

SUPPLEMENT TO “QUASI-BAYESIAN ANALYSIS OF NONPARAMETRIC INSTRUMENTAL VARIABLES MODELS”*

BY KENGO KATO

University of Tokyo

This supplemental file contains the additional technical proofs omitted in the main text, and some technical tools used in the proofs.

APPENDIX A: CDV WAVELET BASES AND BESOV SPACES

A.1. Wavelet bases for $L_2[0, 1]$. We review wavelet theory on the compact interval $[0, 1]$. We refer the reader to [4], [6] and [5, Chapter 7 and Appendix B] as useful general references on wavelet theory in the statistical (and signal processing) context.

Let (φ, ψ) be a Daubechies pair of the scaling function and wavelet of a multiresolution analysis of the space $L_2(\mathbb{R})$ of order N , with ψ having N vanishing moments and support contained in $[-N + 1, N]$, and φ having support contained in $[0, 2N - 1]$ [see 4, Remark 7.1]. We translate φ so that its support is contained in $[-N + 1, N]$. Define

$$\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k), \quad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k).$$

Then, for any fixed $J_0 \geq 0$, it is known that $\{\varphi_{J_0 k}, \psi_{jk}, j \geq J_0, k \in \mathbb{Z}\}$ forms an orthonormal basis for $L_2(\mathbb{R})$. However, we need an orthonormal basis for $L_2[0, 1]$. From the Daubechies pair (φ, ψ) , we wish to construct an orthonormal basis for $L_2[0, 1]$. The construction here is based on Cohen et al. [2, Section 4]. See also Chapter 7.5 of [6] for wavelet bases on $[0, 1]$.

Take a fixed resolution level j such that $2^j \geq 2N$. For $k = N, \dots, 2^j - N - 1$, φ_{jk} are supported in $[0, 1]$ and left unchanged: $\varphi_{jk}^{\text{int}}(x) = \varphi_{jk}(x)$ for $x \in [0, 1]$. At boundaries, $k = 0, \dots, N - 1$, construct suitable functions φ_k^L with support $[0, N + k]$ and φ_k^R with support $[-N - k, 0]$, and define

$$\varphi_{jk}^{\text{int}}(x) = 2^{j/2}\varphi_k^L(2^j x), \quad \varphi_{j, 2^j - k - 1}^{\text{int}}(x) = 2^{j/2}\varphi_k^R(2^j(x - 1)), \quad x \in [0, 1].$$

Note that both φ_k^L and φ_k^R have the same smoothness as φ . Define the multiresolution spaces $V_j = \text{span}\{\varphi_{jk}^{\text{int}}, k = 0, \dots, 2^j - 1\}$, which satisfy the

*Supported by the Grant-in-Aid for Young Scientists (B) (25780152) from the JSPS.

following properties (i) $\dim(V_j) = 2^j$; (ii) $V_j \subset V_{j+1}$; (iii) each V_j contains all polynomials of order at most $N - 1$.

Turning to the wavelet spaces, define W_j by the orthogonal complement of V_j in V_{j+1} . Starting from the Daubechies wavelet ψ , construct ψ_{jk}^{int} similarly to $\varphi_{jk}^{\text{int}}$. Then, we have $W_j = \text{span}\{\psi_{jk}^{\text{int}}, k = 0, \dots, 2^j - 1\}$, and for any $J_0 \geq 1$ with $2^{J_0} \geq 2N$ and $J > J_0$,

$$V_J = V_{J_0} \bigoplus_{j \geq J_0}^{J-1} W_j, \quad L_2[0, 1] = V_{J_0} \bigoplus_{j \geq J_0} W_j.$$

Therefore, $\{\varphi_{J_0 k}^{\text{int}}\}_{k=0}^{2^{J_0}-1} \cup \{\psi_{jk}^{\text{int}}, j \geq J_0, k = 0, \dots, 2^j - 1\}$ forms an orthonormal basis for $L_2[0, 1]$ [see Section 4 of 2, for formal proofs of these results]

DEFINITION 1. Call the so-constructed basis $\{\varphi_{J_0 k}^{\text{int}}\}_{k=0}^{2^{J_0}-1} \cup \{\psi_{jk}^{\text{int}}, j \geq J_0, k = 0, \dots, 2^j - 1\}$ the CDV (Cohen-Daubechies-Vial) wavelet basis for $L_2[0, 1]$ generated from the Daubechies pair (φ, ψ) . If (φ, ψ) is S -regular, i.e., if (φ, ψ) are S -times continuously differentiable, then call the so-generated CDV wavelet basis S -regular.

REMARK 1. For any given positive integer S , there is an S -regular Daubechies pair (φ, ψ) by taking the order N sufficiently large [see 4, Remark 7.1].

A.2. Besov spaces. We recall the definition of Besov spaces.

