Supplement to Bayesian mode and maximum estimation and accelerated rates of contraction

WILLIAM WEIMIN YOO*[1]

[1] *Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands*
E-mail: yooweimin0203@gmail.com
and

SUBHASHIS GHOSAL[2]

[2] *Department of Statistics, North Carolina State University, 4276 SAS Hall, 2311 Stinson Drive, Raleigh, North Carolina 27695-8203, USA*
E-mail: sghosal@ncsu.edu

The supplementary file contains detailed proofs of Corollary 4.2, Proposition 5.1 and Corollary 8.4. in the main paper Yoo and Ghosal [4].

Proof of Corollary 4.2. From the proof of Theorem 4.1 before, we know that $\boldsymbol{\mu}-\boldsymbol{\mu}_0 = \boldsymbol{H}f_0(\boldsymbol{\mu}^*)^{-1}(\nabla f_0(\boldsymbol{\mu}) - \nabla f_0(\boldsymbol{\mu}_0))$. Noting that $\nabla f_0(\boldsymbol{\mu}_0) = \nabla f(\boldsymbol{\mu}) = \boldsymbol{0}$ by Assumption 2, we can use the fact $\|\boldsymbol{A}\boldsymbol{b}\|^2 \geq \lambda_{\min}(\boldsymbol{A}^T\boldsymbol{A})\|\boldsymbol{b}\|^2$ to write

$$\begin{aligned}\|\boldsymbol{\mu}-\boldsymbol{\mu}_0\| &\geq \sqrt{\lambda_{\max}^{-2}\{\boldsymbol{H}f_0(\boldsymbol{\mu}^*)\}}\|\nabla f_0(\boldsymbol{\mu}) - \nabla f_0(\boldsymbol{\mu}_0)\| \\ &\geq \lambda_1^{-1}\|\nabla f_0(\boldsymbol{\mu}) - \nabla f(\boldsymbol{\mu})\|,\end{aligned}$$

by posterior consistency of $\boldsymbol{\mu}^*$ as established in the proof of Theorem 5.2. Let $\delta_n \to 0$ be some sequence. Then for some small enough constant $h > 0$ to be determined below, we have

$$\begin{aligned}\Pi(\|\boldsymbol{\mu}-\boldsymbol{\mu}_0\| \leq h\epsilon_n|\boldsymbol{Y}) \leq{}& \Pi\left(\|\nabla f_0(\boldsymbol{\mu}) - \nabla f(\boldsymbol{\mu})\| \leq \lambda_1 h\epsilon_n, \|\boldsymbol{\mu}-\boldsymbol{\mu}_0\| \leq \delta_n|\boldsymbol{Y}\right) \\ &+ \Pi(\|\boldsymbol{\mu}-\boldsymbol{\mu}_0\| > \delta_n|\boldsymbol{Y}).\end{aligned}$$

Since the posterior of $\boldsymbol{\mu}$ is consistent, the second term is $o_{P_0}(1)$. Using the definition of continuity of $\boldsymbol{x} \mapsto \|\nabla f_0(\boldsymbol{x}) - \nabla f(\boldsymbol{x})\|$ at $\boldsymbol{\mu}_0$ and by taking n large enough (so that δ_n is small enough), we see that

$$\Pi(\|\boldsymbol{\mu}-\boldsymbol{\mu}_0\| \leq h\epsilon_n|\boldsymbol{Y}) \leq \Pi\left[\|\nabla f_0(\boldsymbol{\mu}_0) - \nabla f(\boldsymbol{\mu}_0)\| \leq 2\lambda_1 h\epsilon_n|\boldsymbol{Y}\right] + o_{P_0}(1).$$

To obtain the same rate as the upper bound presented in (4.3) of Theorem 4.1, we then need the lower bound point-wise version of Theorem 9.1, namely for some constant

