Submitted to the Annals of Statistics

FUNCTIONAL REGRESSION FOR GENERAL EXPONENTIAL FAMILIES

BY WEI DOU[*], DAVID POLLARD[†] AND HARRISON H. ZHOU[‡]

Yale University

The paper derives a minimax lower bound for rates of convergence for an infinite-dimensional parameter in an exponential family model. An estimator that achieves the optimal rate is constructed by maximum likelihood on finite-dimensional approximations with parameter dimension that grows with sample size.

1. Introduction. Our main purpose in this paper is to extend the theory developed by Hall and Horowitz (2007)—for regression with mean a linear functional of an unknown square integrable function $\mathbb{B}$ defined on a compact interval of the real line—to observations y_i from an exponential famly whose canonical parameter is of the form $\int_0^1 \mathbb{B}(t)\mathbb{X}_i(t)\,dt$ for observed Gaussian processes $\mathbb{X}_i$.

Our methods introduce several new technical devices. We establish a sharp approximation for maximum likelihood estimators for exponential families parametrized by linear functions of m-dimensional parameters, for an m that grows with sample size. We develop a change of measure argument—inspired by ideas from Le Cam's theory of asymptotic equivalence of models—to eliminate the effect of bias terms from the asymptotics of maximization estimators. And we obtain improved bounds for projections onto subspaces defined by eigenfunctions of perturbations of compact operators, bounds that simplify arguments involving estimates of unknown covariance kernels.

More precisely, we consider problems where the observed data consist of independent, identically distributed pairs $(y_i, \mathbb{X}_i)$ where each $\mathbb{X}_i$ is a Gaussian process indexed by a compact subinterval of the real line, which with no loss of generality we take to be $[0, 1]$. We write $\mathfrak{m}$ for Lebesgue measure on the Borel sigma-field of $[0, 1]$. We denote the corresponding norm and inner product in the space $\mathcal{L}^2(\mathfrak{m})$ by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$.

We assume the conditional distribution of y_i given the process $\mathbb{X}_i$ comes from

[*]Supported in part by NSF FRG grant DMS-0854975

[†]Supported in part by NSF grant MSPA-MCS-0528412

[‡]Supported in part by NSF Career Award DMS-0645676 and NSF FRG grant DMS-0854975

AMS 2000 subject classifications: Primary 62J05, 60K35; secondary 62G20

Keywords and phrases: Functional estimation, exponential families, minimax rates of convergence, approximation of compact operators.

an exponential family $\{Q_\lambda : \lambda \in \mathbb{R}\}$ with parameter

$$\lambda_i = a + \int_0^1 \mathbb{X}_i(t)\mathbb{B}(t)\,dt \tag{1}$$

for an unknown constant a and an unknown $\mathbb{B} \in \mathcal{L}^2(\mathfrak{m})$.

We focus on estimation of $\mathbb{B}$ using integrated squared error loss:

$$L(\mathbb{B}, \widehat{\mathbb{B}}_n) = \|\mathbb{B} - \widehat{\mathbb{B}}_n\|^2 = \int_0^1 \left(\mathbb{B}(t) - \widehat{\mathbb{B}}_n(t)\right)^2 dt.$$

In a companion paper we will show that our methods can be adapted to treat the problem of prediction of a linear functional $\int_0^1 x(t)\mathbb{B}(t)\,dt$ for a known x, extending theory developed by Cai and Hall (2006). In that paper we also consider some of the practical realities in applying the results to the economic problem of predicting occurence of recessions from the U.S. Treasury yield curve.

Our models are indexed by a set $\mathcal{F}$ of parameters $f = (a, \mathbb{B}, K, \mu)$, where μ is the mean and K is the covariance kernel of the Gaussian process. Under assumptions on $\mathcal{F}$ (see Section 3) analogous to the assumptions made by Hall and Horowitz (2007) for a problem of functional linear regression, we find a sequence $\{\rho_n\}$ that decreases to zero for which

$$\liminf_{n\to\infty} \sup_{f\in\mathcal{F}} \mathbb{P}_{n,f}\|\mathbb{B} - \widehat{\mathbb{B}}_n\|^2/\rho_n > 0 \qquad \text{for every estimating sequence } \{\widehat{\mathbb{B}}_n\} \tag{2}$$

and construct one particular estimating sequence of $\widehat{\mathbb{B}}_n$'s for which: for each $\epsilon > 0$ there exists a finite constant C_ϵ such that

$$\sup_{f\in\mathcal{F}} \mathbb{P}_{n,f}\{\|\mathbb{B} - \widehat{\mathbb{B}}_n\|^2 > C_\epsilon \rho_n\} < \epsilon \qquad \text{for large enough } n. \tag{3}$$

For the collection of models $\mathcal{F} = \mathcal{F}(R, \alpha, \beta)$ defined in Section 3, the rate ρ_n equals $n^{(1-2\beta)/(\alpha+2\beta)}$.

In Section 9 we establish a minimax lower bound by means of a variation on Assouad's Lemma.

We begin our analysis of the rate-optimal estimator in Section 4, with an approximation theorem for maximum likelihood estimators in exponential family models for parameters whose dimensions change with sample size. The main result is stated in a form slightly more general than we need for the present paper because we expect the result to find other sieve-like applications. The approximations from this section lie at the heart of our construction of an estimator that achieves the minimax rate from Section 9.

As an aid to the reader, we present our construction of the estimating sequence for (3) in two stages. First (Section 5) we assume that both the mean μ and the covariance kernel K are known. This allows us to emphasize the key ideas in our proofs without the many technical details that need to be handled when μ and K are estimated in the natural way. Many of those details involve the spectral theory of compact operators.

We have found some of the results that we need quite difficult to dig out of the spectral theory literature. In Section 6 we summarize the theory that we use to control errors when approximating K: some of it is a rearrangement of ideas from Hall and Horowitz (2007) and Hall and Hosseini-Nasab (2006); some is adapted from the notes by Bosq (2000) and the monograph by Birman and Solomjak (1987); and some, such as the material in subsection 6.3 on approximation of projections, we believe to be new.

Armed with the spectral theory, we proceed in Section 7 to the case where μ and K are estimated. We emphasize the parallels with the argument for known μ and K, postponing the proofs of the extra approximation arguments (mostly collected together as Lemma 28) to the following section.

The final two sections of the paper establish a bound on the Hellinger distance between members of an exponential family, the key to our change of measure argument, and a maximal inequality for Gaussian processes.

2. Notation. For each matrix A, the spectral norm is defined as $\|A\|_2 := \sup_{|u|\le 1} |Au|$ and the Frobenius norm by $\|A\|_{\mathrm{F}} := \left(\sum_{i,j} A_{i,j}^2\right)^{1/2}$. If A is symmetric, with eigenvalues $\lambda_1, \dots, \lambda_k$, then

$$\|A\|_2 = \max_i |\lambda_i| = \sup_{|u|\le 1} |u'Au| \le \|A\|_{\mathrm{F}}.$$

If A is also positive definite then the absolute values are superfluous for the first two equalities.

When we want to indicate that a bound involving constants c, C, $C_1, \dots$ holds uniformly over all models indexed by a set of parameters $\mathcal{F}$, we write $c(\mathcal{F})$, $C(\mathcal{F})$, $C_1(\mathcal{F}), \dots$. By the usual convention for eliminating subscripts, the values of the constants might change from one paragraph to the next: a constant $C_1(\mathcal{F})$ in one place needn't be the same as a constant $C_1(\mathcal{F})$ in another place.

For sequences of constants c_n that might depend on $\mathcal{F}$, we write $c_n = O_{\mathcal{F}}(1)$ and $o_{\mathcal{F}}(1)$ and so on to show that the asymptotic bounds hold uniformly over $\mathcal{F}$.

We write $h(P, Q)$ for the Hellinger distance between two probability measures P and Q. If both P and Q are dominated by some measure ν, with densities p and q, then $h^2(P, Q) = \nu\left(\sqrt{p} - \sqrt{q}\right)^2$. We use Hellinger distance to bound total variation distance,

$$\|P - Q\|_{\mathrm{TV}} := \sup_A |PA - QA| = \tfrac{1}{2}\nu|p - q| \le h(P, Q).$$

For product measures we use the bound

$$h^2(\otimes_{i\le n}P_i,\otimes_{i\le n}Q_i)\le \sum\nolimits_{i\le n} h^2(P_i,Q_i).$$

To avoid confusion with transposes, we use the dot notation or superscript notation to denote derivatives. For example, $\dddot{\psi}$ or $\psi^{(3)}$ both denote the third derivative of a function ψ,

3. The model. Let $\{Q_\lambda : \lambda\in\mathbb{R}\}$ be an exponential family of probability measures with densities $dQ_\lambda/dQ_0 = f_\lambda(y) = \exp\left(\lambda y-\psi(\lambda)\right)$. Remember that $e^{\psi(\lambda)} = Q_0 e^{\lambda y}$ and that the distribution Q_λ has mean $\psi^{(1)}(\lambda)$ and variance $\psi^{(2)}(\lambda)$.

We assume:

($\psi3$) There exists an increasing real function G on $\mathbb{R}^+$ such that

$$|\psi^{(3)}(\lambda+h)|\le\psi^{(2)}(\lambda)G(|h|)\qquad\text{for all } \lambda \text{ and } h$$

Without loss of generality we assume $G(0)\ge 1$.

($\psi2$) For each $\epsilon>0$ there exists a finite constant C_ϵ for which $\psi^{(2)}(\lambda)\le C_\epsilon\exp(\epsilon\lambda^2)$ for all $\lambda\in\mathbb{R}$. Equivalently, $\psi^{(2)}(\lambda)\le\exp\left(o(\lambda^2)\right)$ as $|\lambda|\to\infty$.

As shown in Section 10, these assumptions on the ψ function imply that

$$h^2(Q_\lambda,Q_{\lambda+\delta})\le\delta^2\psi^{(2)}(\lambda)\,(1+|\delta|)\,G(|\delta|)\qquad\text{for all } \lambda,\delta\in\mathbb{R}. \tag{4}$$

Remark. We may assume that $\psi^{(2)}(\lambda)>0$ for every real λ. Otherwise we would have $0=\psi^{(2)}(\lambda_0)=\text{var}_{\lambda_0}(y)=\nu f_{\lambda_0}(y)(y-\psi^{(1)}(\lambda_0))^2$ for some λ_0, which would make $y=\psi^{(1)}(\lambda_0)$ for ν almost all y and $Q_\lambda\equiv Q_{\lambda_0}$ for every λ.

We assume the observed data are iid pairs $(y_i,\mathbb{X}_i)$ for $i=1,\dots,n$, where:

(a) Each $\{\mathbb{X}_i(t):0\le t\le 1\}$ is distributed like $\{\mathbb{X}(t):0\le t\le 1\}$, a Gaussian process with mean $\mu(t)$ and covariance kernel $K(s,t)$.

(b) $y_i\mid\mathbb{X}_i\sim Q_{\lambda_i}$ with $\lambda_i=a+\langle\mathbb{X}_i,\mathbb{B}\rangle$ for an unknown $\{\mathbb{B}(t):0\le t\le 1\}$ in $\mathcal{L}^2(\mathfrak{m})$ and $a\in\mathbb{R}$.

DEFINITION 5. *For real constants $\alpha>1$ and $\beta>(\alpha+3)/2$ and $R>0$, define $\mathcal{F}=\mathcal{F}(R,\alpha,\beta)$ as the set of all $f=(a,\mathbb{B},\mu,K)$ that satisfy the following conditions.*

(K) The covariance kernel is square integrable with respect to $\mathfrak{m}\otimes\mathfrak{m}$ and has an eigenfunction expansion (as a compact operator on $\mathcal{L}^2(\mathfrak{m})$)

$$K(s,t)=\sum\nolimits_{k\in\mathbb{N}}\theta_k\phi_k(s)\phi_k(t)$$

where the eigenvalues θ_k are decreasing with $Rk^{-\alpha}\ge\theta_k\ge\theta_{k+1}+(\alpha/R)k^{-\alpha-1}$.

(a) $|a| \le R$
(μ) $\|\mu\| \le R$
($\mathbb{B}$) *$\mathbb{B}$ has an expansion $\mathbb{B}(t) = \sum_{k\in\mathbb{N}} b_k\phi_k(t)$ with $|b_k| \le Rk^{-\beta}$, for the eigenfunctions defined by the kernel K.*

Remarks. The awkward lower bound for θ_k in Assumption (K) implies, for all $k < j$,

$$\theta_k - \theta_j \ge R^{-1}\int_k^j \alpha x^{-\alpha-1}dx = R^{-1}\left(k^{-\alpha} - j^{-\alpha}\right). \tag{6}$$

If K and μ were known, we would only need the lower bound $\theta_k \ge R^{-1}k^{-\alpha}$ and not the lower bound for $\theta_k - \theta_{k+1}$. As explained by Hall and Horowitz (2007, page 76), the stronger assumption is needed when one estimates the individual eigenfunctions of K. Note that the subset $\mathcal{B}_K$ of $\mathcal{L}^2(\mathfrak{m})$ in which $\mathbb{B}$ lies depends on K. We regard the need for the stronger assumption on the eigenvalues and the irksome Assumption ($\mathbb{B}$) as artifacts of the method of proof, but we have not yet succeeded in removing either assumption.

More formally, we write $P_{\mu,K}$ for the distribution (a probability measure on $\mathcal{L}^2(\mathfrak{m})$) of each Gaussian process $\mathbb{X}_i$. The joint distribution of $\mathbb{X}_1, \dots, \mathbb{X}_n$ is then $\mathbb{P}_{n,\mu,K} = P^n_{\mu,K}$. We identify the y_i's with the coordinate maps on $\mathbb{R}^n$ equipped with the product measure $\mathbb{Q}_{n,a,\mathbb{B},\mathbb{X}_1,\dots,\mathbb{X}_n} := \otimes_{i\le n} Q_{\lambda_i}$, which can also be thought of as the conditional joint distribution of $(y_1, \dots, y_n)$ given $(\mathbb{X}_1, \dots, \mathbb{X}_n)$. Thus the $\mathbb{P}_{n,f}$ in equations (2) and (3) can be rewritten as an iterated expectation,

$$\mathbb{P}_{n,f} = \mathbb{P}_{n,\mu,K}\mathbb{Q}_{n,a,\mathbb{B},\mathbb{X}_1,\dots,\mathbb{X}_n},$$

the second expectation on the right-hand side averaging out over $y_1, \dots, y_n$ for given $\mathbb{X}_1, \dots, \mathbb{X}_n$, the first averaging out over $\mathbb{X}_1, \dots, \mathbb{X}_n$.

