Bayesian equivalent of the result which Moulines and Soulier obtained for some frequentist estimators.

## 1 Introduction

Let  $X_t, t \in \mathbb{Z}$ , be a stationary Gaussian time series with zero mean and spectral density  $f_o(x), x \in [-\pi, \pi]$ , which takes the form

$$|1 - e^{ix}|^{-2d_o} M_o(x), \quad x \in [-\pi, \pi], \quad (1.1)$$

where  $d_o \in (-\frac{1}{2}, \frac{1}{2})$  is called the long-memory parameter, and  $M$  is a slowly-varying bounded function that describes the short-memory behavior of the series. If  $d_o$  is positive, this makes the autocorrelation function  $\rho(h)$  decay polynomially, at rate  $h^{-(1-2d_o)}$ , and the time series is said to have long-memory. When  $d_o = 0$ ,  $X_t$  has short memory, and the case  $d_o < 0$  is referred to as intermediate memory. Long memory time series models are used in a wide range of applications, such as hydrological or financial time series; see for example Beran (1994) or Robinson (1994). In parametric approaches, a finite dimensional model is used for the short memory part  $M_o$ ; the most well known example is the ARFIMA(p,d,q) model. The asymptotic properties of maximum likelihood estimators (Dahlhaus (1989) or Lieberman et al. (2003)) and Bayesian estimators (Philippe and Rousseau (2002)) have been established in such models and these estimators are consistent and asymptotically normal with a convergence rate of order  $\sqrt{n}$ . However when the model for the short memory part is misspecified, the estimator for  $d$  can be inconsistent, calling for semi-parametric

methods for the estimation of  $d$ . A key feature of semi-parametric estimators of the long-memory parameter is that they converge at a rate which depends on the smoothness of the short-memory part, and apart from the case where  $M_o$  is infinitely smooth, the convergence rate is smaller than  $\sqrt{n}$ . The estimation of the long-memory parameter  $d$  can thus be considered as a non-regular semi-parametric problem. In Moulines and Soulier (2003) (p. 274) it is shown that when  $f_o$  satisfies (1.4), the minimax rate for  $d$  is  $n^{-\frac{2\beta-1}{4\beta}}$ . There are frequentist estimators for  $d$  based on the periodogram that achieve this rate (see Hurvich et al. (2002) and Moulines and Soulier (2003)).

Although Bayesian methods in long-memory models have been widely used (see for instance Ko et al. (2009), Jensen (2004) or Holan and McElroy (2010)), the literature on convergence properties of non- and semi-parametric estimators is sparse. Rousseau et al. (2010) (RCL hereafter) obtain consistency and rates for the  $L_2$ -norm of the log-spectral densities (Theorems 3.1 and 3.2), but for  $d$  they only show consistency (Corollary 1). No results exist on the posterior concentration rate on  $d$ , and thus on the convergence rates of Bayesian semi-parametric estimators of  $d$ . In this paper we aim to fill this gap for a specific family of semi-parametric priors.

We study Bayesian estimation of  $d$  within the FEXP-model (Beran (1993), Robinson (1995)), that contains densities of the form

$$f_{d,k,\theta}(x) = |1 - e^{ix}|^{-2d} \exp \left\{ \sum_{j=0}^k \theta_j \cos(jx) \right\}, \quad (1.2)$$

where  $d \in (-\frac{1}{2}, \frac{1}{2})$ ,  $k$  is a nonnegative integer and  $\theta \in \mathbb{R}^{k+1}$ . The factor  $\exp\{\sum_{j=0}^k \theta_j \cos(jx)\}$  models the function  $M_o$  in (1.1). In contrast to the original finite-dimensional FEXP-model (Beran (1993)), where  $k$  was supposed to be known, or at least bounded,  $f_o$  may have an infinite FEXP-expansion, and we allow  $k$  to increase with the number of observations to obtain approximations  $f$  that are increasingly close to  $f_o$ . Note that the case where the true spectral density satisfies  $f_o = f_{d_o, k_o, \theta_o}$ , is considered in Holan and McElroy (2010). In this paper we will pursue a fully Bayesian semi-parametric estimation of  $d$ , the short memory parameter being considered as an infinite-dimensional nuisance parameter. We obtain results on the convergence rate and asymptotic distribution of the posterior distribution for  $d$ , which we summarize below in section 1.2. These are to our knowledge the first of this kind in the Bayesian literature on semi-parametric time series. First we state the most important assumptions.

## 1.1 Asymptotic framework

For observations  $X = (X_1, \dots, X_n)$  from a Gaussian stationary time series with spectral density  $f$ , let  $T_n(f)$  denote the associated covariance matrix and  $l_n(f)$  denote the log-likelihood

$$l_n(f) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \det(T_n(f)) - \frac{1}{2} X^t T_n^{-1}(f) X.$$

We consider semi-parametric priors on  $f$  based on the FEXP-model defined by (1.2), inducing a parametrization of  $f$  in terms of  $(d, k, \theta)$ . Assuming priors  $\pi_d$  for  $d$ , and, independent of  $d$ ,  $\pi_k$  for  $k$  and  $\pi_{\theta|k}$  for  $\theta|k$ , we study the (marginal) posterior for  $d$ , given by

$$\Pi(d \in D|X) = \frac{\sum_{k=0}^{\infty} \pi_k(k) \int_D \int_{\mathbb{R}^{k+1}} e^{l_n(d,k,\theta)} d\pi_{\theta|k}(\theta) d\pi_d(d)}{\sum_{k=0}^{\infty} \pi_k(k) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\mathbb{R}^{k+1}} e^{l_n(d,k,\theta)} d\pi_{\theta|k}(\theta) d\pi_d(d)}. \quad (1.3)$$

The posterior mean or median can be taken as point-estimates for  $d$ , but we will focuss on the posterior  $\Pi(d|X)$  itself.

It is assumed that the true spectral density is of the form

$$f_o(x) = |1 - e^{ix}|^{-2d_o} \exp \left\{ \sum_{j=0}^{\infty} \theta_{o,j} \cos(jx) \right\}, \quad (1.4)$$

$$\theta_o \in \Theta(\beta, L_o) = \left\{ \theta \in l_2(\mathbb{N}) : \sum_{j=0}^{\infty} \theta_j^2 (1+j)^{2\beta} \leq L_o \right\},$$

for some known  $\beta > 1$ .

In particular, we derive bounds on the rate at which  $\Pi(d \in D|X)$  concentrates at  $d_o$ , together with a Bernstein - von -Mises (BVM) property of this distribution. The posterior concentration rate for  $d$  is defined as the fastest sequence  $\alpha_n$  converging to zero such that

$$\Pi(|d - d_o| < K\alpha_n|X) \xrightarrow{P_o} 0, \quad \text{for a given fixed } K. \quad (1.5)$$

## 1.2 Summary of the results

Under the above assumptions we obtain several results for the asymptotic distribution of  $\Pi(d \in D|X)$ . Our first main result (Theorem 2.1) states that under the sieve prior  $k_n \sim (n/\log n)^{1/(2\beta)}$ ,  $\Pi(d \in D|X)$  is asymptotically Gaussian, and we give expressions for the posterior mean and the posterior variance. A consequence (Corollary 2.1) of this result is that the convergence rate for  $d$  under this prior is at least  $\delta_n = (n/\log n)^{-\frac{2\beta-1}{4\beta}}$ , i.e. in (1.5)  $\alpha_n$  is bounded by  $\delta_n$ . Up to a  $\log n$  term, this is the minimax rate.

By our second main result (Theorem 2.2), the rate for  $d$  is suboptimal when  $k$  is given a a Poisson or a Geometric distribution, or a sieve prior  $k'_n \sim (n/\log n)^{\frac{1}{1+2\beta}}$ . More precisely, there exists  $f_o$  such that the posterior concentration rate  $\alpha_n$  is greater than  $n^{-(\beta-1/2)/(2\beta+1)}$ , and thus suboptimal. Consequently, despite having good frequentist properties for the estimation of the spectral density  $f$  itself (see RCL), these priors are much less suitable for the estimation of  $d$ . This is not a unique phenomenon in (Bayesian) semi-parametric estimation and is encountered for instance in the estimation of a linear functional of the signal in white-noise models, see Li and Zhao (2002) or Arbel (2010).

The BVM property means that asymptotically the posterior distribution of  $d$  behaves like  $\alpha_n^{-1}(d - \hat{d}) \sim \mathcal{N}(0, 1)$ , where  $\hat{d}$  is an estimate whose frequentist distribution (associated to the parameter  $d$ ) is  $\mathcal{N}(d_o, \alpha_n^2)$ . We prove such a property on the posterior distribution of  $d$  given  $k = k_n$ . In regular parametric long-memory models, the BVM property has been established by Philippe and Rousseau (2002). It is however much more difficult to establish BVM theorems in infinite dimensional setups, even for independent and identically distributed models; see for instance Freedman (1999), Castillo (2010) and Rivoirard and Rousseau (2010). In particular it has been proved that the BVM property may not be valid, even for reasonable priors. The BVM property is however very useful since it induces a strong connection between frequentist and Bayesian methods. In particular, it implies that Bayesian credible regions are asymptotically also frequentist confidence regions with the same nominal level. In section 2 we discuss this issue in more detail.

### 1.3 Overview of the paper

In section 2, we present three families of priors based on the sieve model defined by (1.2) with either  $k$  increasing at the rate  $(n/\log n)^{1/(2\beta)}$ ,  $k$  increasing at the rate  $(n/\log n)^{1/(2\beta+1)}$  or with random  $k$ . We study the behavior of the posterior distribution of  $d$  in each case and prove that the former leads to optimal frequentist procedures while the latter two lead to suboptimal procedures. In section 3 we give a decomposition of  $\Pi(d \in D|X)$  defined in (1.3), and obtain bounds for the terms in this decomposition in sections 3.2 and 3.3. Using these results we prove Theorems 2.1 and 2.2 in respectively sections 4 and 5. Conclusions are given in section 6. In the appendices we give the proofs of the lemmas in section 3, as well as some additional results on the derivatives of the log-likelihood. The proofs of various technical results can be found in the supplementary material. We conclude this introduction with an overview of the notation.

### 1.4 Notation

The  $m$ -dimensional identity matrix is denoted  $I_m$ . We write  $|A|$  for the Frobenius or Hilbert-Schmidt norm of a matrix  $A$ , i.e.  $|A| = \sqrt{\text{tr}AA^t}$ , where  $A^t$  denotes the transpose of  $A$ . The operator or spectral norm is denoted  $\|A\|^2 = \sup_{\|x\|=1} x^t A^t A x$ . We also use  $\|\cdot\|$  for the Euclidean norm on  $\mathbb{R}^k$  or  $l^2(\mathbb{N})$ . The inner-product is denoted  $|\cdot|$ . We make frequent use of the relations

$$\begin{aligned} |AB| &= |BA| \leq \|A\| \cdot |B|, & \|AB\| &\leq \|A\| \cdot \|B\|, & \|A\| &\leq |A| \leq \sqrt{n}|A|, \\ |\text{tr}(AB)| &= |\text{tr}(BA)| \leq |A| \cdot |B|, & |x^t Ax| &\leq x^t x |A|, \end{aligned} \quad (1.6)$$

see Dahlhaus (1989), p. 1754. For any function  $h \in L_1([-\pi, \pi])$ ,  $T_n(h)$  is the matrix with entries  $\int_{-\pi}^{\pi} e^{i(l-m)x} h(x) dx$ ,  $l, m = 1, \dots, n$ . For example,  $T_n(f)$  is the covariance matrix of observations  $X = (X_1, \dots, X_n)$  from a time series with

spectral density  $f$ . If  $h$  is square integrable on  $[-\pi, \pi]$  we note

$$\|h\|_2 = \int_{-\pi}^{\pi} h^2(x) dx.$$

The norm  $l$  between spectral densities  $f$  and  $g$  is defined as

$$l(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\log f(x) - \log g(x))^2 dx.$$

Unless stated otherwise, all expectations and probabilities are with respect to  $P_o$ , the law associated with the true spectral density  $f_o$ . To avoid ambiguous notation (e.g.  $\theta_0$  versus  $\theta_{0,0}$ ) we write  $\theta_o$  instead of  $\theta_0$ . Related quantities such as  $f_o$  and  $d_o$  are also denoted with the  $o$ -subscript.

The symbols  $o_P$  and  $O_P$  have their usual meaning. We use boldface when they are uniform over a certain parameter range. Given a probability law  $P$ , a family of random variables  $\{W_d\}_{d \in A}$  and a positive sequence  $a_n$ ,  $W_d = \mathbf{o}_P(a_n, A)$  means that

$$P\left(\sup_{d \in A} |W_d|/a_n > \epsilon\right) \rightarrow 0, (n \rightarrow \infty).$$

When the parameter set is clear from the context we simply write  $\mathbf{o}_P(a_n)$ . In a similar fashion, we write  $\mathbf{o}(a_n)$  when the sequence is deterministic. In conjunction with the  $o_P$  and  $O_P$  notation we use the letters  $\delta$  and  $\epsilon$  as follows. When, for some  $\tau > 0$  and a probability  $P$  we write  $Z = O_P(n^{\tau-\epsilon})$ , this means that  $Z = O(n^{\tau+\epsilon})$  for all  $\epsilon > 0$ . When, on the other hand,  $Z = O_P(n^{\tau-\delta})$ , we mean that this is true for *some*  $\delta > 0$ . If the value of  $\delta$  is of importance it is given a name, for example  $\delta_1$  in Lemma 3.4.

The true spectral density of the process is denoted  $f_o$ . We denote  $k$ -dimensional Sobolev-balls by

$$\Theta_k(\beta, L) = \left\{ \theta \in \mathbb{R}^{k+1} : \sum_{j=0}^k \theta_j^2 (1+j)^{2\beta} \leq L \right\} \subset \mathbb{R}^{k+1}. \quad (1.7)$$

For any real number  $x$ , let  $x_+$  denote  $\max(0, x)$ . The number  $r_k$  denotes the sum  $\sum_{j \geq k+1} j^{-2}$ . Let  $\eta$  be the sequence defined by  $\eta_j = -2/j$ ,  $j \geq 1$  and  $\eta_0 = 0$ . For an infinite sequence  $u = (u_j)_{j \geq 0}$ , let  $u_{[k]}$  denote the vector of the first  $k+1$  elements. In particular,  $\eta_{[k]} = (\eta_0, \dots, \eta_k)$ . The letter  $C$  denotes any generic constant independent of  $L_o$  and  $L$ , which are the constants appearing in the assumptions on  $f_o$  and the definition of the prior.

## 2 Main results

Before stating Theorems 2.1 and 2.2 in section 2.3, we state the assumptions on  $f_o$  and the prior, and give examples of priors satisfying these assumptions.

## 2.1 Assumptions on the prior and the true spectral density

We assume observations  $X = (X_1, \dots, X_n)$  from a stationary Gaussian time series with law  $P_o$ , which is a zero mean Gaussian distribution, whose covariance structure is defined by a spectral density  $f_o$  satisfying (1.4), for known  $\beta > 1$ . It is assumed that for a small constant  $t > 0$ ,  $d_o \in [-\frac{1}{2} + t, \frac{1}{2} - t]$ .

**Assumptions on  $\Pi$ .** We consider different priors, and first state the assumptions that are common to all these priors. The prior on the space of spectral densities consists of independent priors  $\pi_d$ ,  $\pi_k$  and, conditional on  $k$ ,  $\pi_{\theta|k}$ . The prior for  $d$  has density  $\pi_d$  which is strictly positive on  $[-\frac{1}{2} + t, \frac{1}{2} - t]$ , the interval which is assumed to contain  $d_o$ , and zero elsewhere. The prior for  $\theta$  given  $k$  has a density  $\pi_{\theta|k}$  with respect to Lebesgue measure. This density satisfies condition  $\text{Hyp}(\mathcal{K}, c_0, \beta, L_o)$ , by which we mean that for a subset  $\mathcal{K}$  of  $\mathbb{N}$ ,

$$\min_{k \in \mathcal{K}} \inf_{\theta \in \Theta_k(\beta, L_o)} e^{c_0 k \log k} \pi_{\theta|k}(\theta) > 1,$$

where  $L_o$  is as in (1.4). The choice of  $\mathcal{K}$  depends on the prior for  $k$  and  $\theta|k$ . We consider the following classes of priors.

- **Prior A:**  $k$  is deterministic and increasing at rate

$$k_n = \lfloor k_A (n / \log n)^{\frac{1}{2\beta}} \rfloor, \quad (2.1)$$

for a constant  $k_A > 0$ . The prior density for  $\theta|k$  satisfies  $\text{Hyp}(\{k_n\}, c_0, \beta - \frac{1}{2}, L_o)$  for some  $c_0 > 0$  and has support  $\Theta_k(\beta - \frac{1}{2}, L)$ . In addition, for all  $\theta, \theta' \in \Theta_k(\beta - \frac{1}{2}, L)$  such that  $\|\theta - \theta'\| \leq L(n / \log n)^{-\frac{2\beta-1}{4\beta}}$ ,

$$\log \pi_{\theta|k}(\theta) - \log \pi_{\theta|k}(\theta') = h_k^t(\theta - \theta') + o(1), \quad (2.2)$$

for constants  $C, \rho_0 > 0$  and vectors  $h_k$  satisfying  $\|h_k\| \leq C(n/k)^{1-\rho_0}$ . Finally, it is assumed that  $L$  is sufficiently large compared to  $L_o$ .

- **Prior B:**  $k$  is deterministic and increasing at rate

$$k'_n = \lfloor k_B (n / \log n)^{\frac{1}{1+2\beta}} \rfloor,$$

where  $k_B$  is such that  $k'_n < k_n$  for all  $n$ . The prior for  $\theta|k$  has density  $\pi_{\theta|k}$  with respect to Lebesgue measure which satisfies condition  $\text{Hyp}(\{k'_n\}, c_0, \beta, L_o)$  for some  $c_0 > 0$  and is assumed to have support  $\Theta_k(\beta, L)$ . The density also satisfies

$$\log \pi_{\theta|k}(\theta) - \log \pi_{\theta|k}(\theta') = o(1),$$

for all  $\theta, \theta' \in \Theta_k(\beta, L)$  such that  $\|\theta - \theta'\| \leq L(n / \log n)^{-\frac{\beta}{2\beta+1}}$ . This condition is similar to (2.2), but with  $h_k = 0$ , and support  $\Theta_k(\beta, L)$ .

- **Prior C:**  $k \sim \pi_k$  on  $\mathbb{N}$  with  $e^{-c_1 k \log k} \leq \pi_k(k) \leq e^{-c_2 k \log k}$  for  $k$  large enough, where  $0 < c_1 < c_2 < +\infty$ . There exists  $\beta_s > 1$  such that for all

$\beta \geq \beta_s$ , the prior for  $\theta|k$  has density  $\pi_{\theta|k}$  with respect to Lebesgue measure which satisfies condition  $\text{Hyp}(\{k \leq k_0(n/\log n)^{1/(2\beta+1)}\}, c_0, \beta, L_0)$ , for all  $k_0 > 0$  and some  $c_0 > 0$ , as soon as  $n$  is large enough. It has support included in  $\Theta_k(\beta, L)$  and satisfies

$$\log \pi_{\theta|k}(\theta) - \log \pi_{\theta|k}(\theta') = o(1),$$

for all  $\theta, \theta' \in \Theta_k(\beta, L)$  such that  $\|\theta - \theta'\| \leq L(n/\log n)^{-\frac{\beta}{2\beta+1}}$ .

Note that **prior A** is obtained when we take  $\beta' = \beta - \frac{1}{2}$  in prior B.

## 2.2 Examples of priors

The Lipschitz conditions on  $\log \pi_{\theta|k}$  considered for the three types of priors are satisfied for instance for the uniform prior on  $\Theta_k(\beta - \frac{1}{2}, L)$  (resp.  $\Theta_k(\beta, L)$ ), and for the truncated Gaussian prior, where, for some constants  $A$  and  $\alpha > 0$ ,

$$\pi_{\theta|k}(\theta) \propto \mathbb{I}_{\Theta_k(\beta - \frac{1}{2}, L)}(\theta) \exp\left(-A \sum_{j=0}^k j^\alpha \theta_j^2\right).$$

In the case of **Prior A**, the conditions on  $\log \pi_{\theta|k}$  and  $h_k$  in (2.2) are satisfied for  $\alpha < 4\beta - 2$ . To see this, note that for all  $\theta, \theta' \in \Theta_k(\beta - 1/2, L)$ ,

$$\sum_{j=0}^k j^\alpha |\theta_j^2 - (\theta'_j)^2| \leq L^{1/2} \|\theta - \theta'\| k^{\alpha - \beta + 1/2} = o((n/k)^{1-\delta}).$$

In the case of **Prior B** and **Prior C** we may choose  $\alpha < 2\beta$ , since for some positive  $k_0$

$$\sum_{j=0}^k j^\alpha |\theta_j^2 - (\theta'_j)^2| \leq L^{1/2} \|\theta - \theta'\| k^{\alpha - \beta} = o(1),$$

for all  $k \leq k_0(n/\log n)^{1/(2\beta+1)}$  and all  $\theta, \theta' \in \Theta_k(\beta, L)$  such that  $\|\theta - \theta'\| \leq (n/\log n)^{-\beta/(2\beta+1)}$ .

Also a truncated Laplace distribution is possible, in which case

$$\pi_{\theta|k}(\theta) \propto \mathbb{I}_{\Theta_k(\beta - \frac{1}{2}, L)}(\theta) \exp\left(-a \sum_{j=0}^k |\theta_j|\right).$$

The condition on  $\pi_k$  in **Prior C** is satisfied for instance by Poisson distributions.

The restriction of the prior to Sobolev balls is required to obtain a proper concentration rate or even consistency of the posterior of the spectral density  $f$  itself, which is a necessary step in the proof of our results. This is discussed in more detail in section 3.1.



### 2.3 Convergence rates and BVM-results under different priors

Assuming a Poisson prior for  $k$ , RCL (Theorem 4.2) obtain a near-optimal convergence rate for  $l(f, f_o)$ . In Corollary 3.1 below, we show that the optimal rate for  $l$  implies that we have at least a suboptimal rate for  $|d - d_o|$ . Whether this can be improved to the optimal rate critically depends on the prior on  $k$ . By our first main result the answer is positive under **prior A**. The proof is given in section 4.

**Theorem 2.1.** *Under **prior A**, the posterior distribution has the asymptotic expansion*

$$\Pi \left[ \sqrt{\frac{nr_{k_n}}{2}}(d - d_o - b_n(d_o)) \leq z | X \right] = \Phi(z) + o_{P_o}(1), \quad (2.3)$$

where, for  $r_{k_n} = \sum_{j \geq k_n+1} \eta_j^2$  and some small enough  $\delta > 0$ ,

$$b_n(d_o) = \frac{1}{r_{k_n}} \sum_{j=k_n+1}^{\infty} \eta_j \theta_{o,j} + Y_n + o(n^{-1/2-\delta} k_n^{1/2}), \quad Y_n = \frac{\sqrt{2}}{\sqrt{nr_{k_n}}} Z_n,$$

$Z_n$  being a sequence of random variables converging weakly to a Gaussian variable with mean zero and variance 1.

**Corollary 2.1.** *Under **prior A**, the convergence rate for  $d$  is  $\delta_n = (n/\log n)^{-\frac{2\beta-1}{4\beta}}$ , i.e.*

$$\lim_{n \rightarrow \infty} E_0^n [\Pi(d : |d - d_o| > \delta_n | X)] = 0.$$

Equation (2.3) is a Bernstein-von Mises type of result: the posterior distribution is asymptotically normal, centered at a point  $d_o + b_n(d_o)$ , whose distribution is normal with mean  $d_o$  and variance  $2/(nr_{k_n})$ . The expressions for the posterior mean and variance give more insight in how the prior for  $k$  affects the posterior rate for  $d$ . The standard deviation of the limiting normal distribution (2.3) is  $\sqrt{2/(nr_{k_n})} = O(n^{-\frac{2\beta-1}{4\beta}} (\log n)^{\frac{1}{4\beta}})$  and  $b_n(d_o)$  equals

$$\frac{1}{r_{k_n}} \sum_{j=k_n+1}^{\infty} \eta_j \theta_{o,j} + O_{P_o}(k_n^{\frac{1}{2}} n^{\frac{1}{2}}) + o(n^{-1/2-\delta_1} k_n^{1/2}).$$

From the definition of  $\eta_j$ ,  $k_n$  and  $r_{k_n}$  and the assumption on  $\theta_o$ , it follows that

$$\frac{1}{r_{k_n}} \left| \sum_{j=k_n+1}^{\infty} \eta_j \theta_{o,j} \right| \leq \frac{1}{r_{k_n}} \sqrt{\sum_{l>k_n} \theta_{o,l}^2 j^{2\beta}} \sqrt{\sum_{l>k_n} j^{-2\beta-2}} = o(k_n^{-\beta+\frac{1}{2}}). \quad (2.4)$$

See also (1.9) in the supplement. Hence, when the constant  $k_A$  in (2.1) is small enough,

$$|b_n(d_o)| \leq \delta_n, \quad (2.5)$$

and we obtain the  $\delta_n$ -rate of Corollary 2.1. For smaller  $k$ , the standard deviation is smaller but the bias  $b_n(d_o)$  is larger. In Theorem 2.2 below it is shown that this indeed leads to a suboptimal rate.

An important consequence of the BVM-result is that posterior credible regions for  $d$  (HPD or equal-tails for instance) will also be asymptotic frequentist confidence regions. Consider for instance one-sided credible intervals for  $d$  defined by  $P^\pi(d \leq z_n(\alpha)|X) = \alpha$ , so that  $z_n(\alpha)$  is the  $\alpha$ -th quantile of the posterior distribution of  $d$ . Equation (2.3) in Theorem 2.1 then implies that

$$z_n(\alpha) = d_o + b_n(d_o) + \sqrt{\frac{2k_n}{n}}\Phi^{-1}(\alpha)(1 + \mathbf{o}_{\mathbf{P}_o}(1)).$$

As soon as  $\sum_{j \geq k_n} j^{2\beta} \theta_{o,j}^2 = o((\log n)^{-1})$ , we have that

$$z_n(\alpha) = d_o + \sqrt{2/(nr_{k_n})}Z_n + \sqrt{2/(nr_{k_n})}\Phi^{-1}(\alpha)(1 + \mathbf{o}_{\mathbf{P}_o}(1))$$

and

$$P_o^n(d_o \leq z_n(\alpha)) = P(Z_n \leq \Phi^{-1}(\alpha)(1 + o(1))) = \alpha + o(1).$$

Similar computations can be made on equal - tail credible intervals or HPD regions for  $d$ .

Note that in this paper we assume that the smoothness  $\beta$  of  $f_o$  is greater than 1 instead of  $1/2$ , as is required in Moulines and Soulier (2003). This condition is used throughout the proof. Actually had we only assumed that  $\beta > 3/2$ , the proof of Theorem 2.1 would have been greatly simplified as many technicalities in the paper come from controlling terms when  $1 < \beta \leq 3/2$ . We do not believe that it is possible to weaken this constraint to  $\beta > 1/2$  in our setup.

Our second main result states that if  $k$  is increasing at a slower rate than  $k_n$ , the posterior on  $d$  concentrates at a suboptimal rate. The proof is given in section 5.

**Theorem 2.2.** *Given  $\beta > 5/2$ , there exists  $\theta_o \in \Theta(\beta, L_o)$  and a constant  $k_v > 0$  such that under prior  $B$  and  $C$  defined above,*

$$\Pi(|d - d_o| > k_v w_n (\log n)^{-1} | X) \stackrel{P_o}{\rightarrow} 1.$$

with  $w_n = C_w (n/\log n)^{-\frac{2\beta-1}{4\beta+2}}$  and  $C_w = C_1 (L + L_o)^{\frac{1}{4\beta}} l_0^{\frac{2\beta-1}{2\beta}}$ .

The constant  $C_w$  comes from the suboptimal rate for  $|d - d_o|$  derived in Corollary 3.1. Theorem 2.2 is proved by considering the vector  $\theta_o$  defined by  $\theta_{o,j} = c_0 j^{-(\beta+\frac{1}{2})} (\log j)^{-1}$ , for  $j \geq 2$ . This vector is close to the boundary of the Sobolev-ball  $\Theta(\beta, L_o)$ , in the sense that for all  $\beta' > \beta$ ,  $\sum_j j^{2\beta'} \theta_{o,j}^2 = +\infty$ . The proof consists in showing that conditionally on  $k$ , the posterior distribution is asymptotically normal as in (2.3), with  $k$  replacing  $k_n$ , and that the posterior distribution concentrates on values of  $k$  smaller than  $O(n^{1/(2\beta+1)})$ , so that the bias  $b_n(d_o)$  becomes of order  $w_n (\log n)^{-1}$ . The constraint  $\beta > 5/2$  is used to simplify the computations and is not sharp.

It is interesting to note that similar to the frequentist approach, a key issue is a bias-variance trade-off, which is optimized when  $k \sim n^{1/(2\beta)}$ . This choice of  $k$  depends on the smoothness parameter  $\beta$ , and since it is not of the same order as the *optimal* values of  $k$  for the loss  $l(f, f')$  on the spectral densities, the adaptive (near) minimax Bayesian nonparametric procedure proposed in Rousseau and Kruijer (2011) does not lead to optimal posterior concentration rate for  $d$ . While it is quite natural to obtain an adaptive (nearly) minimax Bayesian procedure under the loss  $l(\cdot, \cdot)$  by choosing a random  $k$ , obtaining an adaptive minimax procedure for  $d$  remains an open problem. This dichotomy is found in other semi-parametric Bayesian problems, see for instance Arbel (2010) in the case of the white noise model or Rivoirard and Rousseau (2010) for BVM properties.