DEFINITION 2. Let $0 < s < S, s \in \mathbb{R}, S \in \mathbb{N}$ and $1 \leq p, q \leq \infty$. Let $\{\varphi_{J_0 k}^{\text{int}}\}_{k=0}^{2^{J_0}-1} \cup \{\psi_{jk}^{\text{int}}, j \geq J_0, k = 0, \dots, 2^j - 1\}$ be an S -regular CDV wavelet basis for $L_2[0, 1]$. Let

$$\varphi_{J_0 k}^{\text{int}}(f) = \int_0^1 f(x) \varphi_{J_0 k}^{\text{int}}(x) dx, \quad \psi_{jk}^{\text{int}}(f) = \int_0^1 f(x) \psi_{jk}^{\text{int}}(x) dx.$$

Then the Besov space $B_{p,q}^s$ is defined by the set of functions $\{f \in L_p[0, 1] : \|f\|_{s,p,q} < \infty\}$, where

$$\begin{aligned} \|f\|_{s,p,q} := & \left(\sum_{0 \leq k \leq 2^{J_0}-1} |\varphi_{J_0 k}^{\text{int}}(f)|^p \right)^{1/p} \\ & + \left(\sum_{j \geq J_0} \left(2^{j(s+1/2-1/p)} \left(\sum_{0 \leq k \leq 2^j-1} |\psi_{jk}^{\text{int}}(f)|^p \right)^{1/p} \right)^q \right)^{1/q}, \end{aligned}$$

with the obvious modification in case $p = \infty$ or $q = \infty$.

REMARK 2. Besov spaces cover commonly used smooth function spaces. For example, $B_{\infty,\infty}^s$ is equal to the Hölder-Zygmund space, which coincides with the classical Hölder space for non-integer s . For integer s , they do not coincide but the Hölder-Zygmund space contains the classical Hölder space. Moreover, $B_{2,2}^s$ is equal to the classical L_2 -Sobolev space when s is an integer. See [5], Appendix B.

APPENDIX B: PROOFS OF LEMMAS 1-3

PROOF OF LEMMA 1. For part (ii), the lower bound on $s_{\min}(\mathbb{E}[\phi^J(W)^{\otimes 2}])$ follows from Assumption 1 (iii); the upper bounds on $s_{\max}(\mathbb{E}[\phi^J(W)^{\otimes 2}])$ and $s_{\max}(\mathbb{E}[\phi^J(W)\phi^J(X)^T])$ follow from Assumption 1 (i) and the fact that $\{\phi_l\}$ is an orthonormal basis of $L_2[0, 1]$ (see (8)). Part (iii) follows from Rudelson's [8] inequality and (i). For the reader's convenience, we state Rudelson's inequality in Appendix E. For Part (v), we first note that, by (iii) and Weyl's perturbation theorem [1, Problem III.6.13], $s_{\min}(\hat{\Phi}_{WX}) \geq \tau_{J_n} - O_P(\sqrt{J_n 2^{J_n}/n})$. Since now $\sqrt{J_n 2^{J_n}/n} = o(\tau_{J_n})$, we have $s_{\min}(\hat{\Phi}_{WX}) \geq (1 - o_P(1))\tau_{J_n}$. For the proof of (i), denote by N the order of the Daubechies pair (φ, ψ) generating the CDV wavelet basis $\{\phi_l, l \geq 1\}$. Then, for each $x \in [0, 1]$ and each $j \geq J_0$, the number of nonzero elements in $\phi_{2^j+1}(x), \dots, \phi_{2^{j+1}}(x)$ is bounded by some constant depending only on N , and each $\phi_{2^j+k}(x)$ is bounded by some constant (depending only on ψ) times $2^{j/2}$ for all $k = 1, \dots, 2^j$. Similarly, $\phi_1, \dots, \phi_{2^{J_0}}$ are uniformly bounded. Therefore, there exists a constant D depending only on (φ, ψ) such that $\|\phi^J(x)\|_{\ell^2}^2 \leq D(2^{J_0} + \sum_{j=J_0}^{J-1} 2^j) = D2^J$ for all $x \in [0, 1]$.

Finally, we wish to show Part (iv). First, observe that $\|\mathbb{E}_n[\phi^{J_n}(W_i)R_i]\|_{\ell^2}^2 \leq 2\|\mathbb{E}_n[\phi^{J_n}(W_i)R_i] - \mathbb{E}[\phi^{J_n}(W)R]\|_{\ell^2}^2 + 2\|\mathbb{E}[\phi^{J_n}(W)R]\|_{\ell^2}^2$. By a simple moment calculation, the first term is $O_P(2^{J_n}/n)$. For the second term, by Assumptions 3 and 4 (ii),

$$\begin{aligned} \|\mathbb{E}[\phi^{J_n}(W)R]\|_{\ell^2}^2 &= \|\mathbb{E}[\phi^{J_n}(W)(g_0 - P_J g_0)(X)]\|_{\ell^2}^2 \\ &\lesssim \tau_{J_n}^2 \|g_0 - P_{J_n} g_0\|^2 \\ &\lesssim \tau_{J_n}^2 2^{-2J_n s}. \end{aligned}$$

This completes the proof. $\square$

PROOF OF LEMMA 2. Step 1. We first show that $|\Sigma_n| = 1 + o(1)$ ($|\Sigma_n|$ denotes the determinant of Σ_n). Let $\lambda_{\min,n}$ and $\lambda_{\max,n}$ denote the minimum and maximum eigenvalues of Σ_n , respectively. Then, by Weyl's perturbation

theorem, $1 - o(k_n^{-1}) \leq \lambda_{\min, n} \leq \lambda_{\max, n} \leq 1 + o(k_n^{-1})$, so that $(1 - o(k_n^{-1}))^{k_n} = \lambda_{\min, n}^{k_n} \leq |\Sigma_n| \leq \lambda_{\max, n}^{k_n} = (1 + o(k_n^{-1}))^{k_n}$. Here both sides converge to 1.