To simplify notation, we will often abbreviate $\mathbb{Q}_{n,a,\mathbb{B},\mathbb{X}_1,\dots,\mathbb{X}_n}$ to $\mathbb{Q}_{n,a,\mathbb{B}}$.

4. Maximum likelihood estimation. The theory in this section combine ideas from Portnoy (1988) and from Hjort and Pollard (1993). We write our results in a notation that makes the applications in Section 5 and 7 more straightforward. The notational cost is that the parameters are indexed by $\{0, 1, \dots, N\}$. To avoid an excess of parentheses we write N_+ for $N + 1$. In the applications N changes with the sample size n and $\mathbb{Q}$ is replaced by $\mathbb{Q}_{n,a,\mathbb{B},N}$ or $\widetilde{\mathbb{Q}}_{n,a,\mathbb{B},N}$.

Suppose $\xi_1, \dots, \xi_n$ are (nonrandom) vectors in $\mathbb{R}^{N_+}$. Suppose $\mathbb{Q} = \otimes_{i\le n} Q_{\lambda_i}$ with $\lambda_i = \xi_i'\gamma$ for a fixed $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_N)$ in $\mathbb{R}^{N_+}$. Under $\mathbb{Q}$, the coordinate maps $y_1, \dots, y_n$ are independent random variables with $y_i \sim Q_{\lambda_i}$.

The log-likelihood for fitting the model is

$$L_n(g) = \sum\nolimits_{i\le n} (\xi_i'g)y_i - \psi(\xi_i'g) \qquad \text{for } g \in \mathbb{R}^{N_+},$$

which is maximized (over $\mathbb{R}^{N_+}$) at the MLE $\widehat{g}$ $(= \widehat{g}_n)$.

> **Remark.** As a small amount of extra bookkeeping in the following argument would show, we do not need $\widehat{g}$ to exactly maximize L_n. It would suffice to have $L_n(\widehat{g})$ suitably close to $\sup_g L_n(g)$. In particular, we need not be concerned with questions regarding existence or uniqueness of the argmax.

Define

(i) $J_n = \sum_{i\le n} \xi_i\xi_i'\psi^{(2)}(\lambda_i)$, an $N_+ \times N_+$ matrix
(ii) $w_i := J_n^{-1/2}\xi_i$, an element of $\mathbb{R}^{N_+}$
(iii) $W_n = \sum_{i\le n} w_i \left(y_i - \psi^{(1)}(\lambda_i)\right)$, an element of $\mathbb{R}^{N_+}$

Notice that $\mathbb{Q}W_n = 0$ and $\text{var}_{\mathbb{Q}}(W_n) = \sum_{i\le n} w_i w_i'\psi^{(2)}(\lambda_i) = I_{N_+}$ and

$$\mathbb{Q}|W_n|^2 = \text{trace}\left(\text{var}_{\mathbb{Q}}(W_n)\right) = N_+.$$

LEMMA 7. *Suppose* $0 < \epsilon_1 \le 1/2$ *and* $0 < \epsilon_2 < 1$ *and*

$$\max_{i\le n} |w_i| \le \frac{\epsilon_1\epsilon_2}{2G(1)N_+} \qquad \text{with } G \text{ as in Assumption } (\psi 3).$$

Then $\widehat{g} = \gamma + J_n^{-1/2}(W_n + r_n)$ *with* $|r_n| \le \epsilon_1$ *on the set* $\{|W_n| \le \sqrt{N_+/\epsilon_2}\}$, *which has* $\mathbb{Q}$*-probability greater than* $1 - \epsilon_2$.

PROOF. The equality $\mathbb{Q}|W_n|^2 = N_+$ and Tchebychev give $\mathbb{Q}\{|W_n| > \sqrt{N_+/\epsilon_2}\} \le \epsilon_2$.

Reparametrize by defining $t = J_n^{1/2}(g - \gamma)$. The concave function

$$\mathcal{L}_n(t) := L_n(\gamma + J_n^{-1/2}t) - L_n(\gamma) = \sum\nolimits_{i\le n} y_i w_i't + \psi(\lambda_i) - \psi(\lambda_i + w_i't)$$

is maximized at $\widehat{t}_n = J_n^{1/2}(\widehat{g} - \gamma)$. It has derivative

$$\dot{\mathcal{L}}_n(t) = \sum\nolimits_{i\le n} w_i \left(y_i - \psi^{(1)}(\lambda_i + w_i't)\right).$$

For a fixed unit vector $u \in \mathbb{R}^{N_+}$ and a fixed $t \in \mathbb{R}^{N_+}$, consider the real-valued function of the real variable s,

$$H(s) := u'\dot{\mathcal{L}}_n(st) = \sum\nolimits_{i\le n} u'w_i \left(y_i - \psi^{(1)}(\lambda_i + sw_i't)\right),$$

which has derivatives

$$\dot{H}(s) = -\sum\nolimits_{i\le n} (u'w_i)(w_i't)\psi^{(2)}(\lambda_i + sw_i't)$$
$$\ddot{H}(s) = -\sum\nolimits_{i\le n} (u'w_i)(w_i't)^2\psi^{(3)}(\lambda_i + sw_i't).$$

Notice that $H(0) = u'W_n$ and $\dot{H}(0) = -u' \sum_{i \le n} w_i w_i' \psi^{(2)}(\lambda_i) t = -u't$.

Write M_n for $\max_{i \le n} |w_i|$. By virtue of Assumption ($\psi 3$),

$$\begin{aligned} |\ddot{H}(s)| &\le \sum_{i \le n} |u'w_i| (w_i' t)^2 \psi^{(2)}(\lambda_i) G\left(|s w_i' t|\right) \\ &\le M_n G\left(M_n |st|\right) t' \sum_{i \le n} w_i w_i' \psi^{(2)}(\lambda_i) t \\ &= M_n G\left(M_n |st|\right) |t|^2. \end{aligned}$$

By Taylor expansion, for some $0 < s^* < 1$,

$$|H(1) - H(0) - \dot{H}(0)| \le \tfrac{1}{2} |\ddot{H}(s^*)| \le \tfrac{1}{2} M_n G\left(M_n |t|\right) |t|^2.$$

That is,

$$\left| u' \left(\dot{\mathcal{L}}_n(t) - W_n + t \right) \right| \le \tfrac{1}{2} M_n G\left(M_n |t|\right) |t|^2. \tag{8}$$

Approximation (8) will control the behavior of $\widetilde{\mathcal{L}}(s) := \mathcal{L}_n(W_n + su)$, a concave function of the real argument s, for each unit vector u. By concavity, the derivative

$$\dot{\widetilde{\mathcal{L}}}(s) = u' \dot{\mathcal{L}}_n(W_n + su) = -s + R(s)$$

is a decreasing function of s with

$$|R(s)| \le \tfrac{1}{2} M_n G\left(M_n |W_n + su|\right) |W_n + su|^2$$

On the set $\{|W_n| \le \sqrt{N_+/\epsilon_2}\}$ we have

$$|W_n \pm \epsilon_1 u| \le \sqrt{N_+/\epsilon_2} + \epsilon_1.$$

Thus

$$M_n |W_n \pm \epsilon_1 u| \le \frac{\epsilon_1 \epsilon_2}{2G(1) N_+} \left(\sqrt{N_+/\epsilon_2} + \epsilon_1 \right) < 1,$$

implying

$$\begin{aligned} |R(\pm\epsilon_1)| &\le \tfrac{1}{2} M_n G(1) |W_n \pm \epsilon_1 u|^2 \\ &\le \frac{\epsilon_1 \epsilon_2}{G(1) N_+} \left(N_+/\epsilon_2 + \epsilon_1^2 \right) \\ &\le \epsilon_1 \left(1 + \epsilon_1^2 \epsilon_2 / N_+ \right) < \tfrac{5}{8} \epsilon_1. \end{aligned}$$

Deduce that

$$\begin{aligned} \dot{\widetilde{\mathcal{L}}}(\epsilon_1) &= -\epsilon_1 + R(\epsilon_1) \le -\tfrac{3}{8} \epsilon_1 \\ \dot{\widetilde{\mathcal{L}}}(-\epsilon_1) &= \epsilon_1 + R(-\epsilon_1) \ge \tfrac{3}{8} \epsilon_1 \end{aligned}$$

The concave function $s \mapsto \mathcal{L}_n(W_n + su)$ must achieve its maximum for some s in the interval $[-\epsilon_1, \epsilon_1]$, for each unit vector u. It follows that $|\widehat{t}_n - W_n| \le \epsilon_1$. □

COROLLARY 9. *Suppose $\xi_i = D\eta_i$ for some nonsingular matrix D, so that*

$$J_n = nDA_nD \qquad \text{where} A_n := \frac{1}{n}\sum\nolimits_{i\le n}\eta_i\eta_i'\psi^{(2)}(\lambda_i).$$

If B_n is another nonsingular matrix for which

$$\|A_n - B_n\|_2 \le (2\|B_n^{-1}\|_2)^{-1} \tag{10}$$

and if

$$\max\nolimits_{i\le n}|\eta_i| \le \frac{\epsilon\sqrt{n}/N_+}{G(1)\sqrt{32\|B_n^{-1}\|_2}} \qquad \text{for some } 0<\epsilon<1 \tag{11}$$

then for each set of vectors $\kappa_0,\dots,\kappa_N$ in $\mathbb{R}^{N_+}$ there is a set $\mathcal{Y}_{\kappa,\epsilon}$ with $\mathbb{Q}\mathcal{Y}^c_{\kappa,\epsilon} < 2\epsilon$ on which

$$\sum\nolimits_{0\le j\le N}|\kappa_j'(\widehat{g}-\gamma)|^2 \le \frac{6\|B_n^{-1}\|_2}{n\epsilon}\sum\nolimits_{0\le j\le N}|D^{-1}\kappa_j|^2.$$

Remark. For our applications of the Corollary in Sections 5 and 7, we need $D = \operatorname{diag}(D_0, D_1, \dots, D_N)$ and $\kappa_j = e_j$, the unit vector with a 1 in its jth position, for $j \le m$ and $\kappa_j = 0$ for $j > m$. In our companion paper we will need the more general κ_j's.

PROOF. First we establish a bound on the spectral distance between A_n^{-1} and B_n^{-1}. Define $H = B_n^{-1}A_n - I$. Then $\|H\|_2 \le \|B_n^{-1}\|_2\|A_n - B_n\|_2 \le 1/2$, which justifies the expansion

$$\|A_n^{-1} - B_n^{-1}\|_2 = \|\left((I+H)^{-1} - I\right)B_n^{-1}\|_2 \le \sum\nolimits_{j\ge1}\|H\|_2^k\|B_n^{-1}\|_2 \le \|B_n^{-1}\|_2.$$

As a consequence, $\|A_n^{-1}\|_2 \le 2\|B_n^{-1}\|_2$.

Choose $\epsilon_1 = 1/2$ and $\epsilon_2 = \epsilon$ in Lemma 7. The bound on $\max_{i\le n}|\eta_i|$ gives the bound on $\max_{i\le n}|w_i|$ needed by the Lemma:

$$n|w_i|^2 = \eta_i'D(J_n/n)^{-1}D\eta_i = \eta_i'A_n^{-1}\eta_i \le \|A_n^{-1}\|_2|\eta_i|^2.$$

Define $K_j := J_n^{-1/2}\kappa_j$, so that $|\kappa_j'(\widehat{g}-\gamma)|^2 \le 2(K_j'W_n)^2 + 2(K_j'r_n)^2$. By Cauchy-Schwarz,

$$\sum\nolimits_j (K_j'r_n)^2 \le \sum\nolimits_j |K_j|^2|r_n|^2 = U_\kappa|r_n|^2$$

where

$$\begin{aligned} U_\kappa := \sum\nolimits_j \kappa_j'J_n^{-1}\kappa_j &= \sum\nolimits_j n^{-1}(D^{-1}\kappa_j)'A_n^{-1}D^{-1}\kappa_j \\ &\le 2n^{-1}\|B_n^{-1}\|_2\sum\nolimits_j |D^{-1}\kappa_j|^2. \end{aligned}$$

For the contribution $V_\kappa := \sum_j |K_j' W_n|^2$ the Cauchy-Schwarz bound is too crude. Instead, notice that $\mathbb{Q}V_\kappa = U_\kappa$, which ensures that the complement of the set

$$\mathcal{Y}_{\kappa,\epsilon} := \{|W_n| \le \sqrt{N_+/\epsilon}\} \cap \{V_\kappa \le U_\kappa/\epsilon\}$$

has $\mathbb{Q}$ probability less that 2ϵ. On the set $\mathcal{Y}_{\kappa,\epsilon}$,

$$\sum\nolimits_{0\le j\le N} |\kappa_j'(\widehat{g}-\gamma)|^2 \le 2V_\kappa + 2U_\kappa |r_n|^2 \le 3U_\kappa/\epsilon.$$

The asserted bound follows. □

5. Known Gaussian distribution. Initially we suppose that μ and K are known. We can then calculate all the eigenvalues θ_k, the eigenfunctions ϕ_k for K, and the coefficients $z_{i,k} := \langle \mathbb{Z}_i, \phi_k\rangle$ for the expansion

$$\mathbb{X}_i - \mu = \mathbb{Z}_i = \sum\nolimits_{k\in\mathbb{N}} z_{i,k}\phi_k.$$

The random variables $z_{i,k}$ are independent with $z_{i,k} \sim N(0,\theta_k)$. The random variables $\eta_{i,k} := z_{i,k}/\sqrt{\theta_k}$ are independent standard normals.