### 3 Decomposing the posterior for $d$

To prove Theorems 2.1 and 2.2 we need to take a closer look at (1.3), to understand how the integration over  $\Theta_k$  affects the posterior for  $d$ . We develop  $\theta \rightarrow l_n(d, k, \theta)$  in a point  $\bar{\theta}_{d,k}$  defined below and decompose the likelihood as

$$\exp\{l_n(d, k, \theta)\} = \exp\{l_n(d, k)\} \exp\{l_n(d, k, \theta) - l_n(d, k)\},$$

where  $l_n(d, k)$  is short-hand notation for  $l_n(d, k, \bar{\theta}_{d,k})$ . Define

$$I_n(d, k) = \int_{\Theta_k} e^{l_n(d, k, \theta) - l_n(d, k)} d\pi_{\theta|k}(\theta), \quad (3.1)$$

where  $\Theta_k$  is the generic notation for  $\Theta_k(\beta - \frac{1}{2}, L)$  under **prior A** and  $\Theta_k(\beta, L)$  for priors B and C. The posterior for  $d$  given in (1.3) can be written as

$$\Pi(d \in D | X) = \frac{\sum_{k=0}^{\infty} \pi_k(k) \int_D e^{l_n(d, k) - l_n(d_o, k)} I_n(d, k) d\pi_d(d)}{\sum_{k=0}^{\infty} \pi_k(k) \int_{-\frac{1}{2}+t}^{\frac{1}{2}-t} e^{l_n(d, k) - l_n(d_o, k)} I_n(d, k) d\pi_d(d)}. \quad (3.2)$$

The factor  $\exp\{l_n(d, k) - l_n(d_o, k)\}$  is independent of  $\theta$ , and will under certain conditions dominate the marginal likelihood. In section 3.2 we give a Taylor-approximation which, for given  $k$ , allows for a normal approximation to the marginal posterior. However, to obtain the convergence rates in Theorems 2.1 and 2.2, it also needs to be shown that the integrals  $I_n(d, k)$  with respect to  $\theta$  do not vary too much with  $d$ . This is the most difficult part of the proof of Theorem 2.1 and the argument is presented in section 3.3. Since Theorem 2.2 is essentially a counter-example and it is not aimed to be as general as Theorem 2.1, as far as the range of  $\beta$  is concerned, we can restrict attention to larger  $\beta$ 's, i.e.  $\beta > 5/2$ , for which controlling  $I_n(d, k)$  is much easier.

#### 3.1 Preliminaries

First we define the point  $\bar{\theta}_{d,k}$  in which we develop  $\theta \rightarrow l_n(d, k, \theta)$ . Since the function  $\log(2 - 2\cos(x))$  has Fourier coefficients against  $\cos jx$ ,  $j \in \mathbb{N}$  equal to

$0, 2, \frac{2}{2}, \frac{2}{3}, \dots$ , FEXP-spectral densities can be written as

$$|1 - e^{ix}|^{-2d} \exp \left\{ \sum_{j=0}^{\infty} \theta_j \cos(jx) \right\} = \exp \left\{ \sum_{j=0}^{\infty} (\theta_j + d\eta_j) \cos(jx) \right\}.$$

Given  $f = f_{d,k,\theta}$  and  $f' = f_{d',k',\theta'}$  we can therefore express the norm  $l(f, f')$  in terms of  $(\theta - \theta')$  and  $(d - d')$ :

$$l(f, f') = \frac{1}{2} \sum_{j=0}^{\infty} ((\theta_j - \theta'_j) + \eta_j(d - d'))^2, \quad (3.3)$$

where  $\theta_j$  and  $\theta'_j$  are understood to be zero when  $j$  is larger than  $k$  respectively  $k'$ . Equation (3.3) implies that for given  $d$  and  $k$ ,  $l(f_o, f_{d,k,\theta})$  is minimized by

$$\bar{\theta}_{d,k} := \operatorname{argmin}_{\theta \in \mathbb{R}^{k+1}} \sum_{j=0}^{\infty} (\theta_j - \theta_{o,j} + (d - d_o)\eta_j)^2 = \theta_{o[k]} + (d - d_o)\eta_{[k]}.$$

In particular,  $\theta = \theta_{o[k]}$  minimizes  $l(f_o, f_{d,k,\theta})$  only when  $d = d_o$ ; when  $d \neq d_o$  we need to add  $(d - d_o)\eta_{[k]}$ . The following lemma shows that an upper bound on  $l(f_o, f_{d,k,\theta})$  leads to upper bounds on  $|d - d_o|$  and  $\|\theta - \theta_o\|$ .

**Lemma 3.1.** *Suppose that  $\theta \in \Theta_k(\gamma, L)$  and  $\theta_o \in \Theta_k(\beta, L_o)$ , where  $\gamma \leq \beta$ . Also suppose that for a sequence  $\alpha_n \rightarrow 0$ ,  $l(f_o, f_{d,k,\theta}) \leq \alpha_n^2$  for all  $n$ . Then there are universal constants  $C_1, C_2 > 0$  such that for all  $n$ ,*

$$|d - d_o| \leq C_1(L + L_o)^{\frac{1}{4\gamma}} \alpha_n^{\frac{2\gamma-1}{2\gamma}}, \quad \|\theta - \theta_o\| \leq C_2(L + L_o)^{\frac{1}{4\gamma}} \alpha_n^{\frac{2\gamma-1}{2\gamma}}.$$

*Proof.* For all  $(d, k, \theta)$  such that  $l(f_{d,k,\theta}, f_o) \leq \alpha_n$ , we have, using (3.3),

$$\begin{aligned} 2\alpha_n^2 &\geq 2l(f_{d,k,\theta}, f_o) = 2(\theta_{o,0} - \theta_0)^2 + \sum_{j \geq 1} ((\theta_{o,j} - \theta_j) + \eta_j(d - d_o))^2 \\ &\geq \sum_{j \geq 1} (\theta_{o,j} - \theta_j)^2 + (d - d_o)^2 \sum_{j \geq 1} \eta_j^2 - 2|d - d_o| \sqrt{\sum_{j \geq 1} \eta_j^2} \sqrt{\sum_{j \geq 1} (\theta_{o,j} - \theta_j)^2} \\ &= (\|\theta - \theta_o\| - |d - d_o|\|\eta\|)^2. \end{aligned}$$

The inequalities remain true if we replace all sums over  $j \geq 1$  by sums over  $j \geq m_n$ , for any nondecreasing sequence  $m_n$ . Since  $\|(\eta_j \mathbf{1}_{j > m_n})_{j \geq 1}\|^2$  is of order  $m_n^{-1}$  and  $\|(\theta - \theta_o)_{j > m_n}\|_{j \geq 1}^2 \leq m_n^{-2\gamma} \sum_{j > m_n} (1+j)^{2\beta} (\theta_j - \theta_{o,j})^2 < 2(L + L_o)m_n^{-2\gamma}$ , setting  $m_n = \alpha_n^{-\frac{1}{\gamma}}$  gives the desired rate for  $|d - d_o|$  as well as for  $\|\theta - \theta_o\|$ .  $\square$

The convergence rate for  $l(f_o, f_{d,k,\theta})$  required in Lemma 3.1 can be found in Rousseau and Kruijer (2011). For easy reference we restate it here. Compared to a similar result in RCL, the  $\log n$  factor is improved.

**Lemma 3.2.** Under **prior A**, there exists a constant  $l_0$  depending only on  $L_o$  and  $k_A$  (and not on  $L$ ) such that

$$\Pi((d, k, \theta) : l(f_{d,k,\theta}, f_o) \geq l_0^2 \delta_n^2 | X) \xrightarrow{P_\Sigma} 0,$$

where  $\delta_n = (n/\log n)^{-\frac{2\beta-1}{4\beta}}$ . Under **priors B** and **C**, this statement holds with  $\epsilon_n = (n/\log n)^{-\frac{\beta}{2\beta+1}}$  replacing  $\delta_n$ .

In the proof of Theorem 2.1 (resp. 2.2), this result allows us to restrict attention to the set of spectral densities  $f$  such that  $l(f, f_o) \leq l_0^2 \delta_n^2$  (resp.  $l_0^2 \epsilon_n^2$ ). In addition, by combination with Lemma 3.1 we can now deduce bounds on  $|d - d_o|$  and  $\|\theta - \bar{\theta}_{d,k}\|$ . These bounds, although suboptimal, will be important in the sequel for obtaining the near-optimal rate in Theorem 2.1.

**Corollary 3.1.** Under the result of Lemma 3.2 and **prior A**, we can apply Lemma 3.1 with  $\alpha_n^2 = l_0^2 \delta_n^2$  and  $\gamma = \beta - \frac{1}{2}$ , and obtain

$$\Pi_d(d : |d - d_o| \geq \bar{v}_n | X) \xrightarrow{P_\Sigma} 0, \quad \Pi(\|\theta - \bar{\theta}_{d,k}\| \geq 2l_0 \delta_n | X) \xrightarrow{P_\Sigma} 0,$$

where  $\bar{v}_n = C_1(L + L_o)^{\frac{1}{4\beta-2}} l_0^{\frac{2\beta-2}{2\beta-1}} (n/\log n)^{-\frac{\beta-1}{2\beta}}$ . Under **priors B** and **C** we have  $\gamma = \beta$ ; the rate for  $|d - d_o|$  is then  $w_n = C_w (n/\log n)^{-\frac{2\beta-1}{4\beta+2}}$  and the rate for  $\|\theta - \bar{\theta}_{d,k}\|$  is  $2l_0 \epsilon_n$ . The constant  $C_w = C_1(L + L_o)^{\frac{1}{4\beta}} l_0^{\frac{2\beta-1}{2\beta}}$  is as in Theorem 2.2.

*Proof.* The rate for  $|d - d_o|$  follows directly from Lemma 3.1. To obtain the rate for  $\|\theta - \bar{\theta}_{d,k}\|$ , let  $\alpha_n$  denote either  $l_0 \delta_n$  (the rate for  $l(f_o, f)$  under prior A) or  $l_0 \epsilon_n$  (the rate under priors B and C). Although Lemma 3.1 suggests that the Euclidean distance from  $\theta_o$  to  $\theta$  (contained in  $\Theta_k(\beta, L)$  or  $\Theta_k(\beta - \frac{1}{2}, L)$ ) may be larger than  $\alpha_n$ , the distance from  $\theta$  to  $\bar{\theta}_{d,k}$  is certainly of order  $\alpha_n$ . To see this, note that Lemma 3.2 implies the existence of  $d, k, \theta$  in the model with  $l(f_o, f_{d,k,\theta}) \leq \alpha_n^2$ . From the definition of  $\bar{\theta}_{d,k}$  it follows that  $l(f_o, f_{d,k,\bar{\theta}_{d,k}}) \leq \alpha_n^2$ . The triangle inequality gives  $\|\theta - \bar{\theta}_{d,k}\|^2 = l(f_{d,k,\theta}, f_{d,k,\bar{\theta}_{d,k}}) \leq 4\alpha_n^2$ .  $\square$

The rates  $\bar{v}_n$  and  $w_n$  obtained in Corollary 3.1 are clearly suboptimal; their importance however lies in the fact that they narrow down the set for which we need to prove Theorems 2.1 and 2.2. To prove Theorem 2.2 for example it suffices to show that the posterior mass on  $k_v w_n (\log n)^{-1} < |d - d_o| < w_n$  tends to zero. Note that the lower and the upper bound differ only by a factor  $(\log n)$ . Hence under priors B and C, the combination of Corollary 3.1 and Theorem 2.2 characterizes the posterior concentration rate (up to a  $\log n$  term) for the given  $\theta_o$ . Another consequence of Corollary 3.1 is that we may neglect the posterior mass on all  $(d, k, \theta)$  for which  $\|\theta - \bar{\theta}_{d,k}\|$  is larger than  $2l_0 \delta_n$  (under prior A) or  $2l_0 \epsilon_n$  (under priors B and C).

We conclude this section with a result on  $\bar{\theta}_{d,k}$  and  $\Theta_k(\beta, L)$ . In the definition of  $\bar{\theta}_{d,k}$  we minimize over  $\mathbb{R}^{k+1}$ , whereas the support of priors A-C is the Sobolev

ball  $\Theta_k(\beta, L)$  or  $\Theta_k(\beta - \frac{1}{2}, L)$ . Under the assumptions of Theorems 2.1 and 2.2 however,  $\theta_{d,k}$  is contained in  $\Theta_k(\beta - \frac{1}{2}, L)$  respectively  $\Theta_k(\beta, L)$ . Also the  $l_2$ -ball of radius  $2l_0\delta_n$  (or  $2l_0\epsilon_n$ ) is contained in these Sobolev-balls.

**Lemma 3.3.** *Under the assumptions of Theorem 2.1,  $B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)$  is contained in  $\Theta_k(\beta - \frac{1}{2}, L)$  for all  $d \in [d_o - \bar{v}_n, d_o + \bar{v}_n]$ , if  $L$  is large enough. In particular,  $\bar{\theta}_{d,k} \in \Theta_k(\beta - \frac{1}{2}, L)$ . Similarly, under the assumptions of Theorem 2.2,  $B_k(\bar{\theta}_{d,k}, 2l_0\epsilon_n) \subset \Theta_k(\beta, L)$ , for all  $d \in [d_o - w_n, d_o + w_n]$ .*

*Proof.* Since the constant  $l_0$  is independent of  $L$ ,  $\theta \in B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)$  implies that for  $n$  large enough ,

$$\begin{aligned} \sum_{j=0}^k \theta_j^2 (j+1)^{2\beta-1} &\leq 2 \sum_{j=0}^k (\theta - \bar{\theta}_{d,k})_j^2 (j+1)^{2\beta-1} + 2 \sum_{j=0}^k (\bar{\theta}_{d,k})_j^2 (j+1)^{2\beta-1} \\ &\leq 8\delta^2 (L_o)(n/\log n)^{\frac{2\beta-1}{2\beta}} (k_n + 1)^{2\beta-1} + 4 \sum_{j=0}^{k_n} \theta_{o,j}^2 (j+1)^{2\beta-1} \\ &\quad + 16(d - d_o)^2 \sum_{j=1}^{k_n} j^{2\beta-3}. \end{aligned}$$

The first two terms on the right only depend on  $L_o$ , and are smaller than  $L/4$  when  $L$  is chosen sufficiently large. Because  $\bar{v}_n = C_1(L+L_o)^{\frac{1}{4\beta-2}} l_0^{\frac{2\beta-2}{2\beta-1}} (n/\log n)^{-\frac{\beta-1}{2\beta}}$ , the last term in the preceding display is at most

$$C_1^2 (L + L_o)^{\frac{1}{2\beta-1}} l_0^{\frac{4\beta-4}{2\beta-1}} (n/\log n)^{-\frac{\beta-1}{\beta}} k_A^{2\beta-2} (n/\log n)^{\frac{\beta-1}{\beta}},$$

which, since  $\beta > 1$ , is smaller than  $L/2$  when  $L$  is large enough. We conclude that  $B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)$  is contained in  $\Theta_k(\beta - \frac{1}{2}, L)$  provided  $L$  is chosen sufficiently large. The second statement can be proved similarly.  $\square$

### 3.2 A Taylor approximation for $l_n(d, k)$

Provided that the integrals  $I_n(d, k)$  have negligible impact on the posterior for  $d$ , the conditional distribution of  $d$  given  $k$  will only depend on  $\exp\{l_n(d, k) - l_n(d_o, k)\}$ . Let  $l_n^{(1)}(d, k)$ ,  $l_n^{(2)}(d, k)$  denote the first two derivatives of the map  $d \mapsto l_n(d, k)$ . There exists a  $\bar{d}$  between  $d$  and  $d_o$  such that

$$l_n(d, k) = l_n(d_o, k) + (d - d_o)l_n^{(1)}(d_o, k) + \frac{(d - d_o)^2}{2} l_n^{(2)}(\bar{d}, k). \quad (3.4)$$

Defining

$$b_n(d) = -\frac{l_n^{(1)}(d_o, k)}{l_n^{(2)}(d, k)},$$

which is the  $b_n$  used in Theorem 2.1, we can rewrite (3.4) as

$$l_n(d, k) - l_n(d_o, k) = -\frac{1}{2} \frac{(l_n^{(1)}(d_o, k))^2}{l_n^{(2)}(\bar{d}, k)} + \frac{1}{2} l_n^{(2)}(\bar{d}, k) (d - d_o - b_n(\bar{d}))^2. \quad (3.5)$$

Note that each derivative  $l_n^{(i)}(d, k)$ ,  $i = 1, 2$ , can be decomposed into a centered quadratic form denoted  $\mathcal{S}(l_n^{(i)}(d, k))$  and a deterministic term  $\mathcal{D}(l_n^{(i)}(d, k))$ . In the following lemma we give expressions for  $l_n^{(1)}(d_o, k)$ ,  $l_n^{(2)}(d, k)$  and  $b_n$ , making explicit their dependence on  $k$  and  $\theta_o$ . Since  $k'_n \leq k_n$  and  $w_n < \bar{v}_n$  (see Corollary 3.1) the result is valid for all priors under consideration. The proof is given in appendix A.

**Lemma 3.4.** *Given  $\beta > 1$ , let  $\theta_o \in \Theta(\beta, L_o)$ . If  $k \leq k_n$  and  $|d - d_o| \leq \bar{v}_n$ , then there exists  $\delta_1 > 0$  such that*

$$\begin{aligned} l_n^{(1)}(d_o, k) &:= \mathcal{S}(l_n^{(1)}(d_o, k)) + \mathcal{D}(l_n^{(1)}(d_o, k)) \\ &= \mathcal{S}(l_n^{(1)}(d_o, k)) + \frac{n}{2} \sum_{j=k+1}^{\infty} \theta_{o,j} \eta_j + o(n^\epsilon (k^{-\beta+3/2} + n^{-1/(2\beta)})), \\ l_n^{(2)}(d, k) &= l_n^{(2)}(d_o, k) \left( 1 + \frac{k^{1/2}}{n^{1/2+\epsilon}} + \frac{k^{-2\beta+1+\epsilon}}{n} \right) = -\frac{1}{2} nr_k (1 + \mathbf{O}_{\mathbf{P}_o}(n^{-\delta_1})), \end{aligned}$$

where  $\mathcal{S}(l_n^{(1)}(d_o, k))$  is a centered quadratic form with variance

$$\text{Var}(\mathcal{S}(l_n^{(1)}(d_o, k))) = \frac{n}{2} \sum_{j>k} \eta_j^2 (1 + o(1)) = \frac{nr_k}{2} (1 + o(1)) = O(nk^{-1}).$$

Consequently,

$$\begin{aligned} b_n(d) &= -\frac{l_n^{(1)}(d_o, k)}{l_n^{(2)}(d, k)} = \frac{1}{r_k} \sum_{j=k+1}^{\infty} \theta_{o,j} \eta_j (1 + \mathbf{O}_{\mathbf{P}_o}(n^{-\delta})) \\ &\quad + \frac{2\mathcal{S}(l_n^{(1)}(d_o, k))(1 + \mathbf{O}_{\mathbf{P}_o}(n^{-\delta}))}{nr_k} + \mathbf{O}_{\mathbf{P}_o}(n^{\epsilon-1} k^{-\beta+5/2} + n^{\epsilon-1}), \end{aligned} \quad (3.6)$$

with

$$\frac{2\mathcal{S}(l_n^{(1)}(d_o, k))}{nr_k} = \mathbf{O}_{\mathbf{P}_o}(n^{-\frac{1}{2}} k^{\frac{1}{2}}).$$

**Remark 3.1.** *Recall from (2.4) that  $r_k^{-1} \sum_{j=k+1}^{\infty} \theta_{o,j} \eta_j$  is  $O(k^{-\beta+1/2})$ . The term  $2\mathcal{S}(l_n^{(1)}(d_o, k))/(nr_k)$  is  $O_{\mathbf{P}_o}(k^{-\beta+1/2})$  whenever  $k \sim n^{1/(2\beta)}$ , which is the case under all priors under consideration.*

Substituting the above results on  $l_n^{(1)}$ ,  $l_n^{(2)}$  and  $b_n$  in (3.5), we can give the following informal argument leading to Theorems 2.1 and Theorem 2.2. If we consider  $k$  to be fixed and  $I_n(d, k)$  constant in  $d$ , then (3.5) implies that the posterior distribution for  $d$  is asymptotically normal with mean  $d_o + b_n(d_o)$  and variance of order  $k/n$ .

### 3.3 Integration of the short memory parameter

A key ingredient in the proofs of both Theorems 2.1 and 2.2 is the control of the integral  $I_n(d, k)$  appearing in (1.3), whose dependence on  $d$  should be negligible with respect to  $\exp\{l_n(d, k) - l_n(d_o, k)\}$ . In Lemma 3.5 below we prove this to be the case under the assumptions of Theorems 2.1 and 2.2. For the case of Theorem 2.2 this is fairly simple: the conditional posterior distribution of  $\theta$  given  $(d, k)$  can be proved to be asymptotically Gaussian by a Laplace-approximation. For smaller  $\beta$  and larger  $k$  the control is technically more demanding. In both cases the proof is based on the following Taylor expansion of  $l_n(d, k, \theta)$  around  $\bar{\theta}_{d,k}$ :

$$l_n(d, k, \theta) - l_n(d, k) = \sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d, k)}{j!} + R_{J+1,d}(\theta), \quad (3.7)$$

where

$$\begin{aligned} (\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d, k) &= \sum_{l_1, \dots, l_j=0}^k (\theta - \bar{\theta}_{d,k})_{l_1} \dots (\theta - \bar{\theta}_{d,k})_{l_j} \frac{\partial^j l_n(d, k, \bar{\theta}_{d,k})}{\partial \theta_{l_1} \dots \partial \theta_{l_j}}, \\ R_{J+1,d}(\theta) &= \frac{1}{(J+1)!} \sum_{l_1, \dots, l_{J+1}=0}^k (\theta - \bar{\theta}_{d,k})_{l_1} \dots (\theta - \bar{\theta}_{d,k})_{l_{J+1}} \frac{\partial^{J+1} l_n(d, k, \tilde{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_{J+1}}}. \end{aligned} \quad (3.8)$$

The above expressions are used to derive the following lemma, which gives control of the term  $I_n(d, k)$ .

**Lemma 3.5.** *Under the conditions of Theorem 2.1, the integral  $I_n(d, k)$  defined in (3.1) equals*

$$I_n(d_o, k) \exp \left\{ \mathbf{O}_{\mathbf{P}_o}(1) + \mathbf{O}_{\mathbf{P}_o} \left( \frac{|d - d_o| n^{\frac{1}{2} - \delta_2}}{\sqrt{k}} \right) + \mathbf{O}_{\mathbf{P}_o} \left( (d - d_o)^2 \frac{n^{1 - \delta_2}}{k} \right) \right\},$$

for some  $\delta_2 > 0$ . Under the conditions of Theorem 2.2,

$$I_n(d, k) = I_n(d_o, k) \exp \{ \mathbf{O}_{\mathbf{P}_o}(1) \}.$$

The proof is given in Appendix C, and relies on the expressions for the derivatives  $\nabla^j l_n$  given in Appendix B. Lemma 3.5 should be seen in relation to Lemma 3.4 and the expressions for  $\Pi(d|X)$  and  $l_n(d, k) - l_n(d_o, k)$  in equations (3.2) and (3.4). Lemma 3.5 then shows that the dependence on the integrals  $I_n(d, k)$  on  $d$  is asymptotically negligible with respect to  $l_n(d, k) - l_n(d_o, k)$ . This is made rigorous in the following section.

## 4 Proof of Theorem 2.1

By Lemma 3.2 we may assume posterior convergence of  $l(f_o, f_{d,k,\theta})$  at rate  $l_0^2 \delta_n^2$ , and, by Corollary 3.1, also convergence of  $|d - d_o|$  at rate  $\bar{v}_n$ . By Lemma 3.3,



we may restrict the integration over  $\theta$  to  $B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)$ . Let  $\Gamma_n(z) = \{d : \sqrt{\frac{nr_k}{2}}(d - d_o - b_n(d_o)) \leq z\}$ . Under **prior A**, it suffices to show that for  $k = k_n$ ,

$$\begin{aligned} \frac{N_n}{D_n} &:= \frac{\int_{\Gamma_n(z)} e^{l_n(d,k) - l_n(d_o,k)} \int_{B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)} e^{l_n(d,k,\theta) - l_n(d,k)} d\pi_{\theta|k}(\theta) d\pi_d(d)}{\int_{|d-d_o| < \bar{v}_n} e^{l_n(d,k) - l_n(d_o,k)} \int_{B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)} e^{l_n(d,k,\theta) - l_n(d,k)} d\pi_{\theta|k}(\theta) d\pi_d(d)} \\ &= \frac{\int_{\Gamma_n(z)} \exp\{l_n(d,k) - l_n(d_o,k) + \log I_n(d,k)\} d\pi_d(d)}{\int_{|d-d_o| < \bar{v}_n} \exp\{l_n(d,k) - l_n(d_o,k) + \log I_n(d,k)\} d\pi_d(d)} = \Phi(z) + \mathbf{OP}_o(1). \end{aligned} \quad (4.1)$$

Using the results for  $l_n(d,k) - l_n(d_o,k)$  and  $I_n(d,k)$  given by Lemmas 3.4 and 3.5, we show that for  $A_n \subset \mathbb{R}^n$  defined below such that  $P_o^n(A_n) \rightarrow 1$ ,

$$\frac{N_n}{D_n} \leq \Phi(z) + o(1), \quad \frac{N_n}{D_n} \geq \Phi(z) + o(1), \quad \forall X \in A_n. \quad (4.2)$$

Since  $P_o^n(A_n) \rightarrow 1$  this implies the last equality in (4.1).

Note that Lemmas 3.4 and 3.5 also hold for all  $\delta'_1 < \delta_1$  and  $\delta'_2 < \delta_2$ . In the remainder of the proof, let  $0 < \delta \leq \min(\delta_1, \delta_2)$ . For notational simplicity, let  $\mathcal{D} = \mathcal{D}(l_n^{(1)}(d_o, k))$ , the deterministic part of  $l_n^{(1)}(d_o, k)$ . For a sufficiently large constant  $C_1$  and arbitrary  $\epsilon_1 > 0$ , let  $A_n$  be the set of  $X \in \mathbb{R}^n$  such that

$$\left. \begin{aligned} |\log I_n(d,k) - \log I_n(d_o,k)| &\leq \epsilon_1 + (d - d_o)^2 k^{-1} n^{1-\delta} + |d - d_o| k^{-\frac{1}{2}} n^{\frac{1}{2}-\delta} \\ \left| l_n^{(1)}(d_o, k) - \mathcal{D} \right| &\leq C_1 n^{\frac{1}{2}} k^{-\frac{1}{2}} \sqrt{\log n}, \quad \left| l_n^{(2)}(d,k) + \frac{1}{2} nr_k \right| \leq n^{1-\delta} k^{-1} \end{aligned} \right\}$$

for all  $|d - d_o| \leq \bar{v}_n$ . Since  $k = k_n$  and  $\beta > 1$ , Lemmas 3.4 and 3.5 imply that  $P_o^n(A_n^c) \rightarrow 0$ . We prove the first inequality in (4.2); the second one can be obtained in the same way. Using (3.4) and the definition of  $A_n$ , it follows that for all  $X \in A_n$ ,

$$\begin{aligned} l_n(d,k) - l_n(d_o,k) + \log I_n(d,k) - \log I_n(d_o,k) &\leq \epsilon_1 + (d - d_o)^2 \frac{n^{1-\delta}}{k} \\ &\quad + |d - d_o| \frac{n^{\frac{1}{2}-\delta}}{k^{\frac{1}{2}}} + (d - d_o) l_n^{(1)}(d_o, k) - \frac{nr_k}{4} (d - d_o)^2 (1 - n^{-\delta}) \\ &\leq 2\epsilon_1 - \frac{nr_k}{4} \left(1 - \frac{2}{n^\delta}\right) \left(d - d_o - \frac{2l_n^{(1)}(d_o, k)}{(1 - \frac{2}{n^\delta}) nr_k}\right)^2 + |d - d_o| \frac{n^{\frac{1}{2}-\delta}}{k^{\frac{1}{2}}} + \frac{(l_n^{(1)}(d_o, k))^2}{(1 - \frac{2}{n^\delta}) nr_k} \\ &\leq 3\epsilon_1 - \frac{nr_k}{4} \left(1 - \frac{2}{n^\delta}\right) \left(d - d_o - \frac{b_n(d_o, k)}{1 - \frac{2}{n^\delta}}\right)^2 \\ &\quad + \left| d - d_o - \frac{b_n(d_o, k)}{1 - \frac{2}{n^\delta}} \right| \frac{n^{\frac{1}{2}-\delta}}{k^{\frac{1}{2}}} + \frac{(l_n^{(1)}(d_o, k))^2}{(1 - \frac{2}{n^\delta}) nr_k}, \end{aligned} \quad (4.3)$$

The third inequality follows from (2.5) and Remark 3.1, by which  $b_n(d_o) = O(k^{-\beta + \frac{1}{2}}) = O(\delta_n)$ . This implies that  $|b_n(d_o)| k^{-\frac{1}{2}} n^{\frac{1}{2}-\delta} < \epsilon_1$ , again for large

enough  $n$ . Similar to the preceding display, we have the lower-bound

$$\begin{aligned}
& l_n(d, k) - l_n(d_o, k) + \log I_n(d, k) - \log I_n(d_o, k) \\
& \geq -3\epsilon_1 - \frac{nr_k}{4}(1 + 2n^{-\delta}) \left( d - d_o - \frac{b_n(d_o, k)}{(1 + 2n^{-\delta})} \right)^2 \\
& \quad - \left| d - d_o - \frac{b_n(d_o, k)}{(1 + 2n^{-\delta})} \right| k^{-\frac{1}{2}} n^{\frac{1}{2} - \delta} + \frac{(l_n^{(1)}(d_o, k))^2}{(1 + 2n^{-\delta})nr_k}.
\end{aligned} \tag{4.4}$$

Note that

$$\exp \left\{ \frac{(l_n^{(1)}(d_o, k))^2}{(1 - 2n^{-\delta})nr_k} - \frac{(l_n^{(1)}(d_o, k))^2}{(1 + 2n^{-\delta})nr_k} \right\} = \exp\{o(1)\}, \tag{4.5}$$

which follows from the expression for  $l_n^{(1)}(d_o, k)$  in Lemma 3.4, the definition of  $A_n$  and the assumption that  $X \in A_n$ . Therefore, substituting (4.3) in  $N_n$  and (4.4) in  $D_n$ , the terms  $\frac{(l_n^{(1)}(d_o, k))^2}{4nr_k}$  cancel out and by (4.5) we can neglect the difference between  $\frac{(l_n^{(1)}(d_o, k))^2}{(1 \pm 2n^{-\delta})nr_k}$  and  $\frac{(l_n^{(1)}(d_o, k))^2}{nr_k}$ .

