Supplementary Material for "From Multiple Gaussian Sequences to Functional Data and Beyond: A Stein Estimation Approach"

Mark Koudstaal

Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

Fang Yao

Department of Probability & Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, Beijing, China;
Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

1. Additional Simulation

It is of interest to inspect the performance of recovering new observations that are independent of the sample used for calculating the Stein weights. Since blocking and soft-thresholding methods are not applicable in this context, we compare only to the oracle strategy cross two settings in which new data $\{\boldsymbol{\theta}_i^*\}_{i\leq n}$ of moderate/large sizes with $n = 100, 1000$ and $m = 500, 5000$ are independently generated in each of 1000 Monte Carlo runs, results shown in Table S1 The first setting draws new $\boldsymbol{\theta}_i^*$ from the same GSM (25) with $\lambda_k = 2\alpha k^{-(2\alpha+1)}$, $k = 1, \ldots, m$, which attains comparable recovery errors to the in-sample results in Table 1 in the paper. The second setting is designed to violate the distributional assumption such that $\theta_{ik}^* \sim U\Big(-\sqrt{2\lambda_k \log(mn)}, \sqrt{2\lambda_k \log(mn)}\Big)$ are independent across $k = 1, \ldots, m$ with $\lambda_k = 2\alpha k^{-(2\alpha+1)}$. The increased recovery errors are seen to be bounded within a factor $\log(mn)$ of those from the corresponding Gaussian experiments, supporting the finding in Theorem 4 in the paper.

2. Useful Auxiliary Results

The following Results 1–5 are collected from various references, which are used repeatedly in the proofs of the main lemmas, theorems and propositions. Result 1 below follows from a union bound and the probability assessments of standard normal random variables.

Result 1. *Suppose that $z_1, \ldots, z_N$ are $N(0, 1)$ but not necessarily independent. Then for $N \geq 2$ and $\gamma > \sqrt{2}$,*

$$\mathbf{P}\left\{\max_{i=1,\ldots,N} |z_i| > \gamma(\log N)^{1/2}\right\} \leq N^{1-\gamma^2/2}. \tag{S1}$$

Results 2 and 3 are borrowed from Freedman (1999). The first provides integral approximations and the second gives a concentration for bounding individual and maximal risks.

Table S1. Average and maximal ℓ^2 errors ($\times 10^2$) over n recoveries of new observations $\{\boldsymbol{\theta}_i^*\}_{i \le n}$ independently drawn from Gaussian and uniform distributions, respectively, when the decay parameter $\alpha = 2/3$, the sampling rate $m = 500, 5000$ and the descending decay sequence $\lambda_k = 2\alpha k^{-(2\alpha+1)}$, $k = 1, , \ldots, m$. In the distribution-violated scenario with the uniform distribution, it is notable that, when scaled down by a factor of $\log(mn)$, the errors are bounded by those from the corresponding Gaussian experiments.

Gaussian, $\theta_{ik} \sim N(0, \lambda_k)$					
		$n = 100$		$n = 1000$	
		Avg	Max	Avg	Max
$m = 500$	Oracle $\hat{\boldsymbol{\theta}}_i^{o,a}$	4.36	7.45	4.36	8.60
	$\hat{\boldsymbol{\theta}}_i^{RS}(\sigma^2)$	6.49	11.5	4.98	9.62
	$\hat{\boldsymbol{\theta}}_i^{RS}(\hat{\sigma}^2_{\min})$	5.93	10.3	5.03	9.74
$m = 5000$	Oracle $\hat{\boldsymbol{\theta}}_i^{o,a}$	1.19	1.66	1.19	1.83
	$\hat{\boldsymbol{\theta}}_i^{RS}(\sigma^2)$	1.86	2.69	1.39	2.10
	$\hat{\boldsymbol{\theta}}_i^{RS}(\hat{\sigma}^2_{\min})$	1.62	2.28	1.39	2.11
Uniform, $\theta_{ik} \sim U\Big(-\sqrt{2\lambda_k \log(mn)}, \sqrt{2\lambda_k \log(mn)}\Big)$					
		$n = 100$		$n = 1000$	
		Avg	Max	Avg	Max
$m = 500$	Oracle $\hat{\boldsymbol{\theta}}_i^{o,a}$	16.2	22.1	19.0	27.9
	$\hat{\boldsymbol{\theta}}_i^{RS}(\sigma^2)$	39.3	56.5	30.4	44.7
	$\hat{\boldsymbol{\theta}}_i^{RS}(\hat{\sigma}^2_{\min})$	34.5	48.8	31.1	45.9
$m = 5000$	Oracle $\hat{\boldsymbol{\theta}}_i^{o,a}$	5.17	6.23	5.96	7.52
	$\hat{\boldsymbol{\theta}}_i^{RS}(\sigma^2)$	13.7	17.4	10.1	12.9
	$\hat{\boldsymbol{\theta}}_i^{RS}(\hat{\sigma}^2_{\min})$	11.2	13.9	10.1	13.0

RESULT 2. *Let $1 < a, b < \infty$ and $0 \le c \le \infty$. Suppose that $ab > c+1$ and $\gamma_n \to \infty$. Take $s_k \propto k^a$, $t_k \propto k^c$ and let $g_n = \gamma_n^{b-(1+c)/a}$. Then for any integer q, it follows that*

$$\lim_{n\to\infty} g_n \sum_{k=q}^{\infty} \frac{t_k}{(\gamma_n + s_k)^b} = \lim_{n\to\infty} g_n \sum_{k=q}^{\infty} \frac{k^c}{(\gamma_n + k^a)^b} = \int_0^{\infty} \frac{u^c}{(1+u^a)^b} du$$

and

$$\max_k \frac{t_k}{(\gamma_n + s_k)^b} \propto \gamma_n^{-b+c/a}.$$

RESULT 3. *Suppose that $z_i \overset{i.i.d.}{\sim} N(0,1)$ and $\delta > 0$ satisfies $\delta ||\boldsymbol{c}||_\infty / ||\boldsymbol{c}||_2^2 < 1$ for any $\boldsymbol{c} = (c_1, c_2, \ldots) \in \ell^2$. Define $V = \sum_{k=1}^{\infty} c_k (z_i^2 - 1)$, then*

$$\mathbf{P}(|V| > \delta) < 2 \exp\left(-\delta^2 / 12 ||\boldsymbol{c}||_2^2\right).$$

Further, if $V_i \overset{i.i.d.}{\sim} V$, for δ satisfying $\delta ||\boldsymbol{c}||_1 ||\boldsymbol{c}||_\infty / ||\boldsymbol{c}||_2^2 < 1$,

$$\mathbf{P}(\max_{1\le i\le n} |V_i| > \delta ||\boldsymbol{c}||_1) \le 2 \exp\left(-||\boldsymbol{c}||_1^2 \delta^2 / 12 ||\boldsymbol{c}||_2^2 + \log n\right).$$

Result 4 relates the expectation of a random variable to its expectation taken over a subset. The proof is a simple application of Cauchy-Schwarz inequality, thus is omitted.

RESULT 4. *Suppose g is squared integrable and f is bounded by B on a set A_δ. Then we have that*

$$|\mathbf{E} fg - \mathbf{E} fg \mathbf{1}_{A_\delta^c}| \le B \mathbf{P}^{1/2}(A_\delta)(\mathbf{E} g^2)^{1/2}.$$

The following provides a general result on lower and upper bounds for the "ideal" risk of any block $B \subset \{1, \ldots, m\}$. The upper bound follows from Jeson's inequality and the lower bound is achieved by conditioning and the facts that $\mathbf{E} X^{-1} \ge (\mathbf{E} X)^{-1}$ for positive random variable X and $\mathbf{E} g(\chi_n^2) = n \mathbf{E}\{g(\chi_{n+2}^2)/\chi_{n+2}^2\}$ for Chi-square random variables. The proof is omitted for brevity.

RESULT 5. *For any block B (including singleton), denote the "ideal" block risk by $\mathcal{R}(B) = \mathbf{E}\left\{(||\boldsymbol{\theta}_B||_2^2 |B|/m)/(||\boldsymbol{\theta}_B||_2^2 + |B|/m)\right\}$, then*

$$\frac{||\lambda_B||_1 |B|/m}{2||\lambda_B||_\infty + ||\lambda_B||_1 + |B|/m} \le \mathcal{R}(B) \le \frac{||\lambda_B||_1 |B|/m}{||\lambda_B||_1 + |B|/m}.$$

3. Relation to Weak ℓ_p Spaces

The following result allows us to establish upper bounds on weak ℓ_p norms in terms of simpler maxima of independent variables.

RESULT 6. *Suppose that $x, y \in \mathbb{R}^n$ satisfy $x_1 \ge x_2 \ge \cdots \ge x_n \ge 0$ and $0 \le y_1 \le y_2 \le \cdots \le y_n$. Then for any permutation $\sigma : [n] \to [n]$, it holds that*

$$\max_{i\le n} x_i y_i \le \max_{i\le n} x_{\sigma(i)} y_i.$$

Proof. Given that any permutation may be written as a composition of disjoint cycles, it suffices to show that the maximum value of $x_{\sigma(i)}y_i$ dominates the maximum value of x_iy_i on any cycle. Now a cycle of σ of length r is a collection of points $i_1, \ldots, i_r$ satisfying $\sigma(i_j) = i_{j+1}$ for $j = 1, \ldots, r-1$ and $\sigma(i_r) = i_1$. Now let $\mathcal{S} = \{i_1, \ldots, i_r\}$ and note that $\mathcal{S}$ has minimum and maximum elements, say m, M. Further, set

$$\mathcal{D} = \{(i,j) \in \mathcal{S} : \sigma(i) = j < i\}$$

and observe that we must have $(i,m), (M,j) \in \mathcal{D}$ for some $i, j \in \mathcal{S}$. This seems obvious, but suppose not. Then there is no $j \in \mathcal{S}$ for which $\sigma(M) = j < M$, contradicting that $\mathcal{S}$ is a cycle of σ as we may not cycle back from M. Similarly, there is no $i \in \mathcal{S}$ for which $\sigma(i) = m < i$, again contradicting that $\mathcal{S}$ is a cycle of σ as we may not cycle back to m. Nevertheless, this doesn't come nearly as cleanly as above.

Now notice that since x is decreasing while y is increasing, for any $(i,j) \in \mathcal{D}$ we have that

$$x_{\sigma(i)}y_i \geq \max\{x_jy_j, x_{j+1}y_{j+1}, \ldots, x_iy_i\}.$$

Now we just need to show that $\mathcal{D}$ 'covers' $\{m, m+1, \ldots, M\}$ to get that

$$\max_{i \in \mathcal{S}} x_{\sigma(i)}y_i \geq \max_{m \leq i \leq M} x_iy_i.$$

But this is trivial, as we must return from M to m in order for the cycle property to be maintained. Thus there must be a sequence $(i_k, j_k) \in \mathcal{D}$, $k = 1, \ldots, K$ satisfying $i_1 = M$, $j_K = m$ and $j_{k+1} \geq i_k$ for $k = 1, \ldots, K-1$. If this were not the case, we would not be able to cycle back from some value. This in turn shows that each disjoint cycle of a permutation increases the maximum and establishes the result. □

Next result establishes a precise connection of the decaying condition to the weak ℓ_p spaces. Our model is $\theta_j \sim N(0, \lambda_j)$ with $\lambda_{(j)} \lesssim j^{-(2\alpha+1)}$ for $j \in [m]$ and $\lambda_j \lesssim j^{-(2\alpha+1)}$ for $j > m$. Thus there is some permutation $\sigma : [m] \to [m]$ so that $\lambda_{(j)} = \lambda_{\sigma(j)}$ for $j \in [m]$, and we find that

Result 7. *Let $\sigma : [m] \to [m]$ be the permutation taking $\lambda_{(j)} = \lambda_{\sigma(j)}$ for $j \in [m]$. Then for $\gamma \geq 1/2$ (relaxable) we have that*

$$\max_{i \leq m} \frac{i^\gamma |\theta|_{(i)}}{\sqrt{\log(1+i)}} \leq \max_{i \leq m} \frac{i^\gamma |\theta_{\sigma(i)}|}{\sqrt{\log(1+i)}} \sim \max_{i \leq m} \frac{i^\gamma \lambda_{(i)}^{1/2} z_i}{\sqrt{\log(1+i)}},$$

with $z_i \overset{i.i.d.}{\sim} N(0,1)$. Consequently we find that for $\gamma \leq \alpha + 1/2$ we have

$$\sup_{i \in \mathbb{N}} \frac{i^\gamma |\theta|_{(i)}}{\sqrt{\log(1+i)}} \overset{a.s.}{<} \infty$$

and so the random element θ lies just outside of the weak ℓ_p space $\gamma = \alpha + 1/2$ and in every one for $\gamma < \alpha + 1/2$.

Proof. The first claim follows as an application of the previous theorem combined with the fact that $y_i(\gamma) = i^\gamma / \sqrt{(}\log(1+i))$ is positive and increasing for $\gamma \geq 1/2$ while, by definition, $|\theta|_{(1)} \geq |\theta|_{(2)} \geq \cdots \geq |\theta|_{(m)} \geq 0$.

Now, from the first fact, we have that

$$\sup_{i\in\mathbb{N}} \frac{i^\gamma|\theta|_{(i)}}{\sqrt{\log(1+i)}} \le \sup_{i\in\mathbb{N}} \frac{i^\gamma \lambda_{(i)}^{1/2}|z_i|}{\sqrt{\log(1+i)}}.$$

The final claim of the theorem is now standard. See e.g. result of Durrett p 63 ex 2.3.13: For $X_1, X_2, \dots$ independent, $\sup_n X_n < \infty$ a.s. if and only if $\sum_n \mathbf{P}(X_n > A) < \infty$ for some A. □

4. Proofs of Main Lemmas

We provide the proofs of Lemma 1–3 presented in the paper.

