Supplementary Material for "Bayesian Projected Calibration of Computer Models"

A Proof of Theorem 1

We first present a classic result regarding convergence rate of the Matérn Gaussian process regression from van der Vaart and Zanten (2011).

Theorem A.1. *Suppose η is imposed the Matérn Gaussian process with a smoothness parameter α, and $\eta_0 \in \mathfrak{C}_\alpha(\Omega)\cap\mathcal{H}_\alpha(\Omega)$, where $\alpha > p/2$. Then there exists some constant $C > 0$, such that*

$$\mathbb{E}_0\left\{\int_\Omega \left[\|\eta-\eta_0\|^2_{L_2(\Omega)}\right]\Pi(\mathrm{d}\eta \mid \mathcal{D}_n)\right\} \leq Cn^{-2\alpha/(2\alpha+p)}. \tag{A.1}$$

The first assertion follows immediately from the Markov's inequality:

$$\begin{aligned}
&\mathbb{E}_0\left[\Pi\left(\|\eta-\eta_0\|_{L_2(\Omega)} > M_n n^{-\alpha/(2\alpha+p)} \mid \mathcal{D}_n\right)\right]\\
&\quad\leq \frac{1}{M_n^2 n^{-2\alpha/(2\alpha+p)}}\mathbb{E}_0\left\{\int_\Omega \left[\|\eta-\eta_0\|^2_{L_2(\Omega)}\right]\Pi(\mathrm{d}\eta \mid \mathcal{D}_n)\right\}\\
&\quad\leq \frac{C}{M_n^2} \to 0.
\end{aligned}$$

The posterior distribution of η can be expressed by

$$\Pi(\eta\in\mathcal{U} \mid \mathcal{D}_n) = \left[\int_{\mathcal{U}}\prod_{i=1}^n \frac{p_\eta(y_i,\mathbf{x}_i)}{p_0(y_i,\mathbf{x}_i)}\Pi(\mathrm{d}\eta)\right]\left[\int\prod_{i=1}^n \frac{p_\eta(y_i,\mathbf{x}_i)}{p_0(y_i,\mathbf{x}_i)}\Pi(\mathrm{d}\eta)\right]^{-1},$$

where $p_0(y_i,\mathbf{x}_i) = \psi_\sigma(y_i - \eta_0(\mathbf{x}_i))$ is the density of the true distribution. To prove the second assertion, we need the following result from Xie et al. (2017) to bound the denominator of the preceding display.

Lemma A.1. *Assume the conditions of Theorem 1 hold. For any $D > 0$, define the event*

$$\mathcal{H}_n = \left\{\int\prod_{i=1}^n \frac{p_\eta(y_i,\mathbf{x}_i)}{p_0(y_i,\mathbf{x}_i)}\Pi(\mathrm{d}\eta) \geq \Pi(\|\eta-\eta_0\|_{L_\infty(\Omega)} < \epsilon_n)\exp\left[-\left(D+\frac{1}{\sigma^2}\right)n\epsilon_n^2\right]\right\}.$$

Suppose $(\epsilon_n)_{n=1}^\infty$ is a sequence such that $n\epsilon_n^2\to\infty$ and $\epsilon_n \to 0$. Then $\mathbb{P}_0(\mathcal{H}_n^c)\to 0$.

Since $\alpha > p/2$, there exists some positive β such that $\beta \in (\max\{\underline{\alpha}, p/2\}, \alpha)$. Define $\epsilon_n = n^{-\beta/(2\beta+p)}$. Since the Matérn Gaussian process assigns prior probability one to the space $\mathfrak{C}_\beta(\Omega)$ (see, for example, section 3.1 in van der Vaart and Zanten, 2011), then the Gaussian process prior on η can be regarded as a mean-zero Gaussian random element in the Banach space $\mathfrak{C}_\beta(\Omega)$ equipped with the β-Hölder norm $\|\cdot\|_{\mathfrak{C}_\beta(\Omega)}$. Therefore by the Borell's inequality (see, for example, Ghosal and van der Vaart, 2017) it holds that

$$\Pi\left(\|\eta\|_{\mathfrak{C}_\beta(\Omega)} > 4x\left[\int \|\eta\|^2_{\mathfrak{C}_\beta(\Omega)}\Pi(\mathrm{d}\eta)\right]^{1/2}\right) \leq 2\mathrm{e}^{-2x^2}. \tag{A.2}$$

for any positive x.