To conclude the proof that  $N_n/D_n \leq \Phi(z) + o(1)$  for each  $X \in A_n$ , we make the change of variables

$$u = \sqrt{\frac{nr_k}{2}(1 \pm 2n^{-\delta})} \left( d - d_o - \frac{b_n(d_o)}{1 \pm 2n^{-\delta}} \right),$$

where we take  $+$  in the lower bound for  $D_n$  and  $-$  in the upper-bound for  $N_n$ . Using once more that  $b_n(d_o) = O(\delta_n)$ , we find that for large enough  $n$ ,  $|u| \leq \frac{\bar{v}_n}{4} \sqrt{nr_k}$  implies  $|d - d_o| \leq \bar{v}_n$ . Hence we may integrate over  $|u| \leq \frac{\bar{v}_n}{4} \sqrt{nr_k}$  in the lower-bound for  $D_n$ . In the upper-bound for  $N_n$  we may integrate over  $u \leq z + \epsilon_1$ .

Combining (4.3)-(4.5), it follows that for all  $\epsilon_1$  and all  $X \in A_n$ ,

$$\begin{aligned}
\frac{N_n}{D_n} & \leq e^{7\epsilon_1} \left( \frac{1 + 2n^{-\delta}}{1 - 2n^{-\delta}} \right)^{\frac{1}{2}} \frac{\int_{u < z + \epsilon_1} \exp\{-\frac{1}{2}u^2 + Cn^{-\delta}|u|\} du}{\int_{|u| \leq \frac{\bar{v}_n}{4} \sqrt{nr_k}} \exp\{-\frac{1}{2}u^2 - Cn^{-\delta}|u|\} du} \\
& \leq e^{8\epsilon_1} \frac{\int_{u < z + \epsilon_1} \exp\{-\frac{1}{2}u^2 + Cn^{-\delta}|u|\} du}{\int_{|u| \leq \frac{\bar{v}_n}{4} \sqrt{nr_k}} \exp\{-\frac{1}{2}u^2 - Cn^{-\delta}|u|\} du} \rightarrow \Phi(z + \epsilon_1) e^{8\epsilon_1}.
\end{aligned}$$

Similarly we prove that for all  $\epsilon_1$ ,  $N_n/D_n \geq \Phi(z - \epsilon_1) e^{-8\epsilon_1}$ , when  $n$  is large enough, which terminates the proof of Theorem 2.1.

## 5 Proof of Theorem 2.2

Let  $\beta > 5/2$  and  $\theta_{o,j} = c_0 j^{-(\beta + \frac{1}{2})} (\log j)^{-1}$ . When the constant  $c_0$  is chosen small enough,  $\theta_o \in \Theta(\beta, L_o)$ . In view of Corollary 3.1, the posterior mass on the events  $\{(d, k, \theta) : \|\theta - \theta_{d,k}\| \geq 2l_0\epsilon_n\}$  and  $\{(d, k, \theta) : |d - d_o| \geq w_n\}$  tends to zero

in probability, and may be neglected. Moreover Lemma 3.1 implies that with posterior probability going to 1,  $\|\theta - \theta_0\| \lesssim (n/\log n)^{-(\beta-1/2)/(2\beta+1)}$ . However, within the  $(k+1)$ -dimensional FEXP-model,  $\|\theta - \theta_o\|$  is minimized by setting  $\theta_j = \theta_{o,j}$  ( $j = 0, \dots, k$ ), and for this choice of  $\theta$  we have

$$\|\theta - \theta_o\|^2 = \sum_{l>k} \theta_{o,l}^2 \gtrsim k^{-2\beta} (\log k)^{-2}.$$

Consequently, the fact that  $\|\theta - \theta_0\| \lesssim (n/\log n)^{-(\beta-1/2)/(2\beta+1)}$  implies that  $k > k_n'' := k_l (n/\log n)^{(\beta-1/2)/(\beta(2\beta+1))} (\log n)^{-1/\beta}$ , for some constant  $k_l$ . We conclude that

$$\Pi(k \leq k_n'' | X) = \mathbf{o}_{\mathbf{P}_o}(1),$$

and we can restrict our attention to  $k > k_n''$ .

We decompose  $\Pi_d(|d - d_o| \leq k_v w_n (\log n)^{-1}, k > k_n'' | X)$  as

$$\begin{aligned} & \sum_{m>k_n''} \Pi(|d - d_o| \leq k_v w_n (\log n)^{-1}, k = m | X) \\ &= \sum_{m>k_n''} \Pi(k = m | X) \Pi_m(|d - d_o| \leq k_v w_n (\log n)^{-1} | X), \end{aligned}$$

where  $\Pi_m(|d - d_o| \leq k_v w_n (\log n)^{-1} | X)$  is the posterior for  $d$  within the FEXP-model of dimension  $m+1$ , i.e.  $\Pi_m(|d - d_o| \leq k_v w_n (\log n)^{-1} | X) := \Pi(|d - d_o| \leq k_v w_n (\log n)^{-1} | k = m, X)$ .

To prove Theorem 2.2 it now suffices to show that

$$\sum_{k_n'' \leq m \leq k_n'} \Pi(k = m | X) = \Pi(k_n'' \leq k \leq k_n' | X) \xrightarrow{P_o} 1, \quad (5.1)$$

$$E_o^n \Pi_k(|d - d_o| \leq k_v w_n (\log n)^{-1} | X) \xrightarrow{P_o} 0, \quad \forall k_n'' \leq k \leq k_n'. \quad (5.2)$$

The convergence in (5.1) is a by-product of Theorem 1 in Rousseau and Kruijer (2011). In the remainder we prove (5.2). For every  $k \leq k_n'$  we can write, using the notation of (4.1),

$$\begin{aligned} \Pi_k(|d - d_o| < k_v w_n (\log n)^{-1} | X) &\leq \frac{N_{n,k}}{D_{n,k}} \\ &:= \frac{\int_{|d-d_o| < k_v w_n (\log n)^{-1}} \exp\{l_n(d, k) - l_n(d_o, k) + \log I_n(d, k)\} d\pi_d(d)}{\int_{|d-d_o| < w_n} \exp\{l_n(d, k) - l_n(d_o, k) + \log I_n(d, k)\} d\pi_d(d)}. \end{aligned} \quad (5.3)$$

Let  $\delta_2 > 0$  and  $A_n$  be the set of  $X \in \mathbb{R}^n$  such that

$$\left. \begin{aligned} & |\log I_n(d, k) - \log I_n(d_o, k)| \leq \epsilon_1, \\ & \left| l_n^{(1)}(d_o, k) - \mathcal{D}(l_n^{(1)}(d_o, k)) \right| \leq n^{\frac{1}{2}} k^{-\frac{1}{2}} \sqrt{\log n}, \\ & \left| l_n^{(2)}(d, k) - \mathcal{D}(l_n^{(2)}(d_o, k)) \right| \leq \epsilon_1 n^{-(2+\delta_2)/(2\beta+1)} \end{aligned} \right\}$$

for all  $|d - d_o| \leq w_n$  and  $k_n'' \leq k \leq k_n'$ . Compared to the definition of  $A_n$  in the proof of Theorem 2.1, the constraints on  $l_n^{(2)}(d, k)$  and  $I_n$  are different. For the latter, recall from Lemma 3.5 that  $\log I_n(d, k) = \log I_n(d_o, k) + \mathbf{o}_{\mathbf{P}_o}(1)$ , uniformly over  $d \in (d_o - w_n, d_o + w_n)$ . As in the proof of Theorem 2.1, it now follows from Lemmas 3.4 and 3.5 that  $P_o^n(A_n^c) \rightarrow 0$ . We can write

$$E_0^n \left[ \frac{N_{n,k}}{D_{n,k}} \right] \leq P_o^n(A_n^c) + E_0^n \left[ \frac{N_{n,k}}{D_{n,k}} 1_{A_n} \right],$$

and bound  $N_{n,k}/D_{n,k}$  pointwise for  $X \in A_n$ . Since when  $k \in (k_n'', k_n')$ ,

$$\frac{(l_n^{(1)}(d_o, k))^2}{2|l_n^{(2)}(d_o, k)|} n^{-(2+\delta_2)/(2\beta+1)} = o(1)$$

on  $A_n$ , for all  $\delta_2 > 0$ , analogous to (4.3) and (4.4), we find that for all  $X \in A_n$ , by definition of  $b_n(d_o)$ ,

$$\begin{aligned} l_n(d, k) - l_n(d_o, k) + \log I_n(d, k) &\leq 2\epsilon_1 - \frac{|l_n^{(2)}(d_o, k)|}{2} (d - d_o - b_n(d_o))^2 + \frac{(l_n^{(1)}(d_o, k))^2}{2|l_n^{(2)}(d_o, k)|} \\ l_n(d, k) - l_n(d_o, k) + \log I_n(d, k) &\geq -2\epsilon_1 - \frac{|l_n^{(2)}(d_o, k)|}{2} (d - d_o - b_n(d_o))^2 + \frac{(l_n^{(1)}(d_o, k))^2}{2|l_n^{(2)}(d_o, k)|}, \end{aligned}$$

when  $n$  is large enough since  $k > k_n''$ . We now lower-bound  $b_n(d_o)$  by bounding the terms on the right in (3.6) in Lemma 3.4. By construction of  $\theta_o$  it follows that

$$r_k^{-1} \sum_{j>k} j^{-1} \theta_{o,j} = c_0 r_k^{-1} \sum_{j>k} j^{-\beta-\frac{3}{2}} / (\log j) \geq c k^{-\beta+\frac{1}{2}} (\log k)^{-1},$$

for some  $c > 0$ . Since  $X \in A_n$ ,  $2\mathcal{S}(l_n^{(1)}(d_o, k))/(nr_k) \leq 2\sqrt{k/n}\sqrt{\log n}$ . Since  $k \leq k_n'$ , this bound is  $o(k^{-\beta+\frac{1}{2}}(\log k)^{-1})$ . The last term in (3.6) is  $o(n^{\epsilon-1})$  when  $\beta > 5/2$ , and hence this term is also  $o(k^{-\beta-\frac{1}{2}}(\log k)^{-1})$ . Therefore, the last two terms in (3.6) are negligible with respect to  $r_k^{-1} \sum_{j>k} j^{-1} \theta_{o,j}$ . We deduce that  $b_n(d_o) \geq c k^{-\beta+\frac{1}{2}} (\log k)^{-1} \geq c n^{-(2\beta-1)/(4\beta+2)} (\log n)^{-(2\beta+3)/(4\beta+2)}$  for  $n$  large enough.

Consequently, when the constant  $k_v$  is chosen sufficiently small,  $\sqrt{nr_k'}(b_n(d_o) - k_v w_n (\log n)^{-1}) \geq (c - k_v) n^{1/(4\beta+2)} (\log n)^{-(\beta+1)/(2\beta+1)} := z_n \rightarrow \infty$ . We now substitute the above bounds on  $l_n(d, k) - l_n(d_o, k) + \log I_n(d, k)$  in the right hand side of (5.3), make the change of variables  $u = d - d_o - b_n(d_o)$  and obtain

$$\begin{aligned} \frac{N_{n,k}}{D_{n,k}} &\leq e^{5\epsilon_1} \frac{\int_{u \leq -k_v w_n (\log n)^{-1} - b_n(d_o)} \exp\{-\frac{nr_k u^2}{4}\} du}{\int_{|u| < w_n/2} \exp\{-\frac{nr_k u^2}{4}\} du} \\ &\leq e^{5\epsilon_1} \frac{\int_{v > z_n} \exp\{-\frac{v^2}{2}\} dv}{\int_{|v| < w_n \sqrt{nr_k}/8} \exp\{-\frac{v^2}{2}\} dv} = \mathbf{o}_{\mathbf{P}_o}(1). \end{aligned}$$

This achieves the proof of Theorem 2.2.

## 6 Conclusion

In this paper we have derived conditions leading to a BVM type of result for the long memory parameter  $d \in (-\frac{1}{2}, \frac{1}{2})$  of a stationary Gaussian process, for the class of FEXP-priors. To our knowledge such a result has not been obtained before. The result implies in particular that asymptotically credible intervals for  $d$  have good frequentist coverage.

A by-product of our results is that the *most natural prior* (Prior C) from a Bayesian perspective, which is also the prior leading to adaptive minimax rates under the loss function  $l$  on  $f$ , leads to sub-optimal estimators in terms of  $d$ . Prior A leads to optimal estimators for  $d$  however it is not adaptive. An interesting direction for future work would be to define an adaptive- minimax estimation procedure for  $d$ .

More broadly speaking, the approach considered here to derive the asymptotic posterior distribution of a finite dimensional parameter of interest in a semi-parametric problems could be used in other non - regular models, hence completing (not exhaustively) the recent works of Castillo (2010) and Bickel and Kleijn (2010).

## 7 Acknowledgements

This work was supported by the 800-20072010 grant ANR-07-BLAN-0237-01 SP Bayes.

## References

- Arbel, J. (2010). Bayesian optimal adaptive estimation using a sieve prior, submitted.
- Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika*, 80(4):817–822.
- Beran, J. (1994). *Statistics for long-memory processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.
- Bickel, P. and Kleijn, B. (2010). The semiparametric bernstein-von mises theorem.
- Castillo, I. (2010). A semiparametric bernstein von mises theorem for gaussian process priors. *Probability Theory and Related Fields*.
- Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4):1749–1766.
- Freedman, D. (1999). On the bernstein-von mises theorem with infinite-dimensional parameters. *Ann. Statist.*, 27(4):1119–1140.

- Grenander, U. and Szegő, G. (1958). *Toeplitz forms and their applications*. California Monographs in Mathematical Sciences. University of California Press, Berkeley.
- Holan, S. H. and McElroy, T. S. (2010). Tail exponent estimation via broadband log density-quantile regression. *J. Stat. Plann. Inference*, 140(12):3693–3708.
- Hurvich, C. M., Moulines, E., and Soulier, P. (2002). The FEXP estimator for potentially non-stationary linear time series. *Stochastic Process. Appl.*, 97(2):307–340.
- Jensen, M. J. (2004). Semiparametric Bayesian inference of long-memory stochastic volatility models. *J. Time Ser. Anal.*, 25(6):895–922.
- Ko, K., Qu, L., and Vannucci, M. (2009). Wavelet-based Bayesian estimation of partially linear regression models with long memory errors. *Statist. Sinica*, 19(4):1463–1478.
- Li, X. and Zhao, L. H. (2002). Bayesian nonparametric point estimation under a conjugate prior. *Stat. Probab. Lett.*, 1(4):23–30.
- Lieberman, O. and Phillips, P. C. B. (2004). Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra. *J. Time Ser. Anal.*, 25(5):733–753.
- Lieberman, O., Rosemarin, R., and Rousseau, J. (2011). Asymptotic theory for maximum likelihood estimation in stationary fractional gaussian processes, under short, long and intermediate memory. *Econometric Theory*.
- Lieberman, O., Rousseau, J., and Zucker, D. M. (2003). Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. *Ann. Statist.*, 31(2):586–612. Dedicated to the memory of Herbert E. Robbins.
- Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In *Theory and applications of long-range dependence*, pages 251–301. Birkhäuser Boston, Boston, MA.
- Philippe, A. and Rousseau, J. (2002). Non-informative priors in the case of Gaussian long-memory processes. *Bernoulli*, 8(4):451–473.
- Rivoirard, V. and Rousseau, J. (2010). Bernstein-von mises theorem for linear functionals of the density.
- Robinson, P. M. (1994). Time series with strong dependence. In *Advances in econometrics, Sixth World Congress, Vol. I (Barcelona, 1990)*, volume 23 of *Econom. Soc. Monogr.*, pages 47–95. Cambridge Univ. Press, Cambridge.
- Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23(5):1630–1661.

Rousseau, J., Chopin, N., and Liseo, B. (2010). Bayesian nonparametric estimation of the spectral density of a long memory gaussian process.

Rousseau, J. and Kruijer, W. (2011). Bayesian semi-parametric estimation of the long-memory parameter under fexp-priors.

## A Proof of Lemma 3.4

We decompose the first derivative of  $l_n(d, k)$  as  $l_n^{(1)}(d, k) = \mathcal{S}(l_n^{(1)}(d, k)) + \mathcal{D}(l_n^{(1)}(d, k))$ ,  $\mathcal{S}(l_n^{(1)}(d, k))$  being a centered quadratic form and  $\mathcal{D}(l_n^{(1)}(d, k))$  the remaining deterministic term. To simplify notations, in this proof we write  $\mathcal{S} = \mathcal{S}(l_n^{(1)}(d_o, k))$  and  $\mathcal{D} = \mathcal{D}(l_n^{(1)}(d_o, k))$ . Using (1.6) (supplement) and defining  $A = T_n^{-1}(f_{d_o, k})T_n(H_k f_{d_o, k})T_n^{-1}(f_{d_o, k})$ , we find that

$$\mathcal{D} = -\frac{1}{2}\text{tr}[(T_n(f_{d_o, k}) - T_n(f_o))A], \quad \mathcal{S} = \frac{1}{2}(X^t AX - \text{tr}[T_n(f_o)A]).$$

From (1.4) and (1.8) in the supplement it follows that

$$\begin{aligned} f_o - f_{d_o, k} &= f_{d_o, k}(e^{\Delta_{d_o, k}} - 1) = (\Delta_{d_o, k} + \frac{1}{2}e^\xi \Delta_{d_o, k}^2)f_{d_o, k} \\ &= f_{d_o, k}O(k^{-\beta+1/2}), \quad \xi \in (0, (\Delta_{d_o, k})_+). \end{aligned} \quad (\text{A.1})$$

Consequently, we have

$$\begin{aligned} \mathcal{D} &= \frac{1}{2}\text{tr}[T_n(f_{d_o, k}(\Delta_{d_o, k} + O(\Delta_{d_o, k}^2)))T_n^{-1}(f_{d_o, k})T_n(H_k f_{d_o, k})T_n^{-1}(f_{d_o, k})] \\ &= \frac{n}{4\pi} \int_{-\pi}^{\pi} H_k(x)(\Delta_{d_o, k}(x) + O(\Delta_{d_o, k}^2(x)))dx + \text{error} \\ &= \frac{n}{2} \sum_{j=k+1}^{\infty} \eta_j \theta_{o, j} + O(nk^{-2\beta-1}) + \text{error}. \end{aligned}$$

The last equality follows from (1.9) and (1.11) in the supplement. We bound the error term using Lemma 2.4 (supplement) applied to  $H_k f_{d_o, k}$  and  $f_{d_o, k}$ , whose Lipschitz constants are bounded by  $O(k)$  and  $O(k^{(3/2-\beta)_+})$ , respectively (see Lemma 3.1 in the supplement). Using that  $\|\Delta_{d_o, k}\|_\infty = O(k^{-\beta+1/2})$  (see (1.8) in the the supplement) we then find that the error is  $O(k^{3/2-\beta}n^\epsilon)$ .

The term  $\mathcal{S}$  is a centered quadratic form with variance  $\frac{1}{2}|T_n^{\frac{1}{2}}(f_o)AT_n^{\frac{1}{2}}(f_o)|^2$ . Applying once more (A.1), we find that

$$\begin{aligned} \text{tr}[(T_n(f_o)A)^2] &= \text{tr}\left[(T_n^{-1}(f_{d_o, k})T_n(H_k f_{d_o, k}))^2\right](1 + O(\|\Delta_{d_o, k}\|_\infty)) \\ &= \frac{n}{2\pi} \int_{-\pi}^{\pi} H_k^2(x)dx + O(n^\epsilon k + \|\Delta_{d_o, k}\|_\infty) = nr_k(1 + o(n^{-\delta})), \end{aligned}$$

where the term  $n^\epsilon k$  comes from Lemma 2.4 in the supplement, associated to  $f_{d_o, k}$  and  $f_{d_o, k}H_k$ . This proves the first equality in Lemma 3.4.

Similar to the decomposition of  $l_n^{(1)}(d_o, k)$ , we decompose the second derivative as  $l_n^{(2)}(d, k) = \mathcal{D}(l_n^{(2)}(d, k)) - 2\mathcal{S}_1(l_n^{(2)}(d, k)) + \mathcal{S}_2(l_n^{(2)}(d, k))$ , where

$$\mathcal{S}_1(d) = X^t A_{1,d} X - \text{tr}[T_n(f_o) A_{1,d}], \quad \mathcal{S}_2(d) = X^t A_{2,d} X - \text{tr}[T_n(f_o) A_{2,d}],$$

$$\begin{aligned} \mathcal{D}_2(d) &:= \mathcal{D}(l_n^{(2)}(d, k)) \\ &= -\frac{1}{2} \text{tr}[T_n(f_{d,k}) A_{1,d}] + \text{tr} \left[ (T_n(f_{d,k}) - T_n(f_{d_o,k})) \left( A_{1,d} - \frac{1}{2} A_{2,d} \right) \right] \\ &\quad + \text{tr} \left[ (T_n(f_{d_o,k}) - T_n(f_o)) \left( A_{1,d} - \frac{1}{2} A_{2,d} \right) \right], \end{aligned} \tag{A.2}$$

$$A_{1,d} = T_n^{-1}(f_{d,k}) (T_n(H_k f_{d,k}) T_n^{-1}(f_{d,k}))^2, \quad A_{2,d} = T_n^{-1}(f_{d,k}) T_n(H_k^2 f_{d,k}) T_n^{-1}(f_{d,k}).$$

To control  $\mathcal{D}_2(d)$  we use a first order Taylor expansion around  $d_o$ , implying that  $\mathcal{D}_2(d) = \mathcal{D}_2(d_o) + O(|d - d_o| \sup_{|d' - d_o| \leq \bar{v}_n} |\mathcal{D}'_2(d')|)$ . First we study  $\mathcal{D}_2(d_o)$ . At  $d = d_o$ , the right-hand side of (A.2) equals

$$\begin{aligned} &-\frac{n}{4\pi} \int_{-\pi}^{\pi} H_k^2(x) (1 + (e^{\Delta_{d_o,k}} - 1)) dx + O(kn^\epsilon) \\ &= -\frac{nr_k}{2} \left( 1 + O(k^{-\beta+1/2} + k^2/n^{1-\epsilon}) \right). \end{aligned} \tag{A.3}$$

The  $O(kn^\epsilon)$  term is obtained from Lemma 2.4 (supplement), applied to  $f_{2j} = H_k f_{d_o,k}$  and  $f_{2j-1} = f_{d_o,k}$ , with Lipschitz constants  $O(k)$  for the former and  $O(k^{(3/2-\beta)_+})$  for the latter, together with the bound  $\|\Delta_{d_o,k}\|_\infty = O(k^{-\beta+1/2})$ . Using

$$\begin{aligned} A'_{1,d} &= -3 (T_n^{-1}(f_{d,k}) T_n(H_k f_{d,k}))^3 T_n^{-1}(f_{d,k}) \\ &\quad + 2 T_n^{-1}(f_{d,k}) T_n(H_k^2 f_{d,k}) T_n^{-1}(f_{d,k}) T_n(H_k f_{d,k}) T_n^{-1}(f_{d,k}) \end{aligned} \tag{A.4}$$

and a similar expression for the derivative of  $d \rightarrow A_{2,d}$ , it follows that

$$\begin{aligned} |\mathcal{D}'_2(d')| &\lesssim \text{tr} \left[ (T_n(f_{d_o,k}) T_n^{-1}(f_{d',k}) + I_n) (T_n(|H_k| f_{d',k}) T_n^{-1}(f_{d',k}))^3 \right] \\ &\quad + \text{tr} \left[ (T_n(f_{d_o,k}) T_n^{-1}(f_{d',k}) + I_n) (T_n(|H_k| f_{d',k}) T_n^{-1}(f_{d',k})) T_n(H_k^2 f_{d',k}) T_n^{-1}(f_{d',k}) \right] \\ &\quad + \text{tr} \left[ (T_n(f_{d_o,k}) T_n^{-1}(f_{d',k}) + I_n) T_n(|H_k|^3 f_{d',k}) T_n^{-1}(f_{d',k}) \right] \end{aligned}$$

We control the first term of the right hand side of the above inequality, the second and third terms are controlled similarly. Note first that

$$\begin{aligned} &\text{tr} \left[ T_n(f_{d_o,k}) T_n^{-1}(f_{d',k}) (T_n(|H_k| f_{d',k}) T_n^{-1}(f_{d',k}))^3 \right] \\ &= |T_n^{\frac{1}{2}}(f_{d_o,k}) T_n^{-1}(f_{d',k}) T_n(|H_k| f_{d',k}) T_n^{-1}(f_{d',k}) T_n^{\frac{1}{2}}(|H_k| f_{d',k})|^2 \\ &\leq \|T_n^{\frac{1}{2}}(f_{d_o,k}) T_n^{-\frac{1}{2}}(f_{d',k})\|^2 \\ &\quad \times \|T_n^{-\frac{1}{2}}(f_{d',k}) T_n^{\frac{1}{2}}(|H_k| f_{d',k})\|^2 \|T_n^{-\frac{1}{2}}(f_{d',k}) T_n(|H_k| f_{d',k}) T_n^{-\frac{1}{2}}(f_{d',k})\|^2 \\ &\lesssim n^\epsilon |T_n^{-\frac{1}{2}}(f_{d',k}) T_n(|H_k| f_{d',k}) T_n^{-\frac{1}{2}}(f_{d',k})|^2, \end{aligned} \tag{A.5}$$



where the last inequality comes from Lemma 2.3 in the supplement. Note also that

$$|x|^{-2d'} \lesssim f_{d'}(x) \lesssim |x|^{-2d'} \quad \text{and} \quad T_n(|H_k|f_{d',k}) \lesssim T_n(|H_k||x|^{-2d'}), \quad T_n(|f_{d',k}|) \gtrsim T_n(|x|^{-2d'})$$

and replace  $f_{d'}$  by  $|x|^{-2d'}$  in (A.5), then

$$|T_n^{-1/2}(f_{d',k})T_n(|H_k|f_{d',k})T_n^{-1/2}(f_{d',k})|^2 \lesssim \left(\frac{n}{k} + O(k)\right)$$

using Lemma 2.4 in the supplement associated to  $|H_k||x|^{-2d'}$  which has Lipschitz constant  $k$ . This leads to  $\mathcal{D}_2(d') = O(n^\epsilon \frac{n}{k})$ , which implies that for all  $\beta > 1$ ,

$$\mathcal{D}_2(d) = \mathcal{D}_2(d_o) + \mathbf{o}(|d - d_o|n^{\epsilon+1}k^{-1}) = -\frac{nrk}{2}(1 + \mathbf{o}(n^{-\delta})).$$

For the stochastic terms in  $l_n^{(2)}(d, k)$  we need a chaining argument to control the supremum over  $d \in (d_o - \bar{v}_n, d_o + \bar{v}_n)$ . We show that for all  $\epsilon' > 0$  and  $\gamma_n = n^{\frac{1}{2} + \epsilon'}k^{-\frac{1}{2}}$ ,

$$P_o^n \left( \sup_{|d-d_o| \leq \bar{v}_n} |\mathcal{S}_1(d)| > \gamma_n \right) = o(1), \quad (\text{A.6})$$

i.e. that  $\mathcal{S}_1(d) = \mathbf{o}_{P_o}(\gamma_n)$ . The same can be shown for  $\mathcal{S}_2(d)$  using exactly the same arguments. Consider a covering of  $(d_o - \bar{v}_n, d_o + \bar{v}_n)$  by balls of radius  $n^{-1}$  centered at  $d_j$ ,  $j = 1, \dots, J_n$  with  $J_n \leq 2\bar{v}_n n$ . Then

$$\sup_{|d-d_o| < \bar{v}_n} |\mathcal{S}_1(d)| \leq \max_j |\mathcal{S}_1(d_j)| + \sup_{|d-d'| \leq n^{-1}} |\mathcal{S}_1(d) - \mathcal{S}_1(d')|,$$

and

$$\begin{aligned} P_o^n \left( \sup_{|d-d_o| \leq \bar{v}_n} |\mathcal{S}_1(d)| > \gamma_n \right) &\leq P_o^n \left( \sup_{|d-d'| \leq n^{-1}} |\mathcal{S}_1(d) - \mathcal{S}_1(d')| > \frac{1}{2}\gamma_n \right) \\ &\quad + J_n \max_{1 \leq j \leq J_n} P_o^n \left( |\mathcal{S}_1(d_j)| > \frac{1}{2}\gamma_n \right). \end{aligned} \quad (\text{A.7})$$

To control the first term on the right in (A.7), note that for a standard normal vector  $Z$  and some  $d^* \in (d, d')$ ,

$$\mathcal{S}_1(d) - \mathcal{S}_1(d') = (d - d') \left( Z^t T_n^{\frac{1}{2}}(f_o) A'_{1,d^*} T_n^{\frac{1}{2}}(f_o) Z - \text{tr} [T_n(f_o) A'_{1,d^*}] \right),$$

with  $A'_{1,d}$  as in (A.4). Using Lemma 2.3 (supplement) and the fact that  $\|AB\| \leq \|A\| \|B\|$  for all matrices  $A$  and  $B$ , it follows that  $\|T_n^{\frac{1}{2}}(f_o) A'_{1,d^*} T_n^{\frac{1}{2}}(f_o)\| = O(n^\epsilon)$ , and hence  $Z^t T_n^{\frac{1}{2}}(f_o) A'_{1,d^*} T_n^{\frac{1}{2}}(f_o) Z \leq Z^t Z \|T_n^{\frac{1}{2}}(f_o) A'_{1,d^*} T_n^{\frac{1}{2}}(f_o)\| = O(n^\epsilon) Z^t Z$ . Similarly, it follows that  $\text{tr} [T_n(f_o) A'_{1,d^*}] \lesssim n$ . Consequently, when  $\epsilon = \epsilon'/2$  we have

$$|\mathcal{S}_1(d) - \mathcal{S}_1(d')| \lesssim n^{-1} (Z^t Z n^\epsilon + n),$$

uniformly over all  $d, d'$  such that  $|d - d'| \leq n^{-1}$ . Since  $1 = o(\gamma_n)$ ,

$$P_o^n \left( \sup_{|d-d'| \leq n^{-1}} |\mathcal{S}_1(d) - \mathcal{S}_1(d')| \geq \frac{1}{2}\gamma_n \right) \leq P(Z^t Z > n^{1-\epsilon}\gamma_n/4) = o(1).$$

To bound the last term in (A.7), we apply Lemma 1.3 (supplement) to  $(Z^t AZ - \text{tr}[A])|A|^{-1}$ , with  $A = T_n^{\frac{1}{2}}(f_o)A_{1,d}T_n^{\frac{1}{2}}(f_o)$  since as seen previously  $|A|^2 = \mathbf{O}(n/k) = \mathbf{o}(\gamma_n^2 n^{-2\alpha})$  for  $\alpha$  small enough, it follows that

$$P_o^n \left( \mathcal{S}_1(d) \geq \frac{1}{2}\gamma_n \right) \leq e^{-n^\alpha/8}.$$

Since  $J_n$  increases only polynomially with  $n$ , this finishes the proof of (A.6).