Proof of Lemma 1. Denote $\mathbf{P}_i(\cdot) = \mathbf{P}(\cdot|\boldsymbol{\theta}_i)$ and $\mathbf{E}_i = \mathbf{E}(\cdot|\boldsymbol{\theta}_i)$, let

$$A^-_{k,\delta} = \{||\boldsymbol{Y}_{\cdot k}||^2 < (1-\delta)n(\lambda_k + 1/m)\}$$
$$A^+_{k,\delta} = \{||\boldsymbol{Y}_{\cdot k}||^2 > (1+\delta)n(\lambda_k + 1/m)\}.$$

Since $Y_{\cdot k}$ is Gaussian with independent elements, for any $s > 0$,

$$\begin{aligned}
\mathbf{P}_i(A^+_{k,\delta}) &\le \exp\{-s(1+\delta)n(\lambda_k+1/m)\}\mathbf{E}_i \exp(s||Y_{\cdot k}||^2)\\
&= \exp\{-s(1+\delta)n(\lambda_k+1/m)\}\\
&\quad \times \exp(s\theta_{i,k}^2)(1-2s/m)^{-1/2}\{1-2s(\lambda_k+1/m)\}^{-(n-1)/2}\\
&\le \exp\{-s(1+\delta)n(\lambda_k+1/m)\}\\
&\quad \times \exp(s\theta_{i,k}^2)\{1-2s(\lambda_k+1/m)\}^{-n/2}.
\end{aligned}$$

On choosing s to satisfy $\delta = 2(1+\delta)(\lambda_k+1/m)s$ and $s \le \delta/2\lambda_k$, we find

$$\mathbf{P}_i(A^+_{k,\delta}) \le \exp(\delta\theta_{i,k}^2/2\lambda_k)\{(1+\delta)\exp(-\delta)\}^{n/2}.$$

Taylor approximation to $\log(1+\delta)$ gives $\{(1+\delta)\exp(-\delta)\}^{n/2} \le \exp(-n\delta^2/6)$,

$$\mathbf{P}_i(A^+_{k,\delta}) \le \exp(\delta\theta_{i,k}^2/2\lambda_k)\exp(-n\delta^2/6).$$

Similarly, we calculate that

$$\begin{aligned}
\mathbf{P}_i(A^-_{k,\delta}) &\le \exp\{s(1-\delta)n(\lambda_k+1/m)\}\mathbf{E}_i\exp(-s||Y_{\cdot k}||^2)\\
&\le \exp\{s(1-\delta)n(\lambda_k+1/m)\}\{1+2s(\lambda_k+1/m)\}^{-(n-1)/2}\\
&\le (1-\delta)^{-1/2}\{(1-\delta)\exp(\delta)\}^{n/2},
\end{aligned}$$

by choosing s to satisfy $\delta = 2(1-\delta)(\lambda_k+1/m)s$. Since $\delta \in (0, 1/2)$, $(1-\delta)^{-1/2} \le \sqrt{2}$, while Taylor approximation to $\log(1-\delta)$ gives $\{(1-\delta)\exp(\delta)\}^{n/2} \le \exp(-n\delta^2/4) \le \exp(\delta\theta_{i,k}^2/2\lambda_k)\exp(-n\delta^2/6)$. Combining $\mathbf{P}_i(A^-_{k,\delta})$ and $\mathbf{P}_i(A^+_{k,\delta})$ gives the first result and application of a union bound gives the second. □

To present the proofs for results on equivalence for fixed and random design we need to setup some background and notations on piecewise constant approximations. For a given integer k, let $I_{k,j} = [j/k, (j+1)/k)$ for $j = 0, \dots, k-2$ and $I_{k,k-1} = [1-1/k, 1]$ so that the $I_{k,j}$ form a partition of $[0,1]$ with the measure of each interval, $|I_{k,j}|$, constant at k^{-1}. Let $\phi_{k,j} = k^{1/2}\mathbf{1}_{I_{k,j}}$ so that for a given k these functions form an orthonormal basis for the subspace S_k of $L^2[0,1]$ consisting of functions constant on each of the $I_{k,j}$. Any $f \in L^2[0,1]$ has a projection $P_k f$ onto S_k given by

$$P_k f = \sum_{j=0}^{k-1} \theta_{k,j}\phi_{k,j},$$

where $k^{1/2}\theta_{k,l} = k^{1/2}\langle f, \phi_{k,j}\rangle = |I_{k,j}|^{-1}\int_{I_{k,j}} f$ is the average of f over $I_{k,j}$. Thus $P_k f$ has the interpretation of using the average of f on each $I_{k,j}$ to represent f on that interval and intuition suggests that for a given k this approximation improves as f becomes smoother. Further, for any $J \in \mathbb{N}$ we may expand f in the Haar wavelet expansion

$$f = P_{2^J} f + \sum_{l \geq J}(P_{2^{l+1}} f - P_{2^l} f) = P_{2^J} f + \sum_{l \geq J}\sum_{j=0}^{2^l-1} w_{l,j}(f) H_{l,j}$$

where the $H_{l,j} = 2^{-1/2}(\phi_{2^{l+1},2j} - \phi_{2^{l+1},2j+1})$ form the orthonormal Haar wavelet basis and $w_{l,j}(f) = \langle f, H_{l,j}\rangle$. Setting $w_{0,0}(f) = \langle f, \phi_{1,0}\rangle = \int_0^1 f$, these representations lead to a useful scale of norms which can be used to characterize equivalence for both fixed and random design. For a given $\alpha \geq 1/2$ we define the scale of norms

$$||f||_{(\alpha)} = \left\{\sum_{k=0}^{\infty} 2^{2k\alpha}\sum_{l=0}^{2^k-1} w_{k,l}^2(f)\right\}^{1/2},$$

which are closely related to a specific instance of Besov norms and are easily seen to provide a generalization of various types of smoothness, e.g. Hölder continuity, Sobolev smoothness, in that the norms charachterizing these types of smoothness, for a given $\alpha > 1/2$, may be seen to bound the corresponding $||f||_{(\alpha)}$ norm.

Proof of Lemma 2. From Theorem 2 of Brown *et al.* (2002) we have the bound

$$\Delta(\mathbb{R}_m, \mathbb{W}_m)) \leq 2 \sup_{f \in \Theta_m} D_m(f)$$

where

$$D_m(f) \leq \frac{3}{2}||f - P_{2^J} f||_{(1/2)}^2 + \frac{2^{2J+1}}{m}.$$

Given the assumption that for $\alpha > 1/2$, $||f||_{(\alpha)}^2 < B_m^2$ we calculate that

$$||f - P_{2^J} f||_{(1/2)}^2 = \sum_{l \geq J} 2^l \sum_{j=0}^{2^l-1} w_{l,j}^2(f) \leq 2^{-J(2\alpha-1)} B_m^2.$$

This gives

$$D_m(f) \leq 2B_m^2\left\{2^{-J(2\alpha-1)} + \frac{2^{2J}}{m}\right\}$$

and choosing a sequence $J = J(m)$ to satisfy $2^J \propto m^{1/(2\alpha+1)}$ balances terms and yields the bound for $\Delta(\mathbb{R}_m, \mathbb{W}_m))$.

In Reiß (2008) a bound on Le Cam's distance between fixed design and the white noise model is derived which relies on the distance

$$||f - \mathcal{I}_m f||_2^2 \quad \text{where} \quad \mathcal{I}_m f = \sum_{j=0}^{m-1} f((j+1)/m)\mathbf{1}_{I_{m,j}}.$$

Here $\mathcal{I}_m f$ is a projection onto a design dependent interpolation space.

In what follows, we choose a sequence $J = J(m)$ to satisfy $m/2 \le 2^J < m$. First notice that if $|x-y| < 2^{-J}$ lie in the same dyadic interval at scale J, i.e. $x, y \in [k2^{-J}, (k+1)2^{-J})$ for some $k = 0, \ldots, 2^J - 1$ then we have that $P_{2^J} f(x) = P_{2^J} f(y)$ and so

$$f(x) - f(y) = \sum_{k \ge J} \sum_{l=0}^{2^k - 1} w_{k,l}(f)(H_{k,l}(x) - H_{k,l}(y)),$$

since $||f||_{(\alpha)} < \infty$ for $\alpha > 1/2$ guarantees that the wavelet representation converges uniformly. At each scale k, for each $x \in [0,1]$, $H_{k,l}(x)$ is non-zero for only one index $l = l_k(x)$. Further, noting that $||H_{k,l}||_\infty \le 2^{k/2}$ gives the bound

$$|f(x) - f(y)| \le \sum_{k \ge J} 2^{k/2}(|w_{k,l_k(x)}(f)| + |w_{k,l_k(y)}(f)|).$$

Applying Cauchy-Schwarz to the summand, we calculate that

$$\begin{aligned} |f(x) - f(y)| &\le \sqrt{2} \sum_{k \ge J} 2^{k/2}(w_{k,l_k(x)}^2(f) + w_{k,l_k(y)}^2(f))^{1/2} \\ &= \sqrt{2} \sum_{k \ge J} 2^{-(\alpha-1/2)k} 2^{\alpha k}(w_{k,l_k(x)}^2(f) + w_{k,l_k(y)}^2(f))^{1/2} \\ &\le \sqrt{2} 2^{-(\alpha-1/2)J} \left(\sum_{l \ge J} 2^{2\alpha k}(w_{k,l_k(x)}^2(f) + w_{k,l_k(y)}^2(f)) \right)^{1/2}. \end{aligned}$$

Now the condition $m/2 \le 2^J < m$ guarantees that at most one dyadic $k2^{-J}$ lies in any interval $[i/m, (i+1)/m]$. Let $I_{k,i}^m$ denote the set of indices $l = 0, \ldots, 2^k - 1$ so that $l2^{-k} \in [i/m, (i+1)/m]$. Then, by splitting the bound above where needed, for all x in this interval we arrive at the bound

$$(f(x) - f((i+1)/m))^2 \le 4 \cdot 2^{-(2\alpha-1)J} \sum_{k \ge J} 2^{2\alpha k} \sum_{l \in I_{k,i}^m} w_{k,l}^2(f).$$

This gives

$$\begin{aligned} \int_{i/m}^{(i+1)/m} (f(x) - f((i+1)/m))^2 dx &\le \frac{4 \cdot 2^{-(2\alpha-1)J}}{m} \sum_{k \ge J} 2^{2\alpha k} \sum_{l \in I_{k,i}^m} w_{k,l}^2(f) \\ &\le 8 \cdot 2^{2\alpha} \cdot m^{-2\alpha} \sum_{l \ge J} 2^{2\alpha k} \sum_{l \in I_{k,i}^m} w_{k,l}^2(f). \end{aligned}$$

Noting that $\cup_{i=0}^{m-1} I_{k,i}^m = \{0, \ldots, 2^k - 1\}$ and summing these integral bounds gives

$$||f - \mathcal{I}_m f||_2^2 \le C \cdot m^{-2\alpha} \sum_{k \ge J} 2^{2\alpha k} \sum_{l=0}^{2^k-1} w_{k,l}^2(f) \lesssim m^{-2\alpha} ||f||_{(\alpha)}^2.$$

From this the bounds of Reiß (2008) yield

$$\Delta(\mathbb{F}_m, \mathbb{W}_m) \lesssim m^{1/2-\alpha} \sup_{f \in \Theta_m} ||f||_{(\alpha)} \le m^{-(2\alpha-1)/2} B_m$$

and noting that for $\alpha > 1/2$, $2\alpha + 1 > 2$ and it is assumed that we take $1 \le B_m \to \infty$ concludes proof of the theorem. □

Proof of Lemma 3. We calculate that

$$\begin{aligned} w_{k,l}(f) &= 2^{k/2} \left(\int_{I_{k+1,2l}} f(x)dx - \int_{I_{k+1,2l+1}} f(y)dy \right) \\ &= 2^{1+3k/2} \int_{I_{k+1,2l}} \int_{I_{k+1,2l+1}} (f(x) - f(y))dxdy \\ &= 2^{1+3k/2} \int_{I_{k,l}} \int_{I_{k,l}} (f(x) - f(y)) \mathbf{1}_{I_{k+1,2l}}(x) \mathbf{1}_{I_{k+1,2l+1}}(y)dxdy. \end{aligned}$$

Now if f has a generalized derivative, we may write

$$f(x) - f(y) = \int_x^y f'(s)ds.$$

Then Cauchy-Schwarz gives that

$$w_{k,l}^2(f) \le 2^k \int_{I_{k,l}} \int_{I_{k,l}} \left(\int_x^y f'(s)ds \right)^2 \mathbf{1}_{I_{k+1,2l}}(x) \mathbf{1}_{I_{k+1,2l+1}}(y)dxdy,$$

Another application of Cauchy-Schwarz leads to the bound

$$w_{k,l}^2(f) \le 2^k \int_{I_{k,l}} |f'(s)|^2 ds \int_{I_{k,l}} \int_{I_{k,l}} (y - x) \mathbf{1}_{I_{k+1,2l}}(x) \mathbf{1}_{I_{k+1,2l+1}}(y)dxdy,$$

which gives

$$w_{k,l}^2(f) \le 2^{-2k} \int_{I_{k,l}} |f'(s)|^2 ds.$$

From this, we find that at each scale k,

$$\sum_{l=0}^{2^k-1} w_{k,l}^2(f) \le 2^{-2k} ||f'||_2^2$$

which, for $\alpha \in [1/2, 1)$ leads to the bound $||f||_{(\alpha)}^2 \le 2^{2(1-\alpha)} ||f'||_2^2 / (2^{2(1-\alpha)} - 1)$ and concludes the proof of the lemma. □

5. Proofs of Main Theorems and Propositions

In this section, we present the proofs of Theorem 1–7 and Preposition 1–3 presented in the paper.