By Lemma 15 in van der Vaart and Zanten (2011) there exists a constant $\tilde{M} > 0$ such that $\|f\|_{L_\infty(\Omega)} \leq \tilde{M}\|f\|^{p/(2\beta+p)}_{\mathfrak{C}_\beta(\Omega)}\|f\|^{2\beta/(2\beta+p)}_{L_2(\Omega)}$ for any function $f \in \mathfrak{C}_\beta(\Omega)$. Let $s > 0$ be a constant determined later. Then

$$\begin{aligned}
&\left\{\|\eta - \eta_0\|_{L_2(\Omega)} \leq M_n n^{-\alpha/(2\alpha+p)}\right\} \cap \left\{\|\eta\|_{\mathfrak{C}_\beta(\Omega)} \leq 4s\sqrt{n}\epsilon_n\left[\int \|\eta\|^2_{\mathfrak{C}_\beta(\Omega)}\Pi(\mathrm{d}\eta)\right]^{1/2}\right\}\\
&\quad\subset \left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq \tilde{M}\|\eta - \eta_0\|^{p/(2\beta+p)}_{\mathfrak{C}_\beta(\Omega)} M_n^{2\beta/(2\beta+p)} n^{-(2\alpha\beta)/[(2\alpha+p)(2\beta+p)]}\right\}\\
&\qquad\cap \left\{\|\eta\|_{\mathfrak{C}_\beta(\Omega)} \leq 4s\sqrt{n}\epsilon_n\left[\int \|\eta\|^2_{\mathfrak{C}_\beta(\Omega)}\Pi(\mathrm{d}\eta)\right]^{1/2}\right\}\\
&\quad\subset \left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq \tilde{M}\left(\|\eta\|_{\mathfrak{C}_\beta(\Omega)} + \|\eta_0\|_{\mathfrak{C}_\beta(\Omega)}\right)^{p/(2\beta+p)} M_n^{2\beta/(2\beta+p)} n^{-2\alpha\beta/[(2\alpha+p)(2\beta+p)]}\right\}\\
&\qquad\cap \left\{\|\eta\|_{\mathfrak{C}_\beta(\Omega)} \leq 4s\sqrt{n}\epsilon_n\left[\int \|\eta\|^2_{\mathfrak{C}_\beta(\Omega)}\Pi(\mathrm{d}\eta)\right]^{1/2}\right\}\\
&\quad\subset \left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M_1 M_n^{2\beta/(2\beta+p)} n^{-2\alpha\beta/[(2\alpha+p)(2\beta+p)]} n^{p^2/[2(2\beta+p)^2]}\right\}
\end{aligned}$$

for some constant $M_1 > 0$ depending on η_0 only when n is sufficiently large. Note that $-\alpha/(2\alpha+p) < -\beta/(2\beta+p)$, then taking $M_n = \log n$ yields

$$\begin{aligned}
&\left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M_1 M_n^{2\beta/(2\beta+p)} n^{-2\alpha\beta/[(2\alpha+p)(2\beta+p)]} n^{p^2/[2(2\beta+p)^2]}\right\}\\
&\quad\subset \left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M_1(\log n)^{2\beta/(2\beta+p)} n^{-(2\beta^2-p^2/2)/(2\beta+p)^2}\right\}\\
&\quad\subset \left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M\right\}
\end{aligned}$$

for some constant $M > 0$, where $\beta > p/2$ is applied. Since by the first assertion $\Pi(\|\eta - \eta_0\|_{L_2(\Omega)} \leq M_n n^{-\alpha/(2\alpha+p)} \mid \mathcal{D}_n) = 1 - o_{\mathbb{P}_0}(1)$, it suffices to show that $\mathbb{E}_0\left[\Pi(\mathcal{U}_n \mid \mathcal{D}_n)\right] \to 0$, where $\mathcal{U}_n$ is the event

$$\mathcal{U}_n = \left\{ \|\eta\|_{\mathfrak{C}_\beta(\Omega)} > 4s\sqrt{n}\epsilon_n \left[\int \|\eta\|^2_{\mathfrak{C}_\beta(\Omega)} \Pi(\mathrm{d}\eta)\right]^{1/2} \right\}.$$

The following argument is quite similar to that of Lemma 1 in Ghosal and van der Vaart (2007) and is included here for completeness. Let $\mathcal{H}_n$ be defined as in Lemma A.1 with the constant D be determined later. Then $\mathbb{P}_0(\mathcal{H}_n^c) \to 0$, and we directly compute by Fubini's theorem