Step 2. By Step 1, we have

$$\begin{aligned} & \int |dN(0, \Sigma_n)(x) - dN(0, I_{k_n})(x)| dx \\ &= \frac{1}{(2\pi)^{k_n/2}} \int \left| \frac{1}{|\Sigma_n|^{1/2}} e^{-x^T \Sigma_n^{-1} x/2} - e^{-x^T x/2} \right| dx \\ &\leq \left| \frac{1}{|\Sigma_n|^{1/2}} - 1 \right| + \frac{1}{(2\pi)^{k_n/2} |\Sigma_n|^{1/2}} \int |e^{-x^T \Sigma_n^{-1} x/2} - e^{-x^T x/2}| dx \\ &\leq o(1) + \frac{1}{(2\pi)^{k_n/2} (1 + o(1))} \int e^{-x^T x/2} |e^{-x^T (\Sigma_n^{-1} - I_{k_n}) x/2} - 1| dx \end{aligned}$$

By assumption, we have $\epsilon_n := \|\Sigma_n^{-1} - I_{k_n}\|_{\text{op}} \leq \|\Sigma_n^{-1}\|_{\text{op}} \|I_{k_n} - \Sigma_n\|_{\text{op}} = o(k_n^{-1})$. Now, $|e^{-x^T (\Sigma_n^{-1} - I_{k_n}) x/2} - 1| \leq e^{\epsilon_n x^T x/2} - e^{-\epsilon_n x^T x/2}$. By a direct calculation, the conclusion follows from the fact that $(1 \pm \epsilon_n)^{k_n} = 1 + o(1)$. $\square$

PROOF OF LEMMA 3. The first assertion follows from the assumption. Suppose now that $\hat{A}_n$ is non-singular. Then, $\hat{A}_n^{-1} A_n = (\hat{A}_n - A_n + A_n)^{-1} A_n = (A_n^{-1} \hat{A}_n - I_{k_n} + I_{k_n})^{-1}$. Here, $A_n^{-1} \hat{A}_n - I_{k_n} = A_n^{-1} (\hat{A}_n - A_n)$, so that $\|A_n^{-1} \hat{A}_n - I_{k_n}\|_{\text{op}} \leq \|A_n^{-1}\|_{\text{op}} \|\hat{A}_n - A_n\|_{\text{op}} = s_{\min}^{-1}(A_n) \|A_n - \hat{A}_n\|_{\text{op}} = O_P(\epsilon_n^{-1} \delta_n)$. Let $\hat{\Delta} = I_{k_n} - A_n^{-1} \hat{A}_n$. Then, $\hat{A}_n^{-1} A_n = (I_{k_n} - \hat{\Delta})^{-1} = I_{k_n} + \sum_{m=1}^{\infty} \hat{\Delta}^m$ (Neumann series). Therefore, we conclude that $\|\hat{A}_n^{-1} A_n - I_{k_n}\|_{\text{op}} = \|\sum_{m=1}^{\infty} \hat{\Delta}^m\|_{\text{op}} \leq \sum_{m=1}^{\infty} \|\hat{\Delta}\|_{\text{op}}^m = \|\hat{\Delta}\|_{\text{op}} \cdot \sum_{m=0}^{\infty} \|\hat{\Delta}\|_{\text{op}}^m = O_P(\epsilon_n^{-1} \delta_n)$. $\square$

APPENDIX C: PROOF OF THEOREM 3

For the notational convenience, define

$$\mathbb{E}_{\Pi_n}[\cdot | \mathcal{D}_n] := \int \cdot \Pi_n(dg | \mathcal{D}_n), \quad \mathbb{E}_{\tilde{\Pi}_n}[\cdot | \mathcal{D}_n] := \int \cdot \tilde{\Pi}_n(db^{J_n} | \mathcal{D}_n).$$

PROOF OF THEOREM 3. Define the event

$$\mathcal{E}_{3n} = \{\mathcal{D}_n : \hat{\Phi}_{WX} \text{ and } \hat{\Phi}_{WW} \text{ are non-singular}\}.$$

Then, by Lemma 1, $\mathbb{P}\{1(\mathcal{E}_{3n}) = 1\} = \mathbb{P}(\mathcal{E}_{3n}) \rightarrow 1$. Suppose that $1(\mathcal{E}_{3n}) = 1$. Then, by (16), $\ell_{b^{J_n}}(\mathcal{D}_n)$ defined in the proof of Proposition 4 is bounded from below by

$$\hat{c} \|b^{J_n}\|_{\ell^2}^2 + \text{a term independent of } b^{J_n},$$

for some positive random variable $\hat{c}$. Hence, the integral $\mathbb{E}_{\tilde{\Pi}_n}[\|b^{J_n}\|_{\ell^2}^2 | \mathcal{D}_n]$ is finite as soon as $1(\mathcal{E}_{3n}) = 1$. This proves the first assertion.

In what follows, we wish to prove the convergence rate result (14). First of all, by the triangle inequality and Jensen's inequality,

$$\begin{aligned}
1(\mathcal{E}_{3n})\|\hat{g}_{QB} - g_0\| &\leq 1(\mathcal{E}_{3n})\|\hat{g}_{QB} - P_{J_n}g_0\| + \|g_0 - P_{J_n}g_0\| \\
&= 1(\mathcal{E}_{3n})\|\mathbb{E}_{\Pi_n}[g - P_{J_n}g_0 \mid \mathcal{D}_n]\| + \|g_0 - P_{J_n}g_0\| \\
&= 1(\mathcal{E}_{3n})\|\mathbb{E}_{\tilde{\Pi}_n}[b^{J_n} - b_0^{J_n} \mid \mathcal{D}_n]\|_{\ell^2} + \|g_0 - P_{J_n}g_0\| \\
&\leq 1(\mathcal{E}_{3n})\mathbb{E}_{\tilde{\Pi}_n}[\|b^{J_n} - b_0^{J_n}\|_{\ell^2} \mid \mathcal{D}_n] + \|g_0 - P_{J_n}g_0\|.
\end{aligned}$$