PROOF OF THEOREM 1. For notational convenience, we replace $1/m$ with τ^2, and suppress the superscript in $\hat{\theta}_{ik}^{RS}$ when no confusion arises. We first establish the inequality in the case of $\lambda_k > \delta\tau^2$. Write

$$\hat{\theta}_{ik} = \alpha_{nm,k} Y_{ik} = \frac{\lambda_k}{\lambda_k + \tau^2} Y_{ik} + \left(\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right) Y_{ik}.$$

Using $Y_{ik} = \theta_{ik} + \tau z_{ik}$ with $z_{ik} \sim N(0,1)$ allows us to write $\lambda_k Y_{ik}/(\lambda_k + \tau^2) - \theta_{ik} = (\lambda_k \tau z_{ik} - \tau^2 \theta_{ik})/(\lambda_k + \tau^2)$, then

$$\begin{aligned}\mathbf{E}_i(\hat{\theta}_{ik} - \theta_{ik})^2 &= \mathcal{R}^*_{i,m}(k) + \underbrace{\mathbf{E}_i \left(\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right)^2 Y_{ik}^2}_{I} \\ &\quad + \underbrace{2\,\mathbf{E}_i \left(\frac{\lambda_k \tau z_{ik} - \tau^2 \theta_{ik}}{\lambda_k + \tau^2}\right)\left(\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right) Y_{ik}}_{II}.\end{aligned}$$

We proceed by bounding the terms I and II. For $\delta \in (0, 1/2)$, on the event $A_\delta^{m,c}$, the norm $||Y_{\cdot k}||$ satisfies the bounds $(1-\delta)n(\lambda_k + \tau^2) \le ||Y_{\cdot k}||^2 \le (1+\delta)n(\lambda_k + \tau^2)$ and

$$\left|\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right| \le \frac{3\delta}{1-\delta}\frac{\tau^2}{\lambda_k + \tau^2} = C_\delta \frac{\delta\tau^2}{\lambda_k + \tau^2},$$

where $C_\delta = 3/(1-\delta)$. Since both $\alpha_{n,k}$ and $\lambda_k/(\lambda_k + \tau^2)$ lie in the interval $(0,1)$, this quantity is always bounded by 2. The fact $\tau^2/(\lambda_k + \tau^2) \le 1$ gives

$$\begin{aligned}\mathbf{E}_i \left(\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right)^2 Y_{ik}^2 \mathbf{1}_{A_\delta^c} &\le C_\delta^2 \delta^2 \tau^2 \left(\frac{\theta_{ik}^2 + \tau^2}{\lambda_k + \tau^2}\right) \\ &\le C_\delta^2 \delta^2 \tau^2 \max(1, \theta_{ik}^2/\lambda_k).\end{aligned}$$

Further, $Y_{ik}^4 \le 8(\theta_{ik}^4 + \tau^4 z_{ik}^4)$ gives $(\mathbf{E}_i Y_{ik}^4)^{1/2} \le \{8(\theta_{ik}^4 + 3\tau^4)\}^{1/2} \le \sqrt{24}(\theta_{ik}^2 + \tau^2)$, while writing $\theta_{ik}^2 + \tau^2 = (\theta_{ik}^2/\lambda_k)\lambda_k + \tau^2$ yields $\theta_{ik}^2 + \tau^2 \le (\lambda_k + \tau^2)\max(1, \theta_{ik}^2/\lambda_k)$. In the range under consideration, $\delta\tau^2 = \min(\lambda_k, \delta\tau^2)$, thus an application of Lemma 4 leads to

$$I \le \max\left(1, \theta_{ik}^2/\lambda_k\right)\left\{C_\delta \delta \min(\lambda_k, \delta\tau^2) + \sqrt{24}\mathbf{P}_i^{1/2}(A_\delta^m)(\lambda_k + \tau^2)\right\}.$$

It remains to bound the term II. We begin by writing $(\lambda_k \tau z_{ik} - \tau^2\theta_{ik})Y_{ik} = (\lambda_k \tau z_{ik} - \tau^2 \theta_{ik})(\theta_{ik} + \tau z_{ik})$ and expand to have $(\lambda_k \tau z_{ik} - \tau^2\theta_{ik})Y_{ik} = \lambda_k \tau^2 z_{ik}^2 - \tau^2\theta_{ik}^2 + (\lambda_k\tau - \tau^3) z_{ik}\theta_{ik}$. Hence

$$II = \mathbf{E}_i \left(\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right)\left\{\frac{\lambda_k \tau^2 z_{ik}^2 - \tau^2\theta_{ik}^2 + (\lambda_k \tau - \tau^3) z_{ik}\theta_{ik}}{\lambda_k + \tau^2}\right\}.$$

Pass the expectation through and bound this quantity term by term. For the first term, noting that $\mathbf{E}_i z_i^2 \mathbf{1}_{A_\delta^m} \le \mathbf{E}_i z_i^2 = 1$,

$$\mathbf{E}_i \left(\alpha_{nm,k} - \frac{\lambda_k}{\lambda_k + \tau^2}\right)\frac{\lambda_k \tau^2 z_{ik}^2}{\lambda_k + \tau^2}\mathbf{1}_{A_\delta^m} \le C_\delta \frac{\delta \lambda_k \tau^4}{(\lambda_k + \tau^2)^2} \le C_\delta \frac{\delta \lambda_k \tau^2}{\lambda_k + \tau^2}.$$

Similarly, for the second term,

$$-\frac{\tau^2\theta_{ik}^2}{\lambda_k+\tau^2}\mathbf{E}_i\left(\alpha_{nm,k}-\frac{\lambda_k}{\lambda_k+\tau^2}\right)\mathbf{1}_{A_\delta^m}\le C_\delta\frac{\delta\theta_{ik}^2\tau^4}{(\lambda_k+\tau^2)^2}\le C_\delta\left(\frac{\theta_{ik}^2}{\lambda_k}\right)\frac{\delta\lambda_k\tau^2}{\lambda_k+\tau^2}.$$

Finally we write

$$\mathbf{E}_i\left(\alpha_{nm,k}-\frac{\lambda_k}{\lambda_k+\tau^2}\right)z_{ik}\mathbf{1}_{A_\delta^m}=\mathbf{E}_i\left(\alpha_{nm,k}-\frac{\lambda_k}{\lambda_k+\tau^2}\right)z_{ik}(\mathbf{1}_{z_{ik}<0}+\mathbf{1}_{z_{ik}\ge0})\mathbf{1}_{A_\delta},$$

and use the fact that $\mathbf{E}_i z_{ik}\mathbf{1}_{z_{ik}\ge0}\mathbf{1}_{A_\delta^m}$ and $-\mathbf{E}_i z_{ik}\mathbf{1}_{z_{ik}<0}\mathbf{1}_{A_\delta^m}$ are both bounded by $\mathbf{E}_i z_{ik}\mathbf{1}_{z_{ik}\ge0}=(2\pi)^{-1/2}$ while $2(2\pi)^{-1/2}\le1$ to arrive at

$$\mathbf{E}_i\left(\alpha_{nm,k}-\frac{\lambda_k}{\lambda_k+\tau^2}\right)z_{ik}\mathbf{1}_{A_\delta^m}\le C_\delta\frac{\delta\tau^2}{\lambda_k+\tau^2}.$$

By analogous argument, we can reach a lower bound of $-3\delta\tau^2/(\lambda_k+\tau^2)$. Noting the fact, if a,b are arbitrary numbers with $|b|\le B$ and c and d are positive numbers, then $|a\cdot b\cdot(c-d)|=|a|\cdot|b|\cdot\{\max(c,d)-\min(c,d)\}\le|a|\cdot B\cdot\max(c,d)$. Using this, we find that

$$\frac{(\lambda_k\tau-\tau^3)\theta_{ik}}{\lambda_k+\tau^2}\mathbf{E}_i\left(\alpha_{nm,k}-\frac{\lambda_k}{\lambda_k+\tau^2}\right)z_{ik}\mathbf{1}_{A_\delta^m}\le C_\delta\frac{\delta|\theta_{ik}|\tau^3\max(\lambda_k,\tau^2)}{(\lambda_k+\tau^2)^2}.$$

For any $\alpha\in(0,2)$, since $ab\le(a^2+b^2)/2$, we have

$$2\delta^{1-\alpha/2}\tau\cdot\frac{\delta^{\alpha/2}|\theta_{ik}|\tau^2}{\lambda_k+\tau^2}\le\delta^{2-\alpha}\tau^2+\delta^\alpha\frac{\theta_{ik}^2\tau^4}{(\lambda_k+\tau^2)^2}\le\delta^{2-\alpha}\tau^2+\delta^\alpha\left(\frac{\theta_{ik}^2}{\lambda_k}\right)\frac{\lambda_k\tau^2}{\lambda_k+\tau^2}.$$

We observe that $(\lambda_k\tau z_{ik}-\tau^2\theta_{ik})^2\le2(\lambda_k^2\tau^2z_{ik}^2+\tau^4\theta_{ik}^2)$ and $Y_{ik}^2\le2(\tau^2z_{ik}^2+\theta_{ik}^2)$. Then expanding $(\lambda_k\tau z_{ik}-\tau^2\theta_{ik})^2Y_{ik}^2$ and noting that $\mathbf{E}_i z_{ik}^4=3$, we arrive at the bound $\mathbf{E}_i(\lambda_k\tau z_{ik}-\tau^2\theta_{ik})^2Y_{ik}^2\le12(\lambda_k^2\tau^2+\tau^4\theta_{ik}^2)(\tau^2+\theta_{ik}^2)\le12\lambda_k\tau^2\{\max(1,\theta_{ik}^2/\lambda_k)(\lambda_k+\tau^2)\}^2$, which gives the bound

$$\mathbf{E}_i\left(\frac{\lambda_k\tau z_{ii}-\tau^2\theta_{ik}}{\lambda_k+\tau^2}\right)^2Y_{ik}^2\le12\{\max(1,\theta_{ik}^2/\lambda_k)(\lambda_k+\tau^2)\}^2.$$

Applying Theorem 1 yields

$$\begin{aligned}2II\quad\le\quad&\max\left(1,\theta_{ik}^2/\lambda_k\right)\Bigg\{C_\delta(4\delta+\delta^\alpha)\frac{\lambda_k\tau^2}{\lambda_k+\tau^2}\\&\qquad+C_\delta\delta^{2-\alpha}\tau^2+4\sqrt{12}\mathbf{P}_i^{1/2}(A_\delta^m)(\lambda_k+\tau^2)\Bigg\}\end{aligned}$$

For $\delta\in(0,1/2)$, $\lambda_k\delta\tau^2/(\lambda_k+\tau^2)\le\lambda_k\delta\tau^2/(\lambda_k+\delta\tau^2)\le\min(\lambda_k,\delta\tau^2)$, while in the range under consideration, $\delta\tau^2=\min(\lambda_k,\delta\tau^2)$. Taking $\alpha=1$ and using that $\min(\lambda_k,\tau^2)\le\tau^2$, this reduces to

$$2II\le\max\left(1,\theta_{ik}^2/\lambda_k\right)\left\{6C_\delta\min(\lambda_k,\delta\tau^2)+4\sqrt{12}\mathbf{P}_i^{1/2}(A_\delta^m)\tau^2(\lambda_k+\tau^2)\right\}.$$

Combining bounds for terms I and II gives the bound

$$\begin{aligned}\mathbf{E}_i(\hat\theta_{ik}-\theta_{ik})^2\quad\le\quad&\mathcal{R}_{i,m}^*(k)+\max\left(1,\theta_{ik}^2/\lambda_k\right)\Big\{C'_\delta\min(\lambda_k,\delta\tau^2)\\&\qquad+C\mathbf{P}_i^{1/2}(A_\delta^m)(\lambda_k+\tau^2),\Big\}\end{aligned}$$

where $C'_\delta = (6+\delta)C_\delta$ and $C = (\sqrt{24} + 4\sqrt{12})$, which provides the bound in (15) for the case of $\lambda_k > \delta\tau^2$.

For the case of $\lambda_k \le \delta\tau^2$, $\min(\lambda_k, \delta\tau^2) = \lambda_k$, we have $\alpha_{nm,k} = 0$ on the event A_δ^m which implies that $\mathbf{E}_i(\hat\theta_{ik} - \theta_{ik})^2 \mathbf{1}_{A_\delta^m} \le \theta_{ik}^2$. We also have $\mathbf{E}_i(\hat\theta_{ik} - \theta_{ik})^4 \le 4\mathbf{E}_i(\theta_{ik}^2 + \tau^2 z_{ik}^2)^2 \le 24(\theta_{ik}^4 + \tau^4)$, together with Lemma 4,

$$\begin{aligned}\mathbf{E}_i(\hat\theta_{ik} - \theta_{ik})^2 &\le \theta_{ik}^2 + \sqrt{24}\mathbf{P}_i^{1/2}(A_\delta^m)(\theta_{ik}^2 + \tau^2) \\ &\le \max\left(1, \theta_{ik}^2/\lambda_k\right)\left\{\min(\lambda_k, \delta\tau^2) + \sqrt{24}\mathbf{P}_i^{1/2}(A_\delta^m)(\lambda_k + \tau^2)\right\}\end{aligned}$$

which implies the bound in the second range and completes the proof. □

We state and prove an expanded version of Theorem 2, then Theorem 2 in the paper follows immediately.