$$\begin{aligned}
\mathbb{E}_0\left[\Pi\left(\mathcal{U}_n \mid \mathcal{D}_n\right)\right] &\leq \mathbb{E}_0\left[\mathbb{1}(\mathcal{H}_n)\Pi\left(\mathcal{U}_n \mid \mathcal{D}_n\right)\right] + \mathbb{P}_0(\mathcal{H}_n^c) \\
&= \mathbb{E}_0\left\{\mathbb{1}(\mathcal{H}_n)\left[\int \prod_{i=1}^n \frac{p_\eta(y_i, \mathbf{x}_i)}{p_0(y_i, \mathbf{x}_i)}\Pi(\mathrm{d}\eta)\right]^{-1}\left[\int_{\mathcal{U}_n} \prod_{i=1}^n \frac{p_\eta(y_i, \mathbf{x}_i)}{p_0(y_i, \mathbf{x}_i)}\Pi(\mathrm{d}\eta)\right]\right\} + o(1) \\
&\leq \frac{\exp[(D + 1/\sigma^2)n\epsilon_n^2]}{\Pi(\|\eta - \eta_0\|_{L_\infty(\Omega)} < \epsilon_n)} \int_{\mathcal{U}_n} \mathbb{E}_0\left[\prod_{i=1}^n \frac{p_\eta(y_i, \mathbf{x}_i)}{p_0(y_i, \mathbf{x}_i)}\right]\Pi(\mathrm{d}\eta) + o(1) \\
&\leq \frac{\exp[(D + 1/\sigma^2)n\epsilon_n^2]\Pi(\mathcal{U}_n)}{\Pi(\|\eta - \eta_0\|_{L_\infty(\Omega)} < \epsilon_n)} + o(1).
\end{aligned}$$

By Lemma 3 and Lemma 4 in van der Vaart and Zanten (2011) we have

$$\Pi(\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq \epsilon_n) \geq \exp\left(-C\epsilon_n^{-p/\alpha}\right) \geq \exp\left(-Cn^{p\beta/[\alpha(2\beta+p)]}\right)$$

for some constant $C > 0$. Now take $D = 1/(2\sigma^2)$, $s = 1/\sigma$, and we conclude

$$\begin{aligned}
\mathbb{E}_0\{\Pi(\mathcal{U}_n \mid \mathcal{D}_n)\} &\leq \exp\left(\frac{3}{2\sigma^2}n\epsilon_n^2 + Cn^{p\beta/[\alpha(2\beta+p)]}\right)\Pi(\mathcal{U}_n) + o(1) \\
&\leq 2\exp\left(\frac{3}{2\sigma^2}n\epsilon_n^2 + Cn^{(p\beta/[\alpha(2\beta+p)]} - \frac{2}{\sigma^2}n\epsilon_n^2\right) + o(1) \to 0,
\end{aligned}$$

where the last inequality is due to (A.2) and the fact $\beta < \alpha$.

B Proof of Lemma 1

We first prove the first assertion, *i.e.*, the Taylor's expansion of $\boldsymbol{\theta}_\eta^*$ with respect to η. Recall that $\boldsymbol{\theta}_\eta^* = \arg\min_{\boldsymbol{\theta}\in\Theta} \|\eta(\cdot) - y^s(\cdot, \boldsymbol{\theta})\|^2_{L_2(\Omega)}$. Since by condition A4 it is permitted to interchange the differentiation with respect to $\boldsymbol{\theta}$ and the integral, it follows that

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{\theta}}\|\eta(\cdot) - y^s(\cdot, \boldsymbol{\theta})\|^2_{L_2(\Omega)}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\eta^*} = -2\int_\Omega \left[\eta(\mathbf{x}) - y^s(\mathbf{x}, \boldsymbol{\theta}_\eta^*)\right]\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_\eta^*)\mathrm{d}\mathbf{x}.$$

Now define the function $\mathbf{F} : \mathcal{F} \times \Theta \to \mathbb{R}^q$ by

$$\mathbf{F}(\eta, \boldsymbol{\theta}) = -2 \int_\Omega [\eta(\mathbf{x}) - y^s(\mathbf{x}, \boldsymbol{\theta})] \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \mathrm{d}\mathbf{x}.$$

It follows immediately that $\mathbf{F}(\eta, \boldsymbol{\theta}^*_\eta) = \mathbf{0}$. The partial derivative of F with respect to $\boldsymbol{\theta}$ is given by

$$\mathbf{F}_{\boldsymbol{\theta}}(\eta, \boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{F}(\eta, \boldsymbol{\theta}) = \int_\Omega \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} [\eta(\mathbf{x}) - y^s(\mathbf{x}, \boldsymbol{\theta})]^2 \right\} \mathrm{d}\mathbf{x},$$

and the partial Fréchet derivative of $\mathbf{F}$ with respect to η is a function $\mathbf{F}_\eta : \mathcal{F} \to \mathbb{R}^q$ given by

$$[\mathbf{F}_\eta(\eta, \boldsymbol{\theta})](h) = -2 \int_\Omega h(\mathbf{x}) \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \mathrm{d}\mathbf{x},$$

since F is linear with respect to η. Therefore by the implicit function theorem on Banach space, there exists some $\epsilon > 0$ such that over $\{\eta \in \mathcal{F} : \|\eta - \eta_0\|_{L_2(\Omega)} < \epsilon\}$, the functional $\boldsymbol{\theta}^*_\eta : \eta \mapsto \arg\min_{\boldsymbol{\theta} \in \Theta} \|\eta(\cdot) - y^s(\cdot, \boldsymbol{\theta})\|^2_{L_2(\Omega)}$ is continuous, the Fréchet derivative $\dot{\boldsymbol{\theta}}^*_\eta : \mathcal{F} \to \mathbb{R}^q$ for $\boldsymbol{\theta}^*_\eta$ exists, and can be computed by