## B Control of the derivatives in $\theta$ on the log-likelihood

Before stating Lemma B.1 we first give a general expression for the derivatives of  $l_n(d, k, \theta)$  with respect to  $\theta$ . For all  $j \geq 1$  and  $l = (l_1, \dots, l_j) \in \{0, 1, \dots, k\}^j$ , let  $\sigma = (\sigma(1), \dots, \sigma(|\sigma|))$  be a partition of  $\{1, \dots, j\}$ . Let  $|\sigma|$  be the number of subsets in this partition and  $\sigma(i)$  the  $i$ th subset of  $\{1, \dots, j\}$  in the partition  $\sigma$ . Denoting  $l_{\sigma(i)}$  the vector  $(l_t, t \in \sigma(i))$ , we can write

$$\nabla_{l_{\sigma(i)}} f_{d,k,\theta}(x) = \prod_{t \in \sigma(i)} \cos(l_t x) f_{d,k,\theta}(x).$$

For notational ease we write  $\nabla_{\sigma(i)} f_{d,k,\theta} := \nabla_{l_{\sigma(i)}} f_{d,k,\theta}$ . The derivative  $\frac{\partial^j l_n(d,k,\theta)}{\partial \theta_{l_1} \dots \partial \theta_{l_j}}$  can now be written in terms of the matrices

$$B_\sigma(d, \theta) = \prod_{i=1}^{|\sigma|} B_{\sigma(i)}(d, \theta), \quad B_{\sigma(i)}(d, \theta) = T_n(\nabla_{\sigma(i)} f_{d,k,\theta}) T_n^{-1}(f_{d,k,\theta}). \quad (\text{B.1})$$

There exist constants  $b_\sigma, c_\sigma$  and  $d_\sigma$  such that

$$\begin{aligned} & \frac{\partial^j l_n(d, k, \theta)}{\partial \theta_{l_1} \dots \partial \theta_{l_j}} \\ &= \sum_{\sigma \in \mathcal{S}_j} b_\sigma (X^t T_n^{-1}(f_{d,k,\theta}) B_\sigma(d, \theta) X - \text{tr} [T_n(f_o) T_n^{-1}(f_{d,k,\theta}) B_\sigma(d, \theta)]) \\ &+ \sum_{\sigma \in \mathcal{S}_j} c_\sigma \text{tr} [B_\sigma(d, \theta)] + \sum_{\sigma \in \mathcal{S}_j} d_\sigma \text{tr} [(T_n(f_o) T_n^{-1}(f_{d,k,\theta}) - I_n) B_\sigma(d, \theta)], \end{aligned} \quad (\text{B.2})$$

where  $\mathcal{S}_j$  is the set of partitions of  $\{1, \dots, j\}$ . For the first two derivatives ( $j = 1, 2$ ) the values of the constants  $b_\sigma, c_\sigma$  and  $d_\sigma$  are given below in Lemmas B.4 and B.5. For the higher order derivatives these values are not important for our purpose; we will only need that for any  $j \geq 1$ , the constant  $c_\sigma$  is zero if  $|\sigma| = 1$ .

The following lemma states that  $l_n(d, k, \theta) - l_n(d, k)$  is the sum of a Taylor-approximation  $\sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d_o, k)}{j!}$  and terms whose dependence on  $d$  can be negligible. Since the proof is involved, some of the technical details are treated in Lemmas B.2 and B.3.

**Lemma B.1.** *Given  $\beta > 1$ , let  $k \leq k_n$  and let  $d$  and  $\theta$  be such that  $l(f_o, f_{d,k,\theta}) \leq l_0^2 \delta_n^2$ . Then there exists an integer  $J$  and a constant  $\epsilon > 0$  such that uniformly over  $d \in (d_o - \bar{v}_n, d_o + \bar{v}_n)$  and  $\theta \in B_k(\bar{\theta}_{d,k}, 2l_0 \delta_n)$ ,*

$$\begin{aligned} l_n(d, k, \theta) - l_n(d, k) &= \sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d_o, k)}{j!} \\ &+ (d - d_o) \sum_{j=2}^J \frac{1}{j!} g_{n,j}(\theta - \bar{\theta}_{d,k}) + S_n(d), \end{aligned} \quad (\text{B.3})$$

where, for  $u = \theta - \bar{\theta}_{d,k}$ ,

$$g_{n,j}(u) = \sum_{l_1, \dots, l_j=0}^k u_{l_1} \cdots u_{l_j} \sum_{\sigma \in \mathcal{S}_j} (c_\sigma \text{tr}[T_{1,\sigma}(d_o, k)] + d_\sigma \text{tr}[T_{2,\sigma}(d_o, k)]), \quad (\text{B.4})$$

$$\begin{aligned} T_{1,\sigma}(d_o, k) &= \sum_{i=1}^{|\sigma|} \left( \prod_{l < i} T_n(\nabla_{\sigma(l)} f_{d_o, k}) T_n^{-1}(f_{d_o, k}) \right) \times \\ &\quad [T_n(\nabla_{\sigma(i)} f_{d_o, k} H_k) - T_n(H_k f_{d_o, k}) T_n^{-1}(f_{d_o, k}) T_n(\nabla_{\sigma(i)} f_{d_o, k})] \times \\ &\quad T_n^{-1}(f_{d_o, k}) \left( \prod_{l > i} T_n(\nabla_{\sigma(l)} f_{d_o, k}) T_n^{-1}(f_{d_o, k}) \right), \end{aligned}$$

$$T_{2,\sigma}(d_o, k) = -T_n(H_k f_{d_o, k}) T_n^{-1}(f_{d_o, k}) B_\sigma(d_o, \bar{\theta}_{d_o, k}),$$

and  $S_n(d)$  denotes any term of order

$$S_n(d) = \mathbf{O}_{\mathbf{P}_o}(1) + \mathbf{O}_{\mathbf{P}_o} \left( \frac{|d - d_o| n^{\frac{1}{2} - \delta}}{\sqrt{k}} \right) + \mathbf{O}_{\mathbf{P}_o} \left( (d - d_o)^2 \frac{n^{1 - \delta}}{k} \right). \quad (\text{B.5})$$

When  $\beta > 5/2$  and  $k \leq k'_n$ , we can choose  $J = 2$ , and (B.3) simplifies to

$$l_n(d, k, \theta) - l_n(d, k) = \sum_{j=1}^2 \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d_o, k)}{j!} + \mathbf{O}_{\mathbf{P}_o}(1). \quad (\text{B.6})$$

*Proof.* Recall that by (3.7),

$$\begin{aligned} l_n(d, k, \theta) - l_n(d, k) &= \sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d_o, k)}{j!} \\ &+ \sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j (l_n(d, k) - l_n(d_o, k))}{j!} + R_{J+1, d}(\theta). \end{aligned} \quad (\text{B.7})$$

To prove (B.3) we first show that, writing  $u = \theta - \bar{\theta}_{d,k}$ ,

$$\begin{aligned} & \sum_{j=1}^J \frac{u^{(j)} \nabla^j (l_n(d, k) - l_n(d_o, k))}{j!} \\ &= \sum_{j=1}^J \frac{1}{j!} \sum_{l_1, \dots, l_j=0}^k u_{l_1} \dots u_{l_j} \left( \frac{\partial^j l_n(d, k, \bar{\theta}_{d,k})}{\partial \theta_{l_1} \dots \partial \theta_{l_j}} - \frac{\partial^j l_n(d_o, k, \bar{\theta}_{d_o,k})}{\partial \theta_{l_1} \dots \partial \theta_{l_j}} \right) \quad (\text{B.8}) \\ &= (d - d_o) \sum_{j=1}^J \frac{1}{j!} g_{n,j}(u) + O(S_n(d)). \end{aligned}$$

This result is combined with (B.7) and Lemma B.3 below, by which  $g_{n,1}(u) = O(S_n(d))$ . It then follows that  $l_n(d, k, \theta) - l_n(d, k)$  equals

$$\sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d_o, k)}{j!} + (d - d_o) \sum_{j=2}^J \frac{1}{j!} g_{n,j}(u) + R_{J+1,d}(\theta) + O(S_n(d)).$$

The final step is to prove that  $R_{J+1,d}(\theta)$  is  $\mathbf{op}_{\mathbf{P}_o}(1)$  and hence  $O(S_n(d))$ ; to this end  $J$  needs to be sufficiently large.

First we prove (B.8). For the factors  $\frac{\partial^j l_n(d, k, \bar{\theta}_{d,k})}{\partial \theta} - \frac{\partial^j l_n(d, k, \bar{\theta}_{d_o,k})}{\partial \theta}$  we substitute (B.2). In Lemma B.2 below we give expressions for each of the terms therein, which we substitute in (B.8). The main terms are  $(d - d_o) \text{tr}[T_{1,\sigma}(d_o, k)]$  and  $(d - d_o) \text{tr}[T_{2,\sigma}(d_o, k)]$  in (G.2) and (G.3), which after substitution in (B.8) give the term  $(d - d_o) \sum_{j=1}^J \frac{1}{j!} g_{n,j}(u)$  on the right. The other terms in (G.1)-(G.3) that enter (B.8) through (B.2) are  $O(S_n(d))$ . This is due to the summation over  $u_{l_1}, \dots, u_{l_j}$  in (B.8), and the Cauchy-Schwarz inequality by which

$$\left| \sum_{l_1, \dots, l_j=0}^k u_{l_1} \dots u_{l_j} \right| \leq (\sqrt{k} \|u\|)^j \leq (2l_0 \sqrt{k} \delta_n)^j = o(n^{-\delta}), \quad (\text{B.9})$$

for some  $\delta > 0$ , as  $\|u\| \leq 2l_0 \delta_n$  and (B.8) is proved. We now control  $R_{J+1,d}(\theta)$ .

Combining (3.8) and the first inequality in (B.9), we obtain

$$|R_{J+1,d}(\theta)| \leq \frac{1}{(J+1)!} (\sqrt{k} \delta_n)^{J+1} \max_{l_1, \dots, l_{J+1}} \sup_{\|\bar{\theta} - \bar{\theta}_{d,k}\| \leq 2l_0 \delta_n} \left| \frac{\partial^{J+1} l_n(d, k, \tilde{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_{J+1}}}(x) \right|.$$

We give a direct bound on this derivative using (B.2). For all partitions  $\sigma$  of  $\{l_1, \dots, l_{J+1}\}$  and all  $(l_1, \dots, l_{J+1}) \in \{1, \dots, k\}^{J+1}$ , we bound  $\|B_{\sigma(i)}(d, \theta)\|$ , using  $\|B_{\sigma(i)}(d, \theta)\| \leq \|T_n^{\frac{1}{2}}(\nabla_{\sigma(i)} f_{d,k,\theta}) T_n^{-\frac{1}{2}}(f_{d,k,\theta})\|^2$  (see (1.6)). We bound  $\|T_n^{\frac{1}{2}}(\nabla_{\sigma(i)} f_{d,k,\theta}) T_n^{-\frac{1}{2}}(f_{d,k,\theta})\|$  by application of Lemma 2.3 (supplement) with  $f = f_{d,k,\theta}$  and  $g = \nabla_{\sigma(i)} f_{d,k,\theta}$ . The constant  $M$  in this lemma is bounded by

$$\sum_{j=0}^k |\theta_j| \leq \sum_{j=0}^k |(\bar{\theta}_{d,k})_j| + \sum_{j=0}^k |\theta_j - (\bar{\theta}_{d,k})_j| \leq 2\sqrt{L} + \sqrt{k} \|\theta - \bar{\theta}_{d,k}\| = O(1),$$

since  $\sum_{i=0}^k |(\bar{\theta}_{d,k})_i| \leq 2\sqrt{L}$  (by Lemma 3.3) and  $\|\theta - \bar{\theta}_{d,k}\| \leq \delta_n$ . Consequently, Lemma 2.3 (supplement) implies that

$$\|B_{\sigma(i)}(d, \theta)\| \leq K, \quad (\text{B.10})$$

where  $K$  depends only on  $L, L_o$  and not on  $n, d$  nor  $\theta$ . From the relations in (1.6) and the definition of  $B_\sigma$  it follows that for any  $\sigma, d, \theta$ ,

$$\begin{aligned} |X^t B_\sigma(d, \theta) X| &\leq X^t T_n^{-1}(f_o) X K^{|\sigma|} \|T_n^{\frac{1}{2}}(f_o) T_n^{-\frac{1}{2}}(f_{d,k})\|^2 \leq X^t T_n^{-1}(f_o) X K^{|\sigma|} n^\epsilon, \\ |\text{tr}[B_\sigma(d, \theta)]| &\leq n K^{|\sigma|} \|T_n^{\frac{1}{2}}(f_o) T_n^{-\frac{1}{2}}(f_{d,k})\|^2 \leq n^{1+\epsilon} K^{|\sigma|}. \end{aligned}$$

Therefore we have the bound

$$|R_{J+1}(d, \theta)| \leq C K^{J+1} n^\epsilon (\sqrt{k} \|\theta - \bar{\theta}_{d,k}\|)^{J+1} (X^t T_n^{-1}(f_o) X + n). \quad (\text{B.11})$$

Since  $k \leq k_n$ ,  $\|\theta - \bar{\theta}_{d,k}\| \leq \delta_n$  and the term  $X^t T_n^{-1}(f_o) X$  in (B.11) is the sum of  $n$  independent standard normal variables, there is a constant  $c > 0$  such that

$$P_o \left( \sup_{|d-d_o| \leq \bar{v}_n} \sup_{\|\theta - \bar{\theta}_{d,k}\| \leq 2l_o \delta_n} |R_{J+1}(d, \theta)| > n^{-\epsilon} \right) \leq e^{-cn},$$

provided we choose  $J$  such that  $(J+1)(1-1/\beta) > 2$ . This concludes the proof of (B.3).

To prove (B.6) we first show that for  $J=2$ ,  $|R_{J+1}(d, \theta)| = \mathbf{o}_{\mathbf{P}_o}(1)$ . Since  $k \leq k'_n$ ,  $\beta > 5/2$  and  $\|\theta - \bar{\theta}_{d,k}\| \leq 2l_o \epsilon_n$ , we can choose  $J+1 = 3 > (2\beta+1)/(\beta-\frac{1}{2})$ , and the preceding inequality becomes

$$P_o \left( \sup_{|d-d_o| \leq w_n} \sup_{\|\theta - \bar{\theta}_{d,k}\| \leq 2l_o \epsilon_n} |R_3(d, \theta)| > n^{-\epsilon} \right) \leq e^{-cn}.$$

Combining this result with (B.8), it only remains to be shown that  $(d-d_o)g_{n,1}(u)$  and  $(d-d_o)g_{n,2}(u)$  are  $\mathbf{o}_{\mathbf{P}_o}(1)$ . Recall from Corollary 3.1 that  $|d-d_o| = o(n^{\epsilon-(\beta-1/2)/(2\beta+1)})$  for all  $\epsilon > 0$ . Consequently,

$$|d-d_o| n^{\frac{1}{2}} k^{-\frac{1}{2}} = o(n^{\epsilon+1/(2\beta+1)} k^{-\frac{1}{2}}) = o((\sqrt{k}\epsilon_n)^{-1}), \quad (d-d_o)^2 \frac{n}{k} = o((\sqrt{k}\epsilon_n)^{-1})$$

for all  $\beta > 2$ . This implies that  $S_n(d) = \mathbf{o}_{\mathbf{P}_o}(1)$  and that, by Lemma B.3,  $(d-d_o)g_{n,1}(u) = \mathbf{o}(1)$ . Also, for all  $(l_1, l_2) \in \{1, \dots, k\}^j$  and all partitions  $\sigma$  of  $(l_1, l_2)$ , the limiting integral of  $\text{tr}[T_{1,\sigma}]$  is equal to 0. Since  $\beta > 5/2$  the Lipschitz constants of the functions  $f_{d_o,k}$  or  $f_o$  are  $O(1)$ , so that Lemma 2.4 (supplement) implies  $\text{tr}[T_{1,\sigma}(d_o, k)] = O(n^\epsilon k)$ . Similarly,

$$\text{tr}[T_{2,\sigma}(d_o, k)] = \frac{n}{2\pi} \int_{-\pi}^{\pi} H_k(x) \cos(l_1 x) \cos(l_2 x) dx + O(n^\epsilon k).$$

Thus we have

$$(d-d_o)g_{n,2}(u) = \frac{n(d-d_o)}{2\pi} \sum_{l_1, l_2=0}^k u_{l_1} u_{l_2} \int_{-\pi}^{\pi} H_k(x) \cos(l_1 x) \cos(l_2 x) dx + \mathbf{o}(1),$$

which is  $\mathbf{o}(1)$ . This completes the proof of Lemma B.1.  $\square$

The proof of the following lemma is given in section 4 of the supplement.

**Lemma B.2.** *Let  $W_\sigma(d)$  denote any of the quadratic forms*

$$X^t T_n^{-1}(f_{d,k}) B_\sigma(d, \bar{\theta}_{d,k}) X - \text{tr} [T_n(f_o) T_n^{-1}(f_{d,k}) B_\sigma(d, \bar{\theta}_{d,k})]$$

in (B.2). For any  $j \leq J$ ,  $(l_1, \dots, l_j) \in \{0, \dots, k\}^j$  and  $\sigma \in \mathcal{S}_j$ , we have

$$|W_\sigma(d) - W_\sigma(d_o)| = \mathbf{o}_{\mathbf{P}_o}(|d - d_o| n^{\frac{1}{2} + \epsilon} k^{-\frac{1}{2}}), \quad (\text{B.12})$$

$$\begin{aligned} & \text{tr} [B_\sigma(d, \bar{\theta}_{d,k})] - \text{tr} [B_\sigma(d_o, \bar{\theta}_{d_o})] \\ &= (d - d_o) \text{tr} [T_{1,\sigma}(d_o, k)] + (d - d_o)^2 \mathbf{o}(n^{\epsilon + \frac{1}{2}} k^{-\frac{1}{2} + (1 - \beta/2)_+}) \\ &= (d - d_o) \text{tr} [T_{1,\sigma}(d_o, k)] + (d - d_o)^2 \mathbf{o}(n^{1 - \delta}/k), \end{aligned} \quad (\text{B.13})$$

$$\begin{aligned} & \text{tr} [(T_n(f_o) T_n^{-1}(f_{d,k}) - I_n) B_\sigma(d, \bar{\theta}_{d,k})] - \text{tr} [(T_n(f_o) T_n^{-1}(f_{d_o,k}) - I_n) B_\sigma(d_o, \bar{\theta}_{d_o,k})] \\ &= (d - d_o) \text{tr} [T_{2,\sigma}(d_o, k)] + (d - d_o)^2 \mathbf{o}(n/k) + (d - d_o) \mathbf{o}(n^{\epsilon + \frac{1}{2}} k^{-\frac{1}{2}}). \end{aligned} \quad (\text{B.14})$$

**Lemma B.3.** *For all  $\beta > 1$  there exists a constant  $\delta > 0$  such that uniformly over  $\|\theta - \bar{\theta}_{d,k}\| \leq \delta_n$ ,*

$$|g_{n,1}(\theta - \bar{\theta}_{d,k})| = \mathbf{o}(n^{1/2 - \delta} k^{-1/2}).$$

*Proof.* For  $u = \theta - \bar{\theta}_{d,k}$ , we have

$$g_{n,1}(u) = -\frac{1}{2} \text{tr} \left[ T_n(H_k f_{d_o, \bar{\theta}_{d_o, k}}) T_n^{-1}(f_{d_o, \bar{\theta}_{d_o, k}}) T_n(u^t \nabla f_{d_o, \bar{\theta}_{d_o, k}}) T_n^{-1}(f_{d_o, \bar{\theta}_{d_o, k}}) \right].$$

This follows from (B.4) and Lemma below, by which  $b_\sigma = d_\sigma = \frac{1}{2}$  and  $c_\sigma = 0$  (the only partition for  $j = 1$  being  $\sigma = (\{l\})$ ). By Lemma 2.4 (supplement)  $g_{n,1}(u)$  converges to zero, but at a rate slower than  $n^{1/2 - \delta} k^{-1/2}$ . To obtain the  $\mathbf{o}(n^{1/2 - \delta} k^{-1/2})$  term, we write

$$g_{n,1}(u) = \Delta_1 + \Delta_2 + \Delta_3,$$

and bound the terms on the right using the other lemmas in section 2 of the supplement. We first prove that

$$\begin{aligned} \Delta_1 &= \text{tr} \left[ T_n(H_k f_{d_o, k}) T_n(f_{d_o, k}^{-1}) T_n(u^t \nabla f_{d_o, k}) T_n(f_{d_o, k}^{-1}) \right] - \\ & \quad (16\pi^4) \text{tr} [T_n(H_k) T_n(u^t \mathbf{cos})] = \mathbf{o}(1), \end{aligned}$$

where  $\mathbf{cos}(x) = (1, \cos(x), \dots, \cos(kx))$ . We then prove that

$$\Delta_2 = \text{tr} [T_n(H_k) T_n(u^t \mathbf{cos})] = 0,$$

and finally that

$$\begin{aligned} \Delta_3 &= \text{tr} [T_n(H_k f_{d_o, k}) T_n^{-1}(f_{d_o, k}) T_n(u^t \nabla f_{d_o, k}) T_n^{-1}(f_{d_o, k})] \\ & \quad - \text{tr} \left[ T_n(H_k f_{d_o, k}) T_n \left( \frac{f_{d_o, k}^{-1}}{4\pi^2} \right) T_n(u^t \nabla f_{d_o, k}) T_n \left( \frac{f_{d_o, k}^{-1}}{4\pi^2} \right) \right] \\ &= \mathbf{o}(n^{1/2 - \delta} k^{-1/2}). \end{aligned}$$

To bound  $\Delta_1$  we use Lemma 2.5 (supplement) with  $b_1(x) = H_k(x)$ ,  $b_2(x) = u^t \mathbf{cos}$  and  $L = k^{3/2-\beta}$ . Equation (2.6) then implies that

$$|\Delta_1| \leq C\sqrt{k}\|u\|n^\epsilon \left(1 + k^{3/2-\beta}k^{-1/2}\right) = \mathbf{o}(1).$$

To bound  $\Delta_2$  note that for  $l = 0, \dots, k$  and all  $j_1, j_2 \leq n$ ,

$$(T_n(\cos(lx)))_{j_1, j_2} = \mathbb{I}_{|j_1 - j_2| = l}, \quad (T_n(H_k))_{j_1, j_2} = \sum_{j=k+1}^n \eta_j \mathbb{I}_{j=|j_1 - j_2|}.$$

Therefore

$$\mathrm{tr} [T_n(H_k)T_n(\cos(l.))] = \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{j=k+1}^n \eta_j \mathbb{I}_{j=|j_1 - j_2|} \mathbb{I}_{|j_1 - j_2| = l} = 0,$$

since  $l \leq k$  and  $j > k$ . We now turn to  $\Delta_3$ . Following Lieberman et al. (2011), we consider separately the positive and negative parts of  $H_k$  and of  $u^t \mathbf{cos}$ . Hence we may treat these functions as if they were positive. We first define, for  $\tilde{f}_{d_o, k} = (4\pi^2 f_{d_o, k})^{-1}$ ,

$$\begin{aligned} A_1 &= T_n(H_k f_{d_o, k}) T_n^{-1}(f_{d_o, k}), & B_1 &= T_n(H_k f_{d_o, k}) T_n(\tilde{f}_{d_o, k}), \\ A_2 &= T_n(u^t \nabla f_{d_o, k}) T_n^{-1}(f_{d_o, k}), & B_2 &= T_n(u^t \nabla f_{d_o, k}) T_n(\tilde{f}_{d_o, k}), \\ \tilde{A} &= T_n^{\frac{1}{2}}(H_k f_{d_o, k}) T_n^{-1}(f_{d_o, k}) T_n^{\frac{1}{2}}(u^t \nabla f_{d_o, k}), \\ \tilde{B} &= T_n^{\frac{1}{2}}(H_k f_{d_o, k}) T_n(\tilde{f}_{d_o, k}) T_n^{\frac{1}{2}}(u^t \nabla f_{d_o, k}), \\ \Delta &= I_n - T_n(f_{d_o, k}) T_n(\tilde{f}_{d_o, k}). \end{aligned}$$

Using the same computations as in Lieberman et al. (2011), we find that

$$\begin{aligned} |\Delta_3| &\lesssim |\mathrm{tr} [B_1 B_2 \Delta]| + |\tilde{A} - \tilde{B}| |T_n^{\frac{1}{2}}(H_k f_{d_o, k}) T_n(\tilde{f}_{d_o, k}) \Delta T_n^{\frac{1}{2}}(u^t \nabla f_{d_o, k})| \\ &\quad + |\Delta|^2 \sqrt{k} \|u\| n^\epsilon \\ &\lesssim \sqrt{k} \|u\| n^\epsilon k^{3/2-\beta} + |\mathrm{tr} [B_1 B_2 \Delta]|. \end{aligned}$$

The first term on the right is  $\mathbf{o}(n^{1/2-\delta}k^{-1/2})$ . We bound the last term using Lemma 2.5 (supplement) with  $b_1 = H_k$ ,  $b_2 = u^t \mathbf{cos}$  and  $b_3 = 1$ , which implies that  $\mathrm{tr} [B_1 B_2 \Delta] = 0 + O(\sqrt{k}\|u\|n^\epsilon k^{(3/2-\beta)_+}) = \mathbf{o}(1)$ . This achieves the proof of Lemma B.3.  $\square$

**Lemma B.4.** *Suppose that  $k \leq k_n$  and that  $l(f_o, f_{d_o, k}) \leq l_0^2 \delta_n^2$ . Then all elements of  $\nabla_l l_n(d_o, k)$  ( $l = 0, \dots, k$ ) are the sum of a centered quadratic form,  $\mathcal{S}(\nabla_l l_n(d_o, k))$  with a variance equal to  $\frac{n}{2}(1 + \mathbf{o}(1))$  and a deterministic term,  $\mathcal{D}(\nabla_l l_n(d_o, k))$  which is  $\mathbf{o}(k^{(3/2-\beta)_+} n^\epsilon)$ .*

*Proof.* For all  $l = 0, \dots, k$ , we have

$$\nabla_l l_n(d_o, k) = \mathcal{S}(\nabla_l l_n(d_o, k)) + \mathcal{D}(\nabla_l l_n(d_o, k)),$$

where

$$\begin{aligned}\mathcal{S}(\nabla_l l_n(d_o, k)) &= \frac{1}{2} X^t T_n^{-1}(f_{d_o, k}) T_n(\nabla_l f_{d_o, k}) T_n^{-1}(f_{d_o, k}) X \\ &\quad - \frac{1}{2} \text{tr} [T_n(f_o) T_n^{-1}(f_{d_o, k}) T_n(\nabla_l f_{d_o, k}) T_n^{-1}(f_{d_o, k})], \\ \mathcal{D}(\nabla_l l_n(d_o, k)) &= \frac{1}{2} \text{tr} [(T_n(f_o) T_n^{-1}(f_{d_o, k}) - I_n) T_n(\nabla_l f_{d_o, k}) T_n^{-1}(f_{d_o, k})].\end{aligned}$$

Note that this is a special case of (B.2), with  $j = 1$ ,  $b_\sigma = d_\sigma = \frac{1}{2}$  and  $c_\sigma = 0$ , the only partition being  $\sigma = (\{l\})$ . The variance of  $\mathcal{S}(\nabla_l l_n(d_o, k))$  is equal to

$$\begin{aligned}\text{tr}[(T_n(f_o) T_n^{-1}(f_{d_o, k}) T_n(\nabla_l f_{d_o, k}) T_n^{-1}(f_{d_o, k}))^2] \\ = \frac{n}{2\pi} \int_{-\pi}^{\pi} \left( \frac{f_o}{f_{d_o, k}}(x) \right)^2 \cos^2(lx) dx + \mathbf{O}(n^\epsilon k),\end{aligned}$$

since Lemma 2.4 (supplement) implies that the approximation error of the trace by its limiting integral is of order  $O(n^\epsilon(k + k^{2(3/2-\beta)\vee 0})) = O(n^\epsilon k)$ . Since  $\frac{f_o}{f_{d_o, k}} = e^{\Delta_{d_o, k}}$  (see (A.1)), the integral in the preceding equation is

$$\begin{aligned}\frac{n}{2\pi} \int_{-\pi}^{\pi} (1 + 2\Delta_{d_o, k} + O(\Delta_{d_o, k}^2)) \cos^2(lx) dx \\ = \frac{n}{2} + 2na_{2l}(d_o) + O(n\delta_n^2) = \frac{n}{2}(1 + \mathbf{o}(1)),\end{aligned}$$

where  $a_l$  is defined at the beginning of the supplement. Lemma 1.3 (supplement) then implies that the centered quadratic form is of order  $\mathbf{O}_{\mathbf{P}_o}(n^{\epsilon+1/2})$ . Similarly, Lemma 2.4 (supplement) implies that

$$\begin{aligned}\mathcal{D}(\nabla_l l_n(d_o, k)) &= \frac{n}{2\pi} \int_{-\pi}^{\pi} \frac{(f_o - f_{d_o, k})}{f_{d_o, k}}(x) \cos(lx) dx + \mathbf{O}(n^\epsilon \|\Delta_{d_o, k}\|_1 k) \\ &= \frac{n}{2\pi} \int_{-\pi}^{\pi} \cos(lx) \Delta_{d_o, k}(x) dx + \mathbf{O}(nk^{-2\beta}) + \mathbf{O}(n^\epsilon k^{3/2-\beta}) \\ &= \mathbf{O}(n^\epsilon k^{(3/2-\beta)_+})\end{aligned}$$

which completes the proof of Lemma B.4.  $\square$

**Lemma B.5.** *Let  $A(d)$  be the  $(k+1) \times (k+1)$  matrix with entries  $A_{l_1, l_2}(d) = a_{l_1+l_2}(d)$ , where  $a_l(d) = 1_{l>k}(\theta_{o, l} - 2l^{-1}(d_o - d))$ . Suppose that  $k \leq k_n$  and that  $l(f_o, f_{d, k}) \leq l_0^2 \delta_n^2$ . Then  $J_n(d, k) = -\nabla^2 l_n(d, k, \theta) \Big|_{\theta=\bar{\theta}_{d, k}}$  satisfies*

$$\forall l_1, l_2 \leq k, \quad |(J_n(d, k) - J_n(d_o, k))_{l_1, l_2}| = \mathbf{O}_{\mathbf{P}_o}(|d - d_o| n^\epsilon k + S_n(d)) = \mathbf{O}_{\mathbf{P}_o}(n/k) \quad (\text{B.15})$$

uniformly over  $d \in (d_o - \bar{v}_n, d_o + \bar{v}_n)$  and  $k \leq k_n$ . We also have for all  $l_1, l_2$

$$[J_n(d_o, k) - \frac{n}{2} I_{k+1} - \frac{n}{2} A(d_o)]_{l_1, l_2} := n(R_{2s})_{l_1, l_2} + n(R_{2d})_{l_1, l_2}, \quad (\text{B.16})$$



where  $(R_{2s})_{l_1, l_2}$  is a centered quadratic form of order  $\mathbf{O}_{\mathbf{P}_o}(n^{-1/2+\epsilon})$  and  $(R_{2d})_{l_1, l_2}$  is a deterministic term of order  $o(kn^{\epsilon-1})$ . For the matrix  $A$ , we have  $\|A(d_o)\| = o(1)$  and  $|A(d_o)| = O(1)$ .