Under $\mathbb{Q}_n = \mathbb{Q}_{n,a,\mathbb{B}}$, the y_i's are independent, with $y_i \sim Q_{\lambda_i}$ and

$$\lambda_i = a + \langle \mathbb{X}_i, \mathbb{B}\rangle = b_0 + \sum\nolimits_{k\in\mathbb{N}} z_{i,k}b_k \qquad \text{where } b_0 = a + \langle \mu, \mathbb{B}\rangle.$$

Our task is to estimate the b_k's with sufficient accuracy to be able to estimate $\mathbb{B}(t) = \sum_{k\in\mathbb{N}} b_k\phi_k(t)$ within an error of order $\rho_n = n^{(1-2\beta)/(\alpha+2\beta)}$. In fact it will suffice to estimate the component $H_m\mathbb{B}$ of $\mathbb{B}$ in the subspace spanned by $\{\phi_1,\dots,\phi_m\}$ with $m \sim n^{1/(\alpha+2\beta)}$ because

$$\|H_m^\perp \mathbb{B}\|^2 = \sum\nolimits_{k>m} b_k^2 = O_{\mathcal{F}}(m^{1-2\beta}) = O_{\mathcal{F}}(\rho_n). \tag{12}$$

We might try to estimate the coefficients $(b_0,\dots,b_m)$ by choosing $\widehat{g} = (\widehat{g}_0,\dots,\widehat{g}_m)$ to maximize a conditional log likelihood over all g in $\mathbb{R}^{m+1}$,

$$\sum\nolimits_{i\le n} y_i\lambda_{i,m} - \psi(\lambda_{i,m}) \qquad \text{with } \lambda_{i,m} = g_0 + \sum\nolimits_{1\le k\le m} z_{i,k}g_k.$$

To this end we might try to appeal to Corollary 9 in Section 4, with κ_j equal to the unit vector with a 1 in its jth position for $j \le m$ and $\kappa_j = 0$ otherwise. That would give a bound for $\sum_{j\le m}(\widehat{g}_j - \gamma_j)^2$. Unfortunately, we cannot directly invoke the Corollary with $N = m$ to estimate $\gamma = (b_0, b_1, \dots, b_N)$ when

$$\begin{aligned} &\mathbb{Q} = \mathbb{Q}_{n,a,\mathbb{B}} &&\text{and} && D = \operatorname{diag}(1,\theta_1,\dots,\theta_N)^{1/2} \\ &\xi_i' = (1, z_{i,1},\dots,z_{i,N}) &&\text{and} && \eta_i' = (1,\eta_{i,1},\dots,\eta_{i,N}) \end{aligned} \tag{13}$$

because $\lambda_i \ne \xi_i'\gamma$.

Remark. We could modify Corollary 9 to allow $\ell_i = \xi_i'\gamma + \text{bias}_i$, for a suitably small bias term, but at the cost of extra regularity conditions and a more delicate argument. The same difficulty arises whenever one investigates the asymptotics of maximum likelihood with the true distribution outside the model family.

Instead, we use a two-stage estimation procedure that eliminates the bias term by a change of measure. Condition on the $\mathbb{X}_i$'s. Consider an N much larger than m for which

$$N \sim n^{\zeta} \qquad \text{with } (2+2\alpha)^{-1} > \zeta > (\alpha + 2\beta - 1)^{-1},$$

Such a ζ exists because the assumptions $\alpha > 1$ and $\beta > (\alpha+3)/2$ imply $\alpha + 2\beta - 1 > 2 + 2\alpha$. Define ξ_i, D, and η_i as in equation (13). For $\mathbb{Q}$ use the probability measure

$$\mathbb{Q}_{n,a,\mathbb{B},N} := \otimes_{i\le n} Q_{\lambda_{i,N}} \qquad \text{with } \lambda_{i,N} := \xi_i'\gamma \text{ and } \gamma' = (b_0, b_1, \dots, b_N).$$

Choose $B_n := \mathbb{P}_{n,\mu,K} A_n$. Define $\mathcal{X}_n = \mathcal{X}_{\mathbb{Z},n} \cap \mathcal{X}_{\eta,n} \cap \mathcal{X}_{A,n}$, where

$$\mathcal{X}_{\mathbb{Z},n} := \{\max\nolimits_{i\le n} \|\mathbb{Z}_i\|^2 \le C_0 \log n\} \tag{14}$$

$$\mathcal{X}_{\eta,n} := \{\max\nolimits_{i\le n} |\eta_i|^2 \le C_0 N \log n\} \tag{15}$$

$$\mathcal{X}_{A,n} := \{\|A_n - B_n\|_2 \le (2\|B_n^{-1}\|_2)^{-1}\} \tag{16}$$

If we choose a large enough constant $C_0 = C_0(\mathcal{F})$, Lemma 41 and its Corollary in Section 11 ensure that $\mathbb{P}_{n,\mu,K}\mathcal{X}_{\mathbb{Z},n}^c \le 2/n$ and $\mathbb{P}_{n,\mu,K}\mathcal{X}_{\eta,n}^c \le 2/n$; and in subsection 5.1 we show that

$$\|B_n^{-1}\|_2 = O_{\mathcal{F}}(1) \qquad \text{and} \qquad \mathbb{P}_{n,\mu,K}\|A_n - B_n\|_2^2 = o_{\mathcal{F}}(1).$$

Thus $\mathbb{P}_{n,\mu,K}\mathcal{X}_n^c = o_{\mathcal{F}}(1)$. Moreover, on the set $\mathcal{X}_n$, inequality (10) holds by construction and inequality (11) holds for large enough n because

$$\max\nolimits_{i\le n} |\eta_i|^2 \le O_{\mathcal{F}}(N \log n) = o_{\mathcal{F}}(\sqrt{n}/N).$$

Estimate γ by the $\widehat{g} = (\widehat{g}_0, \dots, \widehat{g}_N)$ defined in Section 4. Then discard most of the estimates by defining $\widehat{\mathbb{B}}_n := \sum_{1\le k\le m} \widehat{g}_k \phi_k$. For each realization of the $\mathbb{X}_i$'s in $\mathcal{X}_n$, the Lemma gives a set $\mathcal{Y}_{m,\epsilon}$ with $\mathbb{Q}_{n,a,\mathbb{B},N}\mathcal{Y}_{m,\epsilon}^c < 2\epsilon$ on which

$$\sum\nolimits_{1\le k\le m} |\widehat{g}_k - \gamma_k|^2 = O_{\mathcal{F}}\left(\sum\nolimits_{1\le k\le m} \theta_k^{-1}\right) = O_{\mathcal{F}}(m^{1+\alpha}/n) = O_{\mathcal{F}}(\rho_n),$$

which implies

$$\|\widehat{\mathbb{B}}_n - \mathbb{B}\|^2 = \sum\nolimits_{1\le k\le m} |\widehat{g}_k - \gamma_k|^2 + \sum\nolimits_{k>m} b_k^2 = O_{\mathcal{F}}(\rho_n).$$

In replacing $\mathbb{Q}_{n,a,\mathbb{B}}$ by $\mathbb{Q}_{n,a,\mathbb{B},N}$ we eliminate the bias problem but now we have to relate the probability bounds for $\mathbb{Q}_{n,a,\mathbb{B},N}$ to bounds involving $\mathbb{Q}_{n,a,\mathbb{B}}$. As we show in subsection 5.2, there exists a sequence of nonnegative constants c_n of order $o_{\mathcal{F}}(\log n)$, such that

$$\|\mathbb{Q}_{n,a,\mathbb{B}} - \mathbb{Q}_{n,a,\mathbb{B},N}\|_{\text{TV}}^2 \le e^{2c_n} \sum\nolimits_{i\le n} |\lambda_i - \lambda_{i,N}|^2 \qquad \text{on } \mathcal{X}_n. \tag{17}$$

From this inequality it follows, for a large enough constant C_ϵ, that

$$\begin{aligned}
&\mathbb{P}_{n,\mu,K}\mathbb{Q}_{n,a,\mathbb{B}}\{\|\widehat{\mathbb{B}}_n - \mathbb{B}\|^2 > C_\epsilon\rho_n\} \\
&\qquad \le \mathbb{P}_{n,\mu,K}\mathcal{X}_n^c + \mathbb{P}_{n,\mu,K}\mathcal{X}_n\left(\|\mathbb{Q}_{n,a,\mathbb{B}} - \mathbb{Q}_{n,a,\mathbb{B},N}\|_{\text{TV}} + \mathbb{Q}_{n,a,\mathbb{B},N}\mathcal{Y}_{m,\epsilon}^c\right) \\
&\qquad \le o_{\mathcal{F}}(1) + 2\epsilon + e^{c_n}\left(\sum\nolimits_{i\le n} \mathbb{P}_{n,\mu,K}|\lambda_i - \lambda_{i,N}|^2\right)^{1/2}.
\end{aligned}$$

By construction,

$$\lambda_i - \lambda_{i,N} = \sum\nolimits_{k>N} z_{i,k}b_k$$

with the $z_{i,k}$'s independent and $z_{i,k} \sim N(0,\theta_k)$. Thus

$$\sum\nolimits_{i\le n} \mathbb{P}_{n,\mu,K}|\lambda_i - \lambda_{i,N}|^2 \le n\sum\nolimits_{k>N} \theta_k b_k^2 = O_{\mathcal{F}}(nN^{1-\alpha-2\beta}) = o_{\mathcal{F}}(e^{-2c_n})$$

because $\zeta > (\alpha + 2\beta - 1)^{-1}$. That is, we have an estimator that achieves the $O_{\mathcal{F}}(\rho_n)$ minimax rate.

5.1. *Approximation of A_n.* Throughout this subsection abbreviate $\mathbb{P}_{n,\mu,K}$ to $\mathbb{P}$. Remember that

$$A_n = n^{-1}\sum\nolimits_{i\le n} \eta_i\eta_i'\psi^{(2)}(\lambda_{i,N}) \qquad \text{with } \lambda_{i,N} = \gamma' D\eta_i,$$

where

$$\begin{aligned}
\gamma' &= (\bar{a}, b_1, \dots, b_N) \\
D &= \text{diag}(1, \sqrt{\theta_1}, \dots, \sqrt{\theta_N}) \\
\eta_i &= (1, \eta_{i,1}, \dots, \eta_{i,N})'
\end{aligned}$$

With $B_n = \mathbb{P}A_n$, we need to show $\|B_n^{-1}\|_2 = O_{\mathcal{F}}(1)$ and $\mathbb{P}\|A_n - B_n\|_2^2 = o_{\mathcal{F}}(1)$.

The matrix A_n is an average of n independent random matrices each of which is distributed like $\mathcal{N}\mathcal{N}'\psi^{(2)}(\gamma' D\mathcal{N})$, where $\mathcal{N}' = (\mathcal{N}_0, \mathcal{N}_1, \dots, \mathcal{N}_N)$ with $\mathcal{N}_0 \equiv 1$ and the other $\mathcal{N}_j$'s are independent $N(0,1)$'s. Moreover, by rotational invariance of the spherical normal, we may assume with no loss of generality that $\gamma' D\mathcal{N} = \bar{a} + \kappa\mathcal{N}_1$, where

$$\kappa^2 = \sum\nolimits_{k=1}^N \theta_k b_k^2 = O_{\mathcal{F}}(1).$$

Thus

$$B_n = \mathbb{P}\mathcal{N}\mathcal{N}'\psi^{(2)}(\bar{a} + \kappa\mathcal{N}_1) = \text{diag}(F, r_0 I_{N-1})$$

where

$$r_j := \mathbb{P}\mathcal{N}_1^j\psi^{(2)}(\bar{a} + \kappa\mathcal{N}_1) \qquad \text{and} \qquad F = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_2 \end{bmatrix}.$$

The block diagonal form of B_n simplifies calculation of spectral norms.

$$\begin{aligned}\|B_n^{-1}\|_2 &= \|\text{diag}(F^{-1}, r_0^{-1} I_{N-1})\|_2 \\ &\leq \max\left(\|F^{-1}\|_2, \|r_0^{-1} I_{N-1}\|_2\right) \leq \max\left(\frac{r_0 + r_2}{r_0 r_2 - r_1^2}, r_0^{-1}\right).\end{aligned}$$

Assumption ($\psi 2$) ensures that both r_0 and r_2 are $O_{\mathcal{F}}(1)$.

Continuity and strict positivity of $\psi^{(2)}$, together with $\max(|\bar{a}|, \kappa) = O_{\mathcal{F}}(1)$, ensure that $c_0 := \inf_{\bar{a},\kappa} \inf_{|x|\leq 1} \psi^{(2)}(\bar{a} + \kappa x) > 0$. Thus

$$\sqrt{2\pi} r_0 \geq c_0 \int_{-1}^{+1} e^{-x^2/2} dx > 0$$

Similarly

$$\begin{aligned}\sqrt{2\pi}(r_0 r_2 - r_1^2) &= \sqrt{2\pi} r_0 \mathbb{P}\psi^{(2)}(\bar{a} + \kappa\mathcal{N}_1)(\mathcal{N}_1 - r_1/r_0)^2 \\ &\geq c_0 r_0 \int_{-1}^{+1} (x - r_1/r_0)^2 e^{-x^2/2} dx \geq c_0 r_0 \int_{-1}^{+1} x^2 e^{-x^2/2} dx.\end{aligned}$$

It follows that $\|B_n^{-1}\|_2 = O_{\mathcal{F}}(1)$.

The random matrix $A_n - B_n$ is an average of n independent random matrices each distributed like $\mathcal{N}\mathcal{N}'\psi^{(2)}(\bar{a} + \kappa\mathcal{N}_1)$ minus its expected value. Thus

$$\mathbb{P}\|A_n - B_n\|_2^2 \leq \mathbb{P}\|A_n - B_n\|_{\text{F}}^2 = n^{-1} \sum\nolimits_{0\leq j,k\leq N} \text{var}\left(\mathcal{N}_j \mathcal{N}_k \psi^{(2)}(\gamma' D\mathcal{N})\right).$$

Assumption ($\psi 2$) ensures that each summand is $O_{\mathcal{F}}(1)$, which leaves us with a $O_{\mathcal{F}}(N^2/n) = o_{\mathcal{F}}(1)$ upper bound.