$$\dot{\boldsymbol{\theta}}^*_\eta(h) = -\left[\mathbf{F}_\theta(\eta, \boldsymbol{\theta}^*_\eta)\right]^{-1} \left[\mathbf{F}_\eta(\eta, \boldsymbol{\theta}^*_\eta)\right](h) = 2\mathbf{V}_\eta^{-1} \int_\Omega h(\mathbf{x}) \frac{\partial y^s}{\partial \boldsymbol{\theta}}(x, \boldsymbol{\theta}^*_\eta) \mathrm{d}\mathbf{x}.$$

Therefore we obtain by the fundamental theorem of calculus and the mean-value theorem that

$$\begin{aligned}
\boldsymbol{\theta}^*_\eta - \boldsymbol{\theta}^*_0 &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}u} \boldsymbol{\theta}^*_{\eta[u]} \mathrm{d}u \\
&= \int_0^1 \dot{\boldsymbol{\theta}}^*_{\eta[u]} \left(\frac{\mathrm{d}}{\mathrm{d}u} \eta[u] \right) \mathrm{d}u \\
&= 2 \int_0^1 \mathbf{V}^{-1}_{\eta[u]} \int_\Omega [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u]}) \mathrm{d}\mathbf{x} \mathrm{d}u \\
&= 2 \int_\Omega [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{V}^{-1}_{\eta[u']} \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u']}) \mathrm{d}\mathbf{x},
\end{aligned}$$

where $\eta[u] = \eta_0 + (\eta - \eta_0)u$ for $0 \le u \le 1$ and $u' \in [0, 1]$. By condition A3, we know that the smallest eigenvalue $\lambda_{\min}(\mathbf{V}_\eta)$ of $\mathbf{V}_\eta$ is strictly positive in an L_2-neighborhood of η_0, and we can without loss of generality require that $\inf_{\|\eta - \eta_0\|_{L_2(\Omega)} \le \epsilon} \lambda_{\min}(\mathbf{V}_\eta) > 0$. Hence we proceed by condition A4 and Jensen's inequality that

$$\|\boldsymbol{\theta}^*_\eta - \boldsymbol{\theta}^*_0\| \le 2 \sup_{\|\eta - \eta_0\|_{L_2(\Omega)} \le \epsilon} \left\|\mathbf{V}_\eta^{-1}\right\| \sup_{(\mathbf{x}, \boldsymbol{\theta}) \in \Omega \times \Theta} \left\| \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \right\| \int_\Omega |\eta(\mathbf{x}) - \eta_0(\mathbf{x})| \mathrm{d}\mathbf{x}$$

$$
\begin{aligned}
&\leq 2\left[\inf_{\|\eta-\eta_0\|_{L_2(\Omega)}\leq\epsilon}\lambda_{\min}(\mathbf{V}_\eta)\right]^{-1}\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\right\|\left\{\int_\Omega[\eta(\mathbf{x})-\eta_0(\mathbf{x})]^2\mathrm{d}\mathbf{x}\right\}^{1/2}\\
&= L_{\eta_0}^{(1)}\|\eta-\eta_0\|_{L_2(\Omega)}
\end{aligned}
$$

for some constant $L_{\eta_0}^{(1)}>0$ depending on η_0 only.

We now analyze the property of $\mathbf{V}_\eta$ as a functional $\{\eta\in\mathcal{F}:\|\eta-\eta_0\|_{L_2(\Omega)}<\epsilon\}\to\in\mathbb{R}^{q\times q}$, $\eta\mapsto\mathbf{V}_\eta$. For a matrix $\mathbf{A}\in\mathbb{R}^{q\times q}$, denote $[\mathbf{A}]_{ij}$ to be the (i,j)-th element of $\mathbf{A}$. Directly compute

$$
\begin{aligned}
[\mathbf{V}_\eta]_{jk}-[\mathbf{V}_0]_{jk} &= 2\int_\Omega\left[\frac{\partial y^s}{\partial\theta_j}(\mathbf{x},\boldsymbol{\theta}_\eta^*)\frac{\partial y^s}{\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_\eta^*)-\frac{\partial y^s}{\partial\theta_j}(\mathbf{x},\boldsymbol{\theta}_0^*)\frac{\partial y^s}{\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_0^*)\right]\mathrm{d}\mathbf{x}\\
&\quad-2\int_\Omega\left\{[\eta(\mathbf{x})-y^s(\mathbf{x},\boldsymbol{\theta}_\eta^*)]\left[\frac{\partial^2 y^s}{\partial\theta_j\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_\eta^*)-\frac{\partial^2 y^s}{\partial\theta_j\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_0^*)\right]\right\}\mathrm{d}\mathbf{x}\\
&\quad-2\int_\Omega\left\{[\eta(\mathbf{x})-\eta_0(\mathbf{x})+y^s(\mathbf{x},\boldsymbol{\theta}_0^*)-y^s(\mathbf{x},\boldsymbol{\theta}_\eta^*)]\frac{\partial^2 y^s}{\partial\theta_j\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_0^*)\right\}\mathrm{d}\mathbf{x}\\
&:= 2V_1-2V_2-2V_3.
\end{aligned}
$$