$m_0 > 0$ and for any $\boldsymbol{x} \in [0,1]^d$,

$$\sup_{\|f_0\|_{\boldsymbol{\alpha},\infty} \leq R} \mathrm{E}_0 \Pi \left(|D^{\boldsymbol{r}} f(\boldsymbol{x}) - D^{\boldsymbol{r}} f_0(\boldsymbol{x})| \leq m_0 n^{-\alpha^* \{1 - \sum_{k=1}^d (r_k/\alpha_k)\}/(2\alpha^*+d)} \middle| \boldsymbol{Y} \right) \to 0. \quad (1)$$

One can proceed to establish such lower bound directly since we have analytical expression for the Gaussian posterior distribution. By taking $\boldsymbol{r} = \boldsymbol{e}_k$ and $h \leq m_0/(2\lambda_1)$, we conclude that $\epsilon_n^2 = \sum_{k=1}^d n^{-2\alpha^*(1-\alpha_k^{-1})/(2\alpha^*+d)} \geq \max_{1\leq k\leq d} n^{-2\alpha^*(1-\alpha_k^{-1})/(2\alpha^*+d)}$. As a result, if one adds an extra lower bound assumption (4.5), we have the lower bound:

$$\mathrm{E}_0 \Pi \left(\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\| \geq h n^{-\alpha^*\{1-(\min_{1\leq k\leq d} \alpha_k)^{-1}\}/(2\alpha^*+d)} \middle| \boldsymbol{Y} \right) \to 1,$$

for a small enough constant $h > 0$. For the posterior lower bound of M, let $\boldsymbol{\mu}^*$ be some point in between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_0$. We Taylor expand f_0 around $\boldsymbol{\mu}_0$, add and subtract M and use the reverse triangle inequality to write

$$\begin{aligned} |M_0 - M| &\geq |f_0(\boldsymbol{\mu}) - f(\boldsymbol{\mu})| + 0.5(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \boldsymbol{H} f_0(\boldsymbol{\mu}^*)(\boldsymbol{\mu} - \boldsymbol{\mu}_0) \\ &\geq |f_0(\boldsymbol{\mu}) - f(\boldsymbol{\mu})| - 0.5\lambda_1 \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2, \end{aligned}$$

by the extra assumption and posterior consistency of $\boldsymbol{\mu}^*$. Choose $m_n = \sqrt{\log\log n}$ and define the set $\mathcal{T} := \{\|\boldsymbol{\mu} - \boldsymbol{\mu}_0\| \leq m_n \epsilon_n\}$. Then for $\omega_n := n^{-\alpha^*/(2\alpha^*+d)}$ and a small enough constant $h > 0$ to be determined below,

$$\begin{aligned} \Pi(|M_0 - M| \leq h\omega_n | \boldsymbol{Y}) &\leq \Pi\left(|f_0(\boldsymbol{\mu}) - f(\boldsymbol{\mu})| - 0.5\lambda_1 \|\boldsymbol{\mu} - \boldsymbol{\mu}_0\|^2 \leq h\omega_n, \mathcal{T} | \boldsymbol{Y} \right) \\ &\quad + \Pi(\mathcal{T}^c | \boldsymbol{Y}) \\ &\leq \Pi(|f_0(\boldsymbol{\mu}) - f(\boldsymbol{\mu})| \leq h\omega_n + 0.5\lambda_1 m_n^2 \epsilon_n^2 | \boldsymbol{Y}) + o_{P_0}(1), \end{aligned}$$

where the last term follows from (4.3) of Theorem 4.1. Using the continuity argument as before for $\boldsymbol{x} \mapsto |f_0(\boldsymbol{x}) - f(\boldsymbol{x})|$ and the fact that $h\omega_n \gg \lambda_1 m_n^2 \epsilon_n^2$ when $\min_{1\leq k\leq d} \alpha_k > 2$, we can further bound the right hand side above by

$$\Pi(|f_0(\boldsymbol{\mu}) - f(\boldsymbol{\mu})| \leq 2h\omega_n | \boldsymbol{Y}) + o_{P_0}(1),$$

for large enough n. By setting $\boldsymbol{r} = \boldsymbol{0}$ in (1) above, we conclude that the first term is $o_{P_0}(1)$ when $h \leq m_0/2$ and the second posterior statement on M is established. □