5.2. *Total variation argument.* To establish inequality (17) we use the bound

$$\|\mathbb{Q}_{n,a,\mathbb{B}} - \widetilde{\mathbb{Q}}_{n,a,\mathbb{B}}\|_{\text{TV}}^2 \leq h^2(\mathbb{Q}_{n,a,\mathbb{B}}, \widetilde{\mathbb{Q}}_{n,a,\mathbb{B}}) \leq \sum\nolimits_{i\leq n} h^2(Q_{\lambda_i}, Q_{\lambda_{i,N}})$$

By Lemma 4

$$h^2(Q_{\lambda_i}, Q_{\lambda_{i,N}}) \leq \delta_i^2 \psi^{(2)}(\lambda_i) \left(1 + |\delta_i|\right) g(|\delta_i|)$$

where

$$
\begin{aligned}
|\delta_i| = |\lambda_i - \lambda_{i,N}| &= |\langle \mathbb{Z}_i, \mathbb{B}\rangle - \langle H_N \mathbb{Z}_i, \mathbb{B}\rangle| \\
&= |\langle \mathbb{Z}_i, H_N^{\perp}\mathbb{B}\rangle| \\
&\le \|\mathbb{Z}_i\| \|H_N^{\perp}\mathbb{B}\| \\
&\le O_{\mathcal{F}}\left(\sqrt{N^{1-2\beta}\log n}\right) \\
&= o_{\mathcal{F}}(1)
\end{aligned}
$$

Thus all the $(1+|\delta_i|)\, g(|\delta_i|)$ factors can be bounded by a single $O_{\mathcal{F}}(1)$ term.

For $(a, \mathbb{B}, \mu, K) \in \mathcal{F}(R, \alpha, \beta)$ and with the $\|\mathbb{Z}_i\|$'s controlled by $\mathcal{X}_n$,

$$
|\lambda_i| \le |a| + (\|\mu\| + \|\mathbb{Z}_i\|)\|\mathbb{B}\| \le C_2\sqrt{\log n}
$$

for some constant $C_2 = C_2(\mathcal{F})$. Assumption ($\psi 2$) then ensures that all the $\psi^{(2)}(\lambda_i)$ are bounded by a single $\exp\left(o_{\mathcal{F}}(\log n)\right)$ term.

6. Approximation of compact operators. Suppose T is a positive, (self-adjoint) compact operator on a Hilbert space $\mathcal{H}$ with eigenvectors $\{e_k\}$ and eigenvalues $\{\theta_k\}$. That is, $Te_i = \theta_i e_i$ with $\theta_1 \ge \theta_2 \ge \cdots \ge 0$. For each x in $\mathcal{H}$,

$$
T = \sum\nolimits_{k\in\mathbb{N}} \theta_k e_k \otimes e_k,
$$

a series that converges in operator norm.

Let $\widetilde{T}$ be another positive, (self-adjoint) compact operator on $\mathcal{H}$ with corresponding representation

$$
\widetilde{T} = \sum\nolimits_{k\in\mathbb{N}} \widetilde{\theta}_k \widetilde{e}_k \otimes \widetilde{e}_k.
$$

Define $\Delta := \widetilde{T} - T$ and $\delta = \|\Delta\|$. The operator $\widetilde{T}$ also has a representation

$$
\widetilde{T} = \sum\nolimits_{j,k\in\mathbb{N}} \widetilde{T}_{j,k} e_j \otimes e_k. \tag{18}
$$

Note that $\widetilde{T}_{j,k} = \widetilde{T}_{j,k}$ because $\widetilde{T}$ is self-adjoint. This representation gives

$$
\Delta = \sum\nolimits_{j,k\in\mathbb{N}} \left(\widetilde{T}_{j,k} - \theta_j\{j=k\}\right) e_j \otimes e_k
$$

and

$$
\|\Delta\|^2 = \sup\nolimits_{\|x\|=1} \langle x, \Delta x\rangle^2 \le \sum\nolimits_{j,k\in\mathbb{N}} \left(\widetilde{T}_{j,k} - \theta_j\{j=k\}\right)^2.
$$

The last inequality will lend itself to the calculation of the expected value of $\|\Delta\|^2$ when $\widetilde{T}$ is random, leading to probabilistic bounds for δ.

In this section we collect some general consequences of δ being small. In the next section we draw probabilistic conclusions when $\widetilde{T}$ is random, for the special case where $T = K$ and $\widetilde{T} = \widetilde{K}$, the usual estimate of the covariance kernel, both acting on $\mathcal{H} = \mathcal{L}^2(\mathfrak{m})$. The eigenvectors will become eigenfunctions $\phi_1, \phi_2, \dots$ and $\widetilde{\phi}_1, \widetilde{\phi}_2, \dots$. We feel this approach makes it easier to follow the overall argument.

Both $\{e_j : j \in \mathbb{N}\}$ and $\{\widetilde{e}_k : k \in \mathbb{N}\}$ are orthonormal bases for $\mathcal{H}$. Define $\sigma_{j,k} := \langle e_j, \widetilde{e}_k \rangle$. Then

$$e_j = \sum\nolimits_{k\in\mathbb{N}} \sigma_{j,k}\widetilde{e}_k \qquad \text{and} \qquad \widetilde{e}_k = \sum\nolimits_{j\in\mathbb{N}} \sigma_{j,k} e_j$$

and

$$\{j = j'\} = \langle e_j, e_{j'} \rangle = \sum\nolimits_{k\in\mathbb{N}} \sigma_{j,k}\sigma_{j',k}.$$

6.1. *Approximation of eigenvalues.* The eigenvalues have a variational characterization (Bosq, 2000, Section 4.2):

$$\theta_j = \inf_{\dim(L)<j} \sup\{\langle x, Tx \rangle : x \perp L \text{ and } \|x\| = 1\}. \tag{19}$$

The first infimum runs over all subspaces L with dimension at most $j-1$. (When j equals 1 the only such subspace is $\emptyset$.) Both the infimum and the supremum are achieved: by $L_{j-1} = \text{span}\{e_i : 1 \le i < j\}$ and $x = e_j$. Similar assertions hold for $\widetilde{T}$ and its eigenvalues.

By the analog of (19) for $\widetilde{T}$,

$$\begin{aligned} \widetilde{\theta}_j &\ge \sup\{\langle x, \widetilde{T}x \rangle : x \perp L_{j-1} \text{ and } \|x\| = 1\} \\ &\ge \sup\{\langle x, Tx \rangle - \delta : x \perp L_{j-1} \text{ and } \|x\| = 1\} = \theta_j - \delta. \end{aligned}$$

Argue similarly with the roles of T and $\widetilde{T}$ reversed to conclude that

$$|\theta_j - \widetilde{\theta}_j| \le \delta \qquad \text{for all } j \in \mathbb{N}. \tag{20}$$

6.2. *Approximation of eigenvectors.* We cannot hope to find a useful bound on $\|\widetilde{e}_k - e_k\|$, because there is no way to decide which of $\pm\widetilde{e}_k$ should be approximating e_k. However, we can bound $\|f_k\|$, where

$$f_k = \sigma_k\widetilde{e}_k - e_k \qquad \text{with } \sigma_k := \text{sign}\,(\sigma_{k,k}) := \begin{cases} +1 & \text{if } \sigma_{k,k} \ge 0 \\ -1 & \text{otherwise} \end{cases},$$

which will be enough for our purposes.

We also need to assume that the eigenvalue θ_k is well separated from the other θ_j's, to avoid the problem that the eigenspace of $\widetilde{T}$ for the eigenvalue $\widetilde{\theta}_k$ might have dimension greater than one. More precisely, we consider a k for which

$$\epsilon_k := \min\{|\theta_j - \theta_k| : j \neq k\} > 5\delta,$$

which implies

$$|\widetilde{\theta}_k - \theta_j| \geq |\theta_k - \theta_j| - \delta \geq \tfrac{4}{5}|\theta_k - \theta_j| \geq \tfrac{4}{5}\epsilon_k.$$

The starting point for our approximations is the equality

$$\langle \Delta\widetilde{e}_k, e_j\rangle = \langle \widetilde{T}\widetilde{e}_k, e_j\rangle - \langle \widetilde{e}_k, Te_j\rangle = (\widetilde{\theta}_k - \theta_j)\sigma_{j,k}. \tag{21}$$

For $j \neq k$ we then have

$$\frac{16}{25}(\theta_k - \theta_j)^2\sigma_{j,k}^2 \leq \langle \sigma_k\Delta\widetilde{e}_k, e_j\rangle^2 \leq 2\langle \Delta f_k, e_j\rangle^2 + 2\langle \Delta e_k, e_j\rangle^2,$$

which implies

$$\sigma_{j,k}^2 \leq \frac{25}{8}\langle \Delta f_k, e_j\rangle^2/\epsilon_k^2 + 2\widetilde{T}_{j,k}^2/(\theta_k - \theta_j)^2 \qquad \text{because } \langle Te_k, e_j\rangle = 0 \text{ for } j \neq k.$$

To simplify notation, write $\sum_j^*$ for $\sum_{j\in\mathbb{N}}\{j \neq k\}$.

The introduction of the σ_k also ensures that

$$\begin{aligned}
\|f_k\|^2 &= \|e_k\|^2 + \|\widetilde{e}_k\|^2 - 2\sigma_k\langle e_k, \widetilde{e}_k\rangle = 2 - 2|\sigma_{k,k}| \\
&\leq 2 - 2\sigma_{k,k}^2 \qquad \text{because } |\sigma_{k,k}| \leq 1 \\
&= 2\sum\nolimits_j^* \sigma_{j,k}^2 \\
&\leq \sum\nolimits_j^* \frac{25}{4}\langle \Delta f_k, e_j\rangle^2/\epsilon_k^2 + \frac{25}{4}\sum\nolimits_j^* \widetilde{T}_{j,k}^2/(\theta_k - \theta_j)^2.
\end{aligned}$$

The first sum on the right-hand side is less than

$$\frac{25}{4}\|\Delta f_k\|^2/\epsilon_k^2 \leq \|\Delta\|^2\|f_k\|^2/(4\delta^2) \leq \|f_k\|^2/4.$$

The second sum can be written as $25\|\Lambda_k\|^2/4$ for

$$\Lambda_k := \sum_{j\in\mathbb{N}} \Lambda_{k,j}e_j \qquad \text{with } \Lambda_{k,j} := \begin{cases} \widetilde{T}_{j,k}/(\theta_k - \theta_j) & \text{if } j \neq k \\ 0 & \text{if } j = k \end{cases}.$$

Our bound for $\|f_k\|^2$ (with an untidy $25/3$ increased to 9) then takes the convenient form

$$\|f_k\|^2 \le 9\|\Lambda_k\|^2 \qquad \text{if } \epsilon_k > 5\|\Delta\|. \tag{22}$$

For our applications, $\mathbb{P}\|\Lambda_k\|^2$ will be of order $O(k^2/n)$.

When δ is much smaller than ϵ_k we can get an even better approximation for f_k itself. Start once more from equality (21), still assuming that $\epsilon_k > 5\delta$. For $j \ne k$,

$$\begin{aligned}
\sigma_k\sigma_{j,k} &= \sigma_k\langle\Delta\widetilde{e}_k, e_j\rangle/(\widetilde{\theta}_k - \theta_j) \\
&= \langle\Delta(e_k + f_k), e_j\rangle/(\theta_k + \gamma_k - \theta_j) \qquad \text{where } \gamma_k = \widetilde{\theta}_k - \theta_k \\
&= \Lambda_{k,j}\left(1 - \frac{\gamma_k}{\theta_j - \theta_k}\right)^{-1} + \frac{\langle\Delta f_k, e_j\rangle}{\widetilde{\theta}_k - \theta_j} \qquad \text{because } \langle Te_k, e_j\rangle = 0 \\
&= \Lambda_{k,j} + r_{k,j} \qquad \text{where } r_{k,j} := \frac{\widetilde{\theta}_k - \theta_k}{\theta_j - \theta_k}\Lambda_{k,j} + \frac{\langle\Delta f_k, e_j\rangle}{\widetilde{\theta}_k - \theta_j}.
\end{aligned}$$

The $r_{k,j}$'s are small:

$$\begin{aligned}
|r_{k,j}| &\le \frac{5}{4}\left(\frac{\delta|\Lambda_{k,j}| + |\langle\Delta f_k, e_j\rangle|}{|\theta_k - \theta_j|}\right) \qquad \text{for } j \ne k, \text{ if } \epsilon_k > 5\delta \\
&\le \frac{5\delta\|\Lambda_k\|}{|\theta_k - \theta_j|} \qquad \text{by inequality (22).}
\end{aligned} \tag{23}$$

Define $r_{k,k} = |\sigma_{k,k}| - 1 = -\frac{1}{2}\|f_k\|^2$ and $r_k = \sum_{j\in\mathbb{N}} r_{k,j}e_j$. We then have a representation (cf. Hall and Hosseini-Nasab, 2006, equation 2.8 and Cai and Hall, 2006, §5.6)

$$f_k = \sigma_k\widetilde{e}_k - e_k = (\sigma_k\langle\widetilde{e}_k, e_k\rangle - 1)\, e_k + \sum\nolimits_j^* \sigma_k\sigma_{j,k}e_j = \Lambda_k + r_k. \tag{24}$$

6.3. *Approximation of projections.* The operator $H_J = \sum_{k\in J} e_k \otimes e_k$ projects elements of $\mathcal{H}$ orthogonally onto span$\{e_k : k \in J\}$; the operator $\widetilde{H}_J = \sum_{k\in J} \widetilde{e}_k \otimes \widetilde{e}_k$ projects elements of $\mathcal{H}$ orthogonally onto span$\{e_k : k \in J\}$. We will be interested in the case $J = \{1, 2, \dots, p\}$ with p equal to either the m or the N from Section 5. In that case, we also write H_p and $\widetilde{H}_p$ for the projection operators.