For V_1, by condition A4 we know that $\partial y^s/\partial\boldsymbol{\theta}$ is Lipschitz continuous on $\Omega\times\Theta$, and therefore

$$
\begin{aligned}
|V_1| &\leq \int_\Omega\left|\frac{\partial y^s}{\partial\theta_j}(\mathbf{x},\boldsymbol{\theta}_\eta^*)\right|\left|\frac{\partial y^s}{\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_\eta^*)-\frac{\partial y^s}{\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_0^*)\right|\mathrm{d}\mathbf{x}\\
&\quad+\int_\Omega\left|\frac{\partial y^s}{\partial\theta_j}(\mathbf{x},\boldsymbol{\theta}_\eta^*)-\frac{\partial y^s}{\partial\theta_j}(\mathbf{x},\boldsymbol{\theta}_0^*)\right|\left|\frac{\partial y^s}{\partial\theta_k}(\mathbf{x},\boldsymbol{\theta}_0^*)\right|\mathrm{d}\mathbf{x}\\
&\leq \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\right\|\left[\left\|\frac{\partial y^s}{\partial\theta_k}(\cdot,\boldsymbol{\theta}_\eta^*)-\frac{\partial y^s}{\partial\theta_k}(\cdot,\boldsymbol{\theta}_0^*)\right\|_{L_1(\Omega)}+\left\|\frac{\partial y^s}{\partial\theta_j}(\cdot,\boldsymbol{\theta}_\eta^*)-\frac{\partial y^s}{\partial\theta_j}(\cdot,\boldsymbol{\theta}_0^*)\right\|_{L_1(\Omega)}\right]\\
&\leq 2\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\right\|\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial^2 y^s}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}(\mathbf{x},\boldsymbol{\theta})\right\|\|\boldsymbol{\theta}_\eta^*-\boldsymbol{\theta}_0^*\|\\
&\lesssim \|\eta-\eta_0\|_{L_2(\Omega)}.
\end{aligned}
$$

Condition A4 also implies that $\partial^2 y^s/(\partial\theta_j\partial\theta_k)$ is Lipschitz continuous on $\Omega\times\Theta$. Hence

$$
\begin{aligned}
|V_2| &\lesssim \int_\Omega[|\eta(\mathbf{x})-\eta_0(\mathbf{x})|+|\eta_0(\mathbf{x})-y^s(\mathbf{x},\boldsymbol{\theta}_\eta^*)]\|\boldsymbol{\theta}_\eta^*-\boldsymbol{\theta}_0^*\|\mathrm{d}\mathbf{x}\\
&\leq L_{\eta_0}^{(1)}\|\eta-\eta_0\|_{L_2(\Omega)}\left\{2\int_\Omega[\eta(\mathbf{x})-\eta_0(\mathbf{x})]^2\mathrm{d}\mathbf{x}+2\int_\Omega[\eta_0(\mathbf{x})-y^s(\mathbf{x},\boldsymbol{\theta}_\eta^*)]^2\mathrm{d}\mathbf{x}\right\}^{1/2}\\
&\leq L_{\eta_0}^{(1)}\|\eta-\eta_0\|_{L_2(\Omega)}\left(2\epsilon^2+4\|\eta_0\|_{L_2(\Omega)}^2+4\sup_{\boldsymbol{\theta}\in\Theta}\|y^s(\cdot,\boldsymbol{\theta})\|_{L_2(\Omega)}^2\right)^{1/2}
\end{aligned}
$$

$$\lesssim L_{\eta_0}^{(1)}\|\eta-\eta_0\|_{L_2(\Omega)}.$$

Now we consider V_3:

$$\begin{aligned}
|V_3| &\leq \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left|\frac{\partial^2 y^s}{\partial\theta_j\partial\theta_k}\right|\int_\Omega[|\eta(\mathbf{x})-\eta_0(\mathbf{x})|+|y^s(\mathbf{x},\boldsymbol{\theta}_0)-y^s(\mathbf{x},\boldsymbol{\theta}_\eta^*)|]\mathrm{d}\mathbf{x}\\
&\leq \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left|\frac{\partial^2 y^s}{\partial\theta_j\partial\theta_k}\right|\left[\|\eta-\eta_0\|_{L_2(\Omega)}+\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\right\|\|\boldsymbol{\theta}_\eta^*-\boldsymbol{\theta}_0^*\|\mathrm{d}x\right]\\
&\leq \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left|\frac{\partial^2 y^s}{\partial\theta_j\partial\theta_k}\right|\left[1+\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta})\right\|L_{\eta_0}^{(1)}\right]\|\eta-\eta_0\|_{L_2(\Omega)}.
\end{aligned}$$