In particular, (B.16) implies that  $|J_n(d_o, k) - \frac{n}{2}I_{k+1} - \frac{n}{2}A(d_o)| = \mathbf{O}_{\mathbf{P}_o}(kn^{1/2+\epsilon}) + o(k^2n^\epsilon)$ .

*Proof.* Let  $d$  and  $k \leq k_n$  be such that  $l(f_o, f_{d,k}) \leq l_0^2 \delta_n^2$  so that  $d \in (d_o - \bar{v}_n, d_o + \bar{v}_n)$  (see Corollary 3.1). Lemma B.1 implies that for all  $l_1, l_2 \leq k$ ,

$$\begin{aligned} & (J_n(d, k))_{l_1, l_2} - (J_n(d_o, k))_{l_1, l_2} \\ & := -(d - d_o) \sum_{\sigma \in \mathcal{S}(l_1, l_2)} (c_\sigma \text{tr}[T_{1, \sigma}(d_o, k)] + d_\sigma \text{tr}[T_{2, \sigma}(d_o, k)]) + \mathbf{O}_{\mathbf{P}_o}(S_n(d)). \end{aligned}$$

Lemma 2.4 (supplement) implies that

$$\text{tr}[T_{i, \sigma}(d_o, k)] = O(kn^\epsilon), \quad i = 1, 2$$

so that (B.15) is satisfied since this term is  $\mathbf{O}_{\mathbf{P}_o}(n^{1-\delta}/k)$ . We then use expression (B.2), with  $\sigma \in \{(\{1\}, \{2\}), (\{1, 2\})\}$  and we denote  $\sigma_1$  and  $\sigma_2$  the first and the second partition respectively. Note that  $c_{\sigma_1} = d_{\sigma_2} = 1/2$ ,  $c_{\sigma_2} = 0$  and  $d_{\sigma_1} = 1$ . From Lemma 2.4 (supplement), the quadratic form in  $(J_n(d_o, k))_{l_1, l_2}$  is associated to a matrix whose Frobenius-norm is  $O(\sqrt{n})$  and whose spectral norm is  $O(n^\epsilon)$ . Hence, this quadratic form is  $\mathbf{O}_{\mathbf{P}_o}(n^{1/2+\epsilon})$ . Also by Lemma 2.4 (supplement), the deterministic terms can be written as

$$\begin{aligned} & \frac{n}{4\pi} \int_{-\pi}^{\pi} \cos(l_1 x) \cos(l_2 x) (1 + \Delta_{d_o, k})(x) dx + o(kn^\epsilon) \\ & = \frac{n}{2} (\mathbb{I}_{l_1=l_2} + a_{l_1+l_2}(d_o)) + o(kn^\epsilon), \end{aligned}$$

and Lemma B.5 is proved.  $\square$

## C Proof of Lemma 3.5

Under the conditions of Theorem 2.1 we have  $k = k_n$  and  $\beta > 1$ , and we may assume (by Lemma 3.2) that  $l(f_o, f_{d,k,\theta}) \leq l_0^2 \delta_n^2$ . Fixing  $d$  and  $k$ , we develop  $\theta \rightarrow l_n(d, k, \theta)$  in  $\bar{\theta}_{d,k}$ . From Lemma B.1 in Appendix B it follows that

$$\begin{aligned} l_n(d, k, \theta) - l_n(d, k) &= \sum_{j=1}^J \frac{(\theta - \bar{\theta}_{d,k})^{(j)} \nabla^j l_n(d_o, k)}{j!} \\ &+ (d - d_o) \sum_{j=2}^J \frac{g_{n,j}(\theta - \bar{\theta}_{d,k})}{j!} + S_n(d), \end{aligned} \tag{C.1}$$

where  $S_n(d)$  is as in (B.5). Substituting (C.1) in the definition of  $I_n(d, k)$  in (3.1), we obtain

$$\begin{aligned}
I_n(d, k) &= \int_{\|\theta - \bar{\theta}_{d,k}\| \leq 2l_0\delta_n} e^{l_n(d,k,\theta) - l_n(d,k)} d\pi_{\theta|k}(\theta) \\
&= e^{S_n(d)} \pi_{\theta|k}(\bar{\theta}_{d,k}) \int_{\|u\| \leq 2l_0\delta_n} e^{\sum_{j=1}^J \frac{1}{j!} u^{(j)} \nabla^j l_n(d_o, k) + (d-d_o) \sum_{j=2}^J \frac{g_{n,j}(u)}{j!} + h_k u} du \\
&= e^{S_n(d)} \pi_{\theta|k}(\bar{\theta}_{d_o, k}) \int_{\|u\| \leq 2l_0\delta_n} e^{\sum_{j=1}^J \frac{1}{j!} u^{(j)} \nabla^j l_n(d_o, k) + (d-d_o) \sum_{j=2}^J \frac{g_{n,j}(u)}{j!} + h_k u} du.
\end{aligned} \tag{C.2}$$

The first equality follows from the definition of  $I_n(d, k)$  and Lemma 3.3, by which we may replace the domain of integration by  $\{\theta : \|\theta - \bar{\theta}_{d,k}\| \leq 2l_0\delta_n\}$ . The second equality follows from the assumptions on  $\pi_{\theta|k}$  in **prior A**, the transformation  $u = \theta - \bar{\theta}_{d,k}$  and substitution of (C.1). Also the third equality follows from the assumptions on  $\pi_{\theta|k}$ : these imply that

$$|\log \pi_{\theta|k}(\bar{\theta}_{d,k}) - \log \pi_{\theta|k}(\bar{\theta}_{d_o, k})| = |d_o - d| |h_k^t \eta_{[k]}| + o(1) = O(|d - d_o| (n/k)^{\frac{1}{2} - \epsilon}) + o(1),$$

for some  $\epsilon > 0$ . Thus, the factor  $e^{S_n(d)} \pi_{\theta|k}(\bar{\theta}_{d,k})$  on the second line of (C.2) may be replaced by  $e^{S_n(d)} \pi_{\theta|k}(\bar{\theta}_{d_o, k})$ . Because  $S_n(d_o) = \mathbf{0}_{\mathbf{P}_o}(1)$ , (C.2) implies that

$$I_n(d_o, k) = (1 + \mathbf{0}_{\mathbf{P}_o}(1)) \int_{\|u\| \leq 2l_0\delta_n} \exp \left\{ h_k u + \sum_{j=1}^J \frac{u^{(j)} \nabla^j l_n(d_o, k)}{j!} \right\} du. \tag{C.3}$$

The most involved part of the proof is to establish the bounds

$$\begin{aligned}
&\pi_{\theta|k}(\bar{\theta}_{d_o, k}) \int_{\|u\| \leq l_0\delta_n} \exp \left\{ h_k u + \sum_{j=1}^J \frac{u^{(j)} \nabla^j l_n(d_o, k)}{j!} \right\} du \\
&\leq I_n(d, k) = e^{S_n(d)} \pi_{\theta|k}(\bar{\theta}_{d_o, k}) \int_{\|u\| \leq 2l_0\delta_n} e^{\sum_{j=1}^J \frac{1}{j!} u^{(j)} \nabla^j l_n(d_o, k) + (d-d_o) \sum_{j=2}^J \frac{g_{n,j}(u)}{j!} + h_k u} du \\
&\leq \pi_{\theta|k}(\bar{\theta}_{d_o, k}) \int_{\|u\| \leq 3l_0\delta_n} \exp \left\{ h_k u + \sum_{j=1}^J \frac{u^{(j)} \nabla^j l_n(d_o, k)}{j!} \right\} du.
\end{aligned} \tag{C.4}$$

Since the posterior distribution of  $\theta$  conditional on  $k = k_n$  and  $d = d_o$  concentrates at  $\bar{\theta}_{d_o, k}$  at a rate bounded by  $l_0\delta_n$  (this follows from Lemma 3.2, with the restriction to  $d = d_o$ ), the left- and right-hand side of (C.4) are asymptotically equal, up to a factor  $(1 + \mathbf{0}_{\mathbf{P}_o}(1))$ . By (C.3), the left- and right-hand side are actually equal to  $I_n(d_o, k)$ . This implies that  $I_n(d, k) = e^{S_n(d)} I_n(d_o, k)$ , which is the required result.

In the remainder we prove (C.4). To do so we construct below a change of variables  $v = \psi(u)$ , which satisfies

$$\begin{aligned} h_k v + \sum_{j=1}^J \frac{v^{(j)} \nabla^j l_n(d_o, k)}{j!} &= h_k u + \sum_{j=1}^J \frac{u^{(j)} \nabla^j l_n(d_o, k)}{j!} \\ &+ (d - d_o) \sum_{j=2}^J \frac{g_{n,j}(u)}{j!} + O(S_n(d)), \end{aligned} \quad (\text{C.5})$$

for all  $\|u\| \leq 2l_0 \delta_n$ . We first define the notation required in the definition of  $\psi$  in (C.8) below. Recall from (B.4) in Lemma B.1 that  $g_{n,j}(u)$  can be decomposed as

$$g_{n,j}(u) = n \sum_{\sigma \in \mathcal{S}_j} (c_\sigma - d_\sigma) \sum_{l_1, \dots, l_j=0}^k u_{l_1} \dots u_{l_j} g_{l_1, \dots, l_j}^{(j)}, \quad j = 2, \dots, J,$$

where  $g_{l_1, \dots, l_j}^{(j)}$  depends on  $\sigma$ . For ease of presentation however we omit this dependence in the notation. Using Lemma 2.4 (supplement) and (B.4) in Lemma B.1, it follows that for all  $j \geq 2$  and  $(l_1, \dots, l_j) \in \{0, \dots, k\}^j$ ,

$$\begin{aligned} g_{l_1, \dots, l_j}^{(j)} &= \gamma_{l_1, \dots, l_j}^{(j)} + r_{l_1, \dots, l_j}^{(j)}, \\ \gamma_{l_1, \dots, l_j}^{(j)} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} H_k(x) \cos(l_1 x) \dots \cos(l_j x) dx. \end{aligned} \quad (\text{C.6})$$

Let  $\bar{G}^{(2)}$  denote the matrix with elements  $\gamma_{l_1, l_2}^{(2)}$ , and  $G^{(2)}$  the matrix with elements  $g_{l_1, l_2}^{(2)}$ . By direct calculation it follows that

$$\gamma_{l_1, l_2}^{(2)} = \mathbb{I}_{l_1 + l_2 > k} \frac{1}{2(l_1 + l_2)}. \quad (\text{C.7})$$

Similarly, for all  $j \geq 3$  and  $l_1, \dots, l_j \in \{0, 1, \dots, k\}$  we define

$$(G^{(j)}(u))_{l_1, l_2} = \sum_{l_3, \dots, l_j=0}^k g_{l_1, \dots, l_j}^{(j)} u_{l_3} \dots u_{l_j}, \quad (\bar{G}^{(j)}(u))_{l_1, l_2} = \sum_{l_3, \dots, l_j=0}^k \gamma_{l_1, \dots, l_j}^{(j)} u_{l_3} \dots u_{l_j}.$$

In contrast to  $G^{(2)}$  and  $\bar{G}^{(2)}$ ,  $G^{(j)}(u)$  and  $\bar{G}^{(j)}(u)$  depend on  $u$ . For notational convenience we will also write  $G^{(2)}(u)$  and  $\bar{G}^{(2)}(u)$ . Finally, let  $\tilde{I}_k = J_n(d_o, k)/n$  be the normalized Fisher information.

We now define the transformation  $\psi$ :

$$\psi(u) = (I_{k+1} - (d - d_o)D(u))u, \quad \text{with} \quad (\text{C.8})$$

$$D(u) = (\tilde{I}_k + L(u))^{-1} G^t(u), \quad G(u) = \sum_{j=2}^J \frac{1}{j!} \sum_{\sigma \in \mathcal{S}_j} (c_\sigma - d_\sigma) G^{(j)}(u),$$

$$(L(u))_{l_1, l_2} = - \sum_{j=3}^J \frac{1}{n(j-1)!} \sum_{l_3, \dots, l_j=0}^k u_{l_3} \dots u_{l_j} \nabla_{l_1, \dots, l_j} l_n(d_o, k). \quad (\text{C.9})$$

The construction of  $G(u)$  is such that

$$nu^t G(u) u = \sum_{j=2}^J g_{n,j}(u). \quad (\text{C.10})$$

Analogous to  $G(u)$  and  $D(u)$  we define  $\bar{G}(u) = \sum_{j=2}^J \frac{1}{j!} \sum_{\sigma \in \mathcal{S}_j} (c_\sigma - d_\sigma) \bar{G}^{(j)}(u)$  and  $\bar{D}(u) = (\tilde{I}_k + L(u))^{-1} \bar{G}^t(u)$ . After substitution of  $v = \psi(u)$ , and using (C.25) in Lemma C.1 it follows that

$$\begin{aligned} \sum_{j=3}^J \frac{(v^{(j)} - u^{(j)}) \nabla^j l_n(d_o, k)}{j!} = \\ - (d - d_o) \sum_{j=3}^J \sum_{l_1, \dots, l_j=0}^k (D(u)u)_{l_1} u_{l_2} \dots u_{l_j} \frac{\nabla_{l_1, \dots, l_j} l_n(d_o, k)}{(j-1)!} + \mathbf{O}(S_n(d)). \end{aligned}$$

The definitions of  $D(u)$  and  $L(u)$  and (C.10) imply that

$$\begin{aligned} -n(v-u)^t \tilde{I}_k u &= n(d-d_o) u^t D^t(u) \tilde{I}_k u = n(d-d_o) u^t G(u) (\tilde{I}_k + L(u))^{-1} \tilde{I}_k u \\ &= n(d-d_o) u^t G(u) \left( I_{k+1} - (\tilde{I}_k + L(u))^{-1} L(u) \right) u \\ &= (d-d_o) \sum_{j=2}^J \frac{1}{j!} g_{n,j}(u) - n(d-d_o) (D(u)u)^t L(u) u \\ &= (d-d_o) \sum_{j=2}^J \frac{1}{j!} g_{n,j}(u) - (d-d_o) \sum_{j=3}^{J-1} \sum_{l_1, \dots, l_j=0}^k (D(u)u)_{l_1} u_{l_2} \dots u_{l_j} \frac{\nabla_{l_1, \dots, l_j} l_n(d_o, k)}{(j-1)!}. \end{aligned}$$

At the same time, the definition of  $\tilde{I}_k$  implies that

$$\frac{1}{2} \left( v^{(2)} - u^{(2)} \right) \nabla^2 l_n(d_o, k) = -n(v-u)^t \tilde{I}_k u - \frac{n(v-u)^t \tilde{I}_k (v-u)}{2}.$$

Combining the preceding results, we find that

$$\begin{aligned} h_k v + \sum_{j=1}^J \frac{v^{(j)} \nabla^j l_n(d_o, k)}{j!} - \left( h_k u + \sum_{j=1}^J \frac{u^{(j)} \nabla^j l_n(d_o, k)}{j!} + (d-d_o) \sum_{j=2}^J \frac{g_{n,j}(u)}{j!} \right) \\ = h_k (v-u) + (v-u)^t \nabla l_n(d_o, k) - \frac{n(v-u)^t \tilde{I}_k (v-u)}{2} + O(S_n(d)) \\ = (v-u)^t \nabla l_n(d_o, k) + O(S_n(d)), \end{aligned}$$

where the last equality follows from (C.24) below in Lemma C.1, together with the assumption on  $h_k$  in **prior A** in (2.2).

Apart from the term  $(v-u)^t \nabla l_n(d_o, k)$  on the last line, the preceding display implies (C.5). Hence, to complete the proof of (C.5) it suffices to show that

$$(v-u)^t \nabla l_n(d_o, k) = -(d-d_o)u^t D^t(u) \nabla l_n(d_o, k) = O(S_n(d)). \quad (\text{C.11})$$

The proof of (C.11) consists of the following steps:

$$|u^t(D(u) - \bar{D}(u))^t \nabla l_n(d_o, k)| = \mathbf{O}_{\mathbf{P}_o}(n^{\frac{1}{2}-\delta} k^{-\frac{1}{2}}), \quad (\text{C.12})$$

$$(d-d_o)u^t \bar{D}^t(u) \mathcal{D}(\nabla l_n(d_o, k)) = O(S_n(d)), \quad (\text{C.13})$$

$$(d-d_o)u^t \bar{D}^t(u) \mathcal{S}(\nabla l_n(d_o, k)) = O(S_n(d)), \quad (\text{C.14})$$

where  $\mathcal{S}(\nabla l_n(d_o, k))$  denotes the centered quadratic form in  $\nabla l_n(d_o, k)$ , and  $\mathcal{D}(\nabla l_n(d_o, k))$  the remaining deterministic term. We will use the same notation below for  $L(u)$ .

Equation (C.12) follows from Lemma B.4 and (C.22) in Lemma C.1 below, which imply that the left-hand side equals  $\mathbf{O}_{\mathbf{P}_o}((\sqrt{k}\|u\|)^2 n^{-1+\epsilon} k \sqrt{n}) = \mathbf{O}_{\mathbf{P}_o}(n^{\frac{1}{2}-\delta} k^{-\frac{1}{2}})$ , for some  $\delta > 0$ . For the proof of (C.13), note that Lemma B.4 implies

$$|u^t \bar{D}^t(u) \mathcal{D}(\nabla l_n(d_o, k))| \lesssim \|\bar{D}(u)u\| \sqrt{k} k^{(3/2-\beta)_++\epsilon}.$$

Combined with Lemma C.1, this implies that the left-hand side is  $O(\sqrt{k} k^{5/2-2\beta+\epsilon})$ , which is  $O(S_n(d))$ . The proof of (C.14) is more involved. Recall that  $\bar{D}(u)$  is defined as  $\bar{D}(u) = (\tilde{I}_k + L(u))^{-1} \bar{G}^t(u)$ . Using (B.16) in Lemma B.5, we obtain

$$(\tilde{I}_k + L(u))^{-1} = 2[I_{k+1} - (A(d_o) + R_{2s} + R_{2d} + L(u))(1 + \mathbf{O}_{\mathbf{P}_o}(1))].$$

Substituting this in  $\bar{D}(u)$ , it follows that (C.14) can be proved by controlling  $\bar{G}^{(j)} \mathcal{S}(\nabla l_n(d_o))$ ,  $\bar{G}^{(j)} A(d_o) \mathcal{S}(\nabla l_n(d_o))$ ,  $\bar{G}^{(j)} R_{2d} \mathcal{S}(\nabla l_n(d_o))$ ,  $\bar{G}^{(j)} \mathcal{D}(L(u)) \mathcal{S}(\nabla l_n(d_o))$ ,  $\bar{G}^{(j)} R_{2s} \mathcal{S}(\nabla l_n(d_o))$  and  $\bar{G}^{(j)} \mathcal{S}(L(u)) \mathcal{S}(\nabla l_n(d_o))$  for all  $j = 3, \dots, J$ . To do so, first note that Lemma B.5 implies that  $\|\bar{G}^{(j)} R_{2s} \mathcal{S}(\nabla l_n(d_o))\| = \mathbf{O}_{\mathbf{P}_o}(n^\epsilon \sqrt{k})$ . Hence,

$$|u^t \bar{G}^{(j)} R_{2s} \mathcal{S}(\nabla l_n(d_o))| = \mathbf{O}_{\mathbf{P}_o}(k^{-\beta+1+\epsilon}) = \mathbf{O}_{\mathbf{P}_o}(1),$$

which clearly is  $O(S_n(d))$ . The terms  $\bar{G}^{(j)} \mathcal{S}(\nabla l_n(d_o))$ ,  $\bar{G}^{(j)} A(d_o) \mathcal{S}(\nabla l_n(d_o))$ ,  $\bar{G}^{(j)} R_{2d} \mathcal{S}(\nabla l_n(d_o))$  and  $\bar{G}^{(j)} \mathcal{D}(L(u)) \mathcal{S}(\nabla l_n(d_o))$  can be written as quadratic forms  $Z^t M Z - \text{tr}[M]$ , where, for a sequence  $(b_l)_{l=0}^k$  and a function  $g$  with  $\|g\|_\infty < \infty$ ,  $M$  is of the form

$$T_n^{\frac{1}{2}}(f_o) T_n^{-1}(f_{d_o, k}) T_n \left( g(x) f_{d_o, k}(x) \sum_l b_l \cos(lx) \right) T_n^{-1}(f_{d_o, k}) T_n^{\frac{1}{2}}(f_o),$$

$Z$  being a vector of  $n$  independent standard Gaussian random variables. Using Lemma 2.4 (supplement) it can be seen that  $|M|^2 \leq n(\sum_l b_l^2 + k/n)$ . Lemma 1.3 (supplement) with  $\alpha = \epsilon + 1/2$  then implies that

$$P_o \left( |Z^t M Z - \text{tr}[M]| > n^{\epsilon+\frac{1}{2}} \left( \sum_l b_l^2 + \frac{k}{n} \right)^{\frac{1}{2}} \right) \leq e^{-cn^\epsilon}. \quad (\text{C.15})$$

For all  $j \in \{3, \dots, J\}$ , the four terms above can now be bounded for a particular choice of  $g$  and  $b_l$ .

- Bound on  $\bar{G}^{(j)}\mathcal{S}(\nabla l_n(d_o))$ . For all  $l_2, \dots, l_j \in \{0, \dots, k\}$ , set  $b_l = \gamma_{l_2, l_4, \dots, l_j}^{(j)}$  and  $g(x) = 1$ . Then we have

$$\sum_{l=0}^k b_l \mathcal{S}(\nabla l_n(d_o))_{l_1} = \mathbf{O}_{\mathbf{P}_o} \left( n^{\frac{1}{2}+\epsilon} \left( \sum_l b_l^2 \right)^{\frac{1}{2}} \right).$$

By induction it can be shown that

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(l_0 x) \cos(l_1 x) \cos(l_2 x) \cdots \cos(l_j x) dx \\ &= 2^{-j} \sum_{\epsilon_1, \dots, \epsilon_j \in \{-1, 1\}} \mathbb{I}_{l_0 + \sum_{i=1}^j \epsilon_i l_i = 0} \end{aligned} \quad (\text{C.16})$$

Consequently,  $\sum_l b_l^2 = O(k^{-1})$  for all  $l_2, l_4, \dots, l_j \in \{0, \dots, k\}$ . Using the fact that  $\sum_l |u_l| = o(1)$ , we obtain that  $(\bar{G}^{(j)}\mathcal{S}(\nabla l_n(d_o)))_{l_2} = \mathbf{O}_{\mathbf{P}_o}(n^{1/2+\epsilon}k^{-1/2})$ , for all  $l_2 \in \{0, \dots, k\}$ . This implies that

$$\|\bar{G}^{(j)}\mathcal{S}(\nabla l_n(d_o))\| = \mathbf{O}_{\mathbf{P}_o}(n^{1/2+\epsilon}) = O(S_n(d)). \quad (\text{C.17})$$

- Bound on  $A(d_o)\mathcal{S}(\nabla l_n(d_o))$ . Set  $b_l = \mathbb{I}_{l+l_1 \geq k} \theta_{o, l+l_1}$  and  $g(x) = 1$ , then

$$(A(d_o)\mathcal{S}(\nabla l_n(d_o)))_{l_1} = \mathbf{O}_{\mathbf{P}_o}(n^{1/2+\epsilon}k^{-\beta}), \quad \forall l_1 \in \{0, \dots, k\}.$$

Combined with Lemma C.1 this implies that

$$\|\bar{G}^{(j)}A(d_o)\mathcal{S}(\nabla l_n(d_o))\| = \mathbf{O}_{\mathbf{P}_o}(n^{1/2+\epsilon}k^{-\beta+1/2}).$$

- Bound on  $\bar{G}^{(j)}R_{2d}\mathcal{S}(\nabla l_n(d_o))$ . Set  $b_l = (R_{2d})_{l_1, l}$  and  $g(x) = 1$ , for all  $l_1 = 0, \dots, k$ , then Lemmas B.5 and C.1 lead to

$$\|\bar{G}^{(j)}R_{2d}\mathcal{S}(\nabla l_n(d_o))\| = \mathbf{O}_{\mathbf{P}_o}(n^{1/2+\epsilon}k^2n^{-1}) = \mathbf{O}_{\mathbf{P}_o}(n^{-1/2+\epsilon}k^2).$$

- Bound on  $\bar{G}^{(j)}\mathcal{D}(L(u))\mathcal{S}(\nabla l_n(d_o))$ . For all  $l_1, l_3, \dots, l_j, l'_3, \dots, l'_j \in \{0, \dots, k\}$ , set

$$\begin{aligned} b_l = \frac{1}{n} \text{tr} & \left[ T_n^{-1}(f_{d_o}) T_n \left( \sum_{l_2=0}^k \gamma_{l_1, l_2, \dots, l_j}^{(j)} \cos(l_2 x) g_1(x) f_{d_o}(x) \right) \times \right. \\ & \left. T_n^{-1}(f_{d_o}) T_n(\cos(lx) g_2(x) f_{d_o}(x)) \cdots T_n^{-1}(f_{d_o}) T_n(g_r(x) f_{d_o}(x)) \right] \end{aligned}$$

where  $g_1(x), \dots, g_r(x)$  are products of functions of the form  $\cos(l'_i x)$  and  $g_1(x) \cdots g_r(x) = \cos(l'_3 x) \cdots \cos(l'_j x)$ . Lemmas 2.1 and 2.6 in the supplement, together with (C.16), imply that

$$\sum_l b_l^2 = O(k^{-1}), \quad \text{and} \quad \|\bar{G}^{(j)}\mathcal{D}(L(u))\mathcal{S}(\nabla l_n(d_o))\| = \mathbf{O}_{\mathbf{P}_o}(n^{\epsilon+1/2}). \quad (\text{C.18})$$

Consequently, the contribution to all these terms in  $(v - u)^t \nabla l_n(d_o, k)$  is of order  $\mathbf{O}(S_n(d))$ .