In this subsection we establish a bound for $\|\widetilde{H}_J\mathbb{B} - H_J\mathbb{B}\|$ for a $\mathbb{B} = \sum_j b_je_j$ in $\mathcal{H}$.

The difference $\widetilde{H}_J - H_J$ equals

$$\begin{aligned}\sum\nolimits_{k\in J}(\sigma_k\widetilde{e}_k)\otimes(\sigma_k\widetilde{e}_k) - e_k\otimes e_k \\ &= \sum\nolimits_{k\in J}\sigma_k\widetilde{e}_k\otimes r_k + \sum\nolimits_{k\in J}(e_k+f_k)\otimes\Lambda_k \\ &\quad + \sum\nolimits_{k\in J}\left((e_k+\Lambda_k+r_k)\otimes e_k - e_k\otimes e_k\right) \\ &= \mathcal{R}_J + \sum\nolimits_{k\in J} e_k\otimes\Lambda_k + \Lambda_k\otimes e_k \\ &\qquad \text{where } \mathcal{R}_J := \sum\nolimits_{k\in J}\sigma_k\widetilde{e}_k\otimes r_k + f_k\otimes\Lambda_k + r_k\otimes e_k.\end{aligned}$$

Self-adjointness of $\widetilde{T}$ implies $\widetilde{T}_{j,k} = \widetilde{T}_{k,j}$ and hence $\Lambda_{j,k} = -\Lambda_{j,k}$. The anti-symmetry eliminates some terms from the main contribution to $\widetilde{H}_J - H_J$:

$$\sum\nolimits_{k\in J} e_k\otimes\Lambda_k + \Lambda_k\otimes e_k = \sum\nolimits_{k\in J}\sum\nolimits_{j\in J^c}\Lambda_{k,j}\left(e_k\otimes e_j + e_j\otimes e_k\right).$$

With this simplification we get the following bound for $\|(\widetilde{H}_J - H_J)\mathbb{B}\|^2$:

$$3\|\sum\nolimits_{k\in J} e_k\sum\nolimits_{j\in J^c}\Lambda_{k,j}b_j\|^2 + 3\|\sum\nolimits_{j\in J^c} e_j\sum\nolimits_{k\in J}\Lambda_{k,j}b_k\|^2 + 3\|\mathcal{R}_J\mathbb{B}\|^2$$

The first two sums contribute

$$3\sum\nolimits_{k\in J}\left(\sum\nolimits_{j\in J^c}\Lambda_{k,j}b_j\right)^2 + 3\sum\nolimits_{j\in J^c}\left(\sum\nolimits_{k\in J}\Lambda_{k,j}b_k\right)^2$$

In the next section the expected value of both sums will simplify because $\mathbb{P}\Lambda_{k,j}\Lambda_{k,j'}$ will be zero if $j\neq j'$.

For the three contributions to the bound for $\|\mathcal{R}_J\mathbb{B}\|^2$ we make repeated use of the inequality, based on equations (22) and (23),

$$|\langle r_k, x\rangle| \le \frac{81}{2}\|\Lambda_k\|^2|x_k| + 5\delta\|\Lambda_k\|\sum\nolimits_j^*\frac{|x_j|}{|\theta_k-\theta_j|},$$

which is valid whenever $\epsilon_k > 5\delta$. To avoid an unnecessary calculation of precise constants, we adopt the convention of the variable constant: we write C for a universal constant whose value might change from one line to the next. The first two contributions are:

$$\begin{aligned}\|\sum\nolimits_{k\in J}\sigma_k\widetilde{e}_k\langle r_k,\mathbb{B}\rangle\|^2 &= \sum\nolimits_{k\in J}\langle r_k,\mathbb{B}\rangle^2 \\ &\le C\sum\nolimits_{k\in J} b_k^2\|\Lambda_k\|^4 + C\delta^2\sum\nolimits_{k\in J}\|\Lambda_k\|^2\left(\sum\nolimits_j^*\frac{|b_j|}{|\theta_k-\theta_j|}\right)^2\end{aligned}$$

and

$$\|\sum_{k\in J} f_k\langle\Lambda_k,\mathbb{B}\rangle\|^2 \le \left(\sum_{k\in J}\|f_k\|\,|\langle\Lambda_k,\mathbb{B}\rangle|\right)^2$$
$$\le C\left(\sum_{k\in J}\|\Lambda_k\|^2\right)\sum_{k\in J}\left(\sum\nolimits_j^* \Lambda_{k,j}b_j\right)^2.$$

For the third contribution, let $x = \sum_j x_j e_j$ be an arbitrary unit vector in $\mathcal{H}$. Then

$$\left(\sum_{k\in J}\langle r_k\otimes e_k\mathbb{B},x\rangle\right)^2 = \left(\sum_{k\in J} b_k\langle r_k,x\rangle\right)^2$$
$$\le C\left(\sum_{k\in J}|b_k x_k|\|\Lambda_k\|^2\right)^2 + C\delta^2\left(\sum_{k\in J}\|\Lambda_k\|b_k\sum\nolimits_j^*\frac{|x_j|}{|\theta_k-\theta_j|}\right)^2$$
$$\le C\left(\sum_{k\in J}|b_k|\|\Lambda_k\|^2\right)^2 + C\delta^2\left(\sum_{k\in J}\|\Lambda_k\|^2\right)\sum_{k\in J}b_k^2\left(\sum\nolimits_j^*\frac{1}{|\theta_k-\theta_j|}\right)^2$$

take the supremum over x, which doesn't even appear in the last line, to get the same bound for $\|\sum_{k\in J} b_k r_k\|^2$.

In summary: if $\min_{k\in J}\epsilon_k > 5\delta$ then $\|(\widetilde{H}_J - H_J)\mathbb{B}\|^2$ is bounded by a universal constant times

$$\sum_{k\in J}\left(\sum_{j\in J^c}\Lambda_{k,j}b_j\right)^2 + \sum_{j\in J^c}\left(\sum_{k\in J}\Lambda_{k,j}b_k\right)^2 + \sum_{k\in J}b_k^2\|\Lambda_k\|^4$$
$$+\left(\sum_{k\in J}|b_k|\|\Lambda_k\|^2\right)^2 + \left(\sum_{k\in J}\|\Lambda_k\|^2\right)\sum_{k\in J}\left(\sum\nolimits_j^*\Lambda_{k,j}b_j\right)^2$$
$$+\delta^2\sum_{k\in J}\|\Lambda_k\|^2\left(\sum\nolimits_j^*\frac{|b_j|}{|\theta_k-\theta_j|}\right)^2$$

(25)

$$+\delta^2\left(\sum_{k\in J}\|\Lambda_k\|^2\right)\sum_{k\in J}b_k^2\left(\sum\nolimits_j^*\frac{1}{|\theta_k-\theta_j|}\right)^2.$$

7. Unknown Gaussian distribution. When μ and K are unknown, we estimate them in the usual way: $\widetilde{\mu}_n(t) = \overline{\mathbb{X}}_n(t) = n^{-1}\sum_{i\le n}\mathbb{X}_i(t)$ and

$$\widetilde{K}(s,t) = (n-1)^{-1}\sum_{i\le n}\left(\mathbb{X}_i(s)-\overline{\mathbb{X}}_n(s)\right)\left(\mathbb{X}_i(t)-\overline{\mathbb{X}}_n(t)\right)$$
$$= (n-1)^{-1}\sum_{i\le n}\left(\mathbb{Z}_i(s)-\overline{\mathbb{Z}}(s)\right)\left(\mathbb{Z}_i(t)-\overline{\mathbb{Z}}(t)\right),$$

which has spectral representation

$$\widetilde{K}(s,t) = \sum_{k\in\mathbb{N}}\widetilde{\theta}_k\widetilde{\phi}_k(s)\widetilde{\phi}_k(t).$$

In fact we must have $\widetilde{\theta}_k = 0$ for $k \geq n$ because all the eigenfunctions $\widetilde{\phi}_k$ corresponding to nonzero $\widetilde{\theta}_k$'s must lie in the $n-1$-dimensional space spanned by $\{\mathbb{Z}_i - \overline{\mathbb{Z}} : i = 1, 2, \ldots, n\}$.

The construction and analysis of the new estimator $\widehat{\mathbb{B}}$ will parallel the method developed in Section 5 for the case of known K and μ. The quantities m and N are the same as before. We write $\widetilde{H}_p$ (for $p = N$ or $p = m$) for the operator that projects orthogonally onto span$\{\widetilde{\phi}_1, \ldots, \widetilde{\phi}_p\}$. Essentially we have only to estimate all the quantities that appeared in the previous proof then show that none of the errors of estimation is large enough to upset analogs of the calculations from Section 5. There is a slight complication caused by the fact that we do not know which of $\pm\widetilde{\phi}_j$ should be used to approximate ϕ_j. At strategic moments we will be forced to multiply by the matrix $\widetilde{S} := \mathrm{diag}(\sigma_0, \ldots, \sigma_N)$ with $\sigma_0 = 1$ and $\sigma_k = \mathrm{sign}\big(\langle \phi_k, \widetilde{\phi}_k \rangle\big)$ for $k \geq 1$. The results from Section 6 will control the difference $f_k := \sigma_k \widetilde{\phi}_k - \phi_k$. The other key quantities are:

(i) $\Delta := \widetilde{K} - K$

(ii) $\widetilde{D} = \mathrm{diag}(1, \widetilde{\theta}_1, \ldots, \widetilde{\theta}_N)^{1/2}$

(iii) $\widetilde{z}_i = (\widetilde{z}_{i,1}, \ldots, \widetilde{z}_{i,N})'$ where $\widetilde{z}_{i,k} = \langle \mathbb{Z}_i, \widetilde{\phi}_k \rangle$

(iv) $\widetilde{z}_{\cdot} = (\widetilde{z}_{\cdot 1}, \ldots, \widetilde{z}_{\cdot N})'$ where $\widetilde{z}_{\cdot k} = \langle \overline{\mathbb{Z}}, \widetilde{\phi}_k \rangle = n^{-1} \sum_{i \leq n} \widetilde{z}_{i,k}$

(v) $\widetilde{\xi}_i = (1, \widetilde{z}_i' - \widetilde{z}_{\cdot}')$ and $\widetilde{\eta}_i = D^{-1} \widetilde{\xi}_i$. [We could define $\widetilde{\eta}_i = \widetilde{D}^{-1} \widetilde{\xi}_i$ but then we would need to show that $\widetilde{D}^{-1} \widetilde{\xi}_i \approx D^{-1} \widetilde{\xi}_i$. Our definition merely rearranges the approximation steps.]

(vi) $\widetilde{\gamma} := (\widetilde{\gamma}_0, \widetilde{b}_1, \ldots, \widetilde{b}_N)'$ where $\mathbb{B} = \sum_{k \in \mathbb{N}} \widetilde{b}_k \widetilde{\phi}_k$ and $\widetilde{\gamma}_0 := a + \langle \mathbb{B}, \overline{\mathbb{X}} \rangle$. [Note that $\lambda_i = \widetilde{\gamma}_0 + \langle \mathbb{B}, \mathbb{Z}_i - \overline{\mathbb{Z}} \rangle$.]

(vii) $\widetilde{\lambda}_{i,N} = \widetilde{\gamma}_0 + \langle \widetilde{H}_N \mathbb{B}, \mathbb{Z}_i - \overline{\mathbb{Z}} \rangle = \widetilde{\xi}_i' \widetilde{\gamma}$.

(viii) $\widehat{g} = \mathrm{argmax}_{g \in \mathbb{R}^{N+1}} \sum_{i \leq n} y_i (\widetilde{\xi}_i' g) - \psi(\widetilde{\xi}_i' g)$ and

$$\widehat{\mathbb{B}} = \sum\nolimits_{1 \leq k \leq m} \widehat{g}_k \widetilde{\phi}_k.$$

[Note that these two quantities differ from the $\widehat{g}$ and $\widehat{\mathbb{B}}$ in Section 5.]

(ix) $\widetilde{A}_n = n^{-1} \sum_{i \leq n} \widetilde{\eta}_i \widetilde{\eta}_i' \psi^{(2)}(\widetilde{\lambda}_{i,N})$

The use of estimated quantities has one simplifying consequence:

$$\mathbb{Z}_i(t) - \overline{\mathbb{Z}}(t) = \sum\nolimits_{k \in \mathbb{N}} (\widetilde{z}_{i,k} - \widetilde{z}_{\cdot k}) \widetilde{\phi}_k(t)$$

so that

$$\widetilde{\theta}_k\{j=k\} = \iint \widetilde{K}(s,t)\widetilde{\phi}_j(s)\widetilde{\phi}_k(t)\,ds\,dt$$
$$= (n-1)^{-1}\sum\nolimits_{i\le n}(\widetilde{z}_{i,j}-\widetilde{z}_{\cdot j})(\widetilde{z}_{i,k}-\widetilde{z}_{\cdot k}),$$

which implies $(n-1)^{-1}\sum_{i\le n}\widetilde{z}_i\widetilde{z}_i' = \widetilde{D}^2$ and

$$(26)\qquad (n-1)^{-1}\sum\nolimits_{i\le n}\widetilde{\eta}_i\widetilde{\eta}_i' = D^{-1}\widetilde{D}^2D^{-1} := \operatorname{diag}(1,\widetilde{\theta}_1/\theta_1,\dots,\widetilde{\theta}_N/\theta_N).$$

We will analyze $\widetilde{K}$ by rewriting it using the eigenfunctions for K. Remember that $z_{i,j} = \langle \mathbb{Z}_i, \phi_j\rangle$ and the standardized variables $\eta_{i,j} = z_{i,j}/\sqrt{\theta_j}$ are independent $N(0,1)$'s. Define $z_{\cdot j} = \langle \overline{\mathbb{Z}}, \phi_j\rangle$ and $\eta_{\cdot j} = n^{-1}\sum_{i\le n}\eta_{i,j}$ and

$$\mathcal{C}_{j,k} := (n-1)^{-1}\sum\nolimits_{i\le n}\left(\eta_{i,j}-\eta_{\cdot j}\right)\left(\eta_{i,k}-\eta_{\cdot k}\right),$$

a sample covariance between two independent $N(0,I_N)$ random vectors. Then

$$\mathbb{Z}_i(t)-\overline{\mathbb{Z}}(t) = \sum\nolimits_{j\in\mathbb{N}}(z_{i,j}-z_{\cdot j})\phi_j(t) = \sum\nolimits_{j\in\mathbb{N}}\sqrt{\theta_j}(\eta_{i,j}-\eta_{\cdot j})\phi_j(t)$$

and

$$(27)\qquad \widetilde{K}(s,t) = \sum\nolimits_{j,k\in\mathbb{N}}\widetilde{K}_{j,k}\phi_j(s)\phi_k(t) \qquad \text{with } \widetilde{K}_{j,k} = \sqrt{\theta_j\theta_k}\,\mathcal{C}_{j,k}$$

Moreover, as shown in Section 6, the main contribution to $f_k = \sigma_k\widetilde{\phi}_k - \phi_k$ is

$$\Lambda_k := \sum\nolimits_{j\in\mathbb{N}}\Lambda_{k,j}\phi_j \qquad \text{with } \Lambda_{k,j} := \begin{cases}\sqrt{\theta_j\theta_k}\,\mathcal{C}_{j,k}/(\theta_k-\theta_j) & \text{if } j\ne k\\ 0 & \text{if } j=k\end{cases}.$$

In fact, most of the inequalities that we need to study the new $\widehat{\mathbb{B}}$ come from simple moment bounds (Lemma 31) for the sample covariances $\mathcal{C}_{j,k}$ and the derived bounds (Lemma 32) for the Λ_k's.