We conclude that $|[\mathbf{V}_\eta]_{jk}-[\mathbf{V}_0]_{jk}|\leq C_{\eta_0}\|\eta-\eta_0\|_{L_2(\Omega)}$ for all $j,k=1,\ldots,q$ for some constant $C_{\eta_0}>0$ depending on η_0 only. By the fact that

$$\sum_{j=1}^q|\lambda_j(\mathbf{A})-\lambda_j(\mathbf{B})|\leq\|\mathbf{A}-\mathbf{B}\|_F^2$$

holds for any positive definite matrices $\mathbf{A},\mathbf{B}\in\mathbb{R}^{q\times q}$ (see, for example, Hoffman and Wielandt, 2003), we obtain

$$|\lambda_{\min}(\mathbf{V}_\eta)-\lambda_{\min}(\mathbf{V}_0)|\leq\|\mathbf{V}_\eta-\mathbf{V}_0\|_F^2=\sum_{j=1}^q\sum_{k=1}^q|[\mathbf{V}_\eta]_{jk}-[\mathbf{V}_0]_{jk}|^2\leq q^2C_{\eta_0}^2\|\eta-\eta_0\|_{L_2(\Omega)}^2.$$

We may also assume without loss of generality that ϵ is sufficiently small such that $|\lambda_{\min}(\mathbf{V}_\eta)-\lambda_{\min}(\mathbf{V}_0)|\leq\lambda_{\min}(\mathbf{V}_0)/2$ whenever $\|\eta-\eta_0\|_{L_2(\Omega)}\leq\epsilon$, in which case it holds that $\|\mathbf{V}_\eta^{-1}\|\geq 2\|\mathbf{V}_0^{-1}\|$. Hence

$$\begin{aligned}
\left\|\mathbf{V}_\eta^{-1}-\mathbf{V}_0^{-1}\right\| &= \left\|\mathbf{V}_0^{-1}(\mathbf{V}_0-\mathbf{V}_\eta)\mathbf{V}_\eta^{-1}\right\|\\
&\leq \left\|\mathbf{V}_0^{-1}\right\|\|\mathbf{V}_0-\mathbf{V}_\eta\|\left\|\mathbf{V}_\eta^{-1}\right\|\\
&\leq 2\left\|\mathbf{V}_0^{-1}\right\|^2\|\mathbf{V}_\eta-\mathbf{V}_0\|_F\\
&\leq 2qC_{\eta_0}\left\|\mathbf{V}_0^{-1}\right\|\|\eta-\eta_0\|_{L_2(\Omega)}
\end{aligned}$$

whenever $\|\eta-\eta_0\|_{L_2(\Omega)}<\epsilon$. Hence

$$\begin{aligned}
\mathbf{r}(\eta,\eta_0) &= \boldsymbol{\theta}_\eta^*-\boldsymbol{\theta}_0^*-2\int_\Omega[\eta(\mathbf{x})-\eta_0(\mathbf{x})]\mathbf{V}_0^{-1}\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}_0^*)\mathrm{d}\mathbf{x}\\
&= 2\int_\Omega[\eta(\mathbf{x})-\eta_0(\mathbf{x})]\left[\mathbf{V}_{\eta[u']}^{-1}\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}_{\eta[u']}^*)-\mathbf{V}_0^{-1}\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}_0^*)\right]\mathrm{d}\mathbf{x}
\end{aligned}$$

$$= 2\int_\Omega [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \left[(\mathbf{V}_{\eta[u']}^{-1} - \mathbf{V}_0^{-1}) \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_{\eta[u']}^*) \right] \mathrm{d}\mathbf{x}$$
$$+ 2\int_\Omega [\eta(\mathbf{x}) - \eta_0(\mathbf{x})] \mathbf{V}_0^{-1} \left[\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_{\eta[u']}^*) - \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_0^*) \right] \mathrm{d}\mathbf{x},$$

and hence,

$$\|\mathbf{r}(\eta, \eta_0)\| \leq 2\int_\Omega |\eta(\mathbf{x}) - \eta_0(\mathbf{x})| \left[\left\| \mathbf{V}_{\eta[u']}^{-1} - \mathbf{V}_0^{-1} \right\| \sup_{(\mathbf{x},\boldsymbol{\theta}) \in \Omega \times \Theta} \left\| \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \right\| \right] \mathrm{d}\mathbf{x}$$
$$+ 2\int_\Omega |\eta(\mathbf{x}) - \eta_0(\mathbf{x})| \left[\|\mathbf{V}_0^{-1}\| \left\| \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_{\eta[u']}^*) - \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_0^*) \right\| \right] \mathrm{d}\mathbf{x}$$
$$\lesssim \|\eta - \eta_0\|_{L_2(\Omega)} q^2 C_{\eta_0}^2 \|\mathbf{V}_0^{-1}\| \|\eta - \eta_0\|_{L_2(\Omega)} + \|\mathbf{V}_0^{-1}\| \|\eta - \eta_0\|_{L_2(\Omega)}^2,$$