Since $\|g_0 - P_{J_n}g_0\| = O(2^{-J_n s})$, it suffices to show that there exists a constant $D > 0$ such that for every $M_n \rightarrow \infty$,

$$\begin{aligned}
&\mathbb{P}\left[1(\mathcal{E}_{3n})\mathbb{E}_{\tilde{\Pi}_n}[\|b^{J_n} - b_0^{J_n}\|_{\ell^2} \mid \mathcal{D}_n] \right. \\
&\quad \left. \leq D \max\{2^{-J_n s}, \tau_{J_n}^{-1} \sqrt{2^{J_n}/n}, \tau_{J_n}^{-1} \epsilon_n \varrho_n M_n\} \right] \rightarrow 1.
\end{aligned}$$

Let $\pi_n^*(\theta^{J_n} \mid \mathcal{D}_n)$ be the (random) density defined in the proof of Theorem 1. Note that $\pi_n^*(\theta^{J_n} \mid \mathcal{D}_n)$ is well-defined as soon as $1(\mathcal{E}_{3n}) = 1$. Let $\delta_n := \epsilon_n + \tau_{J_n} 2^{-J_n s}$. Then we have:

LEMMA 1. *There exists a constant $c_1 > 0$ such that for every sequence $M_n \rightarrow \infty$ with $M_n = o(L_n)$,*

$$\begin{aligned}
&\mathbb{P}\left\{1(\mathcal{E}_{3n}) \int \|\theta^{J_n}\|_{\ell^2} \cdot |\pi_n^*(\theta^{J_n} \mid \mathcal{D}_n) - dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})| d\theta^{J_n} \right. \\
&\quad \left. \leq e^{-c_1 M_n n \delta_n^2} + M_n \sqrt{n} \delta_n \varrho_n \right\} \rightarrow 1,
\end{aligned}$$

where $\Delta_n := \sqrt{n} \mathbb{E}_n[\phi^{J_n}(W_i) R_i]$.

We defer the proof of Lemma 1 after the proof of this theorem. Note that

$$\begin{aligned}
&1(\mathcal{E}_{3n}) \left[\int \|\theta^{J_n}\|_{\ell^2} dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n}) d\theta^{J_n} \right]^2 \\
&\leq 1(\mathcal{E}_{3n}) \int \|\theta^{J_n}\|_{\ell^2}^2 dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n}) d\theta^{J_n} \\
&\leq \|\Delta_n\|_{\ell^2}^2 + \text{tr}(\hat{\Phi}_{WW}).
\end{aligned}$$

By the proof of Theorem 2, there exists a constant $D_1 > 0$ such that $\mathbb{P}\{\|\Delta_n\|_{\ell^2}^2 + \text{tr}(\hat{\Phi}_{WW}) \leq D_1(n\tau_{J_n}^2 2^{-2J_n s} + 2^{J_n})\} \rightarrow 1$. Hence for every sequence $M_n \rightarrow \infty$ with $M_n = o(L_n)$, with probability approaching one,

$$\begin{aligned} & \sqrt{D_1(n\tau_{J_n}^2 2^{-2J_n s} + 2^{J_n})} + e^{-c_1 M_n n \delta_n^2} + M_n \sqrt{n} \delta_n \varrho_n \\ & \geq 1(\mathcal{E}_{3n}) \int \|\theta^{J_n}\|_{\ell^2} \pi_n^*(\theta^{J_n} | \mathcal{D}_n) \\ & = 1(\mathcal{E}_{3n}) \sqrt{n} \int \|\hat{\Phi}_{WX}(b^{J_n} - b_0^{J_n})\|_{\ell^2} \tilde{\pi}_n(b^{J_n} | \mathcal{D}_n) db^{J_n} \\ & \geq 1(\mathcal{E}_{3n}) \sqrt{n} s_{\min}(\hat{\Phi}_{WX}) \mathbb{E}_{\tilde{\Pi}_n} [\|b^{J_n} - b_0^{J_n}\|_{\ell^2} | \mathcal{D}_n]. \end{aligned}$$

Take $M_n \rightarrow \infty$ sufficiently slowly such that $\varrho_n M_n \rightarrow 0$. Since the left side is then $\lesssim \max\{\sqrt{n} \tau_{J_n} 2^{-J_n s}, \sqrt{2^{J_n}}, \sqrt{n} \epsilon_n \varrho_n M_n\}$, there exists a constant $D_2 > 0$ such that

$$\begin{aligned} & \mathbb{P}\left[1(\mathcal{E}_{3n}) s_{\min}(\hat{\Phi}_{WX}) \mathbb{E}_{\tilde{\Pi}_n} [\|b^{J_n} - b_0^{J_n}\|_{\ell^2} | \mathcal{D}_n] \right. \\ & \quad \left. \leq D_2 \max\{\tau_{J_n} 2^{-J_n s}, \sqrt{2^{J_n}/n}, \epsilon_n \varrho_n M_n\} \right] \rightarrow 1. \end{aligned}$$