As before, most of the analysis will be conditional on the $\mathbb{X}_i$'s lying in a set with high probability on which the various estimators and other random quantities are well behaved.

LEMMA 28. *For each $\epsilon>0$ there exists a set $\widetilde{\mathcal{X}}_{\epsilon,n}$, depending on μ and K, with*

$$\sup\nolimits_{\mathcal{F}}\mathbb{P}_{n,\mu,K}\widetilde{\mathcal{X}}_{\epsilon,n}^c < \epsilon \qquad \text{for all large enough } n$$

and on which, for some constant C_ϵ that does not depend on μ or K, *(i)* $\|\Delta\| \le C_\epsilon n^{-1/2}$

(ii) $\max_{i\le n}\|\mathbb{Z}_i\| \le C_\epsilon\sqrt{\log n}$ *and* $\|\overline{\mathbb{Z}}\| \le C_\epsilon n^{-1/2}$

(iii) $\|(\widetilde{H}_m - H_m)\mathbb{B}\|^2 = o_{\mathfrak{F}}(\rho_n)$

(iv) $\|(\widetilde{H}_N - H_N)\mathbb{B}\|^2 = O_{\mathfrak{F}}(n^{-1-\nu})$ *for some* $\nu > 0$ *that depends only on* α *and* β

(v) $\max_{i\le n}|\widetilde{\eta}_i|^2 = o_{\mathfrak{F}}(\sqrt{n}/N)$

(vi) $\|\widetilde{S}\widetilde{A}_n\widetilde{S} - A_n\|_2 = o_{\mathfrak{F}}(1)$

This Lemma (whose proof appears in Section 8) contains everything we need to show that $\|\widehat{\mathbb{B}}-\mathbb{B}\|^2$ has the uniform $O_{\mathfrak{F}}(\rho_n)$ rate of convergence in $\mathbb{P}_{n,f}$ probability, as asserted by equation (3). In what follows, all assertions refer to the numbered parts of Lemma 28.

As before, the component of $\mathbb{B}$ orthogonal to span$\{\widetilde{\phi}_1,\dots,\widetilde{\phi}_m\}$ causes no trouble because

$$\|\widehat{\mathbb{B}} - \mathbb{B}\|^2 = \|\widehat{g} - \widetilde{\gamma}\|_2^2 + \|\widetilde{H}_m^{\perp}\mathbb{B}\|^2$$

and, by (iii),

$$\|\widetilde{H}_m^{\perp}\mathbb{B}\|^2 \le 2\|H_m^{\perp}\mathbb{B}\|^2 + 2\|(\widetilde{H}_m - H_m)\mathbb{B}\|^2 = O_{\mathfrak{F}}(\rho_n) \qquad \text{on } \widetilde{\mathfrak{X}}_{\epsilon,n}.$$

To handle $\|\widehat{g} - \widetilde{\gamma}\|_2$, invoke Corollary 9 for $\mathbb{X}_i$'s in $\widetilde{\mathfrak{X}}_{\epsilon,n}$, with η_i replaced by $\widetilde{\eta}_i$ and A_n replaced by $\widetilde{A}_n$ and B_n replaced by $\widetilde{B}_n = \widetilde{S}B_n\widetilde{S}$, the same B_n and D as before, and $\mathbb{Q}$ equal to

$$\widetilde{\mathbb{Q}}_{n,a,\mathbb{B},N} = \otimes_{i\le n} Q_{\widetilde{\lambda}_{i,N}}.$$

to get a set $\widetilde{\mathcal{Y}}_{m,\epsilon}$ with $\widetilde{\mathbb{Q}}_{n,a,\mathbb{B},N}\widetilde{\mathcal{Y}}_{m,\epsilon}^c < 2\epsilon$ on which $\|\widehat{g} - \widetilde{\gamma}\|_2^2 = O_{\mathfrak{F}}(\rho_n)$. The conditions of the Corollary are satisfied on $\widetilde{\mathfrak{X}}_{\epsilon,n}$, because of (v) and

$$\|\widetilde{A}_n - \widetilde{B}_n\|_2 \le \|\widetilde{A}_n - \widetilde{S}A_n\widetilde{S}\|_2 + \|\widetilde{S}A_n\widetilde{S} - \widetilde{S}B_n\widetilde{S}\|_2 = o_{\mathfrak{F}}(1).$$

To complete the proof it suffices to show that $\|\mathbb{Q}_{n,a,\mathbb{B},N} - \widetilde{\mathbb{Q}}_{n,a,\mathbb{B},N}\|_{\text{TV}}$ tends to zero. First note that

$$\begin{aligned}\widetilde{\lambda}_{i,N} - \lambda_{i,N} &= a + \langle\mathbb{B},\overline{\mathbb{X}}\rangle + \langle\widetilde{H}_N\mathbb{B},\mathbb{Z}_i - \overline{\mathbb{Z}}\rangle - a - \langle\mathbb{B},\mu\rangle - \langle H_N\mathbb{B},\mathbb{Z}_i\rangle \\ &= \langle\widetilde{H}_N^{\perp}\mathbb{B},\overline{\mathbb{Z}}\rangle - \langle H_N^{\perp}\mathbb{B},\overline{\mathbb{Z}}\rangle + \langle H_N^{\perp}\mathbb{B},\overline{\mathbb{Z}}\rangle + \langle\widetilde{H}_N\mathbb{B} - H_N\mathbb{B},\mathbb{Z}_i\rangle\end{aligned}$$

which implies that, on $\widetilde{\mathfrak{X}}_{\epsilon,n}$,

$$\begin{aligned}|\widetilde{\lambda}_{i,N} - \lambda_{i,N}|^2 &\le 2|\langle H_N^{\perp}\mathbb{B},\overline{\mathbb{Z}}\rangle|^2 + 2\|\widetilde{H}_N\mathbb{B} - H_N\mathbb{B}\|^2\left(\|\mathbb{Z}_i\| + \|\overline{\mathbb{Z}}\|\right)^2 \\ &\le O_{\mathfrak{F}}(N^{1-2\beta})C_\epsilon^2 n^{-1} + O_{\mathfrak{F}}(n^{-1-\nu})C_\epsilon^2\left(n^{-1/2} + \sqrt{\log n}\right)^2 \\ &= O_{\mathfrak{F}}(n^{-1-\nu'}) \qquad \text{for some } 0 < \nu' < \nu. \qquad (29)\end{aligned}$$

Now argue as in subsection 5.2: on $\widetilde{\mathcal{X}}_{\epsilon,n}$,

$$\|\widetilde{\mathbb{Q}}_{n,a,\mathbb{B},N} - \mathbb{Q}_{n,a,\mathbb{B},N}\|^2_{\text{TV}} \le \sum\nolimits_{i\le n} h^2\left(Q_{\widetilde{\lambda}_{i,N}}, Q_{\lambda_{i,N}}\right)$$
$$\le \exp\left(o_{\mathcal{F}}(\log n)\right)\sum\nolimits_{i\le n} |\widetilde{\lambda}_{i,N} - \lambda_{i,N}|^2 = o_{\mathcal{F}}(1).$$

Finish the argument as before, by splitting into contributions from $\widetilde{\mathcal{X}}^c_n$ and $\widetilde{\mathcal{X}}_n \cap \widetilde{\mathcal{Y}}^c_{m,\epsilon}$ and $\widetilde{\mathcal{X}}_n \cap \widetilde{\mathcal{Y}}_{m,\epsilon}$.

8. Proofs of unproven assertions from Section 7. Many of the inequalities in this section involve sums of functions of the θ_j's. The following result will save us a lot of repetition. To simplify the notation, we drop the subscripts from $\mathbb{P}_{n,\mu,K}$.

LEMMA 30.

(i) For each $r \ge 1$ there is a constant $C_r = C_r(\mathcal{F})$ for which

$$\kappa_k(r,\gamma) := \sum\nolimits_{j\in\mathbb{N}} \{j \ne k\} \frac{j^{-\gamma}}{|\theta_j - \theta_k|^r} \le \begin{cases} C_r\left(1 + k^{r(1+\alpha)-\gamma}\right) & \text{if } r > 1 \\ C_1\left(1 + k^{1+\alpha-\gamma}\log k\right) & \text{if } r = 1 \end{cases}$$

(ii) For each p,

$$\sum\nolimits_{k\le p}\sum\nolimits_{j>p} \frac{k^{-\alpha-2\beta}j^{-\alpha}}{|\theta_k - \theta_j|^2} = O_{\mathcal{F}}(p^{1-\alpha})$$

PROOF. For (i), argue in the same way as Hall and Horowitz (2007, page 85), using the lower bounds

$$|\theta_j - \theta_k| \ge \begin{cases} c_\alpha j^{-\alpha} & \text{if } j < k/2 \\ c_\alpha |j-k| k^{-\alpha-1} & \text{if } k/2 \le j \le 2k \\ c_\alpha k^{-\alpha} & \text{if } j > 2k \end{cases}$$

where c_α is a positive constant.

For (ii), split the range of summation into two subsets: $\{(k,j) : j > \max(p, 2k)\}$ and $\{(k,j) : p/2 < k \le p < j \le 2k\}$. The first subset contributes at most

$$\sum\nolimits_{k\le p} k^{-\alpha-2\beta} \sum\nolimits_{j>\max(p,2k)} j^{-\alpha}(c_\alpha k^{-\alpha})^{-2} = O_{\mathcal{F}}(p^{1-\alpha})$$

because $\alpha - 2\beta < -3$. The second subset contributes at most

$$\sum\nolimits_{p/2<k\le p} k^{-\alpha-2\beta} c_\alpha^{-2} k^{2\alpha+2} \sum\nolimits_{j>p} j^{-\alpha}(j-k)^{-2} = O_{\mathcal{F}}\left(p.p^{2+\alpha-2\beta}p^{-\alpha}O(1)\right),$$

which is of order $o_{\mathcal{F}}(p^{-\alpha})$. □

The distribution of $\mathcal{C}_{j,k}$ does not depend on the parameters of our model. Indeed, by the usual rotation of axes we can rewrite $(n-1)\mathcal{C}_{j,k}$ as $U_j'U_k$, where $U_1, U_2, \ldots$ are independent $N(0, I_{n-1})$ random vectors. This representation gives some useful equalities and bounds.

LEMMA 31. *Uniformly over distinct j, k, ℓ,*

(i) $\mathbb{P}\mathcal{C}_{j,j} = 1$ *and* $\mathbb{P}\left(\mathcal{C}_{j,j} - 1\right)^2 = 2(n-1)^{-1}$
(ii) $\mathbb{P}\mathcal{C}_{j,k} = \mathbb{P}\mathcal{C}_{j,k}\mathcal{C}_{j,\ell} = 0$
(iii) $\mathbb{P}\mathcal{C}_{j,k}^2 = O(n^{-1})$
(iv) $\mathbb{P}\mathcal{C}_{j,k}^2\mathcal{C}_{\ell,k}^2 = tO(n^{-2})$
(v) $\mathbb{P}\mathcal{C}_{j,k}^4 = O(n^{-2})$

PROOF. Assertion (i) is classical because $|U_j|^2 \sim \chi^2_{n-1}$. For assertion (ii) use $\mathbb{P}(U_1'U_2 \mid U_2) = 0$ and

$$\mathbb{P}(U_1'U_2U_2'U_3 \mid U_2) = \operatorname{trace}\left(U_2U_2'\mathbb{P}(U_3U_1')\right) = 0.$$

For (iii) use $\mathbb{P}(U_1U_1') = I_{n-1}$ and

$$\mathbb{P}(U_1'U_2U_2'U_1 \mid U_2) = \operatorname{trace}\left(U_2U_2'\mathbb{P}(U_1U_1')\right) = \operatorname{trace}(U_2U_2') = |U_2|^2.$$