implying that $\|\mathbf{r}(\eta, \eta_0)\| \leq L_{\eta_0}^{(2)} \|\eta - \eta_0\|_{L_2(\Omega)}^2$ for some constant $L_{\eta_0}^{(2)}$ depending on η_0 only. This completes the proof of the first assertion.

To prove the second assertion, note that if A1 and A3 hold for all η in an L_2-neighborhood $\mathcal{U}$ of η_0, then for any $\eta_1 \in \mathcal{U}$, A1 and A3 hold for all η in an L_2-neighborhood of η_1 inside $\mathcal{U}$. Therefore, the first assertion can be applied to η_1: For all $\eta_1 \in \mathcal{U}$, $\boldsymbol{\theta}_\eta^*$ is a continuous functional of η at $\eta = \eta_1$. Namely, $\boldsymbol{\theta}_\eta^*$ is a continuous functional of $\eta \in \mathcal{U}$. Therefore, $\mathcal{M}(\mathcal{U}) = \{(\eta, \boldsymbol{\theta}_\eta^*) : \eta \in \mathcal{U}\}$ becomes the graph of this continuous functional. It follows directly that the maps $T_1 : \mathcal{M}(\mathcal{U}) \to \mathcal{U} : (\eta, \boldsymbol{\theta}_\eta^*) \mapsto \eta$ and $T_2 : \mathcal{U} \to \mathcal{M}(\mathcal{U}) : \eta \mapsto (\eta, \boldsymbol{\theta}_\eta^*)$ are continuous and invertible to each other. Therefore, the transition map $T_2 \circ T_1^{-1}$ is the identity on $\mathcal{U}$, showing that $\mathcal{M}(\mathcal{U})$ is a Banach manifold.

C Proof of Lemma 2

Before proceeding, we introduce the notion of *covering number* for a metric space $(\mathfrak{X}, d)$. The ϵ-covering number of $(\mathfrak{X}, d)$ for $\epsilon > 0$, is the smallest number of ϵ-balls (with respect to the metric d) that are needed to cover $\mathfrak{X}$.

Since η is imposed the Matérn Gaussian process with a smoothness parameter α, it follows that the concentration function

$$\varphi_{\eta_0}(\epsilon) = \inf_{\eta_1 \in \mathbb{H}_{\Psi_\alpha}(\Omega) : \|\eta_1 - \eta_0\|_{L_\infty(\Omega)} \leq \epsilon} \frac{1}{2} \|\eta_1\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}^2 - \log \Pi(\|\eta\|_{L_\infty(\Omega)} < \epsilon)$$

satisfies $\varphi_{\eta_0}(\epsilon) \leq C\epsilon^{-p/\alpha}$ for some constant $C > 0$ for sufficiently small $\epsilon > 0$. Then by Theorem 2.1 in van der Vaart and van Zanten (2008), it holds that

$$\Pi(\|\eta - \eta_0\|_{L_\infty(\Omega)} < \epsilon_n) \geq \exp(-C^2 n\epsilon_n^2), \tag{C.1}$$

where $\epsilon_n = n^{-\alpha/(2\alpha+p)}$. Pick $\beta > 0$ such that $\beta \in (\max\{\underline{\alpha}, p/2\}, \alpha)$. Then we know that the Matérn Gaussian process $\mathrm{GP}(0, \Psi_\alpha)$ assigns prior probability one to $\mathfrak{C}_\beta(\Omega)$. Now set $\mathcal{B}_n = \epsilon_n \mathfrak{C}_\beta^1(\Omega) + m_n \mathbb{H}_{\Psi_\alpha}^1(\Omega)$, where

$$\mathfrak{C}_\beta^1(\Omega) = \left\{f \in \mathfrak{C}_\beta(\Omega) : \|f\|_{\mathfrak{C}_\beta(\Omega)} \leq 1\right\}, \quad \mathbb{H}_{\Psi_\alpha}^1(\Omega) = \left\{f \in \mathbb{H}_{\Psi_\alpha}(\Omega) : \|f\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)} \leq 1\right\},$$