Proof of Proposition 5.1. By the triangle inequality, $|\widetilde{\sigma}_*^2 - \sigma_0^2| \leq |\widetilde{\sigma}_1^2 - \sigma_0^2| + |\widetilde{\sigma}_2^2 - \sigma_0^2|$. By (a) of Proposition 9.5, the first term is $O_{P_0}(\max\{n^{-1/2}, n^{-2\alpha^*/(2\alpha^*+d)}\})$. To bound the second term, let $\boldsymbol{U} = (\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T + \boldsymbol{I}_n)^{-1}$. By equation (33) of page 355 in Searle [2],

$$\begin{aligned} |\mathrm{E}(\widetilde{\sigma}_2^2 | \boldsymbol{\theta}_0) - \sigma_0^2| &= |n^{-1}\sigma_0^2 \mathrm{tr}(\boldsymbol{U}) - \sigma_0^2| + n^{-1}(\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T \boldsymbol{U} (\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\xi}) \\ &\lesssim n^{-1}[\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{U}) + (\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0)^T \boldsymbol{U} (\boldsymbol{F}_0 - \boldsymbol{Z}\boldsymbol{\theta}_0) \\ &\quad + (\boldsymbol{Z}\boldsymbol{\theta}_0 - \boldsymbol{Z}\boldsymbol{\xi})^T \boldsymbol{U} (\boldsymbol{Z}\boldsymbol{\theta}_0 - \boldsymbol{Z}\boldsymbol{\xi})], \end{aligned} \quad (2)$$

where we have used $(\boldsymbol{x}+\boldsymbol{y})^T\boldsymbol{G}(\boldsymbol{x}+\boldsymbol{y}) \leq 2\boldsymbol{x}^T\boldsymbol{G}\boldsymbol{x}+2\boldsymbol{y}^T\boldsymbol{G}\boldsymbol{y}$ for any matrix $\boldsymbol{G} \geq \boldsymbol{0}$ (Cauchy-Schwarz and the geometric-arithmetic inequalities). Let $\boldsymbol{P}_{\boldsymbol{Z}} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T$ be the orthogonal projection matrix. For matrices $\boldsymbol{Q},\boldsymbol{C},\boldsymbol{T},\boldsymbol{W}$, the binomial inverse theorem (see Theorem 18.2.8 of Harville [1]) says that

$$(\boldsymbol{Q}+\boldsymbol{C}\boldsymbol{T}\boldsymbol{W})^{-1} = \boldsymbol{Q}^{-1} - \boldsymbol{Q}^{-1}\boldsymbol{C}(\boldsymbol{T}^{-1}+\boldsymbol{W}\boldsymbol{Q}^{-1}\boldsymbol{C})^{-1}\boldsymbol{W}\boldsymbol{Q}^{-1}.$$

Applying the above twice to $\boldsymbol{U}$ yields

$$(\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T+\boldsymbol{I}_n)^{-1} = \boldsymbol{I}_n - \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z}+\boldsymbol{V}^{-1})^{-1}\boldsymbol{Z}^T = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{Z}} + \boldsymbol{M}, \tag{3}$$

where $\boldsymbol{M} = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}[\boldsymbol{V}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T \geq \boldsymbol{0}$. Hence the first term in (2) is $n^{-1}\mathrm{tr}(\boldsymbol{P}_{\boldsymbol{Z}}-\boldsymbol{M}) \leq n^{-1}\mathrm{tr}(\boldsymbol{P}_{\boldsymbol{Z}}) = (W+1)/n$. Note that $\boldsymbol{U} \leq \boldsymbol{I}_n$ since $\boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^T \geq \boldsymbol{0}$, and the second term in (2) is bounded by

$$n^{-1}\|\boldsymbol{U}\|_{(2,2)}\|\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\|^2 \leq \|\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\|_\infty^2 \lesssim \sum_{k=1}^d \delta_{n,k}^{2\alpha_k},$$