For (iv) use $\mathbb{P}|U_2|^4 = n^2 - 1$ and

$$\mathbb{P}((U_1'U_2)^2(U_3'U_2)^2 \mid U_2) = |U_2|^4$$

For (v), check that the coefficient of t^4 in the Taylor expansion of

$$\mathbb{P}\exp(tU_1'U_2) = \mathbb{P}\exp\left(\tfrac{1}{2}t^2|U_1|^2\right) = (1-t^2)^{-(n-1)/2}$$

is of order n^2. □

LEMMA 32. *Uniformly over distinct j, k, ℓ,*

(i) $\mathbb{P}\Lambda_{k,j} = \mathbb{P}\Lambda_{k,j}\Lambda_{k,\ell} = 0$
(ii) $\mathbb{P}\Lambda_{k,j}^2 = O_{\mathcal{F}}\left(n^{-1}k^{-\alpha}j^{-\alpha}(\theta_k - \theta_j)^{-2}\right)$
(iii) $\mathbb{P}\Lambda_{k,j}^4 = O_{\mathcal{F}}\left(n^{-2}k^{-2\alpha}j^{-2\alpha}(\theta_k - \theta_j)^{-4}\right)$
(iv) $\mathbb{P}\|\Lambda_k\|^2 = O_{\mathcal{F}}(n^{-1}k^2)$
(v) $\mathbb{P}\|\Lambda_k\|^4 = O_{\mathcal{F}}(n^{-2}k^4)$

PROOF. Assertions (i), (ii), and (iii) follow from Assertions (ii) and (iii) of Lemma 31. For (iv), note that

$$\mathbb{P}\|\Lambda_k\|^2 = \sum\nolimits_j^* \mathbb{P}\Lambda_{j,k}^2 = O_{\mathcal{F}}(n^{-1}k^{-\alpha})\kappa_k(2,\alpha)$$

For (v) note that

$$\begin{aligned}
\mathbb{P}\|\Lambda_k\|^4 &= \mathbb{P}\left(\sum\nolimits_j^* \theta_j\theta_k S_{j,k}^2(\theta_k-\theta_j)^{-2}\right)^2 \\
&= \sum\nolimits_j^* \sum\nolimits_\ell^* \theta_j\theta_\ell\theta_k^2(\theta_k-\theta_j)^{-2}(\theta_k-\theta_\ell)^{-2}\mathbb{P}S_{j,k}^2S_{\ell,k}^2 \\
&= O_{\mathcal{F}}(n^{-2})\left(\sum\nolimits_j^* \theta_j\theta_k(\theta_k-\theta_j)^{-2}\right)^2 \\
&= O_{\mathcal{F}}(n^{-2}k^4).
\end{aligned}$$

□

To prove Lemma 28 we define $\widetilde{\mathfrak{X}}_{\epsilon,n}$ as an intersection of sets chosen to make the six assertions of the Lemma hold,

$$\widetilde{\mathfrak{X}}_{\epsilon,n} := \widetilde{\mathfrak{X}}_{\Delta,n} \cap \widetilde{\mathfrak{X}}_{\mathbb{Z},n} \cap \widetilde{\mathfrak{X}}_{\Lambda,n} \cap \widetilde{\mathfrak{X}}_{\eta,n} \cap \widetilde{\mathfrak{X}}_{A,n},$$

where the complement of each of the five sets appearing on the right-hand side has probability less than $\epsilon/5$. More specifically, for a large enough constant C_ϵ, we define

$$\begin{aligned}
\widetilde{\mathfrak{X}}_{\Delta,n} &= \{\|\Delta\| \le C_\epsilon n^{-1/2}\} \\
\widetilde{\mathfrak{X}}_{\mathbb{Z},n} &= \{\max\nolimits_{i\le n}\|\mathbb{Z}_i\|^2 \le C_\epsilon \log n \text{ and } \|\overline{\mathbb{Z}}\| \le C_\epsilon n^{-1/2}\} \\
\widetilde{\mathfrak{X}}_{\eta,n} &= \{\max\nolimits_{i\le n}|\eta_i|^2 \le C_\epsilon N\log n\} \qquad \text{as in Section 5} \\
\widetilde{\mathfrak{X}}_{A,n} &= \{\|\sum\nolimits_{i\le n}\widetilde{\eta}_i\widetilde{\eta}_i'\|_2 \le C_\epsilon n\}
\end{aligned}$$

The definition of $\widetilde{\mathfrak{X}}_{\Lambda,n}$, in subsection 8.3, is slightly more complicated. It is defined by requiring various functions of the Λ_k's to be smaller than C_ϵ times their expected values.

The set $\widetilde{\mathfrak{X}}_{A,n}$ is almost redundant. From Definition 5 we know that

$$\min_{1\le j<j'\le N}|\theta_j-\theta_{j'}| \ge (\alpha/R)N^{-1-\alpha} \qquad \text{and} \qquad \min_{1\le j\le N}\theta_j \ge R^{-1}N^{-\alpha}.$$

The choice $N \sim n^\zeta$ with $\zeta < (2+2\alpha)^{-1}$ ensures that $n^{1/2}N^{-1-\alpha} \to \infty$. On $\widetilde{\mathfrak{X}}_{\Delta,n}$ the spacing assumption used in Section 6 holds for all n large enough; all the bounds from that Section are avaiable to us on $\widetilde{\mathfrak{X}}_{\epsilon,n}$. In particular,

$$\max\nolimits_{j\le N}|\widetilde{\theta}_j/\theta_j - 1| \le O_{\mathcal{F}}(N^\alpha\|\Delta\|) = o_{\mathcal{F}}(1).$$

Equality (26) shows that $\widetilde{\mathfrak{X}}_{A,n} \subseteq \widetilde{\mathfrak{X}}_{\Delta,n}$ eventually if we make sure $C_\epsilon > 1$.

8.1. *Proof of Lemma 28 part (i).* Observe that

$$\begin{aligned}\mathbb{P}\|\Delta\|^2 &= \sum\nolimits_{j,k} \mathbb{P}\left(\widetilde{K}_{j,k} - \theta_j\{j=k\}\right)^2 \\ &= \sum\nolimits_{j,k} \theta_j\theta_k \mathbb{P}\left(S_{j,k} - \{j=k\}\right)^2 \\ &\le \sum\nolimits_j \theta_j O_{\mathfrak{F}}(n^{-1}) + \sum\nolimits_{j,k} \theta_j\theta_k O_{\mathfrak{F}}(n^{-2}) \\ &= O_{\mathfrak{F}}(n^{-1})\end{aligned}$$

8.2. *Proof of Lemma 28 part (ii).* As before, Corollary 42 controls $\max_{i\le n}\|\mathbb{Z}_i\|^2$. To control the $\overline{\mathbb{Z}}$ contribution, note that $n\|\overline{\mathbb{Z}}\|^2$ has the same distribution as $\|\mathbb{Z}_1\|^2$, which has expected value $\sum_{j\in\mathbb{N}}\theta_j < \infty$.

8.3. *Proof of Lemma 28 parts (iii) and (iv).* Calculate expected values for all the terms that appear in the bound (25) from Section 6.

$$\begin{aligned}&\mathbb{P}_{n,\mu,K}\sum\nolimits_{k\le p}\left(\sum\nolimits_{j>p}\Lambda_{k,j}b_j\right)^2 + \mathbb{P}_{n,\mu,K}\sum\nolimits_{j>p}\left(\sum\nolimits_{k\le p}\Lambda_{k,j}b_k\right)^2 \\ &\quad= \sum\nolimits_{k\le p}\sum\nolimits_{j>p}\mathbb{P}_{n,\mu,K}\Lambda_{k,j}^2\left(b_j^2+b_k^2\right) \qquad \text{by Lemma 32(i)} \\ &\quad= O_{\mathfrak{F}}(n^{-1})\sum\nolimits_{k\le p}\sum\nolimits_{j>p}k^{-\alpha-2\beta}j^{-\alpha}(\theta_k-\theta_j)^{-2} \\ &\quad= O_{\mathfrak{F}}(n^{-1}p^{1-\alpha}) \qquad \text{by Lemma 30}\end{aligned} \tag{33}$$

and

$$\mathbb{P}_{n,\mu,K}\sum\nolimits_{k\le p}b_k^2\|\Lambda_k\|^4 = O_{\mathfrak{F}}(n^{-2})\sum\nolimits_{k\le p}k^{4-2\beta} = O_{\mathfrak{F}}(n^{-2})\left(1+p^{5-2\beta}+\log p\right)$$

and

$$\mathbb{P}_{n,\mu,K}\sum\nolimits_{k\le p}|b_k|\|\Lambda_k\|^2 = O_{\mathfrak{F}}(n^{-1})\sum\nolimits_{k\in J}k^{2-\beta} = O_{\mathfrak{F}}(n^{-1})\left(1+p^{3-\beta}+\log p\right)$$

and

$$\mathbb{P}_{n,\mu,K}\sum\nolimits_{k\le p}\|\Lambda_k\|^2 = O_{\mathfrak{F}}(n^{-1}p^3)$$

and

$$\begin{aligned}\mathbb{P}_{n,\mu,K}\sum\nolimits_{k\le p}\left(\sum\nolimits_j^*\Lambda_{k,j}b_j\right)^2 &= O_{\mathfrak{F}}(n^{-1})\sum\nolimits_{k\le p}\sum\nolimits_j^* k^{-\alpha}j^{-a-2\beta}(\theta_k-\theta_j)^{-2} \\ &= O_{\mathfrak{F}}(n^{-1}) \qquad \text{by Lemma 30}\end{aligned} \tag{34}$$

and

$$\begin{aligned}&\delta^2\mathbb{P}_{n,\mu,K}\sum\nolimits_{k\le p}\|\Lambda_k\|^2\left(\sum\nolimits_j^*\frac{|b_j|}{|\theta_k-\theta_j|}\right)^2 \\ &\qquad= O_{\mathfrak{F}}(n^{-1}\delta^2)\left(p^3+p^{5+2\alpha-2\beta}\log^2 p\right)\end{aligned} \tag{35}$$

and

(36)
$$\sum_{k\le p} b_k^2 \left(\sum_j^* \frac{1}{|\theta_k-\theta_j|}\right)^2 = O_{\mathcal{F}}(1+p^{3+2\alpha-2\beta}\log^2 p) \qquad \text{by Lemma 30.}$$

For some constant $C_\epsilon = C_\epsilon(\mathcal{F})$, on a set $\mathcal{X}_{\Lambda,n}$ with $\mathbb{P}_{n,\mu,K}\mathcal{X}_{\Lambda,n}^c < \epsilon$, each of the random quantities in the previous set of inequalities (for both $p=m$ and $p=N$) is bounded by C_ϵ times its $\mathbb{P}_{n,\mu,K}$ expected value. By virtue of Lemma 32(iv), we may also assume that $\|\Lambda_k\|^2 \le C_\epsilon k^2/n$ on $\mathcal{X}_{\Lambda,n}$.

From inequality (25), it follows that on the set $\mathcal{X}_{\Delta,n}\cap\mathcal{X}_{\Lambda,n}$, for both $p=m$ and $p=N$,

$$\begin{aligned}
&\|(\widetilde{H}_p - H_p)\mathbb{B}\|^2 \\
&\quad \le O_{\mathcal{F}}(n^{-1}p^{1-\alpha}) + O_{\mathcal{F}}(n^{-2})\left(1+p^{5-2\beta}+\log p + p^{6-\beta}+\log^2 p\right) \\
&\qquad + O_{\mathcal{F}}(n^{-1}p^3)O_{\mathcal{F}}(n^{-1}) + O_{\mathcal{F}}(n^{-2})\left(p^3+p^{5+2\alpha-2\beta}\log^2 p\right) \\
&\qquad + O_{\mathcal{F}}(n^{-2}p^3)O_{\mathcal{F}}(1+p^{3+2\alpha-2\beta}\log^2 p) \\
&\quad = O_{\mathcal{F}}(n^{-1}p^{1-\alpha}) \qquad \text{if } p\le N.
\end{aligned}$$

This inequality leads to the asserted conclusions when $p=m$ or $p=N$.

8.4. *Proof of Lemma 28 part (v).* By construction, $\widetilde{\eta}_{i1}=1$ for every i and, for $j\ge 2$,

$$\sqrt{\theta_j}\,\widetilde{\eta}_{i,j} = (\widetilde{z}_{i,j} - \widetilde{z}_{\cdot j}) = \langle \mathbb{Z}_i - \overline{\mathbb{Z}}, \widetilde{\phi}_j\rangle$$

Thus, for $j\ge 2$,

$$\sigma_j\widetilde{\eta}_{i,j} = \theta_j^{-1/2}\langle \mathbb{Z}_i-\overline{\mathbb{Z}}, \phi_j+f_j\rangle = \eta_{i,j}+\widetilde{\delta}_{i,j}$$

with

$$|\delta_{i,j}|^2 \le \theta_j^{-1}\left(\|\mathbb{Z}_i\|+\|\overline{\mathbb{Z}}\|\right)^2\|f_j\|^2 \le O_{\mathcal{F}}\left(\frac{j^{2+\alpha}\log n}{n}\right) \qquad \text{on } \widetilde{\mathcal{X}}_{\epsilon,n}.$$

In vector form,

(37) $$\widetilde{S}\widetilde{\eta}_i = \eta_i + \widetilde{\delta}_i \qquad \text{with } |\widetilde{\delta}_i|^2 = O_{\mathcal{F}}\left(\frac{N^{3+\alpha}\log n}{n}\right) \le o_{\mathcal{F}}(n/N^2) \text{ on } \widetilde{\mathcal{X}}_{\epsilon,n}.$$

It follows that

$$\max\nolimits_{i\le n}|\widetilde{\eta}_i| = \max\nolimits_{i\le n}|\widetilde{S}\widetilde{\eta}_i| \le \max\nolimits_{i\le n}|\eta_i| + o_{\mathcal{F}}(\sqrt{n}/N) = O_{\mathcal{F}}(\sqrt{n}/N) \qquad \text{on } \widetilde{\mathcal{X}}_{\epsilon,n}.$$

8.5. *Proof of Lemma 28 part (vi).* From inequality (29) we know that

$$\epsilon_N := \max_{i\le n}|\widetilde{\lambda}_{i,N}-\lambda_{i,N}| = O_{\mathcal{F}}(n^{-(1+\nu')/2}) \qquad \text{on } \widetilde{\mathfrak{X}}_{\epsilon,n}$$

and from subsection 5.2 we have $\max_{i\le n}|\lambda_{i,N}| = O_{\mathcal{F}}(\sqrt{\log n})$. Assumption ($\psi 3$) in Section 3 and the Mean-Value theorem then give

$$\max_{i\le n}|\psi^{(2)}(\widetilde{\lambda}_{i,N})-\psi^{(2)}(\lambda_{i,N})| \le \epsilon_N\psi^{(2)}(\lambda_{i,N})G(\epsilon_N) = o_{\mathcal{F}}(1).$$

If we replace $\psi^{(2)}(\widetilde{\lambda}_{i,N})$ in the definition of $\widetilde{A}_n$ by $L_i := \psi^{(2)}(\lambda_{i,N})$ we make a change Γ with

$$\|\Gamma\|_2 \le o_{\mathcal{F}}(1)\|(n-1)^{-1}\sum\nolimits_{i\le n}\widetilde{\eta}_i\widetilde{\eta}_i'\|_2,$$

which, by equality (26), is of order $o_{\mathcal{F}}(1)$ on $\widetilde{\mathfrak{X}}_{\epsilon,n}$.