Finally, by Lemma 1, $\mathbb{P}(s_{\min}(\hat{\Phi}_{WX}) \geq 0.5 \tau_{J_n}) \rightarrow 1$, by which we have

$$\begin{aligned} & \mathbb{P}\left[1(\mathcal{E}_{3n}) \mathbb{E}_{\tilde{\Pi}_n} [\|b^{J_n} - b_0^{J_n}\|_{\ell^2} | \mathcal{D}_n] \right. \\ & \quad \left. \leq 2D_2 \max\{2^{-J_n s}, \tau_{J_n}^{-1} \sqrt{2^{J_n}/n}, \tau_{J_n}^{-1} \epsilon_n \varrho_n M_n\} \right] \rightarrow 1. \end{aligned}$$

This leads to the desired conclusion (it is not difficult to see that the final expression holds for every sequence $M_n \rightarrow \infty$). $\square$

PROOF OF LEMMA 1. As before, we say that a sequence of random variables A_n is *eventually* bounded by another sequence of random variables B_n if $\mathbb{P}(A_n \leq B_n) \rightarrow 1$.

Take any $M_n \rightarrow \infty$ with $M_n = o(L_n)$. Then,

$$\begin{aligned} & 1(\mathcal{E}_{3n}) \int \|\theta^{J_n}\|_{\ell^2} \cdot |\pi_n^*(\theta^{J_n} | \mathcal{D}_n) - dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})| d\theta^{J_n} \\ & \leq 1(\mathcal{E}_{3n}) \int_{\|\theta^{J_n}\|_{\ell^2} \leq M_n \sqrt{n} \delta_n} \|\theta^{J_n}\|_{\ell^2} \cdot |\pi_n^*(\theta^{J_n} | \mathcal{D}_n) - dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})| d\theta^{J_n} \\ & \quad + 1(\mathcal{E}_{3n}) \int_{\|\theta^{J_n}\|_{\ell^2} > M_n \sqrt{n} \delta_n} \|\theta^{J_n}\|_{\ell^2} \pi_n^*(\theta^{J_n} | \mathcal{D}_n) d\theta^{J_n} \\ & \quad + 1(\mathcal{E}_{3n}) \int_{\|\theta^{J_n}\|_{\ell^2} > M_n \sqrt{n} \delta_n} \|\theta^{J_n}\|_{\ell^2} dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n}) d\theta^{J_n} \\ & =: I + II + III. \end{aligned}$$

We divide the rest of the proof into three steps.

Step 1. Claim: $\exists c_2 > 0$ such that $\mathbb{P}(II \leq e^{-c_2 M_n^2 n \delta_n^2}) \rightarrow 1$.

(Proof of Step 1): The assertion of Step 1 follows from the same line as in the proof of Proposition 4 by noting that for any $c > 0$, $x e^{-cx^2} \leq e^{-cx^2/2}$ for all $x > 0$ sufficiently large. Hence the proof is omitted.

Step 2. Claim: $\exists c_3 > 0$ such that $\mathbb{P}(III \leq e^{-c_3 M_n n \delta_n^2}) \rightarrow 1$.

(Proof of Step 2): By the Cauchy-Schwarz inequality, the square of III is bounded by $\int \|\theta^{J_n}\|_{\ell^2}^2 dN(\Delta_n, \hat{\Phi}_{WW}) d\theta^{J_n} \int_{\|\theta^{J_n}\|_{\ell^2} > M_n \sqrt{n} \delta_n} dN(\Delta_n, \hat{\Phi}_{WW}) d\theta^{J_n}$. Here the first integral is bounded by

$$\|\Delta_n\|_{\ell^2}^2 + \text{tr}(\hat{\Phi}_{WW}),$$

which is eventually bounded by $D(n\tau_{J_n}^2 2^{-2J_n s} + 2^{J_n})$ for some constant $D > 0$ by the proof of Theorem 2. On the other hand, by the proof of Theorem 1, the second integral is eventually bounded by $\int_{\|\theta^{J_n}\|_{\ell^2} > \sqrt{M_n n} \delta_n} dN(0, I_{2^{J_n}}) d\theta^{J_n}$. By Borell's inequality for Gaussian measures (see (23)), the last integral is bounded by $e^{-c' M_n n \delta_n^2}$ for some small constant $c' > 0$. Taking these together, we obtain the conclusion of Step 2 by choosing the constant $c_3 > 0$ sufficiently small.

Step 3. Claim: $\exists c_4 > 0$ such that $\mathbb{P}(I \leq e^{-c_4 M_n^2 n \delta_n^2} + M_n \sqrt{n} \delta_n \varrho_n) \rightarrow 1$.

(Proof of Step 3): Let $\mathcal{C}_n := \{\theta^{J_n} \in \mathbb{R}^{2^{J_n}} : \|\theta^{J_n}\|_{\ell^2} \leq M_n \sqrt{n} \delta_n\}$. Let $\pi_{n, \mathcal{C}_n}^*(\theta^{J_n} \mid \mathcal{D}_n)$ and $dN^{\mathcal{C}_n}(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})$ denote the probability densities obtained by first restricting $\pi_n^*(\theta^{J_n} \mid \mathcal{D}_n)$ and $dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})$ to the ball $\mathcal{C}_n$ and then renormalizing, respectively. Then, abbreviating $\pi_n^*(d\theta^{J_n} \mid \mathcal{D}_n)$ by π_n^* , $\pi_{n, \mathcal{C}_n}^*(d\theta^{J_n} \mid \mathcal{D}_n)$ by $\pi_{n, \mathcal{C}_n}^*$, $dN(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})$ by dN , and $dN^{\mathcal{C}_n}(\Delta_n, \hat{\Phi}_{WW})(\theta^{J_n})$ by $dN^{\mathcal{C}_n}$ we have