We control  $u^t \bar{G}^{(j)} \mathcal{S}(L(u)) \mathcal{S}(\nabla l_n(d_o, k))$ , by bounding  $\|\bar{G}^{(j)} \mathcal{S}(L(u))\|$  using a similar idea. Indeed, for all  $l_1, l_2 \leq k$ ,  $(\bar{G}^{(j)} \mathcal{S}(L(u)))_{l_1, l_2}$  can be written as a sum of terms of the form  $(Z^t M_{l_1, l_2} Z - \text{tr}(M_{l_1, l_2}))/n$ , where  $Z$  is a vector of  $n$  independent standard Gaussian random variables, and  $M_{l_1, l_2}$  has the form

$$T_n^{\frac{1}{2}}(f_o) T_n^{-1}(f_{d_o, k}) \left( \prod_{i < i_0} T_n(\nabla_{\sigma(i)} f_{d_o, k}) T_n^{-1}(f_{d_o, k}) \right) \times \\ T_n \left( \sum_{l=0}^k \gamma_{l, l_1, l_2, \dots, l_{j-1}}^{(j)} \cos(lx) \nabla_{\sigma(i_0) - \{l\}} f_{d_o, k} \right) T_n^{-1}(f_{d_o, k}) \prod_{i < i_0} T_n(\nabla_{\sigma(i)} f_{d_o, k}) T_n^{\frac{1}{2}}(f_o).$$

We can use the same argument as in (C.15) since for all  $l_1, l_2, \dots, l_{j-1}$

$$|M_{l_1, l_2}| \lesssim |T_n^{-\frac{1}{2}}(f_{d_o, k}) T_n \left( \sum_{l=0}^k \gamma_{l, l_1, l_2, \dots, l_{j-1}}^{(j)} \cos(lx) \nabla_{\sigma(i_0) - \{l\}} f_{d_o, k} \right) T_n^{-\frac{1}{2}}(f_{d_o, k})| \\ = O(n^{1/2+\epsilon} k^{-1/2}).$$

Hence, it follows that  $n^{-1}[Z^t M_{l_1, l_2} Z - \text{tr}(M_{l_1, l_2})] = \mathbf{O}_{\mathbf{P}_o}(n^{-1/2+\epsilon} k^{-1/2})$  and

$$u^t \bar{G} \mathcal{S}(L(u)) \mathcal{S}(\nabla l_n(d_o, k)) = \mathbf{O}_{\mathbf{P}_o}(\|u\| n^\epsilon k) = \mathbf{O}_{\mathbf{P}_o}(n^{1/2-\delta} k^{-1/2}). \quad (\text{C.19})$$

Combining (C.19) and (C.17)-(C.18), we obtain (C.14). This in turn finishes the proof of (C.11), since

$$(v - u)^t \nabla l_n(d_o, k) = \mathbf{O}_{\mathbf{P}_o}(|d - d_o| n^{\frac{1}{2}-\delta} k^{-\frac{1}{2}}) = O(S_n(d)).$$

We now prove that  $\psi(u)$  is a one-to-one transformation. First note that  $\psi(u)$  is continuously differentiable for all  $\|u\| \leq 2l_0 \delta_n$ . This follows from the definition  $\psi(u) = (I_{k+1} - (d - d_o)(\tilde{I}_k + L(u))^{-1} G^t(u))u$ , the fact that  $G(u)$  and  $L(u)$  are polynomial in  $u$  and Lemma C.1, by which  $\|L(u)\| = \mathbf{O}_{\mathbf{P}_o}(1)$ . To prove that  $\psi(u)$  is also one-to-one, we bound the spectral norm of the Jacobian

$$\psi'(u) = I_{k+1} - (d - d_o)D(u) - (d - d_o)(D'(u)u),$$

where  $(D'(u)u)$  is the  $(k+1) \times (k+1)$  matrix with elements

$$\sum_{l=0}^k u_l \frac{\partial (D(u))_{l_1, l}}{\partial u_{l_2}}, \quad l_1, l_2 = 0, \dots, k.$$

For  $\psi(u)$  to be one-to-one, it suffices to have  $\psi'(u) = I_{k+1}(1 + \mathbf{O}_{\mathbf{P}_o}(1))$ .

By (C.24) in Lemma C.1 below, we have  $|d - d_o| \|D(u)\| = \mathbf{O}_{\mathbf{P}_o}(|d - d_o|)$ . Therefore we only need to control the spectral norm of  $D'(u)u$ . For all  $l_1, l_2$ , we have

$$(D'(u)u)_{l_1, l_2} = \left[ -(\tilde{I}_k + L(u))^{-1} \frac{\partial L(u)}{\partial u_{l_2}} (\tilde{I}_k + L(u))^{-1} G^t(u)u + (\tilde{I}_k + L(u))^{-1} \frac{\partial G^t(u)}{\partial u_{l_2}} u \right]_{l_1}. \quad (\text{C.20})$$

Both  $(G(u))_{l_1, l_2}$  and  $(L(u))_{l_1, l_2}$  can be written as

$$F_{l_1, l_2}(u; \tau, b) := \sum_{j=2}^J \tau_j \sum_{l_3, \dots, l_j=0}^k u_{l_3} \cdots u_{l_j} b_{l_1, l_2, \dots, l_j},$$

where the constants  $\tau_j, b_{l_1, \dots, l_j}$  are different for  $G$  and  $L$ , and  $b$  is symmetric in its indices. In particular,  $\tau_2 = 0$  in the case of  $L$ . Using this generic notation for  $G(u)$  and  $L(u)$ , we find that for all  $v \in \mathbb{R}^{k+1}$  and all  $l_1, l_2 \leq k$ ,

$$\left( \frac{\partial F(u; \tau, b)v}{\partial u_{l_2}} \right)_{l_1} = \sum_{j=3}^J \tau_j (j-3+1) \sum_{l_3, \dots, l_j=0}^k v_{l_3} u_{l_4} \cdots u_{l_j} b_{l_1, l_2, \dots, l_j} := F(v, u; \tau', b),$$

where  $\tau'_j = \tau_j(j-3+1)$ ,  $j = 3, \dots, J$ . It therefore has the same form as  $F(u; \tau', b)$ , with  $v$  replacing one of the  $u$ 's. Applying this to the first term of (C.20), with  $v = (\tilde{I}_k + L(u))^{-1} G^t(u)u$ , we find that

$$|(\tilde{I}_k + L(u))^{-1} F(v, u; \tau', b)| \lesssim |F(v, u; \tau', b)| = \mathbf{O}(1),$$

where we used (C.21) and (C.24) from Lemma C.1. The second term of (C.20) is treated similarly with  $v = u$  so that we finally obtain

$$|D'(u)u| = \mathbf{O}(1),$$

and  $\psi$  is one-to-one on  $\{u : \|u\| \leq 2l_0\delta_n\}$ . Using the above bounds we also deduce that the Jacobian is equal to  $\exp(O(S_n(d)))$ , since

$$\begin{aligned} \log \det[\text{Jac}] &= \log \det [I_{k+1} - (d-d_o)D(u) - (d-d_o)D'(u)u] \\ &= O[(d-d_o)(|\text{tr}[D(u)]| + \text{tr}[D'(u)u]) + (d-d_o)^2(|D(u)|^2 + |D'(u)u|^2)] \\ &= O(\sqrt{k}(d-d_o) + k(d-d_o)^2) = O(S_n(d)). \end{aligned}$$

This finishes the proof of (C.4), and hence the proof of Lemma 3.5.

**Lemma C.1.** *Let  $v = \psi(u)$ , with  $\psi$  as in (C.8). Under the conditions of Lemma 3.5, we have*

$$|L(u)| = \mathbf{O}_{\mathbf{P}_o}(n^{-1/2+\epsilon}k) = \mathbf{O}_{\mathbf{P}_o}(1), \quad (\text{C.21})$$

$$|G - \bar{G}| = \mathbf{O}_{\mathbf{P}_o}(n^{-1/2+\epsilon}k) = \mathbf{O}_{\mathbf{P}_o}(1), \quad (\text{C.22})$$

$$|D(u)| = \mathbf{O}_{P_o}(1), \quad (\text{C.23})$$

$$\|u - \psi(u)\| \lesssim |d-d_o| \mathbf{O}_{P_o}(\|u\|), \quad (\text{C.24})$$

and

$$\begin{aligned} & \sum_{j=3}^J \frac{(v^{(j)} - u^{(j)}) \nabla^j l_n(d_o, k)}{j!} \\ &= -(d-d_o) \sum_{j=3}^J \sum_{l_1, \dots, l_j=0}^k (D(u)u)_{l_1} u_{l_2} \cdots u_{l_j} \frac{\nabla_{l_1, \dots, l_j} l_n(d_o, k)}{(j-1)!} + O(S_n(d)), \end{aligned} \quad (\text{C.25})$$

uniformly over  $\|u\| \leq 2l_0\delta_n$ .



*Proof.* We first prove (C.21). From (B.2), we recall that  $\nabla_{l_1, \dots, l_j} l_n(d_o, k)$  is the sum of a centered quadratic form  $\mathcal{S}(\nabla_{l_1, \dots, l_j} l_n(d_o, k))$  and a deterministic term  $\mathcal{D}(\nabla_{l_1, \dots, l_j} l_n(d_o, k))$ . For all  $l_1, \dots, l_j$ ,  $\mathcal{S}(\nabla_{l_1, \dots, l_j} l_n(d_o, k))$  equals

$$X^t \left( T_n^{-1}(f_{d_o, k}) \sum_{\sigma \in \mathcal{S}_j} b_\sigma B_\sigma(d_o) \right) X - \text{tr} \left[ T_n(f_o) T_n^{-1}(f_{d_o, k}) \sum_{\sigma \in \mathcal{S}_j} b_\sigma B_\sigma(d_o) \right],$$

with  $B_\sigma(d_o) := B_\sigma(d_o, \bar{\theta}_{d_o, k})$  as defined in (B.1). Using Lemma 1.3 (supplement) together with (B.10) we obtain that for all  $l_1, \dots, l_j$ ,  $\mathcal{S}(\nabla_{l_1, \dots, l_j} l_n(d_o, k)) = \mathbf{o}_{\mathbf{P}_o}(n^{\frac{1}{2}+\epsilon})$ , and its contribution to  $|L(u)|$  is  $\mathbf{o}_{\mathbf{P}_o}(k(\sqrt{k}\|u\|)^{j-2}n^{-\frac{1}{2}+\epsilon}) = \mathbf{o}_{\mathbf{P}_o}(n^{-1/2-\delta}k)$ . The deterministic term in (B.2) is

$$\mathcal{D}(\nabla_{l_1, \dots, l_j} l_n(d_o, k)) = \sum_{\sigma} c_\sigma \text{tr}[B_\sigma(d_o)] + \sum_{\sigma} d_\sigma \text{tr}[(T_n(f_o) T_n^{-1}(f_{d_o, k}) - I_n) B_\sigma(d_o)].$$

We bound the contribution of the first term to  $|L(u)|$ ; the second term can be treated similarly. Let  $\tilde{L}(u)$  be the matrix when in (C.9) we replace  $\nabla_{l_1, \dots, l_j} l_n(d_o, k)$  by  $\sum_{\sigma} c_\sigma \text{tr}[B_\sigma(d_o)]$ . Hence,

$$(\tilde{L}(u))_{l_1, l_2} = - \sum_{j=3}^J \frac{1}{(j-1)!} \sum_{\sigma \in \mathcal{S}_j} c_\sigma \sum_{l_3, \dots, l_j=0}^k u_{l_3} \dots u_{l_j} \frac{\text{tr}[B_\sigma(d_o)]}{n}, \quad (\text{C.26})$$

$$\text{where } \frac{1}{n} \text{tr}[B_\sigma(d_o)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(l_1 x) \dots \cos(l_j x) dx + E_\sigma, \quad (\text{C.27})$$

$E_\sigma$  being the approximation error. For each  $\sigma$  and  $j \geq 4$ , the contribution of the integral in (C.27) to  $(\tilde{L}(u))_{l_1, l_2}$  is  $O(\int_{-\pi}^{\pi} |u^t \mathbf{cos}|^{j-2}(x) dx) = O(\|u\|^2)$ ; hence its contribution to  $|\tilde{L}(u)|$  is  $k\|u\|^2 = o(n^{-1/2-\delta}k)$ . For  $j = 3$ , we have

$$\frac{1}{4\pi} \sum_{l_3=1}^k u_{l_3} \int_{-\pi}^{\pi} \cos(l_1 x) \cos(l_2 x) \cos(l_3 x) dx = \frac{1}{2} (u_{l_1+l_2} \mathbb{I}_{l_3=l_1+l_2} + u_{|l_1-l_2|} \mathbb{I}_{l_3=|l_1-l_2|}),$$

and the contribution of this term to  $|\tilde{L}(u)|$  is of order  $\sqrt{k}\|u\| = o(n^{-1/2+\epsilon}k)$ . Next we bound the contribution to  $|\tilde{L}(u)|$  of the error term  $E_\sigma$  in (C.27). Note that we can write the last sum in (C.26) as

$$\sum_{l_3, \dots, l_j=0}^k u_{l_3} \dots u_{l_j} \frac{\text{tr}[B_\sigma(d_o)]}{n} = \frac{1}{n} \text{tr} \left[ \prod_{i=1}^p T_n(b_i(x) f_{d_o, k}) T_n^{-1}(f_{d_o, k}) \right], \quad (\text{C.28})$$

where

$$b_i(x) = (u^t \mathbf{cos}(x))^{|\sigma(i)| - \delta_1(i) - \delta_2(i)} \cos(l_1 \cdot)^{\delta_1(i)} \cos(l_2 \cdot)^{\delta_2(i)}, \quad (\text{C.29})$$

$\delta_1(i) = \mathbb{I}_{1 \in \sigma(i)}$ ,  $\delta_2(i) = \mathbb{I}_{2 \in \sigma(i)}$  and  $p = |\sigma|$ . If  $p \leq 3$ , then Lemma 2.4 (supplement) implies that

$$E_\sigma = O((\sqrt{k}\|u\|)^{j-2} n^{\epsilon-1} [k^{2(3/2-\beta)+} + k]) = o(n^{-1-\delta}k). \quad (\text{C.30})$$

If  $p \geq 4$ , then Lemma 2.6 (supplement) together with (C.28), with

$$f = f_{d_o, k}, \quad f_{2i} = b_i f_{d_o, k}, \quad i \leq |\sigma|,$$

$L = k^{(3/2-\beta)_+}$ ,  $M, m^{-1} = O(1)$ ,  $M^{(i)} = O((\sqrt{k}\|u\|)^{|\sigma(i)|-\delta_1(i)-\delta_2(i)})$  and  $L^{(i)} = O(k(\sqrt{k}\|u\|)^{|\sigma(i)|-\delta_1(i)-\delta_2(i)})$ , leads to the bound

$$\begin{aligned} & \sum_{l_3, \dots, l_j=0}^k u_{l_3} \dots u_{l_j} \frac{\text{tr}[B_\sigma(d_o)]}{n} - \frac{1}{n} \text{tr} \left[ \prod_{i=1}^{|\sigma|} T_n(b_i f_{d_o, k}) T_n \left( \frac{1}{4\pi^2 f_{d_o, k}} \right) \right] \\ & = \mathbf{o}(k^{(3/4-\beta/2)_+} n^{-1/2+\epsilon} \|u\| (\sqrt{k}\|u\|)^{j-3}) = \mathbf{o}(n^{-1/2-\delta}). \end{aligned}$$

Using Lemma 2.1 (supplement) we finally obtain that

$$\frac{1}{n} \text{tr} \left[ \prod_{i=1}^{|\sigma|} T_n(b_i f_{d_o, k}) \right] - \frac{1}{2\pi} \int_{-\pi}^{\pi} (u^t \mathbf{cos})^{j-2}(x) \cos(l_1 x) \cos(l_2 x) dx = \mathbf{o}(n^{-1-\delta} k).$$

Therefore the contribution of the approximation error  $E_\sigma$  in  $|\tilde{L}(u)|$  is of order  $\mathbf{o}(n^{-1/2-\delta} k)$ . Using a similar argument we control the terms in the form  $\text{tr} [T_n(f_o)(T_n^{-1}(d_o, k) - I_n) B_\sigma(d_o)]$  and (C.21) is proved.

We now prove (C.22) and bound

$$(G - \bar{G})_{l_1, l_2} = \sum_{j=2}^J \frac{1}{j!} \sum_{\sigma \in \mathcal{S}_j} (c_\sigma - d_\sigma) \sum_{l_3, \dots, l_j=0}^k r_{l_1, \dots, l_j}^{(j)} u_{l_3} \dots u_{l_j},$$

with  $r_{l_1, \dots, l_j}^{(j)}$  as in (C.6). These are the approximation errors which occur when replacing  $\frac{1}{n} \text{tr}[T_{1, \sigma}(d_o, k)]$  and  $\frac{1}{n} \text{tr}[T_{2, \sigma}(d_o, k)]$  by their limiting integrals (see also (B.4)). Therefore, for each  $\sigma \in \mathcal{S}_j$ ,  $\sum_{l_3, \dots, l_j=0}^k r_{l_1, \dots, l_j}^{(j)} u_{l_3} \dots u_{l_j}$  is a combination of terms of the form

$$\frac{1}{n} \text{tr} \left[ \prod_{i=1}^p T_n(b_i(x) f_{d_o, k}) T_n^{-1}(f_{d_o, k}) \right] - \frac{1}{2\pi} \int_{-\pi}^{\pi} (u^t \mathbf{cos}(x))^{j-2} H_k(x) \cos(l_1 x) \cos(l_2 x) dx,$$

with  $p \in \{|\sigma|, |\sigma| + 1\}$  and the functions  $b_i$  defined as in (C.29) apart from  $b_1(x) = H_k(x) (\sum_{l=0}^k u_l \cos(lx))^{|\sigma(1)|-\delta_1(1)-\delta_2(1)} \cos(l_1 \cdot)^{\delta_1(1)} \cos(l_2 \cdot)^{\delta_2(1)}$ . Therefore, using the same construction as in (C.28)-(C.30), we obtain that

$$|(G - \bar{G})_{l_1, l_2}| = O(n^{-1/2+\epsilon}), \quad |G - \bar{G}| = O(n^{-1/2+\epsilon} k) = o(1).$$

To prove (C.23), we use the just obtained bound on  $|G - \bar{G}|$ , and in addition establish a bound  $|\bar{G}|$ . We treat each term  $G^{(j)}$  in  $G(u) = \sum_{j=2}^J \frac{1}{j!} \sum_{\sigma \in \mathcal{S}_j} (c_\sigma - d_\sigma) G^{(j)}(u)$  separately. First we show that  $|\bar{G}^{(2)}| = O(1)$ , which follows from definition (C.7), by which

$$\text{tr} \left[ (\bar{G}^{(2)})^2 \right] = \sum_{l_1, l_2, l_1+l_2 \geq k}^k \frac{1}{(l_1 + l_2)^2} \leq 1.$$

Consequently,  $|\bar{G}^{(2)}| \leq 1$ . For  $j \geq 3$ , note that for all  $0 \leq l_1, l_2 \leq k$ ,

$$\begin{aligned} |\bar{G}_{l_1, l_2}^{(j)}(u)| &= \left| \sum_{l_3, \dots, l_j=0}^k \gamma_{l_1, l_2, \dots, l_j}^{(j)} u_{l_3} \dots u_{l_j} \right| \\ &\leq \sum_{l_3, \dots, l_j=0}^k |u_{l_3} \dots u_{l_j}| \int_{-\pi}^{\pi} |H_k(x)| \left| \sum_{l_3=0}^k \cos(l_3 x) u_{l_3} \right| dx \\ &\leq (\sqrt{k} \|u\|)^{j-3} \frac{\|u\|}{\sqrt{k}} = (\sqrt{k})^{j-4} (\|u\|)^{j-2}. \end{aligned}$$

Therefore,  $|\bar{G}^{(j)}(u)| \leq k(\sqrt{k})^{j-4} (\|u\|)^{j-2} = (\sqrt{k} \|u\|)^{j-2} = o(1)$ , for all  $j \geq 3$ . Hence  $|\bar{G}(u)| = \mathbf{O}(1)$ , which combined with  $\|(\tilde{I}_k^+ L(u)) - 1\| = \mathbf{O}_{P_o}(1)$  (see Lemma B.5) and (C.21), imply that

$$|\bar{D}(u)| = |(\tilde{I}_k + L(u))^{-1} \bar{G}^t(u)| = O(1),$$

uniformly over  $\|u\| \leq 2l_0 \delta_n$ . It follows that

$$|D(u)| \leq \|(\tilde{I}_k + L(u))^{-1}\| (|\bar{G}| + |G - \bar{G}|) = \mathbf{O}_{P_o}(1).$$

This concludes the proof of (C.23); (C.24) directly follows from this result since  $\|u - \psi(u)\| \leq |d - d_o| |D(u)| \|u\|$ . Finally, we prove (C.25). We have

$$\begin{aligned} \sum_{j=3}^J \frac{(v^{(j)} - u^{(j)}) \nabla^j l_n(d_o, k)}{j!} &= -(d - d_o) \sum_{j=3}^J \sum_{l_1, \dots, l_j=0}^k (D(u)u)_{l_1} u_{l_2} \dots u_{l_j} \frac{\nabla_{l_1, \dots, l_j} l_n(d_o, k)}{(j-1)!} \\ &+ (d - d_o)^2 \sum_{j=3}^J \binom{j}{2} \sum_{l_1, \dots, l_j=0}^k (D(u)u)_{l_1} (D(u)u)_{l_2} \dots u_{l_j} \frac{\nabla_{l_1, \dots, l_j} l_n(d_o, k)}{(j-1)!} \\ &+ \dots + (-1)^J \sum_{l_1, \dots, l_J=0}^k (D(u)u)_{l_1} (D(u)u)_{l_2} \dots (D(u)u)_{l_J} \frac{\nabla_{l_1, \dots, l_J} l_n(d_o, k)}{(J-1)!} \end{aligned}$$

Using the same argument as in the proof of (C.21), we find that for all for all  $j \geq 3$

$$\begin{aligned} &\sum_{l_1, \dots, l_j=0}^k (D(u)u)_{l_1} (D(u)u)_{l_2} u_{l_3} \dots u_{l_j} \frac{\nabla_{l_1, \dots, l_j} l_n(d_o, k)}{j!} \\ &= n \int_{-\pi}^{\pi} (D(u)u)^t \cos(x)^2 (u^t \cos(x))^{j-2} dx \\ &+ (\sqrt{k} \|u\|)^{j-1} O(\sqrt{n} n^\epsilon (\sqrt{k} \|u\|) + k + \sqrt{n} \|u\| k^{1/2(3/2-\beta)_+}) \\ &= \mathbf{o}(n^{1-\delta} k^{-1}), \quad \text{for some } \delta > 0. \end{aligned}$$

Similarly, the higher-order terms in the above expression for  $\sum_{j=3}^J \frac{(v^{(j)} - u^{(j)}) \nabla^j l_n(d_o, k)}{j!}$  can be shown to be  $O(S_n(d))$ , which terminates the proof of Lemma C.1.  $\square$

The rest of the paper corresponds to the supplementary material

## D Technical results

Let  $\eta_j = -1_{j>0}2/j$  and recall that  $\bar{\theta}_{d,k} = \theta_{o[k]} + (d_o - d)\eta_{[k]}$ . Let the sequence  $\{a_j\}$  be defined as  $a_j = \theta_{o,j} + (d_o - d)\eta_j$  when  $j > k$  and  $a_j = 0$  when  $j \leq k$ . In addition, define

$$H_k(x) = \sum_{j=k+1}^{\infty} \eta_j \cos(jx), \quad G_k(x) = \sum_{j=1}^k \eta_j \cos(jx), \quad (\text{D.1})$$

$$\Delta_{d,k}(x) = \sum_{j=k+1}^{\infty} (\theta_{o,j} + (d_o - d)\eta_j) \cos(jx) = \sum_{j=k+1}^{\infty} a_j \cos(jx). \quad (\text{D.2})$$

Using this notation we can write

$$-2 \log |1 - e^{ix}| = -\log(2 - 2 \cos(x)) = G_k(x) + H_k(x), \quad (\text{D.3})$$

$$\begin{aligned} f_{d,k}(x) &= f_{d,k,\bar{\theta}_{d,k}}(x) = f_o(x) \exp \left\{ - \sum_{j=k+1}^{\infty} a_j \cos(jx) \right\} \\ &= f_o(x) e^{-\Delta_{d,k}(x)} = f_o(x) e^{(d-d_o)H_k(x) - \Delta_{d_o,k}(x)}. \end{aligned} \quad (\text{D.4})$$

Given  $d, k$  and  $\theta_o$ , the sequence  $\{a_j\}$  represents the closest possible distance between  $f_o$  and  $f_{d,k,\theta}$ , since

$$l(f_o, f_{d,k}) = l(f_o, f_{d,k,\bar{\theta}_{d,k}}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Delta_{d,k}^2(x) dx = \sum_{j>k} a_j^2. \quad (\text{D.5})$$

From (D.4) it also follows that for all  $d$ ,

$$\frac{\partial}{\partial d} f_{d,k} = H_k f_{d,k}. \quad (\text{D.6})$$

**Lemma D.1.** *When  $\theta_o \in \Theta(\beta, L_o)$ , there exist constants such that for any positive integer  $k$ ,*

$$k^{-1} \lesssim \int_{-\pi}^{\pi} H_k^2(x) dx \lesssim k^{-1}, \quad (\text{D.7})$$

$$\sum_{l>k} |\theta_{o,l}| = O(k^{-\beta+\frac{1}{2}}), \quad \sum_{l \geq 0} |\theta_{o,l}| = O(1), \quad (\text{D.8})$$

$$\int_{-\pi}^{\pi} \Delta_{d_o,k}(x) H_k(x) dx = \sum_{j>k} \eta_j \theta_{o,j} = O\left(k^{-\frac{1+2\beta}{2}}\right), \quad (\text{D.9})$$

$$\int_{-\pi}^{\pi} \Delta_{d_o,k}^2(x) dx = \sum_{l>k} \theta_{o,l}^2 = O(k^{-2\beta}), \quad (\text{D.10})$$

$$\int_{-\pi}^{\pi} \Delta_{d_o,k}^2(x) H_k(x) dx = O(k^{-2\beta-1}), \quad (\text{D.11})$$

$$\int_{-\pi}^{\pi} H_k^4(x) dx \lesssim \frac{\log k}{k}. \quad (\text{D.12})$$

When  $k \rightarrow \infty$ , the big- $O$  in (D.8)-(D.11) may be replaced by a small- $o$ , since  $\sum_{l>k} \theta_{o,l}^2 l^{2\beta}$  then tends to zero.

*Proof.* The result for  $\int H_k^2(x) dx$  follows directly from the definition of  $H_k$ . The assumption that  $\theta_o \in \Theta(\beta, L_o)$  and the Cauchy-Schwarz inequality imply that

$$\sum_{l>k} |\theta_{o,l}| \leq \sqrt{\sum_{l>k} \theta_{o,l}^2 l^{2\beta}} \sqrt{\sum_{l>k} l^{-2\beta}} = O(k^{-\beta+\frac{1}{2}}),$$

proving the first result in (D.8). Similarly, one can prove (D.9). For (D.10), note that  $\sum_{l>k} \theta_{o,l}^2 \leq k^{-2\beta} \sum_{l>k} \theta_{o,l}^2 l^{2\beta}$ . For the other bounds we omit the details of the proof. They follow from the fact that for all sequences  $a, b$  and  $c$ ,

$$\begin{aligned} & 2 \sum_{l,m,n>k} a_l b_m c_n \int_{-\pi}^{\pi} \cos(lx) \cos(mx) \cos(nx) dx \\ &= \sum_{m,n>k} b_m c_n \sum_{l>k} a_l \int_{-\pi}^{\pi} \cos(lx) (\cos((m+n)x) + \cos((m-n)x)) dx \\ &= \sum_{m,n>k} a_{m+n} b_m c_n + \sum_{m,n>k; m-n>k} a_{m-n} b_m c_n. \end{aligned}$$

□

Before stating the next lemma we give bounds for the functions  $H_k$  and  $G_k$ . Since  $-2 \log |1 - e^{ix}| = -\log(x^2 + O(x^4))$ , there exist positive constants  $c, B_0, B_1$  and  $B_2$  such that

$$|H_k(x)| \geq B_0 |\log x|, \quad |x| \leq ck^{-1}, \quad (\text{D.13})$$

$$|H_k(x)| \leq B_1 |\log x| + B_2 \log k, \quad x \in [-\pi, \pi]. \quad (\text{D.14})$$

**Lemma D.2.** Let  $a_j = (\theta_{o,j} - (d - d_o)\eta_j) 1_{j>k}$ , as in (D.2). Then for  $p \geq 1$  and  $q = 2, 3, 4$  there exist constants  $c(p, q)$  such that for all  $d \in (-\frac{1}{2}, \frac{1}{2})$  and  $k \leq \exp(|d - d_o|^{-1})$ ,

$$\begin{aligned} \int_{-\pi}^{\pi} \left( \frac{f_o(x)}{f_{d,k}(x)} \right)^p |H_k|^q(x) dx &= O\left( \frac{(\log k)^{c(p,q)}}{k} \right) \\ &+ O((\log k)^{q+pB_2|d-d_o|} |d - d_o|^{-\frac{q}{2}} e^{-|d-d_o|^{-1}}), \end{aligned} \quad (\text{D.15})$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{f_o}{f_{d,k}}(x) - 1 \right) \cos(ix) \cos(jx) dx = \frac{1}{2} a_{i+j} 1_{i+j>k} + O\left( \sum_{j>k} a_j^2 \right), \quad (\text{D.16})$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{f_o}{f_{d,k}}(x) - 1 \right) H_k^2(x) dx = O(|d - d_o| k^{-1} \log k), \quad (\text{D.17})$$

where the constant  $B_2$  in (D.15) is as in (D.14), and the constants in (D.16) and (D.17) are uniform in  $d$ . The constant  $c(p, q)$  in (D.15) equals  $0, \frac{1}{2}, 1$  when respectively  $q = 2, 3, 4$ .