Expanded Theorem 2. *Consider multiple GSMs (4) with the decay assumption (6), and suppose that $n, m \to \infty$ with $m^{\gamma_1} \lesssim n \lesssim m^{\gamma_2}$ for any $\gamma_2 \ge \gamma_1 > 0$. Then the terms bounding the conditional risks $\mathcal{R}_{i,m}(\hat{\boldsymbol\theta}_i^{RS}) = \mathbf{E}_{\boldsymbol\theta_i}||\boldsymbol\theta_i - \hat{\boldsymbol\theta}_i^{RS}||_{\ell_2}^2$ satisfy the following inequalities for*

$$\delta = \sqrt{4(\kappa+1)\log m/n} \tag{S2}$$

on a set of probability at least $\{1 - O(n^{-2})\}$, where C_δ and C are constants as in Theorem 1, and $\mathcal{R}_{i,m}^*$ are the conditional oracle risk.

(i)

$$\max_{i\le n} \mathcal{R}_{i,m}^* = \{1 + o(1)\}\sum_{k=1}^\infty \frac{\lambda_k/m}{\lambda_k + 1/m} \propto m^{-2\alpha/(2\alpha+1)}.$$

(ii)

$$\max_{i\le n} \mathbf{P}_i^{1/2}(A_\delta^m) \le \exp\{o(1)\} m^{-\kappa/2}.$$

(iii)

$$\begin{aligned}\max_{i\le n} e_i &\le \log(nm)\left\{C_\delta \sum_{k=1}^\infty \min(\lambda_k, \delta/m)\right. \\ &\qquad \left. + C(1 + ||\boldsymbol\lambda||_{\ell_2}^2)\max_{i\le n}\mathbf{P}_i^{1/2}(A_\delta^m)\right\} \\ &\propto \log(mn)\left\{\delta^{2\alpha/(2\alpha+1)} m^{-2\alpha/(2\alpha+1)} + m^{-\kappa/2}\right\}.\end{aligned}$$

(iv)

$$\max_{i\le n}\sum_{k>m}\theta_{ik}^2 = \{1 + o(1)\}\sum_{k>m}\lambda_k \propto m^{-2\alpha}.$$

It follows that $m^{-\kappa/2}\log(mn) = o\{m^{-2\alpha/(2\alpha+1)}\}$ for $\kappa \ge 2$, and

$$\max_{i\le n} \mathbf{E}_{\boldsymbol\theta_i}||\boldsymbol\theta_i - \hat{\boldsymbol\theta}_i^{RS}||_{\ell_2}^2 = \{1 + o_{a.s.}(1)\}\sum_{k=1}^\infty \frac{\lambda_k/m}{\lambda_k + 1/m} \propto m^{-2\alpha/(2\alpha+1)},$$

which achieves the optimal oracle risk.

Proof of Expanded Theorem 2. In the sequel, denote $\|\cdot\|_1 \triangleq \|\cdot\|_{\ell_1} = \sum_k |x_k|$ for $x \in \ell^1$ and $\|\cdot\|_2 \triangleq \|\cdot\|_{\ell_2}$. Define $\zeta_{i,m}$ and $\eta_{i,m}$ as

$$\zeta_{i,m} = \sum_{k=1}^{\infty} \left(\frac{1/m}{\lambda_k + 1/m} \right)^2 (\theta_{ik}^2 - \lambda_k), \quad \eta_{i,m} = \sum_{k>m} (\theta_{ik}^2 - \lambda_k)$$

and introduce sequences $\boldsymbol{q}_\zeta = (q_{\zeta,1}, q_{\zeta,2}, \ldots), \boldsymbol{q}_\eta = (q_{\eta,1}, q_{\eta,2}, \ldots) \in \ell_2$ with

$$q_{\zeta,k} = \frac{\lambda_k/m^2}{(\lambda_k + 1/m)^2}, \quad q_{\eta,k} = \lambda_k \mathbf{1}(k > m).$$

Set $\gamma(\boldsymbol{q}.) = \|\boldsymbol{q}.\|_{\ell_1}^2 / \|\boldsymbol{q}.\|_{\ell_2}^2$, where $\|x\|_{\ell_1} = \sum_k |x_k|$ for $x \in \ell^1$, the decay condition (6) together with Lemma 2 guarantees $\gamma(\boldsymbol{q}_\zeta) \propto m^{1/(2\alpha+1)}$ and $\gamma(\boldsymbol{q}_\eta) \propto m$. Further, Lemma 2 gives $||\boldsymbol{q}_\zeta||_1 \propto m^{-2\alpha/(2\alpha+1)}$ and $||\boldsymbol{q}_\eta||_1 \propto m^{-2\alpha}$. Then, for sets D_{δ_ζ} and E_{δ_η} defined by

$$D_{\delta_\zeta} = \left\{ \max_{i \le n} |\zeta_{i,m}| > \delta_\zeta \cdot ||\boldsymbol{q}_\zeta||_1 \right\}, \quad E_{\delta_\eta} = \left\{ \max_{i \le n} |\eta_{i,m}| > \delta_\eta \cdot ||\boldsymbol{q}_\eta||_1 \right\},$$

we apply the concentration results of Lemma 3 to conclude that, for $\delta_\zeta \propto m^{-1/(2\alpha+1)} \log n$ and $\delta_\eta \propto m^{-1} \log n$, by adjusting constants, we have $\mathbf{P}(D_{\delta_\zeta}) \le 2n^{-p}$ and $\mathbf{P}(E_{\delta_\eta}) \le 2n^{-p}$ for any $p > 0$. Further, for $\gamma > \sqrt{2}$, we have

$$\mathbf{P} \left(\max_{i \le n,\ k \le m} |\lambda_k^{-1/2} \theta_{ik}| > \gamma \{\log(nm)\}^{1/2} \right\} \le (nm)^{1-\gamma^2/2}.$$

Employing a union bound and choosing constants so that $p \ge 2$, then we have the the following bounds hold simultaneously, with probability at least $\{1 - 4n^{-2} - (nm)^{1-\gamma^2/2}\}$ that reduces to $\{1 - O(n^{-2})\}$ when $\gamma \ge \sqrt{6}$,

$$\max_{i \le n} |\zeta_{i,m}| \le \delta_\zeta \cdot ||\boldsymbol{q}_\zeta||_1, \quad \max_{i \le n} |\eta_{i,m}| \le \delta_\eta \cdot ||\boldsymbol{q}_\eta||_1 \tag{S3}$$
$$\text{and} \quad \max_{i \le n, k \le m} |\lambda_k^{-1/2} \theta_{ik}| \le \gamma \{\log(nm)\}^{1/2}.$$

From Lemma 2, $\sum_{k=1}^{\infty} \lambda_k/m/(\lambda_k + 1/m) \propto ||\boldsymbol{q}_\zeta||_1 \propto m^{-2\alpha/(2\alpha+1)}$ and the conditions on $m, n \to \infty$ implies $\delta_\zeta \propto m^{-1/(2\alpha+1)} \log n = o(1)$, which yields the assertion (i).

One can show that, with the sets $A_{k,\delta}^c$ (respectively, $A_\delta^{m,c}$) amended to

$$A_{k,\delta}^c = \{(1-\delta) n(\lambda_k + 1/m) \le ||Y_{\cdot k}||^2 \le n(\lambda_k + 1/m)/(1-\delta)\},$$

the concentration factor in Theorem 1 can be improved to $\exp(-n\delta^2/4)$, and this change affects only constants in the proof of Theorem 1, as $(1+\delta)$ and $1/(1-\delta)$ are of the same order as $\delta \to 0$. Further, with this concentration factor and $\delta = (12 \log m/n)^{1/2}$, we have $\exp(-n\delta^2/4 + \log m) = m^{-2}$. Then, on the set considered, $\max_{i \le n} ||\boldsymbol{\theta}_i^2/\boldsymbol{\lambda}||_{m,\infty} \le \gamma^2 \log(nm)$ and $\delta \log(nm) = o(1)$, which proves the assertion (ii).

Assertion (iii) follows from applying the bounds in (S3) to Theorem 1 and assertion (iv) holds due to (S3) together with $\|\boldsymbol{q}_\eta\|_1 \propto m^{-2\alpha}$. Combining (i)–(iv) leads to an upper bound for the final conclusion. To show the lower bound, we see from the proof of Theorem 1 that $|e_{ik}|$ satisfy the bound in the theorem in the range $\lambda_k \ge \delta\tau^2$. Adding risks over this range leads to $\mathbf{E}_{\boldsymbol{\theta}_i}||\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i^{RS}||_{\ell_2}^2 \geq \sum_{k:\lambda_k \geq \delta\tau^2} \mathcal{R}^*_{i,m}(k) - \max_{i \leq n}|e_i|$. By what has been shown, the lower bound now follows by noting that

$$\sum_{k:\lambda_k \geq \delta\tau^2} \mathcal{R}^*_{i,m}(k) \geq \sum_{k:\lambda_k \geq \delta\tau^2} \frac{\lambda_k/m}{\lambda_k + 1/m} - \max_{i \leq n}|\zeta_{i,m}|,$$

combined with the fact that for $\delta = o(1)$,

$$\sum_{k:\lambda_k \geq \delta\tau^2} \frac{\lambda_k/m}{\lambda_k/m + 1/m} = \{1 + o(1)\} \sum_{k=1}^{\infty} \frac{\lambda_k/m}{\lambda_k + 1/m}. \qquad \square$$

Proof of Theorem 3. The seminal paper Beran and Dümbgen (1998) has shown that, for any blocking estimator $\hat{\theta}_{i,\mathcal{B}}$ (18), the risk satisfies

$$\mathcal{R}_i(\hat{\theta}_{i,\mathcal{B}}) \triangleq \mathbf{E}_{\boldsymbol{\theta}_i}||\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_{i,\mathcal{B}}||_2^2 = \underbrace{\sum_{k=1}^{K_m} \frac{||\boldsymbol{\theta}_{i,B_k}||_2^2|B_k|/m}{||\boldsymbol{\theta}_{i,B_k}||_2^2 + |B_k|/m}}_{I} + \underbrace{O\left(\frac{K_m^{1/2}||\boldsymbol{\theta}_i||_2}{m}\right)}_{II}.$$

Lemma 3 entails that

$$\mathbf{P}(|||\boldsymbol{\theta}_i||_2^2 - ||\boldsymbol{\lambda}||_1| > \delta||\boldsymbol{\lambda}||_1) \leq 2\exp(-c\delta^2)$$

where $c = ||\boldsymbol{\lambda}||_1^2/||\boldsymbol{\lambda}||_2^2 \propto (4\alpha+1)/4\alpha^2$ is constant and bounded, given the assumption on λ_k. This in turn implies that, under the condition that n grows polynomially in m, $\max_{i \leq n}||\boldsymbol{\theta}_i||_2 = O_{a.s.}(\sqrt{||\boldsymbol{\lambda}||_1 \log m})$. Then, under the condition $K_m = o\{m^{1/(\alpha+1/2)}/\log m\}$, the second term of the block risk satisfies $II = o_{a.s.}\{m^{-2\alpha/(2\alpha+1)}\}$.

For now we drop the subscript i and control the first term for a single realization. Let $\mathcal{B}$ denote the collection of blocks. First we discard $O(\log m)$ blocks are of size $< \log m$ to retain the collection of blocks $\mathcal{GB}$ composed of the remaining blocks. Notice that, with slight abuse of notation,

$$\mathcal{R}(\mathcal{B}) \triangleq \sum_{B \in \mathcal{B}} \frac{||\theta_B||_2^2|B|/m}{||\theta_B||_2^2 + |B|/m} \geq \sum_{B \in \mathcal{GB}} \frac{||\theta_B||_2^2|B|/m}{||\theta_B||_2^2 + |B|/m} = \mathcal{R}(\mathcal{GB}),$$

and all blocks $B \in \mathcal{GB}$ satisfy $|B| \geq \log m$. Thus we can split each $B \in \mathcal{GB}$ into a collection of blocks of size $\{\log m, \log m + 1, \ldots, 2\log m\}$ to form a refinement of $\mathcal{GB}$, say $\mathcal{GB}^*$. We see that these risks are decreasing over refinements, since they correspond to solutions of the same minimization problem with an increasing number of parameters (Tsybakov, 2009; Johnstone, 2015), i.e., $\mathcal{R}(\mathcal{GB}) \geq \mathcal{R}(\mathcal{GB}^*)$.