$$\begin{aligned} I &\leq 1(\mathcal{E}_{3n}) \int \|\theta^{J_n}\|_{\ell^2} \cdot |\pi_{n, \mathcal{C}_n}^* - dN^{\mathcal{C}_n}| \\ &\quad + 1(\mathcal{E}_{3n}) \int_{\|\theta^{J_n}\|_{\ell^2} \leq M_n \sqrt{n} \delta_n} \|\theta^{J_n}\|_{\ell^2} \cdot |\pi_{n, \mathcal{C}_n}^* - \pi_n^*| \\ &\quad + 1(\mathcal{E}_{3n}) \int_{\|\theta^{J_n}\|_{\ell^2} \leq M_n \sqrt{n} \delta_n} \|\theta^{J_n}\|_{\ell^2} \cdot |dN^{\mathcal{C}_n} - dN| \\ &=: IV + V + IV. \end{aligned}$$

By the proof of Theorem 1, the term IV is eventually bounded by

$$1(\mathcal{E}_{3n}) M_n \sqrt{n} \delta_n \int |\pi_{n, \mathcal{C}_n}^* - dN^{\mathcal{C}_n}| \leq M_n \sqrt{n} \delta_n \varrho_n.$$

For the term V , we have

$$\begin{aligned} V &\leq 1(\mathcal{E}_{3n})M_n\sqrt{n}\delta_n \int_{\|\theta^{J_n}\|_{\ell^2} \leq M_n\sqrt{n}\delta_n} |\pi_{n,\mathcal{C}_n}^* - \pi_n^*| \\ &= 1(\mathcal{E}_{3n})M_n\sqrt{n}\delta_n \int_{\|\theta^{J_n}\|_{\ell^2} > M_n\sqrt{n}\delta_n} \pi_n^*. \end{aligned}$$

By the proof of Proposition 4, there exists a constant $c_5 > 0$ such that the integral on the right side is eventually bounded by $e^{-c_5 M_n^2 n \delta_n^2}$, so that $\mathbb{P}(V \leq e^{-c_5 M_n^2 n \delta_n^2 / 2}) \rightarrow 1$. Likewise, by Borell's inequality for Gaussian measures, there exists a constant $c_6 > 0$ such that $\mathbb{P}(VI \leq e^{-c_6 M_n n \delta_n^2}) \rightarrow 1$. Taking these together, we obtain the conclusion of Step 3 by choosing the constant $c_4 > 0$ sufficiently small.

Finally, Steps 1-3 lead to the conclusion of Lemma 1. $\square$

APPENDIX D: PROOFS FOR SECTION 4

PROOF OF PROPOSITION 2. We only consider the mildly ill-posed case. The proof for the severely ill-posed case is similar. For either case of product or isotropic priors, it suffices to check conditions P1) and P2) in Theorem 1. We shall do this with the choice $\epsilon_n = \sqrt{2^{J_n}(\log n)/n} \sim (\log n)^{1/2} n^{-(r+s)/(2r+2s+1)}$.

Case of product priors: Let $c_{\min} := \min_{x \in [-A, A]} q(x) > 0$. Since $\|b^{J_n} - b_0^{J_n}\|_{\ell^2}^2 = \sum_{l=1}^{2^{J_n}} (b_l - b_{0l})^2 \leq 2^{J_n} \max_{1 \leq l \leq 2^{J_n}} (b_l - b_{0l})^2$, we have

$$\begin{aligned} \tilde{\Pi}_n(b^{J_n} : \|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq \epsilon_n) &\geq \tilde{\Pi}_n \left(b^{J_n} : \max_{1 \leq l \leq 2^{J_n}} |b_l - b_{0l}| \leq \epsilon_n / \sqrt{2^{J_n}} \right) \\ &\geq \prod_{l=1}^{2^{J_n}} \tilde{\Pi}_n(b^{J_n} : |b_l - b_{0l}| \leq \epsilon_n / \sqrt{2^{J_n}}). \end{aligned}$$

Since $\exists \epsilon \in (0, A)$, $b_{0l} \in [-A + \epsilon, A - \epsilon]$ for all $l \geq 1$, for all n sufficiently large, the last expression is bounded from below by

$$\left(\frac{c_{\min} \epsilon_n}{\sqrt{2^{J_n}}} \right)^{2^{J_n}} = e^{-2^{J_n} \log(\sqrt{2^{J_n}} / (c_{\min} \epsilon_n))} \geq e^{-C n \epsilon_n^2},$$

where $C > 0$ is a sufficiently large constant, which verifies condition P1).

Second, with this ϵ_n , γ_n in condition P2) is $\sim (\log n)^{1/2} n^{-s/(2r+2s+1)}$. Let, say, $L_n \sim (\log n)^{1/2}$ so that $L_n \gamma_n \sim (\log n) n^{-s/(2r+2s+1)}$. Then, $\{b^{J_n} : \|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq L_n \gamma_n\} \subset [-A, A]^{2^{J_n}}$ for all n sufficiently large, so that