From Assumption ($\psi 2$) we have $c_n := \log\max_{i\le n}L_i = o_{\mathcal{F}}(\log n)$. Uniformly over all unit vectors u in $\mathbb{R}^{N+1}$ we therefore have

$$\begin{aligned}
u'\widetilde{S}\widetilde{A}_n\widetilde{S}u &= o_{\mathcal{F}}(1) + (n-1)^{-1}\sum\nolimits_{i\le n}L_iu'(\eta_i+\widetilde{\delta}_i)(\eta_i+\widetilde{\delta}_i)'u\\
&= o_{\mathcal{F}}(1) + \left(1+O(n^{-1})\right)u'A_nu\\
&\qquad + O_{\mathcal{F}}\left(n^{-1}\right)\sum\nolimits_{i\le n}L_i\left((u'\widetilde{\delta}_i)^2+2(u'\eta_i)(u'\widetilde{\delta}_i)\right)
\end{aligned}$$

Rearrange then take a supremum over u to conclude that

$$\|\widetilde{S}\widetilde{A}_n\widetilde{S}-A_n\|_2 \le o_{\mathcal{F}}(1) + O_{\mathcal{F}}(e^{c_n})\max_{i\le n}\left(|\widetilde{\delta}_i|^2+2|\widetilde{\delta}_i|\,|\eta_i|\right)$$

Representation (37) and the defining property of $\widetilde{\mathfrak{X}}_{\eta,n}$ then ensure that the upper bound is of order $o_{\mathcal{F}}(1)$ on $\widetilde{\mathfrak{X}}_{\epsilon,n}$.

9. The minimax lower bound.

We will apply a slight variation on Assouad's Lemma—combining ideas from Yu (1997) and from van der Vaart (1998, Section 24.3)—to establish inequality (2).

We consider behavior only for $\mu = 0$ and $a = 0$, for a fixed K with spectral decomposition $\sum_{j\in\mathbb{N}}\theta_j\phi_j\otimes\phi_j$. For simplicity we abbreviate $\mathbb{P}_{n,0,K}$ to $\mathbb{P}$. Let $J = \{m+1, m+2, \dots, 2m\}$ and $\Gamma = \{0,1\}^J$. Let $\beta_j = Rj^{-\beta}$. For each γ in Γ define $\mathbb{B}_\gamma = \epsilon\sum_{j\in J}\gamma_j\beta_j\phi_j$, for a small $\epsilon > 0$ to be specified, and write $\mathbb{Q}_\gamma$ for the product measure $\otimes_{i\le n}Q_{\lambda_i(\gamma)}$ with

$$\lambda_i(\gamma) = \langle\mathbb{B}_\gamma, \mathbb{Z}_i\rangle = \epsilon\sum\nolimits_{j\in J}\gamma_j\beta_jz_{i,j}.$$

For each j let $\Gamma_j = \{\gamma\in\Gamma : \gamma_j = 1\}$ and let ψ_j be the bijection on Γ that flips the jth coordinate but leaves all other coordinates unchanged. Let π be the uniform distribution on Γ, that is, $\pi_\gamma = 2^{-m}$ for each γ.

For each estimator $\widehat{\mathbb{B}} = \sum_{j\in\mathbb{N}} \widehat{b}_j\phi_j$ we have $\|\mathbb{B}_\gamma - \widehat{\mathbb{B}}\|^2 \geq \sum_{j\in J}\left(\gamma_j\beta_j - \widehat{b}_j\right)^2$ and so

$$\sup_{\mathcal{F}} \mathbb{P}_{n,f}\|\mathbb{B} - \widehat{\mathbb{B}}\|^2 \geq \sum\nolimits_{\gamma\in\Gamma}\pi_\gamma \sum\nolimits_{j\in J}\mathbb{P}\mathbb{Q}_\gamma\left(\epsilon\gamma_j\beta_j - \widehat{b}_j\right)^2$$
$$= 2^{-m}\sum\nolimits_{j\in J}\sum\nolimits_{\gamma\in\Gamma_j}\mathbb{P}\left(\mathbb{Q}_\gamma(\epsilon\beta_j - \widehat{b}_j)^2 + \mathbb{Q}_{\psi_j(\gamma)}(0-\widehat{b}_j)^2\right)$$
$$\geq 2^{-m}\sum\nolimits_{j\in J}\sum\nolimits_{\gamma\in\Gamma_j}\tfrac{1}{4}(\epsilon\beta_j)^2\mathbb{P}\|\mathbb{Q}_\gamma \wedge \mathbb{Q}_{\psi_j(\gamma)}\|, \qquad (38)$$

the last lower bound coming from the fact that

$$(\epsilon\beta_j - \widehat{b}_j)^2 + (0-\widehat{b}_j)^2 \geq \tfrac{1}{4}(\epsilon\beta_j)^2 \qquad \text{for all } \widehat{b}_j.$$

We assert that, if ϵ is chosen appropriately,

$$\min\nolimits_{j,\gamma}\mathbb{P}\|\mathbb{Q}_\gamma \wedge \mathbb{Q}_{\psi_j(\gamma)}\| \text{ stays bounded away from zero as } n\to\infty, \qquad (39)$$

which will ensure that the lower bound in (38) is eventually larger than a constant multiple of $\sum_{j\in J}\beta_j^2 \geq c\rho_n$ for some constant $c > 0$. Inequality (2) will then follow.

To prove (39), consider a γ in Γ and the corresponding $\gamma' = \psi_j(\gamma)$. By virtue of the inequality

$$\|\mathbb{Q}_\gamma \wedge \mathbb{Q}_{\gamma'}\| = 1 - \|\mathbb{Q}_\gamma - \mathbb{Q}_{\gamma'}\|_{\text{TV}} \geq 1 - \left(2\wedge\sum\nolimits_{i\leq n} h^2(Q_{\lambda_i(\gamma)}, Q_{\lambda_i(\gamma')})\right)^{1/2}$$

it is enough to show that

$$\limsup\nolimits_{n\to\infty}\max\nolimits_{j,\gamma}\mathbb{P}\left(2\wedge\sum\nolimits_{i\leq n} h^2(Q_{\lambda_i(\gamma)}, Q_{\lambda_i(\gamma')})\right) < 1. \qquad (40)$$

Define $\mathcal{X}_n = \{\max_{i\leq n}\|\mathbb{Z}_i\|^2 \leq C_0\log n\}$, with the constant C_0 large enough that $\mathbb{P}\mathcal{X}_n^c = o_(1)$. On $\mathcal{X}_n$ we have

$$|\lambda_i(\gamma)|^2 \leq \sum\nolimits_{j\in J}\beta_j^2\|Z_i\|^2 = O(\rho_n)\log n = o(1)$$

and, by inequality (4),

$$h^2(Q_{\lambda_i(\gamma)}, Q_{\lambda_i(\gamma')}) \leq O_{\mathcal{F}}(1)|\lambda_i(\gamma) - \lambda_i(\gamma')|^2 \leq \epsilon^2 O_{\mathcal{F}}(1)\beta_j^2 z_{i,j}^2.$$

We deduce that

$$\mathbb{P}\left(2\wedge\sum\nolimits_{i\leq n} h^2(Q_{\lambda_i(\gamma)}, Q_{\lambda_i(\gamma')})\right) \leq 2\mathbb{P}\mathcal{X}_n^c + \sum\nolimits_{i\leq n}\epsilon^2 O_{\mathcal{F}}(1)\beta_j^2\mathbb{P}\mathcal{X}_n z_{i,j}^2$$
$$\leq o(1) + \epsilon^2 O(1) n\beta_j^2\theta_j.$$

The choice of J makes $\beta_j^2\theta_j \leq R^2 m^{-\alpha-2\beta} \sim R^2/n$. Assertion (40) follows.

10. Hellinger distances in an exponential family. We need to show that $h^2(Q_\lambda, Q_{\lambda+\delta}) \le \delta^2\psi^{(2)}(\lambda)\,(1+|\delta|)\,G(|\delta|)$ for all real λ and δ.

Temporarily write λ' for $\lambda+\delta$ and $\overline{\lambda}$ for $(\lambda+\lambda')/2 = \lambda+\delta/2$.

$$\begin{aligned} 1-\tfrac{1}{2}h^2(Q_\lambda,Q_{\lambda'}) &= \int\sqrt{f_\lambda(y)f_{\lambda'}(y)} \\ &= \int \exp\left(\overline{\lambda}y - \tfrac{1}{2}\psi(\lambda) - \tfrac{1}{2}\psi(\lambda')\right) \\ &= \exp\left(\psi(\overline{\lambda}) - \tfrac{1}{2}\psi(\lambda) - \tfrac{1}{2}\psi(\lambda')\right) \\ &\ge 1 + \psi(\overline{\lambda}) - \tfrac{1}{2}\psi(\lambda) - \tfrac{1}{2}\psi(\lambda') \end{aligned}$$

That is,

$$h^2(Q_\lambda,Q_{\lambda'}) \le \psi(\lambda) + \psi(\lambda+\delta) - 2\psi(\lambda+\delta/2).$$

By Taylor expansion in δ around 0, the right-hand side is less than

$$\tfrac{1}{4}\delta^2\psi^{(2)}(\lambda) + \tfrac{1}{6}\delta^3\left(\psi^{(3)}(\lambda+\delta^*) - \tfrac{1}{8}\psi^{(3)}(\lambda-\delta^*/2)\right)$$

where $0<|\delta^*|<|\delta|$. Invoke inequality (3) twice to bound the coefficient of $\delta^3/6$ in absolute value by

$$\psi^{(2)}(\lambda)\left(G(|\delta|) + \tfrac{1}{8}G(|\delta|/2)\right) \le \tfrac{9}{8}\psi^{(2)}(\lambda)G(|\delta|).$$

The stated bound simplifies some unimportant constants.

11. Bounds for Gaussian processes. As a consequence of defining property (K), the centered process $\mathbb{Z} := \mathbb{X}-\mu$ has an expansion $\mathbb{Z}(t) = \sum_{k\in\mathbb{N}}\sqrt{\theta_k}\eta_k\phi_k(t)$ where the η_k's are independent $N(0,1)$'s, implying

$$\|\mathbb{Z}\|^2 = \iint \sum\nolimits_{k,k'\in\mathbb{N}}\sqrt{\theta_k\theta_{k'}}\eta_k\eta_{k'}\phi_k(t)\phi_{k'}(s)\,dt\,ds = \sum\nolimits_{k\in\mathbb{N}}\theta_k\eta_k^2.$$

LEMMA 41. *Suppose $W_i = \sum_{k\in\mathbb{N}}\tau_{i,k}\eta_{i,k}^2$ for $i=1,\dots,n$, where the $\eta_{i,k}$'s are independent standard normals and the $\tau_{i,k}$'s are nonnegative constants with $\infty > T := \max_{i\le n}\sum_{k\in\mathbb{N}}\tau_{i,k}$. Then*

$$\mathbb{P}\{\max\nolimits_{i\le n} W_i > 4T(\log n + x)\} < 2e^{-x} \qquad \text{for each } x\ge 0.$$

PROOF. Without loss of generality suppose $T=1$. For $s=1/4$, note that

$$\mathbb{P}\exp(sW_i) = \prod\nolimits_{k\in\mathbb{N}}(1-2s\tau_{i,k})^{-1/2} \le \exp\left(\sum\nolimits_{k\in\mathbb{N}} s\tau_{i,k}\right) \le e^{1/4}$$

by virtue of the inequality $-\log(1-t) \leq 2t$ for $|t| \leq 1/2$. With the same s, it then follows that

$$\begin{aligned}\mathbb{P}\{\max_{i\leq n} W_i > 4(\log n + x)\} \\ &\leq \exp\left(-4s(\log n + x)\right)\mathbb{P}\exp\left(\max_{i\leq n} sW_i\right) \\ &\leq e^{-x}\frac{1}{n}\sum\nolimits_{i\leq n}\mathbb{P}\exp(sW_i).\end{aligned}$$

The 2 is just a clean upper bound for $e^{1/4}$. □

COROLLARY 42.

$$\mathbb{P}_n\{\max_{i\leq n}\|\mathbb{Z}_i\|^2 > C'(\log n + x)\} \leq 2e^{-x}$$

where $C' = 4C\sum_{k\in\mathbb{N}} k^{-\alpha} < \infty$.

References.

BIRMAN, M. S. and SOLOMJAK, M. Z. (1987). *Spectral Theory of Self-Adjoint Operators in Hilbert Space*. Reidel Translation from the 1980 Russian edition.

BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. *Lecture Notes in Statistics* **149**. Springer-Verlag.

CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Annals of Statistics* **34** 2159–2179.

HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35** 70-91.

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Statist. Soc. B* **68** 109–126.

HJORT, N. L. and POLLARD, D. (1993). Asymptotics for minimisers of convex processes Technical Report.

PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Annals of Statistics* **16** 356-366.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

YU, B. (1997). Assouad, Fano, and Le Cam. In *A Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G. L. Yang, eds.) 423–435. Springer-Verlag, New York.

STATISTICS DEPARTMENT
YALE UNIVERSITY
E-MAIL: Wei.Dou@yale.edu
E-MAIL: David.Pollard@yale.edu
E-MAIL: Huibin.Zhou@yale.edu
URL: http://www.stat.yale.edu/