Define $Z = m\mathcal{R}(\mathcal{GB}^*)/2\log m$, and we can show that it is a self-bounded function, meaning that it does not change much if we vary the individual inputs, so that it satisfies desirable concentration properties. Precisely, for $k \in B$, where $B \in \mathcal{GB}^*$, set

$$Z_k = Z - \frac{m}{2\log m} \frac{\theta_k^2|B|/m}{||\theta_B||_2^2 + |B|/m}.$$

Then, given the condition on the block sizes in $\mathcal{GB}^*$, $|B|/2\log m \leq 1$ and consequently $0 \leq Z - Z_k \leq \theta_k^2/(||\theta_B||_2^2 + |B|/m) \leq 1$. Similarly, taking summation, we find the identity

$Z = \sum_{B\in\mathcal{GB}^*}\sum_{k\in B}(Z - Z_k)$, which verifies that Z is self-bounded (see Theorem 6.12 and the following discussion in Boucheron *et al.*, 2013). Then, for $\delta \in (0,1)$,

$$\mathbf{P}\{|\mathcal{R}(\mathcal{GB}^*) - \mathbf{E}\mathcal{R}(\mathcal{GB}^*)| > \delta\mathbf{E}\mathcal{R}(\mathcal{GB}^*)\} \le 2\exp\left\{-\frac{m\mathbf{E}\mathcal{R}(\mathcal{GB}^*)}{2\log m}\frac{\delta^2}{2+2\delta/3}\right\}.$$

From decreasing error under refinement, and Lemma 5,

$$\mathbf{E}\mathcal{R}(\mathcal{GB}^*) \ge \mathbf{E}\sum_{B\in\mathcal{GB}}\sum_{k\in B}\frac{\theta_k^2/m}{\theta_k^2+m} \ge \frac{1}{3}\sum_{B\in\mathcal{GB}}\sum_{k\in B}\frac{\lambda_k/m}{\lambda_k+m}.$$

Given that at most $O(\log m)$ blocks are of size $\le \log m$, we have

$$\mathbf{E}\mathcal{R}(\mathcal{GB}^*) \ge 6^{-1}[Cm^{1/(2\alpha+1)} - O\{(\log m)^2\}]/m \gtrsim m^{-2\alpha/(2\alpha+1)},$$

since $(\lambda_k/m)/(\lambda_k+1/m) \ge \min(\lambda_k, 1/m)/2$ and the decay assumption guarantees that the remaining λ_k satisfying $\lambda_k > 1/m$ are on the order of $\left[m^{1/(2\alpha+1)} - O\{(\log m)^2\}\right]$. This establishes that there exists $C > 0$ such that

$$\mathbf{P}\{|\mathcal{R}(\mathcal{GB}^*) - \mathbf{E}\mathcal{R}(\mathcal{GB}^*)| > \delta\mathbf{E}\mathcal{R}(\mathcal{GB}^*)\} \le 2\exp\left\{-C\frac{m^{1/(2\alpha+1)}\delta^2}{\log m}\right\},$$

which in turn guarantees that $\mathcal{R}(\mathcal{GB}^*) = \{1 + o_{a.s.}(1)\}\mathbf{E}\mathcal{R}(\mathcal{GB}^*)$. Further, a union bound based on $\delta \triangleq \delta_m \to 0$ such that the r.h.s. is summable yields that this holds simultaneously for n independent risks satisfying the growth constraints in m. Thus, letting $\mathcal{R}_i(\mathcal{GB}^*)$ denote the corresponding risk of the refinement of the i observation, eventually for $i = 1, \ldots, n$, we have $|\mathcal{R}_i(\mathcal{GB}^*) - \mathbf{E}\mathcal{R}_i(\mathcal{GB}^*)| \le \delta\mathbf{E}\mathcal{R}_i(\mathcal{GB}^*)$. Hence

$$\min_{i\le n}\mathcal{R}_i(\mathcal{G}) \ge \min_{i\le n}\mathcal{R}_i(\mathcal{GB}^*) \ge \{1 + o_{a.s.}(1)\}\mathbf{E}\mathcal{R}(\mathcal{GB}^*).$$

We consider the lower bounds of Lemma 5 for a given block B. First consider that when (we assume throughout w.l.o.g. that $\lambda_1 = ||\lambda_B||_\infty$, for simplicity)

$$\lambda_1 \le |B|^{1/2}\max\left(\frac{1}{|B|}\sum_{j>1}\lambda_j, 1/m\right)$$

we have that

$$2\lambda_1 + \sum_{j\in B}\lambda_j + |B|/m \le (1 + 2|B|^{-1/2})\left(\sum_{j\in B}\lambda_j + |B|/m\right).$$

This shows that for λ_1 in this range,

$$\frac{||\lambda_B||_1|B|/m}{2||\lambda_B||_\infty + ||\lambda_B||_1 + |B|/m} \ge \frac{|B|^{1/2}}{2+|B|^{1/2}}\frac{||\lambda_B||_1|B|/m}{||\lambda_B||_1 + |B|/m} \ge \frac{|B|^{1/2}}{2+|B|^{1/2}}\sum_{j\in B}\frac{\lambda_i/m}{\lambda_i + 1/m}.$$

Now noting that for positive a_i, b_i

$$\frac{a_1 + \cdots + a_n}{b_1 + \cdots + b_n} \ge \min_{i\le n}\frac{a_i}{b_i}$$

shows that

$$\frac{||\lambda_B||_1|B|/m}{2||\lambda_B||_\infty + ||\lambda_B||_1 + B/m} \geq \min\left(\frac{||\lambda_B||_\infty|B|/m}{3||\lambda_B||_\infty + 1/m}, \sum_{j>1}\frac{\lambda_j/m}{\lambda_j + 1/m}\right).$$

Further, when

$$\lambda_1 > |B|^{1/2}\max\left(\frac{1}{|B|}\sum_{j>1}\lambda_j, 1/m\right)$$

we have

$$\frac{||\lambda_B||_\infty|B|/m}{3||\lambda_B||_\infty + 1/m} \geq \frac{|B|^{3/2}}{3}\frac{\frac{1}{|B|}\sum_{j>1}\lambda_j/m}{\frac{1}{|B|-1}\sum_{j>1}\lambda_j + 1/m} = \frac{|B|^{1/2}}{3}\frac{(|B|-1)\sum_{j>1}\lambda_j/m}{\sum_{j>1}\lambda_j + (|B|-1)/m}.$$

Then given that

$$\frac{(|B|-1)\sum_{j>1}\lambda_j/m}{\sum_{j>1}\lambda_j + (|B|-1)/m} \geq \sum_{j>1}\frac{\lambda_j/m}{\lambda_j + 1/m},$$

we find that in this range we have

$$\frac{||\lambda_B||_1|B|/m}{2||\lambda_B||_\infty + ||\lambda_B||_1 + B/m} \geq \min\left(1, \frac{|B|^{1/2}}{3}\right)\sum_{j>1}\frac{\lambda_j/m}{\lambda_j + 1/m}.$$

Now note that the decay assumptions imply there are at most $\sim (m/|B|^{1/2})^{1/(2\alpha+1)}$ blocks for which it holds that $\lambda_1 > |B|^{1/2}\max\left(\frac{1}{|B|}\sum_{j>1}\lambda_j, 1/m\right)$. From this, it follows that the risk $\mathbf{E}\mathcal{R}(\mathcal{GB}^*)$ has a lower bound of

$$\min_{B\in\mathcal{GB}^*}\min\left(\frac{|B|^{1/2}}{|B|^{1/2}+2}, \min\left(1, \frac{|B|^{1/2}}{3}\right)\right)\sum_{B\in\mathcal{GB}^*}\sum_{j\in B}\frac{\lambda_j/m}{\lambda_j + 1/m} - C\frac{(m/|B|^{1/2})^{1/(2\alpha+1)}}{m}.$$

Here the term we subtract off comes from the at most $\sim (m/|B|^{1/2})^{1/(2\alpha+1)}$ blocks for which $||\lambda_B||_\infty/m(||\lambda_B||_\infty + 1/m) \leq 1/m$ was both added and subtracted to arrive at the factor of $\sum_B\sum_{j\in B}\frac{\lambda_j/m}{\lambda_j+1/m}$. Now note that $(m/|B|^{1/2})^{1/(2\alpha+1)}/m = m^{-2\alpha/(2\alpha+1)}/|B|^{1/(4\alpha+2)} = o(m^{-2\alpha/(2\alpha+1)})$ since $|B| \asymp \log m$. Then as the decay $m^{-2\alpha/(2\alpha+1)} = O(\sum_B\sum_{j\in B}\lambda_j/(1+m\lambda_j))$ while for block sizes $|B| \geq \log m$,

$$\min\left(\frac{|B|^{1/2}}{|B|^{1/2}+2}, \min\left(1, \frac{|B|^{1/2}}{3}\right)\right) \geq \min\left(\frac{(\log m)^{1/2}}{(\log m)^{1/2}+2}, \min\left(1, \frac{(\log m)^{1/2}}{3}\right)\right) \sim 1,$$

we see that

$$\mathbf{E}\mathcal{R}(\mathcal{GB}^*) \geq (1+o(1))\sum_{B\in\mathcal{GB}^*}\sum_{j\in B}\frac{\lambda_j/m}{\lambda_j + 1/m}.$$

As we throw away at most $O(\log m)$ blocks of size $\lesssim \log m$ to arrive at $\mathcal{GB}^*$, the risk unaccounted for in $\mathcal{GB}^*$ is at most $O((\log m)^2/m)$ so that

$$\sum_{B\in\mathcal{GB}^*}\sum_{j\in B}\frac{\lambda_j/m}{\lambda_j + 1/m} = (1+o(1))\sum_{j=1}^{m}\frac{\lambda_j/m}{\lambda_j + 1/m} = (1+o(1))\sum_{j=1}^{\infty}\frac{\lambda_j/m}{\lambda_j + 1/m}.$$

Thus combining the above, we arrive at

$$\min_{i\le n}\mathcal{R}_i(\mathcal{G})\ge(1+o_{a.s.}(1))\sum_{j=1}^{\infty}\frac{\lambda_j/m}{\lambda_j+1/m}.$$

To show that the risk can be much worse for any realization $\boldsymbol{\theta}_i$, the condition $K_m = o\{m^{1/(2\alpha+1)}/\log m\}$ implies that there must be some block of size $|B|\gtrsim m^{2\alpha/(\alpha+1)}\log m$, otherwise we would have $o(m)$ effects. Fix $\gamma>0$ such that $\mathbf{P}(|Z|\le\gamma)=\exp(-1)$ for $Z\sim N(0,1)$. Drop the subscript i,

$$\mathbf{P}\left\{\max_{k\le\log K_m}\theta_{(k)}^2/\lambda_{(k)}\le\gamma\right\}=\exp(-\log K_m)=1/K_m.$$

Thus, with probability $(1-1/K_m)$, at least one of these coefficients satisfies

$$\theta_{(k)}^2>\gamma\lambda_{(k)}>Ck^{-(2\alpha+1)}\gtrsim(\log K_m)^{-(2\alpha+1)}\ge(\log m)^{-(2\alpha+1)}.$$

Now suppose the permutation places the coefficients corresponding to the largest $\log K_m$ coefficients in $|B|$. Eventually, for the block under consideration, we have $|B|/m\le\theta_{(k)}^2$ and when this happens,

$$\min(\theta_{(k)}^2,|B|/m)\gtrsim\min(\theta_{(k)}^2,m^{-1/(2\alpha+1)}\log m)\ge m^{-1/(2\alpha+1)}\log m.$$

Using the characterization of $\mathcal{R}(\hat\theta_{\mathcal{B}})$ from the first part of the proof, we have that, on the event under consideration,

$$\mathcal{R}(\hat\theta_{\mathcal{B}})+o_{a.s.}\{m^{-2\alpha/(2\alpha+1)}\}\ge\frac{||\boldsymbol{\theta}_B||_2^2|B|/m}{||\boldsymbol{\theta}_B||_2^2+|B|/m}\ge\min(\theta_{(k)}^2,|B|/m)/2.$$

Given the constraint on α, $1\le2\alpha$. This establishes that, with probability at least $(1-1/K_m)$,

$$\mathcal{R}(\hat\theta_{\mathcal{B}})\ge\{1+o_{a.s.}(1)\}m^{-1/(\alpha+1)}\log m\gg m^{-2\alpha/(2\alpha+1)}.\qquad\square$$

Proof of Theorem 4. (i) A standard argument using Borel-Cantelli Lemma leads to $\max_{i\le n}\|\boldsymbol{\theta}_i^2/\boldsymbol{\lambda}\|_{m,\infty}=\{1+o_{a.s.}(1)\}2\log(nm)$. Recall the definitions of hyper-rectangles $A_{mn,k}(\lambda_k)$, $B_{mn,k}(\lambda_k)$, $A_{mn}(\boldsymbol{\lambda})$, $B_{mn}(\boldsymbol{\lambda})$ and $\boldsymbol{\Theta}_{mn}(\boldsymbol{\lambda})$ from the paper. Lemma 1 entails that $\mathbf{P}\big(\cup_{i\le n}\{\boldsymbol{\theta}_i\notin A_{mn,k}(\boldsymbol{\lambda})\}\big)\le(nm)^{1-a/2}$ and $\mathbf{P}\big(\cup_{i\le n}\{\boldsymbol{\theta}_i\notin B_{mn,k}(\lambda_k)\}\big)\le n(nk)^{1-b/2}$. When $a>(\gamma_1+2)/(\gamma_1+1)$ and $b>(2\gamma_1+3)/(\gamma_1+1)$, these events are summable, and Borel-Cantelli Lemma implies they occur only finitely often. As a consequence, eventually $\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_n\in\boldsymbol{\Theta}_{mn}(\boldsymbol{\lambda})$.