$\tilde{\pi}_n(b^{J_n}) = \prod_{l=1}^{2^{J_n}} q(b_l)$ is positive for all $\|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq L_n \gamma_n$. Let $\|b^{J_n}\|_{\ell^2} \leq L_n \gamma_n$ and $\|\tilde{b}^{J_n}\|_{\ell^2} \leq L_n \gamma_n$. Then,

$$\begin{aligned} \frac{\tilde{\pi}_n(b_0^{J_n} + b^{J_n})}{\tilde{\pi}_n(b_0^{J_n} + \tilde{b}^{J_n})} &= \exp \left[\sum_{l=1}^{2^{J_n}} \{ \log q(b_{0l} + b_l) - \log q(b_{0l} + \tilde{b}_l) \} \right] \\ &\leq \exp \left\{ L \sum_{l=1}^{2^{J_n}} |b_l - \tilde{b}_l| \right\} \leq \exp \left\{ L \sqrt{2^{J_n}} \|b^{J_n} - \tilde{b}^{J_n}\|_{\ell^2} \right\} \\ &\leq e^{2L\sqrt{2^{J_n}} L_n \gamma_n} = e^{o(1)}, \end{aligned}$$

where the last step is due to $s > 1/2$. Likewise, we have

$$\frac{\tilde{\pi}_n(b_0^{J_n} + b^{J_n})}{\tilde{\pi}_n(b_0^{J_n} + \tilde{b}^{J_n})} \geq e^{-2L\sqrt{2^{J_n}} L_n \gamma_n} = e^{-o(1)}.$$

Therefore, condition P2) is verified.

Case of isotropic priors: Let $c_{\min} := \min_{x \in [0, A]} r(x) > 0$. Then, for all n sufficiently large,

$$\begin{aligned} \tilde{\Pi}_n(b^{J_n} : \|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq \epsilon_n) &= \frac{\int_{\|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq \epsilon_n} r(\|b^{J_n}\|_{\ell^2}) db^{J_n}}{\int r(\|b^{J_n}\|_{\ell^2}) db^{J_n}} \\ &= \frac{\int_{\|b^{J_n}\|_{\ell^2} \leq \epsilon_n} r(\|b^{J_n} + b_0^{J_n}\|_{\ell^2}) db^{J_n}}{\int r(\|b^{J_n}\|_{\ell^2}) db^{J_n}} \geq \frac{c_{\min} \int_{\|b^{J_n}\|_{\ell^2} \leq \epsilon_n} db^{J_n}}{\int r(\|b^{J_n}\|_{\ell^2}) db^{J_n}} \\ &= c_{\min} \frac{\int_{x \in [0, \epsilon_n]} x^{2^{J_n}-1} dx}{\int_0^\infty x^{2^{J_n}-1} r(x) dx} \geq c_{\min} \left(\frac{\epsilon_n}{2^{J_n}} \right)^{2^{J_n}} \times e^{-c'' 2^{J_n} \log(2^{J_n})} \\ &= c_{\min} e^{-2^{J_n} \log(2^{J_n}/\epsilon_n) - c'' 2^{J_n} \log(2^{J_n})} \geq e^{-C n \epsilon_n^2}, \end{aligned}$$

where $C > 0$ is a sufficiently large constant, which verifies condition P1).

Second, with this ϵ_n , γ_n in condition P2) is $\sim (\log n)^{1/2} n^{-s/(2r+2s+1)}$. Let $L_n \sim (\log n)^{1/2}$ so that $L_n \gamma_n \sim (\log n) n^{-s/(2r+2s+1)}$. Since $\|b_0^{J_n}\|_{\ell^2} \leq \|g_0\| < A$ and $L_n \gamma_n \rightarrow 0$, $\{b^{J_n} : \|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq L_n \gamma_n\} \subset \{b^{J_n} : \|b^{J_n}\|_{\ell^2} \leq A\}$ for all n sufficiently large, so that $\tilde{\pi}_n(b^{J_n}) \propto r(\|b^{J_n}\|_{\ell^2})$ is positive for all $\|b^{J_n} - b_0^{J_n}\|_{\ell^2} \leq L_n \gamma_n$. Let $\|b^{J_n}\|_{\ell^2} \leq L_n \gamma_n$ and $\|\tilde{b}^{J_n}\|_{\ell^2} \leq L_n \gamma_n$. Then by Parseval's identity,

$$\|b_0^{J_n} + b^{J_n}\|_{\ell^2} \leq \|b_0^{J_n}\|_{\ell^2} + L_n \gamma_n \rightarrow \|g_0\|,$$

and likewise we have

$$\|b_0^{J_n} + b^{J_n}\|_{\ell^2} \geq \|b_0^{J_n}\|_{\ell^2} - L_n \gamma_n \rightarrow \|g_0\|.$$

Therefore, we conclude that, uniformly in $\|b^{J_n}\|_{\ell^2} \leq L_n \gamma_n$, $\|\tilde{b}^{J_n}\|_{\ell^2} \leq L_n \gamma_n$,

$$\frac{\tilde{\pi}_n(b_0^{J_n} + b^{J_n})}{\tilde{\pi}_n(b_0^{J_n} + \tilde{b}^{J_n})} = \frac{r(\|b_0^{J_n} + b^{J_n}\|_{\ell^2})}{r(\|b_0^{J_n} + \tilde{b}^{J_n}\|_{\ell^2})} \rightarrow \frac{r(\|g_0\|)}{r(\|g_0\|)} = 1.$$

Hence condition P2) is verified. $\square$

PROOF OF PROPOSITION 3. Given the proof of Proposition 2 and the discussion following Theorem 3, it is sufficient to verify that ϱ_n is $o((\log n)^{-1/2})$ in the mildly ill-posed case (in the severely ill-posed case, it is sufficient to verify that $\varrho_n = o(1)$). However, this is readily verified by tracking the proof of Proposition 2. $\square$

APPENDIX E: TECHNICAL TOOLS

We state here Rudelson's inequality for the reader's convenience.