*Proof.* When  $d = d_o$ , (D.15) directly follows from (D.7) and (D.12), because of the boundedness of  $(f_o/f_{d_o,k})^p = \exp\{p\Delta_{d_o,k}\}$ . Now suppose  $d \neq d_o$ . Let  $C_k = \max_{x \in [-\pi, \pi]} \exp\{|\Delta_{d_o,k}(x)|\}$  and  $b_m = \max_{x \in [m, \pi]} |(d - d_o)H_k(x)|$ , for  $m = e^{-\frac{1}{|d-d_o|}} < e^{-1}$ . Since  $\sum_{j=0}^{\infty} |\theta_{o,j}| < \infty$ , the sequence  $C_k$  is bounded by some constant  $C$ . To prove (D.15) we write

$$\begin{aligned} & \frac{1}{2} \int_{-\pi}^{\pi} \left( \frac{f_o(x)}{f_{d,k}(x)} \right)^p |H_k|^q(x) dx \\ &= \int_0^m \left( \frac{f_o(x)}{f_{d,k}(x)} \right)^p |H_k|^q(x) dx + \int_m^{\pi} \left( \frac{f_o(x)}{f_{d,k}(x)} \right)^p |H_k|^q(x) dx. \end{aligned} \quad (\text{D.18})$$

We first bound the last integral in the preceding display, by substitution of  $(f_o/f_{d,k})^p = \exp\{p\Delta_{d,k}\} = \exp\{-p(d-d_o)H_k + p\Delta_{d_o,k}\}$ . From (D.14) it follows that

$$b_m \leq |d - d_o|(B_1|d - d_o|^{-1} + B_2 \log k) \leq B_1 + B_2,$$

as  $k \leq \exp(|d - d_o|^{-1})$ . Hence we obtain  $(f_o/f_{d,k})^p \leq Ce^{b_m}$  on  $(m, \pi)$ . For  $q = 2$  and  $q = 4$  the bound on the last integral in (D.18) therefore follows from (D.7) and (D.12); for  $q = 3$  the bound follows from the Cauchy-Schwarz inequality.

Next we bound the first integral in (D.18). Because the function  $x^{|d-d_o|}(\log x)^2$  has a local maximum of  $4|d-d_o|^{-2}e^{-2}$  at  $x = e^{-2/|d-d_o|}$ ,  $(\log x)^2 \leq 4x^{-|d-d_o|}|d-d_o|^{-2}e^{-2}$  for all  $x \in [0, m]$ . Again using (D.14) we find that

$$\begin{aligned} & \int_0^m \left( \frac{f_o(x)}{f_{d,k}(x)} \right)^p |H_k|^q(x) dx \lesssim \sum_{j=0}^q \binom{q}{j} \int_0^m (B_1 |\log x|)^j (B_2 \log k)^{q-j} e^{-p(d-d_o)H_k(x)} dx \\ & \lesssim \sum_{j=0}^q \binom{q}{j} (\log k)^{q-j+pB_2|d-d_o|} \int_0^m (B_1 |\log x|)^j x^{-pB_1|d-d_o|} dx \\ & \leq \sum_{j=0}^q \binom{q}{j} (\log k)^{q-j+pB_2|d-d_o|} \left( \frac{2B_1^2}{e|d-d_o|} \right)^{\frac{j}{2}} \int_0^m x^{-(j/2+pB_1)|d-d_o|} dx \\ & \lesssim (\log k)^{q+pB_2|d-d_o|} |d-d_o|^{-\frac{q}{2}} e^{-1/|d-d_o|}. \end{aligned}$$

We now prove (D.16).

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{f_o}{f_{d,k}}(x) - 1 \right) \cos(ix) \cos(jx) dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( e^{\Delta_{d,k}(x)} - 1 \right) \cos(ix) \cos(jx) dx \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \Delta_{d,k}(x) + \frac{1}{2} \Delta_{d,k}^2(x) e^{(\Delta_{d,k}(x))_+} \right) \cos(ix) \cos(jx) dx. \end{aligned}$$

The linear term equals

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\pi}^{\pi} \Delta_{d,k}(x) \cos(ix) \cos(jx) dx \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left( \sum_{l>k} a_l \cos(lx) \right) (\cos((i+j)x) + \cos((i-j)x)) dx = \frac{1}{2} a_{i+j} 1_{i+j>k}. \end{aligned}$$

For the quadratic term we have

$$\begin{aligned} & \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \Delta_{d,k}^2(x) e^{(\Delta_{d,k}(x))^+} \cos(ix) \cos(jx) dx \right| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \Delta_{d,k}^2(x) e^{(\Delta_{d,k}(x))^+} dx \\ & \leq \frac{1}{2\pi} \int_0^m \Delta_{d,k}^2(x) e^{-\Delta_{d,k}(x)} dx + \frac{(1 + Ce^{b_m})}{2\pi} \int_{-\pi}^{\pi} \Delta_{d,k}^2(x) dx. \end{aligned} \tag{D.19}$$

This is  $O(\sum_{j>k} a_j^2)$ , which follows from (D.5) and integration over  $(0, e^{-\frac{1}{\bar{v}_n}})$  and  $(e^{-\frac{1}{\bar{v}_n}}, \pi)$  as above.

To prove (D.17), write  $\exp(\Delta_{d,k}) - 1 = \Delta_{d,k} + \Delta_{d,k}^2 e^\xi$  with  $\Delta_{d,k} = -(d - d_o)H_k(x) + \Delta_{d_o,k}(x)$  and  $|d - d_o| \leq \bar{v}_n$ , substitute (D.14) and proceed as in the proof of (D.15) above. The biggest term is a multiple of  $|d - d_o| \int_{-\pi}^{\pi} |H_k(x)|^3 dx$ , which is  $O(\bar{v}_n k^{-1})$ . This is larger than the approximation error when  $\beta > \frac{1}{2}(1 + \sqrt{2})$ .  $\square$

**Lemma D.3.** *Let  $A$  be a symmetric matrix matrix such that  $|A| = 1$  and let  $Y = (Y_1, \dots, Y_n)$  be a vector of independent standard normal random variables. Then for any  $\alpha > 0$ ,*

$$P(Y^t A Y - \text{tr}(A) > n^\alpha) \leq \exp\{-n^\alpha/8\}.$$

*Proof.* Note that  $\|A\| \leq |A| = 1$  so that for all  $s \leq 1/4$ ,  $sy^t A y \leq s_0 y^t y \|A\| \leq y^t y/4$  and  $\exp\{sY^t A Y\}$  has finite expectation. Choose  $s = 1/4$ , then by Markov's inequality,

$$\begin{aligned} P(Y^t A Y - \text{tr}(A) > n^\alpha) & \leq e^{-n^\alpha/4} E e^{(Y^t A Y - \text{tr}(A))/4} \\ & = \exp\left\{-n^\alpha/4 - \frac{1}{2} \log \det[I_n - A/2] - \text{tr}(A)/4\right\} \\ & \leq \exp\{-n^\alpha/4 + \text{tr}(A^2)/4\}. \end{aligned}$$

The last inequality follows from the fact that  $A(I_n - \tau A/2)^{-1}$  has eigenvalues  $\lambda_j(1 - \tau\lambda_j/2)^{-1}$ , where  $\lambda_j$  are the eigenvalues of  $A$  for all  $\tau \in (0, 1)$ . Hence,  $\text{tr}(A^2(I_n - \tau A/2)^{-2})$  is bounded by  $4\text{tr}(A^2)$ . The result follows from the fact that when  $n$  is large enough  $n^\alpha > 2\text{tr}(A^2) = 2$ .  $\square$

## E Convergence of the trace of a product of Toeplitz matrices

Suppose  $T_n(f_j)$  ( $j = 1, \dots, p$ ) are covariance matrices associated with spectral densities  $f_j$ . According to a classical result by Grenander and Szegö (Grenander and Szegö (1958)),

$$\frac{1}{n} \operatorname{tr} \left[ \prod_{j=1}^p T_n(f_j) \right] \rightarrow (2\pi)^{2p-1} \int_{-\pi}^{\pi} \prod_{j=1}^p f_j(x) dx.$$

In this section we give a series of related results. We first recall a result from Rousseau et al. (2010).

**Lemma E.1.** *Let  $1/2 > t > 0$  and  $L^{(i)}, M^{(i)} > 0$ ,  $\rho_i \in (0, 1]$ ,  $d_i \in [-1/2 + t, 1/2 - t]$  for all  $i = 1, \dots, 2p$  and let  $f_i$ , ( $i \leq 2p$ ) be functions on  $[-\pi, \pi]$  satisfying*

$$|f_i(x)| = |x|^{-2d_i} g_i(x), \quad |g_i(x)| \leq M^{(i)}, \quad |g_i(x) - g_i(y)| \leq \frac{M^{(i)} |x - y|}{|x| \wedge |y|} + L^{(i)} |x - y|^{\rho_i} \quad (\text{E.1})$$

and assume that  $\sum_{i=1}^p (d_{2i-1} + d_{2i}) < \frac{1}{2}$ . Then for all  $\epsilon > 0$  there exists a constant  $K$  depending only on  $\epsilon, t$  and  $q = \sum_{j=1}^p (d_{2j-1} + d_{2j})_+$  such that

$$\begin{aligned} & \left| \frac{1}{n} \operatorname{tr} \left[ \prod_{j=1}^p T_n(f_{2j-1}) T_n(f_{2j}) \right] - (2\pi)^{2p-1} \int_{-\pi}^{\pi} \prod_{j=1}^{2p} f_j(x) dx \right| \\ & \leq K \sum_{j=2}^{2p} \left( \prod_{i \neq j} M^{(i)} \right) L^{(j)} n^{-\rho_j + \epsilon + 2q} + K \prod_{i=1}^{2p} M^{(i)} n^{-1+q+\epsilon}. \end{aligned}$$

To prove a similar result involving also inverses of matrices, we need the following two lemmas. They can be found elsewhere, but as we make frequent use of them they are included for easy reference and are formulated in a way better suited to our purpose. The first lemma can be found on p.19 of Rousseau et al. (2010), and is an extension of Lemma 5.2 in Dahlhaus (1989).

**Lemma E.2.** *Suppose that for  $0 < t < 1/2$  and  $d \in [-1/2 + t, 1/2 - t]$*

$$|f(x)| = |x|^{-2d} g(x), \quad m \leq |g(x)| \leq M, \quad |g(x) - g(y)| \leq L |x - y|^\rho \quad (\text{E.2})$$

and assume that  $0 < m \leq 1 \leq M < +\infty$  and  $L \geq 1$ . Then, for all  $\epsilon > 0$ , there exists a constant  $K$  depending on  $t$  and  $\epsilon$  only such that

$$|I_n - T_n^{\frac{1}{2}}(f) T_n \left( \frac{1}{4\pi^2 f} \right) T_n^{\frac{1}{2}}(f)|^2 \leq KL \frac{M^2}{m^2} n^{1-\rho+\epsilon}.$$



*Proof.* By Lemma E.1,

$$\begin{aligned} |I_n - T_n^{\frac{1}{2}}(f)T_n \left( \frac{1}{4\pi^2 f} \right) T_n^{\frac{1}{2}}(f)|^2 &= \text{tr} \left\{ I_n - 2T_n^{\frac{1}{2}}(f)T_n \left( \frac{1}{4\pi^2 f} \right) T_n^{\frac{1}{2}}(f) \right. \\ &\quad \left. + T_n^{\frac{1}{2}}(f)T_n \left( \frac{1}{4\pi^2 f} \right) T_n(f)T_n \left( \frac{1}{4\pi^2 f} \right) T_n^{\frac{1}{2}}(f) \right\} \end{aligned}$$

converges to zero, the approximation error being bounded by  $K[L(1+M^2/m^2)+M^2/m^2]$ .  $\square$

The next result can be found as Lemma 3 in Lieberman et al. (2011), and is an extension of Lemma 5.3 in Dahlhaus (1989).

**Lemma E.3.** *Suppose that  $f_1$  and  $f_2$  are such that  $|f_1(x)| \geq m|x|^{-2d_1}$  and  $|f_2(x)| \leq M|x|^{-2d_2}$  for constants  $d_1, d_2 \in (-\frac{1}{2}, \frac{1}{2})$  and  $m, M > 0$ . Then*

$$\|T_n^{-\frac{1}{2}}(f_1)T_n^{\frac{1}{2}}(f_2)\| \leq C \frac{M}{m} n^{(d_2-d_1)_+ + \epsilon}.$$

*Proof.* In the proof of Lemma 5.3 on p. 1761 in Dahlhaus (1989), the first inequality only depends on the upper and lower bounds  $m$  and  $M$ .  $\square$

Using the preceding lemmas, the approximation result given in Lemma E.1 for traces of matrix products can be extended to include matrix inverses.

**Lemma E.4.** *Suppose that  $f$  satisfies (E.2) with constants  $d, \rho, L, m$  and  $M$ . For  $f_{2j}, j = 1, \dots, p$ , assume that (E.1) holds with constants  $d_{2j}, \rho_{2j}, L^{(2j)}$  and  $M^{(2j)}$  ( $j = 1, \dots, p$ ). For convenience, we denote  $M^{(2j-1)} = m^{-1}$ ,  $\rho_{2j-1} = \rho$  and  $L^{(2j-1)} = L$  ( $j = 1, \dots, p$ ). Suppose in addition that  $d, d_{2j} \in [-\frac{1}{2} + t, \frac{1}{2} - t]$  satisfy  $\sum_{j=1}^p (d_{2j} - d)_+ < \frac{1}{2}(\rho - \frac{1}{2})$ , and let  $q = \sum_{j=1}^p (d_{2j} - d)_+$ . Then for all  $\epsilon > 0$  there exists a constant  $K$  such that*

$$\begin{aligned} &\left| \frac{1}{n} \text{tr} \left\{ \prod_{j=1}^p T_n^{-1}(f)T_n(f_{2j}) \right\} - \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{j=1}^p \frac{f_{2j}(x)}{f(x)} dx \right| \\ &\leq K \left[ \sum_{j=2}^{2p} \left( \prod_{i \neq j}^{2p} M^{(i)} \right) L^{(j)} n^{-\rho_j} + n^{-1} \prod_{i \leq 2p} M^{(i)} \right] n^{\epsilon+2q} \\ &\quad + \left( \prod_{j=1}^p M^{(2j)} \right) \left( L \frac{M}{m} \right)^{\frac{(p+1)}{2}} n^{(1-\rho)\frac{(p+1)}{2} - 1 + \epsilon + 2q}, \end{aligned} \quad (\text{E.3})$$

and setting  $\tilde{f} = 1/(4\pi^2 f)$ ,

$$\begin{aligned} &\left| \frac{1}{n} \text{tr} \left\{ \prod_{j=1}^p T_n^{-1}(f)T_n(f_{2j}) \right\} - \text{tr} \left\{ \prod_{j=1}^p T_n(\tilde{f})T_n(f_j) \right\} \right| \\ &\leq \left( \prod_{j=1}^p M^{(2j)} \right) \left( L \frac{M}{m} \right)^{\frac{(p+1)}{2}} n^{(1-\rho)\frac{(p+1)}{2} - 1 + \epsilon + 2q}. \end{aligned} \quad (\text{E.4})$$

*Proof.* Without loss of generality, we consider the  $f_{2j}$ 's to be nonnegative. When this is not the case, we write  $f_{2j} = f_{2j}^+ - f_{2j}^-$  and treat the positive and negative part separately; see also Dahlhaus (1989), p. 1755-56. To prove (E.4), we use the construction of Lemma 5 from Lieberman et al. (2011), who treat the case  $\rho = 1$  and  $d_{2j} = d'$ . Inspection of their proof shows that this extends to  $\rho \neq 1$  and  $d_{2j}$  that differ with  $j$ . To prove (E.3), we use the construction of Dahlhaus' Theorem 5.1 (see also the remark on p. 744 of Lieberman and Phillips (2004), after (28)), and apply Lemma E.1 with  $f_{2j-1} = \tilde{f} = \frac{1}{4\pi^2 f}$ ,  $j = 1, \dots, p$ . This gives the first term on the right in (E.3). The last term in (E.3) follows from (E.4).  $\square$

Although the bound provided by Lemma E.4 is sufficiently tight for most purposes, certain applications require sharper bounds. These can only be obtained if we exploit specific properties of  $f$  and  $f_{2j}$ . In Lemma E.5 below we improve on the first term on the right in (E.3). This is useful when for example  $b_i(x) = \cos(jx)$ ; the Lipschitz constant  $L$  is then of order  $O(k)$ , but the boundedness of  $b_i$  actually allows a better result. In Lemma E.6 we improve on the last term of (E.3).

**Lemma E.5.** *Let  $f(x) = |x|^{-2d}g(x)$  with  $-1/2 < d < 1/2$  and  $g$  a bounded Lipschitz function satisfying  $m < g < M$ , with Lipschitz constant  $L$ .*

- *Let  $b_1, \dots, b_p$  be bounded functions and let  $\|b\|_\infty$  denote a common upper bound for these functions. Then for all  $\epsilon > 0$ ,*

$$\begin{aligned} & \left| \operatorname{tr} \left[ \prod_{i=1}^p T_n(b_i f) T_n(f^{-1}) \right] - (2\pi)^p \operatorname{tr} \left[ \prod_{i=1}^p T_n(b_i) \right] \right| \\ & \leq C n^\epsilon \left( \frac{M}{m} \right)^p \|b\|_\infty^{p-1} \left( \|b\|_\infty + L \sum_{j=1}^p \|b_j\|_2 \right). \end{aligned} \quad (\text{E.5})$$

- *Let  $b_j$  ( $j \geq 2$ ) be bounded functions. Let  $b_1$  be such that  $\|b_1\|_2 < +\infty$ , and assume that for all  $a > 0$  there exists  $M'(a) > 0$  such that*

$$\int_{-\pi}^{\pi} |b_1(x)| |x|^{-1+a} dx \leq M'(a).$$

*Then for all  $a > 0$*

$$\begin{aligned} & \left| \operatorname{tr} \left[ \prod_{i=1}^p T_n(b_i f) T_n(f^{-1}) \right] - (2\pi)^p \operatorname{tr} \left[ \prod_{i=1}^p T_n(b_i) \right] \right| \\ & \leq C \left( \frac{M}{m} \right)^p \prod_{i \geq 2} \|b_i\|_\infty (n^{3pa} M'(a) + L(\log n)^{2p-1} \|b_1\|_2). \end{aligned} \quad (\text{E.6})$$

*Proof.* We prove (E.5); the proof of (E.6) follows exactly the same lines. We define  $\Delta_n(x) = e^{ix}$  and  $L_n(x) = n \wedge |x|^{-1}$  where the latter is an upper bound

of the former. Using the decomposition as on p. 1761 in Dahlhaus (1989) or as in the proof of we find that

$$\begin{aligned}
& \left| \operatorname{tr} \left[ \prod_{i=1}^p T_n(b_i f) T_n(f^{-1}) \right] - (2\pi)^p \operatorname{tr} \left[ \prod_{i=1}^p T_n(b_i) \right] \right| \\
& \leq C \left| \int_{[-\pi, \pi]^{2p}} \prod_{i=1}^p b_i(x_{2i-1}) \left( \prod_{i=1}^p \frac{f(x_{2i-1})}{f(x_{2i})} - 1 \right) \Delta_n(x_1 - x_2) \dots \Delta_n(x_{2p} - x_1) dx \right| \\
& \leq C \left| \int_{[-\pi, \pi]^{2p}} \prod_{i=1}^p b_i(x_{2i-1}) \frac{g(x_{2i-1})}{g(x_{2i})} \left( \prod_{i=1}^p \frac{|x_{2i-1}|^{-2d}}{|x_{2i}|^{-2d}} - 1 \right) \Delta_n(x_1 - x_2) \dots \Delta_n(x_{2p} - x_1) dx \right| \\
& \quad + C \left| \int_{[-\pi, \pi]^{2p}} \prod_{i=1}^p b_i(x_{2i-1}) \left( \prod_{i=1}^p \frac{g(x_{2i-1})}{g(x_{2i})} - 1 \right) \Delta_n(x_1 - x_2) \dots \Delta_n(x_{2p} - x_1) dx \right| \\
& \leq C \left( \frac{M \|b\|_\infty}{m} \right)^p \sum_{j=1}^p \int_{[-\pi, \pi]^{2p}} \prod_{i=1}^j \frac{|x_{2i-1} - x_{2i}|^{1-3a}}{(|x_{2i}| \wedge |x_{2i-1}|)^{1-a}} L_n(x_1 - x_2) \dots L_n(x_{2p} - x_1) dx \\
& \quad + CL \left( \frac{M \|b\|_\infty}{m} \right)^{p-1} \sum_{j=1}^p \int_{[-\pi, \pi]^{2p}} |b_j(x_{2j-1})| |x_{2j-1} - x_{2j}| L_n(x_1 - x_2) \dots L_n(x_{2p} - x_1) dx \\
& \leq C \left( \frac{M \|b\|_\infty}{m} \right)^p n^{3pa} \left( \int_{[-\pi, \pi]} |x|^{-1+a} dx \right)^p + CL \left( \frac{M \|b\|_\infty}{m} \log n \right)^{p-1} (\log n)^{2p-1} \sum_{j=1}^p \|b_j\|_2.
\end{aligned}$$

□

**Lemma E.6.** *Let  $\tilde{f} = 1/(4\pi^2 f)$ , and let  $\rho > 1/2$  and  $L > 1$ , then under the conditions of Lemma E.4 we have the following alternative bound for (E.4):*

$$\begin{aligned}
& \left| \operatorname{tr} \left\{ \prod_{j=1}^p T_n^{-1}(f) T_n(f_{2j}) \right\} - \operatorname{tr} \left\{ \prod_{j=1}^p T_n(\tilde{f}) T_n(f_{2j}) \right\} \right| \\
& \lesssim \sqrt{L} n^{(1-\rho/2)+2q+\epsilon} \left\{ \sum_{j=1}^{p-1} \left( \sqrt{M_{2p}} \prod_{l=j+1}^{p-1} M^{(2l)} \right) \times \left( \int_{-\pi}^{\pi} \frac{|f_{2p}(x)|}{f(x)} \prod_{l=1}^j \frac{f_{2l}^2}{f^2}(x) dx \right)^{\frac{1}{2}} \right. \\
& \quad \left. + \prod_{l=2}^p M^{(2l)} \left( \int_{-\pi}^{\pi} \frac{f_{2l}^2}{f^2}(x) dx \right)^{\frac{1}{2}} + \text{error} \right\}
\end{aligned} \tag{E.7}$$

where

$$\text{error} \leq L^{3/4} n^{(1-3\rho)/4} \prod_{l=1}^p M^{(2l)} + \sum_{j=1}^p \sqrt{L^{(2j)} n^{-\rho_{2j}} M^{(2j)}} \prod_{l \neq j} M^{(2l)}$$

**Remark E.1.** *The constant appearing on the right hand side of (E.7) depends on  $M$  and  $m$ , but in all our applications of Lemma E.6, the constants  $M$  and  $m$  will be bounded and of no consequence.*

*Proof.* Following the construction of Dahlhaus (1989), equation (13), we write  $|\operatorname{tr}\{\prod_{j=1}^p T_n^{-1}(f)T_n(f_{2j})\} - \operatorname{tr}\{\prod_{j=1}^p T_n(\tilde{f})T_n(f_{2j})\}|$  as

$$\begin{aligned} & \left| \operatorname{tr} \left\{ \prod_{j=1}^p A_j - \prod_{j=1}^p B_j \right\} \right| \\ &= \left| \operatorname{tr} \left\{ (A_1 - B_1) \prod_{l=2}^p A_l + \sum_{j=2}^p \left( \prod_{l=1}^{j-1} B_l \right) (A_j - B_j) \prod_{l=j+1}^p A_l \right\} \right|, \end{aligned} \quad (\text{E.8})$$

where  $A_j = T_n^{\frac{1}{2}}(f_{2j-2})T_n^{-1}(f)T_n^{\frac{1}{2}}(f_{2j})$ ,  $B_j = T_n^{\frac{1}{2}}(f_{2j-2})T_n(\tilde{f})T_n^{\frac{1}{2}}(f_{2j})$  and  $f_0 := f_{2p}$  (similarly for  $\rho_0$ ,  $L^{(0)}$  and  $M^{(0)}$ ). When  $j = p$ , the factor  $\prod_{l=j+1}^p A_l$  is understood to be the identity. Without loss of generality, the functions  $f_{2j}$  are assumed to be positive (it suffices to write  $f_{2j} = f_{2j+} - f_{2j-}$ ). Lemma E.3 implies that for each  $j$ ,

$$\|T_n^{-\frac{1}{2}}(f)T_n^{\frac{1}{2}}(f_{2j})\| \lesssim \frac{M^{(2j)}}{m} n^{(d_{2j}-d)_++\epsilon}. \quad (\text{E.9})$$

Using the relations in (1.6) (main paper) it then follows that

$$\begin{aligned} \left\| \prod_{l=j+1}^p A_l \right\| &\leq \prod_{l=j+1}^p \|T_n^{\frac{1}{2}}(f_{2l})T_n^{-\frac{1}{2}}(f)\|^2 \\ &\lesssim \left( \prod_{l=j+1}^{p-1} M^{(2l)} \right) n^{2\sum_{l=j+1}^{p-1} (d_{2l}-d)_++(d_{2j}-d)_++(d_{2p}-d)_+} \sqrt{M^{(2p)}M^{(2j)}}. \end{aligned} \quad (\text{E.10})$$

First we treat the term  $(A_1 - B_1) \prod_{l=2}^p A_l$  on the right in (E.8). Writing  $R =$

$I_n - T_n^{\frac{1}{2}}(f)T_n(\tilde{f})T_n^{\frac{1}{2}}(f)$ , it follows that

$$\begin{aligned}
& \left| \operatorname{tr} \left[ (A_1 - B_1) \prod_{l=2}^p A_l \right] \right| \\
&= \left| \operatorname{tr} \left[ T_n^{\frac{1}{2}}(f_{2p})T_n^{-\frac{1}{2}}(f)RT_n^{-\frac{1}{2}}(f)T_n(f_2)T_n^{-\frac{1}{2}}(f)T_n^{-\frac{1}{2}}(f)T_n^{\frac{1}{2}}(f_4) \prod_{l=3}^p A_l \right] \right| \\
&\leq |R| \|T_n^{-\frac{1}{2}}(f)T_n(f_2)T_n^{-\frac{1}{2}}(f)\| \|T_n^{\frac{1}{2}}(f_{2p})T_n^{-\frac{1}{2}}(f)\| \|T_n^{-\frac{1}{2}}(f)T_n^{\frac{1}{2}}(f_4)\| \left\| \prod_{l=3}^p A_l \right\| \\
&\lesssim L^{\frac{1}{2}} n^{(1-\rho)/2+\epsilon+\frac{1}{2}+2q} \prod_{l=2}^p M^{(2l)} \left( \int_{-\pi}^{\pi} \frac{f_2^2(x)}{f^2(x)} dx + \text{error} \right)^{\frac{1}{2}} \\
&\lesssim L^{\frac{1}{2}} n^{(1-\rho)/2+\epsilon+\frac{1}{2}+2q} \prod_{l=2}^p M^{(2l)} \times \\
&\quad \left( \int_{-\pi}^{\pi} \frac{f_2^2(x)}{f^2(x)} dx + n^{\epsilon+2q} \left( L^{3/2} \left( M^{(2)} \right)^2 n^{(1-3\rho)/2} + M^{(2)} L^{(2)} n^{-\rho_2} \right) \right)^{\frac{1}{2}}.
\end{aligned} \tag{E.11}$$

The first inequality follows from the relations in (1.6) (main paper). The second inequality follows after writing  $|T_n^{-\frac{1}{2}}(f)T_n(f_2)T_n^{-\frac{1}{2}}(f)|$  as the sum of a limiting integral and an approximation error; in addition we use (E.9) and Lemma E.2, by which

$$|R|^2 \leq KL(M/m)^2 n^{1-\rho+\epsilon} \lesssim Ln^{1-\rho+\epsilon}. \tag{E.12}$$

This follows from Lemma E.4, which we use to bound the approximation error. The second term within the brackets in (E.11) constitutes part of the term *error*.