Next we prove the robustness guarantee (19) for recovering a new fixed $\boldsymbol{\theta}^*$. We first look at risk for a single frequency component, dropping the subscript k, and then lift the results. As in the proof of Theorem 1, replace $1/m$ by τ^2, i.e., $\hat\theta^*=\alpha Y$, where $Y\sim N(\theta^*,\tau^2)$, $|\theta^*|\le\{C\lambda\log(mn)\}^{1/2}$ and α is the Stein coefficient formed from the data. Further, α is independent of Y and enjoys the concentration results used in the proof of Theorem 1. We are interested in the risk $\mathcal{R}(\hat\theta^*)=\mathbf{E}_{\theta^*}(\alpha Y-\theta^*)^2$ that can be expanded as

$$\begin{aligned}\mathcal{R}(\hat\theta^*)&=\mathbf{E}_{\theta^*}(\alpha Y-\theta^*)^2=\mathbf{E}_{\theta^*}\left\{(\alpha-1)\,\theta^*+\alpha z\right\}^2\\&=\theta^{*2}\mathbf{E}\,(\alpha-1)^2+\tau^2\mathbf{E}\alpha^2.\end{aligned}$$

Consider the regime where $\lambda > \delta\tau^2$. Using the probability inequalities in Lemma 1 and Theorem 1, we have

$$\begin{aligned}\mathbf{E}\alpha^2 &\leq \left(\frac{\lambda}{\lambda+\tau^2}\right)^2 + \mathbf{P}(A_\delta^m)\\ \mathbf{E}(\alpha-1)^2 &\leq \left(\frac{1+2\delta}{1-\delta}\right)^2\left(\frac{\tau^2}{\lambda+\tau^2}\right)^2 + \mathbf{P}(A_\delta^m).\end{aligned}$$

Combining these inequalities and employing some algebra yield, for $\lambda \geq \delta\tau^2$,

$$\mathcal{R}(\hat{\theta}^*) \leq \left(\frac{1+2\delta}{1-\delta}\right)^2\left\{\frac{\lambda\tau^2}{\lambda+\tau^2} + \left(\frac{\tau^2}{\lambda+\tau^2}\right)^2(\theta^{*2}-\lambda)\right\} + \mathbf{P}(A_\delta^m)(\theta^{*2}+\tau^2).$$

Similarly, when $\lambda < \delta\tau^2$, we find that

$$\mathcal{R}(\hat{\theta}^*) \leq \{1+\mathbf{P}(A_\delta^m)\}\theta^{*2} + \mathbf{P}(A_\delta^m)\tau^2.$$

For $q \in \mathbb{N}$ and $x \in \ell_2$ let P_q denote the projection, $P_q x = (x_1, \ldots, x_q, 0, 0, \ldots)$, and $\boldsymbol{I}$ the identity $\boldsymbol{I}x = x$. Reintroducing indices k and taking summation, these inequalities provide an upper bound on $\mathcal{R}_m(\hat{\boldsymbol{\theta}}^{*RS})$ as follows, substituting $\tau^2 = 1/m$,

$$\begin{aligned}\left(\frac{1+2\delta}{1-\delta}\right)^2\left\{\sum_{k=1}^{\infty}\frac{\lambda_k/m}{\lambda_k+1/m} + \sum_{k=1}^{m}\left(\frac{1/m}{\lambda_k+1/m}\right)^2(\theta_k^{*2}-\lambda_k)\right\}\\ +(1+\|\boldsymbol{\theta}^*\|_2^2)\mathbf{P}(A_\delta^m) + \|(\boldsymbol{I}-P_m)\boldsymbol{\theta}^*\|_2^2.\end{aligned}$$

Distributing $\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}_{mn}(\boldsymbol{\lambda})}$ through the above expression and using integral approximations, we completes the proof by noting that all terms are $o[\mathcal{R}_m\{\boldsymbol{\Theta}_{mn}(\boldsymbol{\lambda})\}]$ with the exception of

$$\sup_{\boldsymbol{\theta}^*\in\boldsymbol{\Theta}_{mn}(\boldsymbol{\lambda})}\sum_{k=1}^{m}\left(\frac{1/m}{\lambda_k+1/m}\right)^2(\theta_k^{*2}-\lambda_k) \lesssim \log(nm)\sum_{k=1}^{m}\frac{\lambda_k/m}{\lambda_k+1/m}.$$

To see that this is also a lower bound, note first that for $x \in [0,1]$ we have

$$|\mathbf{E}x - \mathbf{E}x\mathbf{1}_A| = \mathbf{E}x\mathbf{1}_{A^c} \leq \mathbf{P}(A^c),$$

which gives $\mathbf{E}x \geq \mathbf{E}x\mathbf{1}_{A^c} - \mathbf{P}(A^c)$. In the situation where we apply our estimator to a given frequency (suppressing k) the risk of estimating θ, $\mathcal{R}(\theta)$, satisfies

$$\mathcal{R}(\theta) = \theta^2\mathbf{E}(1-\alpha)^2 + \tau^2\mathbf{E}\alpha^2 \geq \theta^2\mathbf{E}(1-\alpha)^2.$$

We may choose $\delta = o(1)$ and set A_δ, with $\mathbf{P}(A_\delta^c) \leq Cm^{-2}$ (with C independent of k), on which $(1-\delta)n(\lambda+\tau^2) \leq \|Y\|^2 \leq (1+\delta)n(\lambda+\tau^2)$. On A_δ we have $1 \geq 1-\alpha \geq (1-\delta)/m/(\lambda+1/m)$ which gives that

$$\mathbf{E}(1-\alpha)^2 \geq \left(\frac{1-\delta}{1+m\lambda}\right)^2 - \mathbf{P}(A_\delta^c) \geq \left(\frac{1-\delta}{1+m\lambda}\right)^2 - Cm^{-2}.$$

Hence adding a subset of frequencies and taking superium over Θ we find that

$$\sup_{\boldsymbol{\theta}\in\Theta}\mathcal{R}(\boldsymbol{\theta}) \gtrsim \sum_{\lambda_k<1/m,k<m}\left\{\left(\frac{1-\delta}{1+m\lambda_k}\right)^2 - m^{-2}\right\}\lambda_k\log(nm) \gtrsim m^{-2\alpha/(2\alpha+1)}\log(nm).$$

This follows since $\lambda_k < 1/m$ happens when $k > m^{1/(2\alpha+1)}$ and in this range $m\lambda_k \le 1$ so that $(1-\delta)/(1+m\lambda_k) \ge (1-\delta)/2$. This establishes that the logarithmic factor cannot be eliminated. □

Proof of Theorem 5. From the proof of Theorem 1, the derivation of the key oracle inequality relies on sets $A^c_{k,\delta}$ containing most of the probability mass. In the case of unknown variance these generalize to

$$A^c_{k,\delta} = \{(1-\delta)(\lambda_k + \sigma^2/m) \le ||\boldsymbol{Y}_k||^2/n \le (1+\delta)(\lambda_k + \sigma^2/m)\}.$$

Theorem 1 continues to hold in this setting with

$$\mathbf{P}_i(A^m_\delta) \le 3\exp(\delta \max_{i\le n} ||\boldsymbol{\theta}^2_i/\boldsymbol{\lambda}||_{m,\infty}) \exp(-n\delta^2/6 + \log m),$$

thus the conditional concentration of measure continues to guarantees that these sets capture "most" realizations. We now show that, for realizations in this range, a small amendment as in (20) guarantees that they remain close to the optimal linear factors for estimation of θ_{ik} from $Y_{\cdot k}$.

Recall the amended formula of $\alpha_{nm,k}(p)$ presented in the corresponding section of the paper, and denote $\tau^2 = 1/m$ and take $q_\delta = (1+2\delta)/(1-\delta) \gtrsim 1$. After some algebra, on the event $A^{m,c}_\delta$, we have $\alpha_{nm,k}(p) = 0$ when $\lambda_k \le \delta\tau^2/(1+\delta) + \kappa_m$, where $\kappa_m \propto Q^m_p(\boldsymbol{\lambda}_m) \propto \{(1-p)m\}^{-(2\alpha+1)}$. This means that, as in the proof of Theorem 1, we have $\alpha_{nm,k}(p) = 0$ when $\lambda_k \le \delta\tau^2/(1+\delta)$. Further, with this choice of q_δ, it holds on $A^{m,c}_\delta$ that

$$\left|\alpha_{n,k}(p) - \frac{\lambda_k}{\lambda_k + \tau^2}\right| \le \left\{\frac{q_\delta + 4 + (q_\delta + 2)\delta}{1-\delta}\right\} \frac{\delta\tau^2 + \kappa_m}{\lambda_k + \tau^2} = C_\delta \frac{\delta\tau^2 + \kappa_m}{\lambda_k + \tau^2}.$$

Adjusting the constants in the proof of Theorem 1 appropriately and using $(\delta + \tau^{-2}\kappa_m) \le 2(\delta^2 + \tau^{-4}\kappa^2_m)$, when $\lambda_k > \delta\tau^2/(1+\delta)$, similar arguments give the bounds, with I and II as in the proof of Theorem 1,

$$\begin{aligned} I &\le \max\left(1, \theta^2_{ik}/\lambda_k\right) \left[2C^2_\delta \left\{\delta(1+\delta)\min(\lambda_k, \delta\tau^2) + \tau^{-4}\kappa^2_m\right\}\right. \\ &\qquad \left. + 2\sqrt{24}\mathbf{P}^{1/2}_i(A_\delta)(\lambda_k + \tau^2)\right], \\ 2II &\le \max\left(1, \theta^2_{ik}/\lambda_k\right) \left[6C_\delta(1+\delta)\{\min(\lambda_k, \delta\tau^2) + \kappa_m\}\right. \\ &\qquad \left. + 4\sqrt{12}\mathbf{P}^{1/2}_i(A_\delta)(\lambda_k + \tau^2)\right]. \end{aligned}$$

Therefore, with slightly larger constants, the conclusion of Theorem 1 continues to hold with the e_{ik} adjusted up by an additive term

$$\kappa_m(1 + \tau^{-4}\kappa_m)\max\left(1, \theta^2_{ik}/\lambda_k\right) \lesssim \kappa_m \max\left(1, \theta^2_{ik}/\lambda_k\right).$$

Since $\kappa_m \propto m^{-(2\alpha+1)}$, multiplying m entails the additional risks of the order $m^{-2\alpha}\log(mn)$ that is negligible. Thus the results in Theorem 2 hold, similarly for the robustness guarantee in minimax sense of Theorem 4.

Revisiting the arguments, we see that the oracle inequality continues to hold, and thus the estimator will be adaptive for all $p \le p^*$, where p^* is the largest value such that the additive term $m\kappa_m \log(mn) \propto m\{(1-p)m\}^{-(2\alpha+1)}\log(mn) = o\{m^{-2\alpha/(2\alpha+1)}\}$. In other words, changing p below p^* does not change the rest of the oracle inequality which has been shown to attain the proper rate. Thus we have an adaptive estimator of σ^2, $\hat{\sigma}^2(p) = mQ_p^m(||\boldsymbol{Y}_m||^2/n)$, when $p \leq p^*$ and all estimators in this range are simultaneously adaptive to the oracle. □

Proof of Theorem 6. The proof follows the method deriving the examples in the paper. The condition on the norm guarantees that for a sample $f_1, \dots, f_n \overset{i.i.d.}{\sim} f$ we have

$$\max_{i \leq n} ||f|| \lesssim n^{1/\beta} \log n,$$

a.s. while the condition on β guarantees that $m^{-(2\alpha-1)/(2\alpha+1)}(n^{1/\beta}\log n)^2 = o(1)$. Then by theorem (2) we may form a parameter space, containing the f_i a.s., over which we have Le Cam equivalence in both the fixed and random design cases. This proves the theorem. □

Proof of Theorem 7. Part (iv) of the expanded Theorem 2, blown up by a $\log(n)$ factor, follows from a crude bound on the Hanson-wright inequality (Rudelson and Vershynin, 2013; Hanson and Wright, 1971). This is seen to control the tails of all quantities in the proof of expanded Theorem 2, i.e., sums over terms $> m$. The following result then extends the proof of the expanded Theorem 2 to the case where Conditions (A) in the paper are satisfied, by controlling the sums over terms $\leq m$, which implies Theorem 7. □

Supplementary Lemma 1. *Suppose that $\xi \sim N_m(0, \Sigma)$ and $Q = \xi^T D \xi$, where D is $m \times m$ symmetric and both matrices are positive definite and non-degenerate. Then we have that*

$$\mathbf{P}\left(|Q - \mathbf{Tr}(D\Sigma)| > \delta \mathbf{Tr}(D\Sigma)\right) \leq 2\exp\left(-c\delta^2 \frac{(\mathbf{Tr}(D\Sigma))^2}{||\Sigma^{1/2}D\Sigma^{1/2}||_{HS}^2}\right).$$

In particular, setting $D = diag(d_1, \dots, d_m)$ with $d_k = \tau^4/(\lambda_k + \tau^2)^2$, $\tau^2 = m^{-1}$, where $\lambda_k = \Sigma_{kk} = \mathbf{Cov}(\theta_i, \theta_i)$ and $\Sigma_{jk} = \mathbf{Cov}(\theta_k, \theta_k)$ satisfy the conditions (A), we find that

$$\mathbf{P}\left(|Q - \mathbf{Tr}(D\Sigma)| > \delta \mathbf{Tr}(D\Sigma)\right) \leq 2\exp\left(-c\frac{m^{1/(2r+2)}\delta^2}{B_m^2}\right).$$

Proof of Supplementary Lemma 1. The proof of the first inequality is a basic application of the Hanson-Wright inequality (see, e.g., Rudelson and Vershynin, 2013; Hanson and Wright, 1971) and will be ommitted for brevity. Setting $\Sigma^{1/2}D\Sigma^{1/2}$ and using that A is symmetric, we have $\mathbf{Tr}(A^2) = ||A||_{HS}^2$. Then owing to $\mathbf{Tr}(PQ) = \mathbf{Tr}(QP)$,

$$\begin{aligned}\mathbf{Tr}(A^2) &= \mathbf{Tr}(\Sigma^{1/2}D\Sigma D\Sigma^{1/2}) = \mathbf{Tr}(D^{1/2}\Sigma D\Sigma D^{1/2}) \\ &= \mathbf{Tr}((D^{1/2}\Sigma D^{1/2})^2) = ||D^{1/2}\Sigma D^{1/2}||_{HS}^2,\end{aligned}$$

thus $||\Sigma^{1/2}D\Sigma^{1/2}||_{HS}^2 = ||D^{1/2}\Sigma D^{1/2}||_{HS}^2$. Now let $\Delta = \mathrm{diag}(\Sigma)$ be the diagonal matrix formed from the diagonal elements of Σ. Then we may write

$$\mathbf{Tr}(A^2) = ||(\Delta D)^{1/2}\Delta^{-1/2}\Sigma\Delta^{-1/2}(\Delta D)^{1/2}||_{HS}^2.$$

Now using that $||PQ||_{HS}^2 \leq ||P||_{op}^2||Q||_{HS}^2$, we find that

$$\mathbf{Tr}(A^2) \leq ||(\Delta D)^{1/2}||_{op}^2||(\Delta D)^{1/2}||_{HS}^2||\Delta^{-1/2}\Sigma\Delta^{-1/2}||_{op}^2.$$