in view of (8.3). By (3) and $(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{Z} = \boldsymbol{0}$, the last term in (2) is $n^{-1}(\boldsymbol{\theta}_0-\boldsymbol{\xi})^T[\boldsymbol{V}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1}(\boldsymbol{\theta}_0-\boldsymbol{\xi}) \leq n^{-1}\sum_{j=0}^W \boldsymbol{\delta}_n^{\boldsymbol{i}_j}(\theta_{0,\boldsymbol{i}_j}-\xi_{\boldsymbol{i}_j})^2 = O_{P_0}(n^{-1})$, since $\delta_{n,k}=o(1), k=1,\ldots,d$, $\theta_{0,\boldsymbol{i}_j} = O_{P_0}(1)$ and $\xi_{\boldsymbol{i}_j} = O(1)$ by assumption on the prior for any $0 \leq j \leq W$. Combining the three bounds established into (2), we obtain $|\mathrm{E}(\widetilde{\sigma}_2^2|\boldsymbol{\theta}_0)-\sigma_0^2| \lesssim n^{-1}+\sum_{k=1}^d \delta_{n,k}^{2\alpha_k}$.

We write $n\widetilde{\sigma}_2^2 = (\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\xi}) + 2(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{U}\boldsymbol{\varepsilon}$ by substituting $\boldsymbol{Y} = \boldsymbol{F}_0+\boldsymbol{\varepsilon}$. Observe that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\theta}_0$ are independent by definition. Using the inequality $\mathrm{Var}(A_1+A_2) \leq 2\mathrm{Var}(A_1)+2\mathrm{Var}(A_2)$ (from Cauchy-Schwarz and geometric-arithmetic inequalities), we conclude that $\mathrm{Var}(\widetilde{\sigma}_2^2|\boldsymbol{\theta}_0)$ is bounded up to a constant multiple by

$$n^{-2}[(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0)^T\boldsymbol{U}^2(\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0) + (\boldsymbol{Z}\boldsymbol{\theta}_0-\boldsymbol{Z}\boldsymbol{\xi})^T\boldsymbol{U}^2(\boldsymbol{Z}\boldsymbol{\theta}_0-\boldsymbol{Z}\boldsymbol{\xi}) + \mathrm{Var}(\boldsymbol{\varepsilon}^T\boldsymbol{U}\boldsymbol{\varepsilon})]. \tag{4}$$

In view of (8.3) and $\boldsymbol{U} \leq \boldsymbol{I}_n$, the first term is bounded by $n^{-2}\|\boldsymbol{U}\|_{(2,2)}^2\|\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\|^2 \leq n^{-1}\|\boldsymbol{F}_0-\boldsymbol{Z}\boldsymbol{\theta}_0\|_\infty^2 \lesssim n^{-1}\sum_{k=1}^d \delta_{n,k}^{2\alpha_k}$. Observe that since $\boldsymbol{V} \geq \boldsymbol{0}$,

$$\begin{aligned}\boldsymbol{Z}^T\boldsymbol{M}^2\boldsymbol{Z} &= [\boldsymbol{V}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}[\boldsymbol{V}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1} \\ &\leq [\boldsymbol{V}+(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}]^{-1} \leq \boldsymbol{Z}^T\boldsymbol{Z}.\end{aligned} \tag{5}$$