Next we bound the term  $(\prod_{l=1}^{j-1} B_l)(A_j - B_j) \prod_{l=j+1}^p A_l$  in (E.8) for  $j = 2$ . Similar to the preceding decomposition, we have

$$\begin{aligned}
& \left| \operatorname{tr} \left[ B_1(A_2 - B_2) \prod_{l=3}^p A_l \right] \right| = \left| \operatorname{tr} \left[ B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f) R T_n^{-\frac{1}{2}}(f) T_n^{\frac{1}{2}}(f_4) \prod_{l=3}^p A_l \right] \right| \\
&\leq |B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f)| |R| \|T_n^{-\frac{1}{2}}(f) T_n^{\frac{1}{2}}(f_4)\| \left\| \prod_{l=3}^p A_l \right\|.
\end{aligned}$$

The terms  $|R|$ ,  $\|T_n^{-\frac{1}{2}}(f) T_n^{\frac{1}{2}}(f_4)\|$  and  $\|\prod_{l=3}^p A_l\|$  are bounded as in (E.9), (E.10)

and (E.12). For the term  $|B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f)|$  we have the decomposition

$$\begin{aligned} |B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f)|^2 &= \text{tr} \left[ T_n^{-\frac{1}{2}}(f) T_n^{\frac{1}{2}}(f_2) B_1^t B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f) \right] \\ &= \text{tr} \left[ B_1^t B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-1}(f) T_n^{\frac{1}{2}}(f_2) \right] = \text{tr} \left[ B_1^t B_1 T_n^{\frac{1}{2}}(f_2) T_n(\tilde{f}) T_n^{\frac{1}{2}}(f_2) \right] \\ &\quad + \text{tr} \left[ B_1^t B_1 T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f) R T_n^{-\frac{1}{2}}(f) T_n^{\frac{1}{2}}(f_2) \right] \\ &\leq |B_1 T_n^{\frac{1}{2}}(f_2) T_n^{\frac{1}{2}}(\tilde{f})|^2 + |B_1^t B_1| \|R\| \|T_n^{\frac{1}{2}}(f_2) T_n^{-\frac{1}{2}}(f)\|^2. \end{aligned}$$

Using again Lemmas E.1, E.2 and E.3, we find that the first term on the right is bounded by

$$n \left\{ \int_{-\pi}^{\pi} \frac{|f_{2p}(x)| f_2^2(x)}{f^3(x)} dx + M^{(2p)} M^{(2)} n^{4(d_2-d)_++2(d_{2p}-d)_++\epsilon} \left( M^{(2)} L n^{-\rho} + L^{(2)} n^{-\rho_2} \right) \right\}$$

and the second term by

$$\begin{aligned} n^{\frac{1}{2}-\frac{\rho}{2}+(d_2-d)_++\epsilon} \sqrt{L} M^{(2)} \left[ n \int_{-\pi}^{\pi} \frac{f_{2p}^2(x) f_2^2(x)}{f^4(x)} dx + M^{(2p)} M^{(2)} n^{4(d_2-d)_++2(d_{2p}-d)_++\epsilon} \times \right. \\ \left. \times \left( (M^{(2)})^2 (M^{(2p)})^2 L n^{1-\rho} + (M^{(2p)})^2 M^{(2)} L^{(2)} n^{1-\rho_2} + (M^{(2)})^2 M^{(2p)} L^{(2p)} n^{1-\rho_{2p}} \right) \right]^{\frac{1}{2}}. \end{aligned}$$

Consequently,

$$\begin{aligned} &\left| \text{tr} \left[ B_1 (A_2 - B_2) \prod_{l=3}^p A_l \right] \right| \\ &\lesssim L^{\frac{1}{2}} n^{\frac{1}{2}+(1-\rho)/2+\epsilon+2q} \sqrt{M^{(2p)}} \prod_{l=2}^{p-1} M^{(2l)} \left[ \int_{-\pi}^{\pi} \frac{f_2^2(x) |f_{2p}(x)|}{f^3(x)} dx + M^{(2p)} (M^{(2)})^2 L n^{-\rho} \right. \\ &\quad \left. + n^{-\rho/2} (L)^{1/2} \sqrt{M^{(2)}} \left( \int_{-\pi}^{\pi} \frac{f_2^2(x) f_{2p}^2(x)}{f^4(x)} dx \right)^{\frac{1}{2}} + L^{(2)} M^{(2)} M^{(2p)} n^{-\rho_2} \right. \\ &\quad \left. + (LL^{(2)})^{1/2} M^{(2p)} (M^{(2)})^{3/2} n^{-(\rho+\rho_2)/2} + (LL^{(2p)} M^{(2p)})^{1/2} (M^{(2)})^2 n^{-(\rho+\rho_{2p})/2} \right]^{\frac{1}{2}}. \end{aligned}$$

Note that

$$\left( \int_{-\pi}^{\pi} \frac{f_2^2(x) f_{2p}^2(x)}{f^4(x)} dx \right)^{\frac{1}{2}} \lesssim M^{(2)} M^{(2p)},$$

$$(LL^{(2)} M^{(2p)})^{1/2} (M^{(2)})^{3/2} n^{-(\rho+\rho_2)/2} \leq L (M^{(2)})^2 n^{-\rho} + L^{(2)} M^{(2)} M^{(2p)} n^{-\rho_2}$$

and  $L n^{-\rho} \lesssim L^{3/2} n^{(1-3\rho)/2}$ . Therefore the terms on the right are of the same order as the right hand side of (E.11). A similar argument applies to the term  $(LL^{(2p)} M^{(2p)})^{1/2} (M^{(2)})^2 n^{-(\rho+\rho_{2p})/2}$ .

Finally, we bound the term  $(\prod_{l=1}^{j-1} B_l) (A_j - B_j) \prod_{l=j+1}^p A_l$  in (E.8) for  $j \geq 3$ . For  $j \geq 3$ , Lemma E.1 implies that

$$\left| \prod_{l=1}^{j-1} B_l \right|^2 = n \left( \int_{-\pi}^{\pi} \frac{f_{2p}(x) f_{2j-2}(x)}{f^2(x)} \prod_{l=1}^{j-2} \frac{f_{2l}^2(x)}{f^2(x)} dx + error_j \right),$$

where

$$error_j \lesssim n^{\epsilon+2\sum_{l=1}^{j-1}(d_{2l}-d)_+} \prod_{l=1}^{j-1} M^{(2l)} M^{(2l-2)} \left( Ln^{-\rho} + \sum_{l=1}^{j-1} \frac{L^{(2l)}}{M^{(2l)}} n^{-\rho_{2l}} \right).$$

Consequently, we have for all  $j \geq 2$

$$\begin{aligned} & \left| \text{tr} \left[ \left( \prod_{l=1}^{j-1} B_l \right) (A_j - B_j) \prod_{l=j+1}^p A_l \right] \right| \\ & \leq \left| \prod_{l=1}^{j-1} B_l \right| \|R\| \prod_{l=j+1}^p \|A_l\| \|T_n^{\frac{1}{2}}(f_{2j}) T_n^{-\frac{1}{2}}(f)\| \|T_n^{\frac{1}{2}}(f_{2j-2}) T_n^{-\frac{1}{2}}(f)\| \\ & \lesssim L^{\frac{1}{2}} n^{\frac{1}{2} + (1-\rho)/2 + 2q + \epsilon} \sqrt{M^{(2p)}} \prod_{l=j+1}^{p-1} M^{(2l)} \left( \int_{-\pi}^{\pi} \frac{f_{2p}(x) f_{2j}(x)}{f^2(x)} \prod_{l=1}^{j-1} \frac{f_{2l}^2(x)}{f^2(x)} dx + error_j \right)^{\frac{1}{2}}. \end{aligned}$$

for all  $j \geq 3$ , which finishes the proof of Lemma E.6.  $\square$

## F Hölder constants of various functions

**Lemma F.1.** *Let  $\theta_o \in \Theta(\beta, L_o)$ . Then  $f_o$  satisfies condition (E.2) with  $\rho = 1$  when  $\beta > \frac{3}{2}$ , and with any  $\rho < \beta - \frac{1}{2}$  when  $\beta \leq \frac{3}{2}$ . The Hölder-constant only depends on  $L_o$ . When  $\theta \in \Theta_k(\beta, L)$ ,  $f_{d,k,\theta}$  satisfies (E.2) with  $\rho = 1$ , regardless of  $\beta$ . The Hölder-constant is of order  $k^{\frac{3}{2}-\beta}$ . The function  $-\log(2 - 2\cos(x)) f_{d,k,\theta}$  satisfies condition (E.1) with  $\rho = 1$  and Hölder-constant of order  $k^{\frac{3}{2}-\beta}$ . The functions  $G_k f_{d,k,\theta}$  and  $H_k f_{d,k,\theta}$ , with  $G_k$  and  $H_k$  as in (D.1), satisfy (E.1) with  $\rho = 1$  and Hölder-constant of order  $k$ .*

*Proof.* The function  $\sum_{j=0}^{\infty} \theta_{o,j} \cos(jx)$  (i.e. the logarithm of the short-memory part of  $f_o$ ), has smoothness  $\rho < \beta - \frac{1}{2}$ , since

$$\begin{aligned} \sum_{j=0}^{\infty} |\theta_{o,j}| |\cos(jx) - \cos(jy)| & \leq \left( \sum_{j=0}^{\infty} |\theta_{o,j}| j^{\rho} \right) |x - y|^{\rho} \\ & \leq \left( \sum_{j=0}^{\infty} \theta_{o,j}^2 j^{2\beta} \right)^{\frac{1}{2}} \left( \sum_{j=0}^{\infty} j^{-2(\beta-\rho)} \right)^{\frac{1}{2}} |x - y|^{\rho}, \end{aligned}$$

which is finite only when  $\rho < \beta - \frac{1}{2}$ . Since  $\sum_{j=0}^{\infty} |\theta_j| \lesssim \sqrt{L}$  when  $\theta \in \Theta(\beta - 1/2, L)$  and  $\beta > 1$ , the functions  $\sum_{j=0}^{\infty} \theta_j \cos(jx)$  and  $\exp\{\sum_{j=0}^{\infty} \theta_j \cos(jx)\}$  have the same smoothness; only the values of  $L$  and  $M$  differ. The same calculation can be made when the FEXP-expansion is finite: when  $\theta \in \Theta_k(\beta, L)$ , then for all  $x, y \in [-\pi, \pi]$ ,

$$\left| \sum_{j=0}^k \theta_j (\cos(jx) - \cos(jy)) \right| \leq |x - y| \sum_{j=0}^k j |\theta_j| \lesssim \sqrt{L} k^{\frac{3}{2}-\beta} |x - y|. \quad (\text{F.1})$$

Since

$$|G_k(x) - G_k(y)| \leq 2 \sum_{j=1}^k \eta_j |\cos(jx) - \cos(jy)| = O(k)|x - y|, \quad (\text{F.2})$$

$G_k f_{d,k,\theta}$  has Hölder-smoothness  $\rho = 1$ , its Hölder-constant being  $O(k)$ . The same result holds for  $H_k f_{d,k,\theta}$ , since  $H_k(x) = -\log(2 - 2\cos(x)) - G_k(x)$  (see (D.3)) and  $k^{\frac{3}{2}-\beta} = o(k)$  for all  $\beta > 1$ .  $\square$

## G Proof of Lemma B.2

For easy reference we first restate the result. Let  $W_\sigma(d)$  denote any of the quadratic forms

$$X^t T_n^{-1}(f_{d,k}) B_\sigma(d, \bar{\theta}_{d,k}) X - \text{tr} [T_n(f_o) T_n^{-1}(f_{d,k}) B_\sigma(d, \bar{\theta}_{d,k})]$$

in (B.2) (in the main paper). Then for any  $j \leq J$ ,  $(l_1, \dots, l_j) \in \{0, \dots, k\}^j$  and  $\sigma \in \mathcal{S}(l_1, \dots, l_j)$ , we have

$$|W_\sigma(d) - W_\sigma(d_o)| = \mathbf{o}_{\mathbf{P}_o}(|d - d_o| n^{\frac{1}{2}+\epsilon} k^{-\frac{1}{2}}), \quad (\text{G.1})$$

$$\begin{aligned} & \text{tr} [B_\sigma(d, \bar{\theta}_{d,k})] - \text{tr} [B_\sigma(d_o, \bar{\theta}_{d_o})] \\ &= (d - d_o) \text{tr} [T_{1,\sigma}(d_o, k)] + (d - d_o)^2 \mathbf{o}(n^{\epsilon+\frac{1}{2}} k^{-\frac{1}{2}+(1-\beta/2)+}) \\ &= (d - d_o) \text{tr} [T_{1,\sigma}(d_o, k)] + (d - d_o)^2 \mathbf{o}(n^{1-\delta}/k), \end{aligned} \quad (\text{G.2})$$

$$\begin{aligned} & \text{tr} [(T_n(f_o) T_n^{-1}(f_{d,k}) - I_n) B_\sigma(d, \bar{\theta}_{d,k})] - \text{tr} [(T_n(f_o) T_n^{-1}(f_{d_o,k}) - I_n) B_\sigma(d_o, \bar{\theta}_{d_o,k})] \\ &= (d - d_o) \text{tr} [T_{2,\sigma}(d_o, k)] + (d - d_o)^2 \mathbf{o}(n/k) + (d - d_o) \mathbf{o}(n^{\epsilon+\frac{1}{2}} k^{-\frac{1}{2}}). \end{aligned} \quad (\text{G.3})$$

*Proof of Lemma B.2.* We first prove (G.2). Developing the left-hand side in  $d$  we obtain, for all  $j$ ,  $(l_1, \dots, l_j) \in \{0, \dots, k\}^j$  and  $\sigma \in \mathcal{S}_j$ ,

$$\begin{aligned} & \text{tr} [B_\sigma(d, \bar{\theta}_{d,k})] - \text{tr} [B_\sigma(d_o, \bar{\theta}_{d_o})] \\ &= (d - d_o) \text{tr} [B'_\sigma(d_o, \bar{\theta}_{d_o,k})] + \frac{(d - d_o)^2}{2} \text{tr} [B''_\sigma(\bar{d}, \bar{\theta}_{\bar{d}})], \end{aligned} \quad (\text{G.4})$$

where  $\bar{d} \in (d, d_o)$ , and  $B'$  and  $B''$  denote the first and second derivative with respect to  $d$ , respectively. Writing

$$\begin{aligned} \tilde{B}_{\sigma(i)}(d, k) &= T_n(H_k \nabla_{\sigma(i)} f_{d,k}) T_n^{-1}(f_{d,k}) \\ &\quad - T_n(\nabla_{\sigma(i)} f_{d,k}) T_n^{-1}(f_{d,k}) T_n(H_k f_{d,k}) T_n^{-1}(f_{d,k}), \end{aligned}$$

it follows that  $B'_\sigma(d, \bar{\theta}_{d,k})$  equals

$$B'_\sigma(d, \bar{\theta}_{d,k}) = \sum_{i=1}^{|\sigma|} \prod_{j < i} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}) \tilde{B}_{\sigma(i)}(d, k) \prod_{j > i} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}).$$



We recall the definition of  $T_{1,\sigma}$  in Lemma B.1 (main paper), and conclude that  $B'_\sigma(d, \bar{\theta}_{d,k}) = T_{1,\sigma}(d, k)$ . Consequently, the first term on the right in (G.4) equals  $(d - d_o)\text{tr}[T_{1,\sigma}(d_o, k)]$ .

The second derivative  $B''_\sigma(d, \bar{\theta}_{d,k})$  equals

$$\begin{aligned} & 2 \sum_{i_1 < i_2}^{\lfloor \sigma \rfloor} \prod_{j < i_1} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}) \tilde{B}_{\sigma(i_1)}(d, k) \prod_{i_1 < j < i_2} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}) \\ & \quad \times \tilde{B}_{\sigma(i_2)}(d, k) \prod_{i_2 < j} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}) \\ & + \sum_{i=1}^{\lfloor \sigma \rfloor} \prod_{j < i} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}) \tilde{B}'_{\sigma(i)}(d, k) \prod_{i < j} T_n(\nabla_{\sigma(j)} f_{d,k}) T_n^{-1}(f_{d,k}). \end{aligned}$$

We now show that  $\text{tr} [B''_\sigma(d, \bar{\theta}_{d,k})] = \mathbf{o}(n^{\epsilon + \frac{1}{2}} k^{-\frac{1}{2} + (1-\beta/2)_+})$ . From Lemma E.4 and the above expression for  $B''_\sigma(d, \bar{\theta}_{d,k})$ , it can be seen that  $\text{tr} [B''_\sigma(d, \bar{\theta}_{d,k})]$  converges to zero. To bound the approximation error, we cannot use directly Lemma E.4 because the bound in (E.4) becomes too large when  $\beta < 2$  and  $|\sigma|$  is larger than 1. We therefore use Lemmas E.1 and E.6. Let  $A''_\sigma(d, \bar{\theta}_{d,k})$  be the matrix obtained after replacing every factor  $T_n^{-1}(f_{d,k})$  in  $B''_\sigma(d, \bar{\theta}_{d,k})$  by  $T_n(\tilde{f}_{d,k})$ , for  $\tilde{f}_{d,k} = f_{d,k}^{-1}/(4\pi^2)$ . We recall from Lemma F.1 that the Lipschitz constant of  $f_{d,k}$  is  $O(k^{(2-\beta)_+})$ , and for  $H_k^j f_{d,k}$  and  $H_k^j \nabla_{\sigma(m)} f_{d,k}$  ( $m \leq |\sigma|$ ,  $j = 1, 2$ ) it is  $O(k \log k)$ . Consequently, Lemma E.1 implies that

$$\left| \text{tr} [A''_\sigma(d, \bar{\theta}_{d,k})] \right| = O(kn^\epsilon) = \mathbf{o}(n^{\epsilon + \frac{1}{2}} k^{-\frac{1}{2} + (1-\beta/2)_+})$$

when  $k \leq k_n$  and  $\beta > 1$ . It follows from Lemma E.6 that

$$\begin{aligned} \left| \text{tr} [A''_\sigma(d, \bar{\theta}_{d,k})] - \text{tr} [B''_\sigma(d, \bar{\theta}_{d,k})] \right| &= O(n^{1/2 + \epsilon} k^{(1-\beta/2)_+}) \left( \int_{-\pi}^{\pi} H_k^2(x) dx \right)^{\frac{1}{2}} \\ &= \mathbf{o}(n^{\epsilon + \frac{1}{2}} k^{-\frac{1}{2} + (1-\beta/2)_+}). \end{aligned}$$

Note that in the case where  $B''_\sigma(d, \bar{\theta}_{d,k})$  contains a Toeplitz matrix of the form  $T_n(H_k^2 f_{d,k})$  or  $T_n(H_k^2 \nabla_{\sigma(m)} f_{d,k})$  then it contains no other Toeplitz matrix involving  $H_k$  and we can set  $f_2 = H_k^2 f_{d,k}$  or  $f_2 = H_k^2 \nabla_{\sigma(m)} f_{d,k}$  and use Remark 2.1; this leads to the above error rate. Combining the preceding results for  $|\text{tr}[A''_\sigma(d, \bar{\theta}_{d,k})]|$  and  $|\text{tr}[A''_\sigma(d, \bar{\theta}_{d,k})] - \text{tr}[B''_\sigma(d, \bar{\theta}_{d,k})]|$  we obtain that

$$\left| \text{tr} [B''_\sigma(d, \bar{\theta}_{d,k})] \right| = \mathbf{o}(n^{\epsilon + \frac{1}{2}} k^{-\frac{1}{2} + (1-\beta/2)_+}) = \mathbf{o}(n^{1-\delta}/k),$$

which completes the proof of (G.2).

Next, we prove (G.3). Writing  $f_o - f_{d,k} = f_o - f_{d_o,k} + f_{d_o,k} - f_{d,k}$ , it follows

that the left-hand side of (G.3) equals

$$\begin{aligned}
& \operatorname{tr} [T_n(f_o - f_{d,k})T_n^{-1}(f_{d,k})B_\sigma(d, \bar{\theta}_{d,k})] - \operatorname{tr} [T_n(f_o - f_{d_o,k})T_n^{-1}(f_{d_o,k})B_\sigma(d_o, \bar{\theta}_{d_o,k})] \\
&= \operatorname{tr} [T_n(f_o - f_{d_o,k}) \{T_n^{-1}(f_{d,k})B_\sigma(d, \bar{\theta}_{d,k}) - T_n^{-1}(f_{d_o,k})B_\sigma(d_o, \bar{\theta}_{d_o,k})\}] \\
&+ \operatorname{tr} [T_n(f_{d_o,k} - f_{d,k})T_n^{-1}(f_{d,k})B_\sigma(d, \bar{\theta}_{d,k})] \\
&:= C_1 + C_2.
\end{aligned}$$

Using (D.4) we write  $f_{d,k} = f_{d_o,k}e^{(d-d_o)H_k}$  and  $f_o = f_{d_o,k}e^{\Delta_{d_o,k}}$ , and we develop  $C_\sigma(d, \bar{\theta}_{d,k}) = T_n^{-1}(f_{d,k})B_\sigma(d, \bar{\theta}_{d,k})$  around  $d = d_o$ . It follows that

$$\begin{aligned}
C_1 &= \operatorname{tr} [T_n(f_o - f_{d_o,k}) \{T_n^{-1}(f_{d,k})B_\sigma(d, \bar{\theta}_{d,k}) - T_n^{-1}(f_{d_o,k})B_\sigma(d_o, \bar{\theta}_{d_o,k})\}] \\
&= (d - d_o)\operatorname{tr} [T_n(f_o - f_{d_o,k})C'_\sigma(d_o, \bar{\theta}_{d_o,k})] \\
&\quad + (d - d_o)^2 \int_0^1 (1-u)\operatorname{tr} [T_n(f_o - f_{d_o,k})C''_\sigma(d_u, \bar{\theta}_{d_u,k})] du,
\end{aligned}$$

with  $d_u = ud + (1-u)d_o$ . For the first term on the right, we write, using Lemmas E.1 and E.6,

$$\begin{aligned}
& \operatorname{tr} [T_n(f_o - f_{d_o,k})C'_\sigma(d_o, \bar{\theta}_{d_o,k})] \\
&= \frac{n}{2\pi} \int_{-\pi}^{\pi} \frac{f_o - d_{d_o,k}}{f_{d_o,k}} H_k(x) \cos(l_1 x) \dots \cos(l_{|\sigma|} x) dx + \text{error},
\end{aligned}$$

where  $\sigma$  is a partition of  $\{1, \dots, j\}$  and the error term is

$$O\left(\|\Delta_{d_o,k}\|_\infty n^\epsilon \left(k + k^{0.5(3/2-\beta)+\frac{\sqrt{n}}{\sqrt{k}}}\right) + k^{0.5(3/2-\beta)+\left(\frac{n}{k^{2\beta}} + k\|\Delta_{d_o,k}\|_\infty\right)^{\frac{1}{2}}}\right),$$

which is  $o(k^{-1/2}n^{1/2-\delta})$ . Similarly, Lemmas E.1 and E.6 imply that there exists  $c \in \mathbb{R}$  such that for all  $d \in (d_o - \bar{v}_n, d_o + \bar{v}_n)$

$$\begin{aligned}
& \operatorname{tr} [T_n(f_o - f_{d_o,k})C''_\sigma(d, \bar{\theta}_{d,k})] \\
&= \frac{cn}{2\pi} \int_{-\pi}^{\pi} \frac{f_o - d_{d_o,k}}{f_{d_o,k}} H_k^2(x) \cos(l_1 x) \dots \cos(l_{|\sigma|} x) dx + \text{error},
\end{aligned}$$

where the error term is of order

$$O\left(\|\Delta_{d_o,k}\|_\infty n^\epsilon \left(k + k^{\frac{1}{2}(2-\beta)+\frac{\sqrt{n}}{\sqrt{k}}}\right) + k^{\frac{1}{2}(2-\beta)+\left(\frac{n}{k^{2\beta}} + k\|\Delta_{d_o,k}\|_\infty\right)^{\frac{1}{2}}}\right) = o\left(\frac{n^{1-\delta}}{k}\right).$$

This implies that  $C_1 = O(S_n(d))$ .

Using a Taylor expansion of  $C_\sigma(d, \bar{\theta}_{d,k})$  and of  $e^{-(d-d_o)H_k}$  around  $d_o$ , it follows that

$$\begin{aligned}
C_2 &= -(d - d_o)\operatorname{tr} [T_n(f_{d_o,k}H_k)T_n^{-1}(f_{d_o,k})B_\sigma(d_o, \bar{\theta}_{d_o,k})] \\
&\quad - \frac{1}{2}(d - d_o)^2\operatorname{tr} [T_n(f_{d_o,k}H_k^2e^{-t(d'-d_o)H_k})C_\sigma(d', \bar{\theta}_{d',k}) + 2T_n(f_{d_o,k}H_k)C'_\sigma(d', \bar{\theta}_{d',k})],
\end{aligned}$$

for some  $d'$  between  $d$  and  $d_o$ . The first term equals  $\text{tr}[T_{2,\sigma}]$ . The second equals

$$\begin{aligned} & -\frac{1}{2}(d-d_o)^2 \text{tr} \left[ T_n(f_{d',k} H_k^2) C_\sigma(d', \bar{\theta}_{d',k}) + 2T_n(f_{d_o,k} H_k) C'_\sigma(d', \bar{\theta}_{d',k}) \right] \\ & = -\frac{n(d-d_o)^2}{2\pi} \int_{-\pi}^{\pi} H_k^2(x) \cos(l_1 x) \dots \cos(l_{|\sigma|} x) dx + \text{error}, \end{aligned}$$

where the error term is  $O(n^\epsilon (k + k^{0.5(2-\beta)} + (nk^{-1} + kn^\epsilon)^{1/2})) = o(k^{-1}n^{1-\delta})$ . Therefore

$$C_2 = (d-d_o) \text{tr}[T_{2,\sigma}] + O(n/k).$$

Finally, to prove (G.1), let  $Z = T_n^{-\frac{1}{2}}(f_o)X$  and let  $A_d = T_n^{\frac{1}{2}}(f_o)T_n^{-1}(f_{d,k})B_\sigma(d, \bar{\theta}_{d,k})T_n^{\frac{1}{2}}(f_o)$ . Then for any  $|d-d_o| \leq \bar{v}_n$ , we have

$$W_\sigma(d) - W_\sigma(d_o) = Z^t(A_d - A_{d_o})Z - \text{tr}(A_d - A_{d_o}).$$

Writing  $A'_d$  for the derivative of  $A_d$  with respect to  $d$ , it follows that

$$A_d - A_{d_o} = (d-d_o)A'_{\bar{d}}, \quad (\text{G.5})$$

for some  $\bar{d}$  between  $d$  and  $d_o$ . Using (D.6), we find that

$$\begin{aligned} A'_d &= T_n^{\frac{1}{2}}(f_o)T_n^{-1}(f_{d,k})T_n(H_k f_{d,k})T_n^{-1}(f_{d,k})T_n(B_\sigma(d, \bar{\theta}_{d,k}))T_n^{\frac{1}{2}}(f_o) \\ &+ T_n^{\frac{1}{2}}(f_o)T_n^{-1}(f_{d,k})B'_\sigma(d, \bar{\theta}_{d,k})T_n^{\frac{1}{2}}(f_o). \end{aligned}$$

Therefore, Lemma 2 of Lieberman et al. (2011) and the inequalities in (1.6) (main paper) imply that

$$\begin{aligned} |A_d - A_{d_o}| &\leq |d-d_o| |A'_{\bar{d}}| \\ &\leq C|d-d_o| \|T_n^{\frac{1}{2}}(f_o)T_n^{-\frac{1}{2}}(f_{\bar{d},k})\|^2 \prod_{i=1}^{|\sigma|} \|T_n^{-\frac{1}{2}}(f_{\bar{d},k})B_{\sigma(i)}(\bar{d}, \bar{\theta}_{\bar{d},k})T_n^{\frac{1}{2}}(f_{\bar{d},k})\| \\ &+ |T_n^{-\frac{1}{2}}(f_{\bar{d},k})T_n(H_k \nabla_{\sigma(i)} f_{\bar{d},k})T_n^{-\frac{1}{2}}(f_{\bar{d},k})| \\ &= |d-d_o| n^\epsilon O\left(|T_n^{-\frac{1}{2}}(f_{\bar{d},k})T_n(H_k f_{\bar{d},k})T_n^{-\frac{1}{2}}(f_{\bar{d},k})| + |T_n^{-\frac{1}{2}}(f_{\bar{d},k})T_n(H_k \nabla_{\sigma(i)} f_{\bar{d},k})T_n^{-\frac{1}{2}}(f_{\bar{d},k})|\right), \end{aligned} \quad (\text{G.6})$$

where  $\sigma(i)$  can also be the empty set, in which case  $\nabla_{\sigma(i)} f_{\bar{d},k} = f_{\bar{d},k}$ . We bound the terms between brackets using Lemma E.4, with  $p=2$ ,  $f = f_{\bar{d},k}$  and  $g_1 = g_2$  equalling either  $H_k f_{\bar{d},k}$  or  $H_k \nabla_{\sigma(i)} f_{\bar{d},k}$ . The Hölder constants of these functions are given by Lemma F.1. Hence we find that

$$\begin{aligned} |T_n^{-\frac{1}{2}}(f_{\bar{d},k})T_n(H_k f_{\bar{d},k})T_n^{-\frac{1}{2}}(f_{\bar{d},k})|^2 &= \text{tr} \left[ (T_n^{-1}(f_{\bar{d},k})T_n(H_k f_{\bar{d},k}))^2 \right] \\ &= \frac{n}{2\pi} \int_{-\pi}^{\pi} H_k^2(x) dx + O(n^\epsilon (k + k^{2-\beta})) = O(n^{1-1/(2\beta)} (\log n)^{1/(2\beta)}). \end{aligned} \quad (\text{G.7})$$

The last inequality follows from equation (D.7) in Lemma D.1 and the fact that  $k = k_n$  and  $\beta > 1$ . Similarly, it follows that

$$|T_n^{-\frac{1}{2}}(f_{\bar{d},k})T_n(H_k \nabla_{\sigma(i)} f_{\bar{d},k})T_n^{-\frac{1}{2}}(f_{\bar{d},k})|^2 = O(n^{1-1/(2\beta)}(\log n)^{1/(2\beta)}). \quad (\text{G.8})$$

Inserting (G.6), (G.7) and (G.8) in (G.5), we find that  $|A_d - A_{d_o}| \leq |d - d_o|n^{1/2-1/(4\beta)+\epsilon}$ , for all  $|d - d_o| \leq \bar{v}_n$  and all  $\epsilon > 0$ , when  $n$  is large enough. Consequently, we can apply Lemma 1.3 with  $A = (A_d - A_{d_o})/|A_d - A_{d_o}|$ , so that when  $n$  is large enough

$$\sup_{|d-d_o| \leq \bar{v}_n} P_o \left( |W_\sigma(d) - W_\sigma(d_o)| > |d - d_o|n^{2\epsilon + \frac{1}{2} - \frac{1}{4\beta}} \right) \leq e^{-n^\epsilon/8}. \quad (\text{G.9})$$

Using the above computations with  $|d - d'| \leq n^{-2}$ , we obtain

$$|W_\sigma(d) - W_\sigma(d')| \leq n^{-2+\epsilon} (n + Z^t Z).$$

Hence, for all  $\epsilon < \frac{1}{2}$  and  $c > 0$ ,

$$P_o \left( \sup_{|d'-d| \leq n^{-2}} |W_\sigma(d) - W_\sigma(d')| > n^{-\epsilon} \right) \leq P_o (Z^t Z > n^{2-2\epsilon}) \leq e^{-cn}, \quad (\text{G.10})$$

provided  $n$  is large enough. Hence, we obtain (G.1) by combining (G.9) and (G.10) in a simple chaining argument over the interval  $(d_o - \bar{v}_n, d_o + \bar{v}_n)$ .  $\square$