Now for the application under consideration, we have $D = \operatorname{diag}(d_1, \ldots, d_m)$ with $d_k = \tau^4/(\lambda_k + \tau^2)^2$ where $\lambda_k = \Sigma_{kk} = \langle \psi_k, C\psi_k \rangle$ and $\Sigma_{jk} = \langle \psi_j, C\psi_k \rangle$. Further, under Conditions (A), $\lambda_{(k)} \propto k^{-(2r+2)}$, Lemma (2) gives that $||(\Delta D)^{1/2}||_{op}^2 \lesssim \tau^2$ and $||(\Delta D)^{1/2}||_{HS}^2 \lesssim \tau^{2-1/(r+1)} \lesssim \mathbf{Tr}(D\Sigma)$. This implies that

$$\frac{(\mathbf{Tr}(D\Sigma))^2}{||(\Delta D)^{1/2}||_{op}^2 ||(\Delta D)^{1/2}||_{HS}^2} \lesssim \tau^{-1/(r+1)} \propto m^{1/(2r+2)}$$

and this leads to the concentration

$$\mathbf{P}\left(|Q - \mathbf{Tr}(D\Sigma)| > \delta \mathbf{Tr}(D\Sigma)\right) \leq 2\exp\left(-c\frac{m^{1/(2r+2)}\delta^2}{||\Gamma||_{op}^2}\right),$$

with $\Gamma = \Delta^{-1/2}\Sigma\Delta^{-1/2}$ denoting the correlation matrix, as in Condition (A). Now the Gershgorin disk theorem, together with the condition from condition (A) implies that

$$||\Gamma||_{op} \leq \max_{i \leq m} \sum_{i=1}^{m} |\Gamma_{ij}| \leq B_m$$

and this gives that

$$\mathbf{P}\left(|Q - \mathbf{Tr}(D\Sigma)| > \delta \mathbf{Tr}(D\Sigma)\right) \leq 2\exp\left(-c\frac{m^{1/(2r+2)}\delta^2}{B_m^2}\right)$$

which concludes the proof of this lemma and thus Theorem 7. □

PROOF OF THEOREM 8. It suffices to show the following two supplementary lemmas that extend the key concentration inequalities to the general situation $\sum_{k=1}^{\infty} \lambda_k < \infty$. These are easily seen to extend the proof of Theorem 2 under the Conditions (B) listed in Section 3.3, and the Theorem 8 follows. □

As was done there, we take

$$\mathcal{R} = \sum_{k=1}^{\infty} \frac{\lambda_k/m}{\lambda_k + 1/m}.$$

SUPPLEMENTARY LEMMA 2. *Suppose that $\xi \sim N_m(0, \Sigma)$ and $Q = \xi^T D\xi$, where D is $m \times m$ symmetric and both matrices are positive definite and non-degenerate. Then it is immediate from Hanson-Wright that for some $c > 0$*

$$\mathbf{P}\left(|Q - \mathbf{Tr}(D\Sigma)| > \delta\right) \leq 2\exp\left\{-c \min\left(\frac{\delta^2}{||\Sigma^{1/2}D\Sigma^{1/2}||_{HS}^2}, \frac{\delta}{||\Sigma^{1/2}D\Sigma^{1/2}||}\right)\right\}.$$

Set $D = \mathbf{diag}(d_1, \ldots, d_m)$ with $d_k = \tau^4/(\lambda_k + \tau^2)^2$, $\tau^2 = m^{-1}$ and $\lambda_k = \Sigma_{kk} = \mathbf{Cov}(\theta_i, \theta_i)$ and $\Sigma_{jk} = \mathbf{Cov}(\theta_k, \theta_k)$. Further, let $\Delta = \mathbf{diag}(\delta_1, \ldots, \delta_m)$ where $\delta_j = \lambda_j d_j$. Then with $\mathcal{R}$ as in the discussion above, this inequality yields

$$\mathbf{P}\left(|Q - \mathbf{Tr}(\Delta)| > \delta\mathcal{R}\right) \leq 2\exp\left\{-c\frac{\delta\mathcal{R}}{||\Delta||_\infty ||\Gamma_m||} \min\left(1, \frac{\delta}{||\Gamma_m||}\right)\right\}.$$

Thus, as long as $\delta \leq ||\Gamma_m||$, we have

$$\mathbf{P}\left(|Q - \mathbf{Tr}(\Delta)| > \delta\mathcal{R}\right) \leq 2\exp\left(-c\frac{\mathcal{R}\delta^2}{||\Delta||_\infty ||\Gamma_m||^2}\right).$$

Noting that $||\Delta||_\infty \leq 1/m$ gives the useful result $\mathbf{P}\left(|Q - \mathbf{Tr}(\Delta)| > \delta\mathcal{R}\right) \leq 2\exp\left(-c\frac{m\mathcal{R}\delta^2}{||\Gamma_m||^2}\right)$.

Proof of Supplementary Lemma 2. The first inequality is a direct application of the Hanson-Wright (see, e.g., Rudelson and Vershynin, 2013; Hanson and Wright, 1971). Setting $A = \Sigma^{1/2} D \Sigma^{1/2}$ and using that A is symmetric, we have $\mathbf{Tr}(A^2) = ||A||_{HS}^2$. Then owing to $\mathbf{Tr}(PQ) = \mathbf{Tr}(QP)$,

$$\begin{aligned} \mathbf{Tr}(A^2) &= \mathbf{Tr}(\Sigma^{1/2} D \Sigma D \Sigma^{1/2}) = \mathbf{Tr}(D^{1/2} \Sigma D \Sigma D^{1/2}) \\ &= \mathbf{Tr}((D^{1/2} \Sigma D^{1/2})^2) = ||D^{1/2} \Sigma D^{1/2}||_{HS}^2, \end{aligned}$$

thus $||\Sigma^{1/2} D \Sigma^{1/2}||_{HS}^2 = ||D^{1/2} \Sigma D^{1/2}||_{HS}^2$. Now using that $\Delta = D\mathbf{diag}(\Sigma)$ and $\Gamma = \mathbf{diag}(\Sigma)^{-1/2} \Sigma \mathbf{diag}(\Sigma)^{-1/2}$ we arrive at

$$\mathbf{Tr}(A^2) = ||]\Delta^{1/2} \Gamma \Delta^{1/2}||_{HS}^2.$$

since for diagonal matrices U and V, $(UV)^{1/2} = U^{1/2} V^{1/2}$. Now using that $||PQ||_{HS}^2 \leq ||P||^2 ||Q||_{HS}^2$, we find that

$$\mathbf{Tr}(A^2) \leq ||\Delta^{1/2}||^2 ||\Delta^{1/2}||_{HS}^2 ||\Gamma||^2.$$

Using that for a diagonal matrix U (with slight abuse of notation), $||U|| = ||U||_\infty = \max_i |u_{ii}|$ and $||U^{1/2}||_{HS}^2 = \sum_i |u_{ii}| = ||U||_1$ we arrive at $\mathbf{Tr}(A^2) \leq ||\Delta||_\infty ||\Delta||_1 ||\Gamma||^2$.

Similarly, setting $B = D^{1/2} \Sigma^{1/2}$ we have $||A|| = ||B'B||$. Then using that for any matrix C, $||C|| = ||C'||$ and $||C'C|| = ||C||^2$ (which gives $||C'C|| = ||CC'||$) together with $\Delta = D\mathbf{diag}(\Sigma)$ and $\Gamma = \mathbf{diag}(\Sigma)^{-1/2} \Sigma \mathbf{diag}(\Sigma)^{-1/2}$

$$||A|| = ||B'B|| = ||BB'|| = ||D^{1/2} \Sigma D^{1/2}|| = ||\Delta^{1/2} \Gamma \Delta^{1/2}|| \leq ||\Delta^{1/2}||^2 ||\Gamma|| = ||\Delta||_\infty ||\Gamma||.$$

Based on these two inequalities and the fact that $||\Delta||_1 = \sum_i \delta_i \leq \mathcal{R}$, we find

$$\min\left(\frac{\delta^2 \mathcal{R}^2}{||A||_{HS}^2}, \frac{\delta \mathcal{R}}{||A||}\right) \geq \min\left(\frac{\delta^2 \mathcal{R}^2}{||\Delta||_\infty ||\Delta||_1 ||\Gamma||^2}, \frac{\delta \mathcal{R}}{||\Delta||_\infty ||\Gamma||}\right) \geq \frac{\delta \mathcal{R}}{||\Delta||_\infty ||\Gamma||} \min\left(1, \frac{\delta}{||\Gamma||}\right).$$

Noting that for D diagonal $(D\Sigma)_{ii} = \sum_k D_{ik} \Sigma_{ki} = D_{ii} \Sigma_{ii}$ gives $\mathbf{Tr}(D\Sigma) = \mathbf{Tr}(\Delta)$ and we see the two bounds for $\mathbf{P}\left(|Q - \mathbf{Tr}(\Delta)| > \delta \mathcal{R}\right)$ now follow on applying the fact that for $c > 0$

$$\exp\left\{-c \min\left(\frac{\delta^2 \mathcal{R}^2}{||A||_{HS}^2}, \frac{\delta \mathcal{R}}{||A||}\right)\right\} \leq \exp\left\{-c \frac{\delta \mathcal{R}}{||\Delta||_\infty ||\Gamma||} \min\left(1, \frac{\delta}{||\Gamma||}\right)\right\}.$$

to the initial Hanson-Wright based inequality of the theorem. □

For the next lemma and its proof, we take $\xi_> = \xi_>(m) = (\underbrace{0, \ldots, 0}_{m}, \xi_{m+1}, \xi_{m+1}, \ldots)'$, which corresponds to the tail bias of our estimators.

Supplementary Lemma 3. *(Hanson-Wright for the tails) Let $Q_> = \xi_>^T \xi_>$ denote the tail bias and $\Sigma_>$ the covariance matrix corresponding to $\xi_>$. Then using $||\Sigma_>|| \leq ||\Sigma_>||_{HS}$ we find*

$$\mathbf{P}\left(|Q_> - \mathbf{Tr}(\Sigma_>)| > \delta \mathcal{R}\right) \leq 2 \exp\left\{-c \min\left(\frac{\delta^2 \mathcal{R}^2}{||\Sigma_>||_{HS}^2}, \frac{\delta \mathcal{R}}{||\Sigma_>||_{HS}}\right)\right\}.$$

From the final assumptions of the section, it follows that $Q_> = o_{a.s.}(\mathcal{R})$.

Proof of Supplementary Lemma 3. Noting that $||\Sigma_>|| \leq ||\Sigma_>||_{HS}$, we see the inequality is a direct application of the Hanson-Wright (see, e.g., Rudelson and Vershynin, 2013; Hanson and Wright, 1971), as above.

Further,

$$\min\left(\frac{\delta^2\mathcal{R}^2}{||\Sigma_>||_{HS}^2}, \frac{\delta\mathcal{R}}{||\Sigma_>||_{HS}}\right) = \frac{\delta\mathcal{R}}{||\Sigma_>||_{HS}} \min\left(1, \frac{\delta\mathcal{R}}{||\Sigma_>||_{HS}}\right).$$

Thus taking $\delta = 2||\Sigma_>||_{HS}\log m/\mathcal{R}$, we find that

$$\mathbf{P}\left(|Q_> - \mathbf{Tr}(\Sigma_>)| > 2||\Sigma_>||_{HS}\log m\right) \leq 2m^{-2}$$

and so with high probabilty (summable) we have $|Q_> - \mathbf{Tr}(\Sigma_>)| \leq 2||\Sigma_>||_{HS}\log m$. Now by Cauchy-Schwarz, $\Sigma_{jk}^2 = (\mathbf{E}\theta_j\theta_k)^2 \leq \mathbf{E}\theta_j^2\mathbf{E}\theta_k^2 = \lambda_j\lambda_k$. Thus, by definition of the Hilbert-Schmidt norm,

$$||\Sigma_>||_{HS}^2 = \sum_{j,k>m} \Sigma_{jk}^2 \leq \left(\sum_{k>m}\lambda_k\right)^2 = \left(\mathbf{Tr}(\Sigma_>)\right)^2.$$

Hence, by the assumptions of the section we find that $Q_> = O(\log m \sum_{k>m}\lambda_k) = o(\mathcal{R})$. □

Proof of Proposition 1. Let Ψ_{jk} represent the Meyer type wavelets on $\mathbb{R}$ that the ψ_{jk} are a periodization of. Further, with slight abuse of notation, we let $\widehat{K}$, $\widehat{\Psi}_{jk}$ denote the continuous Fourier transforms of K and Ψ_{jk}, while $\widehat{k}$ and $\widehat{\psi}_{jk}$ denote the discrete Fourier transforms of k and ψ_{jk}. Owing to the periodization, we have $\widehat{k}(n) = \widehat{K}(n)$ and $\widehat{\psi}_{jk}(n) = \widehat{\Psi}_{jk}(n)$ for $n \in \mathbb{Z}$. Due to stationarity, the covariance operator is convolution with k, then Parseval equality and standard properties of the Fourier transform, together with the fact that $\hat{\Psi}$ is real (the Ψ for which $\Psi_{jk}(\cdot) = 2^{j/2}\Psi(2^j\cdot - k)$), allow us to calculate that

$$\begin{aligned}\mathbf{Cov}(\theta_{jk}, \theta_{j'k'}) &= \langle\psi_{jk}, k * \psi_{j'k'}\rangle = \sum_{n\in\mathbb{Z}} \overline{\widehat{\Psi}}_{jk}(n)\widehat{K}(n)\widehat{\Psi}_{j'k'}(n) \\ &= 2^{-(j+j')/2}\sum_{n\in\mathbb{Z}} e^{2\pi i n(2^{-j}k - 2^{-j'}k')}\widehat{\Psi}(-2^{-j}n)\widehat{K}(n)\widehat{\Psi}(2^{-j'}n).\end{aligned}$$

The localization of $\hat{\Psi}$, $\widehat{\Psi}(-2^{-j}n)\widehat{\Psi}(2^{-j'}n) = 0$ for all $n \in \mathbb{Z}$ if $|j - j'| > 1$, proves (i).