Using (3), idempotency of $\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{Z}}$ and $(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{Z}})\boldsymbol{Z} = \boldsymbol{0}$, the second term in (4) is $n^{-2}(\boldsymbol{\theta}_0-\boldsymbol{\xi})^T\boldsymbol{Z}^T(\boldsymbol{I}_n-\boldsymbol{P}_{\boldsymbol{Z}}+\boldsymbol{M})^2\boldsymbol{Z}(\boldsymbol{\theta}_0-\boldsymbol{\xi})$, which is

$$n^{-2}(\boldsymbol{\theta}_0-\boldsymbol{\xi})^T\boldsymbol{Z}^T\boldsymbol{M}^2\boldsymbol{Z}(\boldsymbol{\theta}_0-\boldsymbol{\xi}) \leq n^{-2}(\boldsymbol{\theta}_0-\boldsymbol{\xi})^T\boldsymbol{Z}^T\boldsymbol{Z}(\boldsymbol{\theta}_0-\boldsymbol{\xi}), \tag{6}$$

in view of (5). By (8.4) in the proof of Lemma 8.1, we can write $\boldsymbol{Z}^T\boldsymbol{Z} = n_2\boldsymbol{\Delta}\boldsymbol{A}\boldsymbol{\Delta}$ where $\boldsymbol{\Delta} = \mathrm{diag}\{\boldsymbol{\delta}_n^{\boldsymbol{i}_j} : j=0,\ldots,W\}$ and $\boldsymbol{A} \to \mathrm{E}\mathbb{U}\mathbb{U}^T$ in probability entry-wise, where $\mathbb{U} = (\boldsymbol{U}^{\boldsymbol{i}_0},\ldots,\boldsymbol{U}^{\boldsymbol{i}_W})^T$ for $\boldsymbol{U} = (U_1,\ldots,U_d)^T \sim \mathrm{Uniform}[-1,1]^d$. This gives $\|\boldsymbol{A}\|_{(2,2)} \to$

$\|\mathrm{E}\mathbb{U}\mathbb{U}^T\|_{(2,2)}$ in probability. The entries of $\mathrm{E}\mathbb{U}\mathbb{U}^T$ are mixed moments of Uniform$[-1,1]$ and hence the matrix is nonsingular with $\|\mathrm{E}\mathbb{U}\mathbb{U}^T\|_{(2,2)} < \infty$. Since $\|\boldsymbol{\Delta}\|_{(2,2)} = 1$ and $n_2 \leq n$, the right hand side of (6) is bounded by

$$n_2 n^{-2}\|\boldsymbol{A}\|_{(2,2)}\|\boldsymbol{\Delta}\|^2_{(2,2)}\|\boldsymbol{\theta}_0 - \boldsymbol{\xi}\|^2 = O_{P_0}(n^{-1}),$$

because $\|\boldsymbol{\theta}_0 - \boldsymbol{\xi}\| \leq \|\boldsymbol{\theta}_0\| + \|\boldsymbol{\xi}\| = O_{P_0}(1)$. By Lemma A.10 of Yoo and Ghosal [3] with $\|\boldsymbol{U}\|_{(2,2)} \leq 1$ and Gaussian errors by Assumption 1, the last term in (4) is $O(1/n)$. Combining this with the three bounds established above, we obtain $\mathrm{Var}(\widetilde{\sigma}_2^2|\boldsymbol{\theta}_0) = O_{P_0}(1/n)$. Therefore, the mean square error is $\mathrm{E}_0(\widetilde{\sigma}_2^2 - \sigma_0^2)^2 = \mathrm{E}\{\mathrm{E}[(\widetilde{\sigma}_2^2 - \sigma_0^2)^2|\boldsymbol{\theta}_0]\} \lesssim n^{-1} + \sum_{k=1}^d \delta_{n,k}^{4\alpha_k}$.

To prove (b), observe that $\mathrm{E}(\sigma^2|\boldsymbol{Y}) \lesssim n^{-1} + \widetilde{\sigma}_*^2$ and $\mathrm{Var}(\sigma^2|\boldsymbol{Y}) \lesssim n^{-3} + n^{-1}\widetilde{\sigma}_*^4$. Therefore by Markov's inequality, the second stage posterior of σ^2 concentrates around the second stage empirical Bayes estimator $\widetilde{\sigma}_*^2$, and thus (b) will inherit the rate from (a) as established above. □