THEOREM 1 (Rudelson's [8] inequality). *Let $Z_1, \dots, Z_n$ be i.i.d. random vectors in $\mathbb{R}^k$ with $\Sigma := E[Z_1^{\otimes 2}]$. Then for every $k \geq 2$,*

$$\mathbb{E} \left[\left\| \frac{1}{n} \sum_{i=1}^n Z_i^{\otimes 2} - \Sigma \right\|_{\text{op}} \right] \leq \max\{\|\Sigma\|_{\text{op}}^{1/2} \delta, \delta^2\}, \quad \delta = D \sqrt{\frac{\log k}{n} \mathbb{E}[\max_{1 \leq i \leq n} \|Z_i\|_{\ell^2}^2]},$$

where D is a universal constant.

Rudelson's inequality implies the following corollary useful in our application.

COROLLARY 1. *Let $(X_1, Y_1^T)^T, \dots, (X_n, Y_n^T)^T$ be i.i.d. random vectors with $X_i \in \mathbb{R}^{k_1}, Y_i \in \mathbb{R}^{k_2}$, and $k_1 + k_2 \geq 2$. Let $\Sigma_X := E[X_1^{\otimes 2}]$, $\Sigma_Y := E[Y_1^{\otimes 2}]$ and $\Sigma_{XY} := E[X_1 Y_1^T]$. Suppose that there exists a finite number m such that $\mathbb{E}[\max_{1 \leq i \leq n} \|X_i\|_{\ell^2}^2] \vee \mathbb{E}[\max_{1 \leq i \leq n} \|Y_i\|_{\ell^2}^2] \leq m$. Then*

$$\mathbb{E} \left[\left\| \frac{1}{n} \sum_{i=1}^n X_i Y_i^T - \Sigma_{XY} \right\|_{\text{op}} \right] \leq \max\{(\|\Sigma_X\|_{\text{op}}^{1/2} \vee \|\Sigma_Y\|_{\text{op}}^{1/2}) \delta, \delta^2\},$$

with $\delta = D \sqrt{\frac{m \log(k_1 + k_2)}{n}},$

where D is a universal constant.

PROOF. Let $Z_i = (X_i, Y_i^T)^T$, and apply Rudelson's inequality to $Z_1, \dots, Z_n$. Note that by the variational characterization of the operator norm, we have $\|n^{-1} \sum_{i=1}^n X_i Y_i^T - \Sigma_{XY}\|_{\text{op}} \leq \|n^{-1} \sum_{i=1}^n Z_i^{\otimes 2} - \mathbb{E}[Z_1^{\otimes 2}]\|_{\text{op}}$, and by the Cauchy-Schwarz inequality, $\|\mathbb{E}[Z_1^{\otimes 2}]\|_{\text{op}} \leq 2\|\Sigma_X\|_{\text{op}} + 2\|\Sigma_Y\|_{\text{op}}$. $\square$

Lastly, we shall recall Talagrand's [9] concentration inequality for general empirical processes. The following version is due to [7]. Here, for a generic class $\mathcal{F}$ of measurable functions on some measurable space $\mathcal{X}$, we say that $\mathcal{F}$ is pointwise measurable if there exists a countable subclass $\mathcal{G} \subset \mathcal{F}$ such that for any $f \in \mathcal{F}$, there exists a sequence $\{g_m\} \subset \mathcal{G}$ with $g_m(x) \rightarrow f(x)$ for every $x \in \mathcal{X}$. See Chapter 2.3 of [10].

THEOREM 2 (Massart's form of Talagrand's inequality). *Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables taking values in some measurable space $\mathcal{X}$. Let $\mathcal{F}$ be a pointwise measurable class of functions on $\mathcal{X}$ such that $\mathbb{E}[f(\xi_1)] = 0$ for all $f \in \mathcal{F}$ and $\sup_{f \in \mathcal{F}} \sup_{x \in \mathcal{X}} |f(x)| \leq B$ for some constant $B > 0$. Let σ^2 be any positive constant such that $\sigma^2 \geq \sup_{f \in \mathcal{F}} \mathbb{E}[f^2(\xi_1)]$. Let $Z := \sup_{f \in \mathcal{F}} |\sum_{i=1}^n f(\xi_i)|$. Then for every $x > 0$,*

$$\mathbb{P}\{Z \geq C(\mathbb{E}[Z] + \sigma\sqrt{nx} + Bx)\} \leq e^{-x},$$

where $C > 0$ is a universal constant.

REFERENCES

- [1] Bhatia, R. (1997). *Matrix Analysis*. Springer.
- [2] Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54-81.
- [3] Ghosal, S., Ghosh, J. K. and van der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500-531.
- [4] Härdle, W., Kerkycharian, F., Picard, D., and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. Springer.
- [5] Johnstone, I.M. (2011). *Gaussian Estimation: Sequence and Multiresolution Models*. Unpublished draft.
- [6] Mallat, S. (2009). *A Wavelet Tour of Signal Processing*. Third Edition. Academic Press.
- [7] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863-884.
- [8] Rudelson, M. (1999). Random vectors in the isotropic position. *J. Functional Anal.* **164** 60-72.
- [9] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505-563.
- [10] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

GRADUATE SCHOOL OF ECONOMICS, UNIVERSITY OF TOKYO
7-3-1 HONGO, BUNKYO-KU, TOKYO 113-0033, JAPAN.
E-MAIL: kkato@e.u-tokyo.ac.jp