Now assume $|j - j'| \leq 1$, let T_s denote the translation operator, $T_s f(x) = f(x - s)$, and D_a, $a > 0$, denote the dilation operator, $D_a f(x) = f(ax)$. Further, as defined in the paper,

$$|p - q|_j = \in \theta_{n\in\mathbb{Z}}|p - q + n2^j|.$$

First note that the decay conditions on $\widehat{K}$ imposed by Sobolev regularity guarantee that $\widehat{K}(n) = \widehat{K}(2^j 2^{-j}n) \propto 2^{-2(r+1)j}\widehat{K}(2^{-j}n)$. Further, properties of Fourier transform give that

$$(T_{k-2^{j-j'}k'}K)^\wedge(2^{-j}n) = e^{2\pi i n(2^{-j}k - 2^{-j'}k')}\widehat{K}(2^{-j}n),$$

and

$$\widehat{\Psi}(2^{-j'}n) = \widehat{\Psi}(2^{-j}\cdot 2^{-(j'-j)}n) = (D_{2^{j-j'}}\Psi)^\wedge(2^{-j}n).$$

Then as the product of Fourier tranforms is the Fourier transform of convolution, piecing these observations together and applying the Poisson summation formula gives

$$\mathbf{Cov}(\theta_{jk}, \theta_{j'k'}) \propto 2^{-2(r+1)j} \sum_{n \in \mathbb{Z}} \Psi * (T_{k-2^{j-j'}k'} K) * (D_{2^{j-j'}} \Psi)(n2^j).$$

Meyer-type wavelets are in the Schwartz space of functions, being infinitely differentiable and decaying (as do all derivatives) faster than any polynomial. In particular, we have assumed $|K(x)| \lesssim (1+|x|)^{-l}$, $l > 1$. Given that Ψ is Schwartz, we also have $|\Psi(x)|, |D_a\Psi(x)| \lesssim (1+|x|)^{-l}$. It is easy to verify that for two functions h, g satisfying this type of decay, one has $|h * g(x)| \lesssim (1+|x|)^{-l}$, from which we deduce that

$$\begin{aligned} \mathbf{Cov}(\theta_{jk}, \theta_{j'k'}) &\lesssim 2^{-2(r+1)j} \sum_{n \in \mathbb{Z}} \frac{1}{(1+|n2^j + k - 2^{j-j'}k'|)^l} \\ &= \frac{2^{-2(r+1)j}}{(1+|k-2^{j-j'}k'|_j)^l} \sum_{n \in \mathbb{Z}} \left(\frac{1+|k-2^{j-j'}k'|_j}{1+|n2^j + k - 2^{j-j'}k'|} \right)^l . \end{aligned}$$

Now as $c^j_{kk'} \triangleq k - 2^{j-j'}k' \in \{-2^j+1, \ldots, -1, 0, 1, \ldots, 2^j - 1\}$, we have that the infemum in the definition of $|k - 2^{j-j'}k'|_j = |c^j_{kk'}|_j$ is attained at one of $n^* = -1, 0, 1$ with n^* having the opposite sign of $c^j_{kk'}$. Further, when $|n| > 1$, by definition of $|\cdot|_j$, we have

$$\frac{1+|c^j_{kk'}|_j}{1+|n2^j + c^j_{kk'}|} = \frac{1+|n^*2^j + c^j_{kk'}|}{1+|n2^j + c^j_{kk'}|} \le \frac{1}{|n|}$$

from which we conclude that

$$\mathbf{Cov}(\theta_{jk}, \theta_{j'k'}) \lesssim \frac{2 \cdot 2^{-2(r+1)j}}{(1+|k-2^{j-j'}k'|_j)^l} \left(1 + \sum_{n>1} \frac{1}{n^l} \right) \lesssim \frac{2^{-2(r+1)j}}{(1+|k-2^{j-j'}k'|_j)^l}.$$

This proves (ii) and establishes the assertions regarding the correlations.

Given these results and with p, q fixed, we find that

$$\sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \frac{\mathbf{Cov}(\theta_{jk}, \theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk})\mathbf{Var}(\theta_{pq})}} \lesssim \sum_{j=\min(0,p-1)}^{p+1} \sum_{k=0}^{2^j-1} \frac{\mathbf{Cov}(\theta_{jk}, \theta_{pq})}{2^{-2(r+1)p}}$$

where

$$\sum_{k=0}^{2^j-1} \frac{\mathbf{Cov}(\theta_{jk}, \theta_{pq})}{2^{-2(r+1)p}} \lesssim \sum_{k=0}^{2^j-1} \frac{1}{(1+|k-2^{p-j}q|_p)^l}.$$

At each of the scales $j = p-1, p, p+1$, $|k - 2^{p-j}q|_p$ takes values in $0, 1, 2, \ldots, 2^p - 1$. As p, q are fixed, it can take each value at most twice as we cycle through k. This gives that

$$\sum_{k=0}^{2^j-1} \frac{1}{(1+|k-2^{p-j}q|_p)^l} \le 2 \sum_{m=0}^{2^p} \frac{1}{(1+m)^l} \le 2 + 2 \int_0^{\infty} \frac{dx}{(1+x)^l} \le \frac{2l}{l-1}.$$

This establishes the claim of the proposition with $B \propto 6l/(l-1)$. □

Proof of Proposition 2. By specification, ψ_{jk} satisfy $\int x^p \psi_{jk} = 0$ for $p = 0, 1, \ldots, 2r+1$, which gives that $A\psi_{jk} = 0$, hence $\mathbf{Cov}(\theta_{jk}, \theta_{j'k'}) = \langle \psi_{jk}, K_r \psi_{j'k'} \rangle = \langle \psi_{jk}, B\psi_{j'k'} \rangle$. Given that B is a polynomial of degree $(2r+1)$, and $\psi_{j'k'}$ is compactly supported,

$$\begin{aligned} B\psi_{j'k'}(x) &= \int_0^1 \psi_{j'k'}(y) \left(\int_0^{\min(x,y)} \frac{(x-u)^r (y-u)^r}{(r!)^2} du \right) dy \\ &= \underbrace{\int_0^x \psi_{j'k'}(y) \left(\int_0^y \frac{(x-u)^r (y-u)^r}{(r!)^2} du \right) dy}_{I} \\ &\qquad \underbrace{\int_x^1 \psi_{j'k'}(y) \left(\int_0^x \frac{(x-u)^r (y-u)^r}{(r!)^2} du \right) dy}_{II} . \end{aligned}$$

Both terms I and II integrate $\psi_{j'k'}$ against polynomials of degree at most $(2r+1)$ in y over the ranges $[0, x]$ and $[x, 1]$, respectively. Thus if, e.g., x is less than the minimum in $\mathbf{Supp}(\psi_{j'k'})$, we have $I = 0$ as the integrand is 0, while $II = 0$ as the $\psi_{j'k'}$ integrates the polynomial to 0 over its support. By a similar argument, we see that, if x is greater than the minimum in $\mathbf{Supp}(\psi_{j'k'})$, we have both $I = 0$ and $II = 0$. This shows that $\mathbf{Supp}(B\psi_{j'k'}) \subseteq \mathbf{Supp}(\psi_{j'k'})$, which proves that $\psi_{jk}(x)(B\psi_{j'k'})(x) = 0$ for all x if $\mathbf{Supp}(\psi_{jk}) \cap \mathbf{Supp}(\psi_{j'k'}) = \emptyset$.

Notice that $|\partial_x^p B(x,y)|, |\partial_y^p B(x,y)| \le 1$ for all $p = 0, \ldots, 2r+1$, we may employ techniques in Cohen (2003); Escande and Weiss (2015) to show that, for $j > j'$

$$|\langle \psi_{jk}, B\psi_{j'k'} \rangle| \lesssim 2^{-(2r+1)j} \cdot 2^{-(j+j')/2} = 2^{-(r+1)(j+j')} 2^{-(r+1/2)(j-j')}$$

and by symmetry, this gives that for all j, j'

$$|\mathbf{Cov}(\theta_{jk}, \theta_{j'k'})| = |\langle \psi_{jk}, B\psi_{j'k'} \rangle| \lesssim 2^{-(r+1)(j+j')} 2^{-(r+1/2)|j-j'|},$$

which, combined with what has been shown, establishes the bounds on correlations in the proposition.

For the last assertion, with p, q fixed, first split

$$\begin{aligned} \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \frac{\mathbf{Cov}(\theta_{jk}, \theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk}) \mathbf{Var}(\theta_{pq})}} &= \underbrace{\sum_{j=0}^{p} \sum_{k=0}^{2^j-1} \frac{\mathbf{Cov}(\theta_{jk}, \theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk}) \mathbf{Var}(\theta_{pq})}}}_{I} \\ &\quad + \underbrace{\sum_{j=p+1}^{\infty} \sum_{k=0}^{2^j-1} \frac{\mathbf{Cov}(\theta_{jk}, \theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk}) \mathbf{Var}(\theta_{pq})}}}_{II} . \end{aligned}$$

Now in sum I, for each $j < p$, there is only one ψ_{jk} with $\mathbf{Supp}(\psi_{jk}) \cap \mathbf{Supp}(\psi_{j'k'}) \neq \emptyset$. For $j = p$, there are a finite number depending on $(2r+1)$, or the number of moments we integrate to 0. This gives that $I \lesssim p$. Similarly, for II at each $j > p$, there are at most $O(2^{j-p})$ of ψ_{jk} for which $\mathbf{Supp}(\psi_{jk}) \cap \mathbf{Supp}(\psi_{j'k'}) \neq \emptyset$. Given the bounds derived on correlations, this gives

$$II \lesssim \sum_{j=p+1}^{\infty} 2^{-(r-1/2)(j-p)} \le \frac{1}{2^{r-1/2} - 1},$$

which, combined with the bound for I, proves the last bound in this proposition. $\square$

PROOF OF PROPOSITION 3. We fix p and q, as in the proof of the previous proposition, split the sum

$$\sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1}\frac{\mathbf{Cov}(\theta_{jk},\theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk})\mathbf{Var}(\theta_{pq})}} = \underbrace{\sum_{j=0}^{p}\sum_{k=0}^{2^j-1}\frac{\mathbf{Cov}(\theta_{jk},\theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk})\mathbf{Var}(\theta_{pq})}}}_{I} + \underbrace{\sum_{j=p+1}^{\infty}\sum_{k=0}^{2^j-1}\frac{\mathbf{Cov}(\theta_{jk},\theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk})\mathbf{Var}(\theta_{pq})}}}_{II}.$$

for $j \leq p$, $d((j,k),(p,q))$ takes at most 2^j values as we range through k, each of which is boundable by a number in $\{0,1,\ldots,2^j\}$. Further, it takes any value at most twice. Thus we find that in this range

$$\sum_{k=0}^{2^j-1}\frac{\mathbf{Cov}(\theta_{jk},\theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk})\mathbf{Var}(\theta_{pq})}} \lesssim 2^{-p+j}\sum_{k=0}^{2^j}\frac{1}{(1+k)^\gamma} \lesssim 2^{-p+j}\log 2^j = j2^j/2^p,$$

which gives that $I \lesssim 2^{-p}\sum_{j=0}^{p} j2^j \lesssim p$. Similarly, for $j > p$, $d((j,k),(p,q))$ takes at most 2^p values as we range through k, each of which is boundable by a number in $\{0,1,\ldots,2^p\}$. Further, it may take any value at most $2\cdot 2^{j-p}$ times. Thus we find that in this range

$$\sum_{k=0}^{2^j-1}\frac{\mathbf{Cov}(\theta_{jk},\theta_{pq})}{\sqrt{\mathbf{Var}(\theta_{jk})\mathbf{Var}(\theta_{pq})}} \lesssim 2^{-(\kappa-1)(j-p)}\sum_{k=0}^{2^p}\frac{1}{(1+k)^\gamma} \lesssim 2^{-(\kappa-1)(j-p)}\log 2^p,$$

and thus $II \lesssim p\sum_{j=p+1}^{\infty} 2^{-(\kappa-1)(j-p)} \lesssim p$. Combining bounds proves the result. $\square$

References

Beran, R. and Dümbgen, L. (1998) Modulation of estimators and confidence sets. *Annals of Statistics*, **26**, 1826–1856.

Boucheron, S., Lugosi, G. and Massart, P. (2013) *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C.-H. (2002) Asymptotic equivalence theory for nonparametric regression with random design. *Annals of Statistics*, **30**, 688–707.

Cohen, A. (2003) *Numerical Analysis of Wavelet Methods*. Elsevier, first edn.

Escande, P. and Weiss, P. (2015) Sparse wavelet representations of spatially varying blurring operators. *SIAM Journal on Imaging Sciences*, **8**, 2976–3014.

Freedman, D. (1999) Wald lecture: On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Annals of Statistics*, **27**, 1119–1141.

Hanson, D. L. and Wright, F. T. (1971) A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.*, **42**, 1079–1083.

Johnstone, I. M. (2015) *Gaussian Estimation: Sequence and Multiresolution Models.* Unpublished Mongraph.

Reiß, M. (2008) Asymptotic equivalence for nonparametric regression with multivariate and random design. *Annals of Statistics*, **36**, 1957–1982.

Rudelson, M. and Vershynin, R. (2013) Hanson-wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, **18**, 1–9.

Tsybakov, A. B. (2009) *Introduction to Nonparametric Estimation.* Springer.