Proof of Corollary 8.4. By (8.7), we have

$$\begin{aligned}\|D^{\boldsymbol{r}} f_{\boldsymbol{\theta}} - D^{\boldsymbol{r}} f_{\boldsymbol{\theta}_0}\|_\infty &= \sup_{\boldsymbol{x}\in\mathcal{Q}} |D^{\boldsymbol{r}} f_{\boldsymbol{\theta}}(\boldsymbol{x}) - D^{\boldsymbol{r}} f_{\boldsymbol{\theta}_0}(\boldsymbol{x})| \\ &\lesssim |\theta_{\boldsymbol{r}} - \theta_{0,\boldsymbol{r}}| + \sum_{\boldsymbol{r}\leq \boldsymbol{i}\leq \boldsymbol{m_\alpha}, \boldsymbol{i}\neq \boldsymbol{r}} |\theta_{\boldsymbol{i}} - \theta_{0,\boldsymbol{i}}|\boldsymbol{\delta}_n^{\boldsymbol{i}-\boldsymbol{r}}. \end{aligned} \tag{7}$$

Hence, the upper bound (8.8) is applicable and uniformly over $\|f_0\|_{\boldsymbol{\alpha},\infty} \leq R$, we will have $\mathrm{E}_0 \sup_{\sigma^2\in\mathcal{K}_n} \mathrm{E}[\|D^{\boldsymbol{r}} f_{\boldsymbol{\theta}} - D^{\boldsymbol{r}} f_{\boldsymbol{\theta}_0}\|_\infty^2|\boldsymbol{Y}, \sigma^2] \lesssim \delta_n^{-2\boldsymbol{r}}(n^{-1} + \sum_{k=1}^d \delta_{n,k}^{2\alpha_k})$. Moreover, since the bound in (8.9) is uniform for all $\boldsymbol{x} \in \mathcal{Q}$, this implies that $\mathrm{E}_0\|D^{\boldsymbol{r}} f_{\boldsymbol{\theta}_0} - D^{\boldsymbol{r}} f_{0,\boldsymbol{z}}\|_\infty^2 \lesssim \sum_{k=1}^d \delta_{n,k}^{2\alpha_k - 2r_k}$. Therefore, we conclude that uniformly over $\|f_0\|_{\boldsymbol{\alpha},\infty} \leq R$,

$$\begin{aligned}&\mathrm{E}_0 \sup_{\sigma^2\in\mathcal{K}_n} \mathrm{E}[\|D^{\boldsymbol{r}} f_{\boldsymbol{\theta}} - D^{\boldsymbol{r}} f_{0,\boldsymbol{z}}\|_\infty^2|\boldsymbol{Y}, \sigma^2] \\ &\quad\lesssim \mathrm{E}_0 \sup_{\sigma^2\in\mathcal{K}_n} \mathrm{E}[\|D^{\boldsymbol{r}} f_{\boldsymbol{\theta}} - D^{\boldsymbol{r}} f_{\boldsymbol{\theta}_0}\|_\infty^2|\boldsymbol{Y}, \sigma^2] + \mathrm{E}_0\|D^{\boldsymbol{r}} f_{\boldsymbol{\theta}_0} - D^{\boldsymbol{r}} f_{0,\boldsymbol{z}}\|_\infty^2 \\ &\quad\lesssim \boldsymbol{\delta}_n^{-2\boldsymbol{r}}\left(\frac{1}{n} + \sum_{k=1}^d \delta_{n,k}^{2\alpha_k}\right).\end{aligned}$$

The empirical and hierarchical posterior contraction rates then follow from (8.10) and (8.11) with absolute values replaced by sup-norms. □

References

[1] Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag New York, Inc.

[2] Searle, S. R. (1982). *Matrix Algebra Useful for Statistics.* John Wiley and Sons, Inc.

[3] Yoo, W. W. and Ghosal, S. (2016). Supremum norm posterior contraction and credible sets for nonparametric multivariate regression. *Ann. Statist.*, 44(3):1069–1102.

[4] Yoo, W. W. and Ghosal, S. (2018). Bayesian mode and maximum estimation and accelerated rates of contraction. To apppear in Bernoulli.