m_n is some sequence determined later, and $\mathbb{H}_{\Psi_\alpha}(\Omega)$ is the reproducing kernel Hilbert space (abbreviated as RKHS) associated with the Matérn covariance function Ψ_α. Denote Φ to be the cumulative distribution function of the standard normal distribution and set $m_n = -2\Phi^{-1}(\exp[-(2C + 1/\sigma^2)n\epsilon_n^2])$. Since $\eta \sim \mathrm{GP}(0, \Psi_\alpha)$ can be viewed as a Gaussian random element in the Banach space $\mathfrak{C}_\beta(\Omega)$ with the norm $\|\cdot\|_{\mathfrak{C}_\beta(\Omega)}$, then by the Borell's inequality (van der Vaart and van Zanten, 2008) we have

$$\begin{aligned}
\Pi(\mathcal{B}_n) &\geq \Phi\left(\Phi^{-1}\left(\exp\left(-Cn\epsilon_n^2\right)\right) + m_n\right) \\
&= \Phi\left(\Phi^{-1}\left(\exp\left(-Cn\epsilon_n^2\right)\right) - 2\Phi^{-1}\left(\exp\left[-\left(2C + \frac{1}{\sigma^2}\right)n\epsilon_n^2\right]\right)\right) \\
&\geq \Phi\left(-\Phi^{-1}\left(\exp\left[-\left(2C + \frac{1}{\sigma^2}\right)n\epsilon_n^2\right]\right)\right) \\
&= 1 - \exp\left[-\left(2C + \frac{1}{\sigma^2}\right)n\epsilon_n^2\right].
\end{aligned}$$

Hence

$$\Pi(\eta \in \mathcal{B}_n^c) \leq \exp\left[-\left(2C + \frac{1}{\sigma^2}\right)n\epsilon_n^2\right]. \tag{C.2}$$

Now we prove the first inequality using (C.1) and (C.2). Let $\mathcal{H}_n$ be defined as in Lemma A.1. Denote $M_n = \log n$. Then

$$\begin{aligned}
\mathbb{E}_0[\Pi(\mathcal{B}_n^c \mid \mathcal{D}_n)] &\leq \mathbb{E}_0[\mathbb{1}(\mathcal{H}_n)\Pi(\mathcal{B}_n^c \mid \mathcal{D}_n)] + \mathbb{P}_0(\mathcal{H}_n^c) \\
&= \mathbb{E}_0\left\{\mathbb{1}(\mathcal{H}_n)\left[\int \prod_{i=1}^n \frac{p_\eta(y_i, \mathbf{x}_i)}{p_0(y_i, \mathbf{x}_i)}\Pi(\mathrm{d}\eta)\right]^{-1}\left[\int_{\mathcal{B}_n^c} \prod_{i=1}^n \frac{p_\eta(y_i, \mathbf{x}_i)}{p_0(y_i, \mathbf{x}_i)}\Pi(\mathrm{d}\eta)\right]\right\} + o(1)
\end{aligned}$$

$$\leq \frac{\exp[(D+\sigma^{-2})n\epsilon_n^2]}{\Pi(\|\eta-\eta_0\|_{L_\infty(\Omega)}<\epsilon_n)}\Pi(\mathcal{B}_n^c)+o(1)$$
$$\leq \exp\left[\left(D+\frac{1}{\sigma^2}\right)n\epsilon_n^2+Cn\epsilon_n^2-\left(2C+\frac{1}{\sigma^2}\right)n\epsilon_n^2\right]+o(1)$$
$$\leq \exp\left[(D-C)\,n\epsilon_n^2\right]+o(1).$$

Hence taking $D=C/2$ yields $\mathbb{E}_0[\Pi(\mathcal{B}_n^c\mid\mathcal{D}_n)]\to 0$.

Finally we prove the second inequality involving the bracketing integral. Since $\mathbb{H}_{\Psi_\alpha}(\Omega)$ is the RKHS of the Matérn covariance function with a smoothness parameter α, then $\mathbb{H}_{\Psi_\alpha}(\Omega)$ coincides with the Sobolev space $\mathcal{H}_{\alpha+p/2}(\Omega)$ (see, for example, Corollary 1 of Tuo and Wu, 2016). The logarithm of the covering number of $\rho\mathbb{H}^1_{\Psi_\alpha}(\Omega)$ is bounded by (Edmunds and Triebel, 2008)

$$\log\mathcal{N}\left(\epsilon,\rho\mathbb{H}^1_{\Psi_\alpha}(\Omega),\|\cdot\|_{L_\infty(\Omega)}\right)\lesssim\left(\frac{\rho}{\epsilon}\right)^{2p/(2\alpha+p)}$$

for sufficiently small $\epsilon>0$. The metric entropy for the α-Hölder space $\epsilon_n\mathfrak{C}^1_\alpha(\Omega)$ is also known in the literature (see, for example, van der Vaart and Wellner, 1996):

$$\log\mathcal{N}\left(\epsilon,\epsilon_n\mathfrak{C}^1_\beta(\Omega),\|\cdot\|_{L_\infty(\Omega)}\right)\lesssim\left(\frac{\epsilon_n}{\epsilon}\right)^{p/\beta}.$$

Hence for sufficiently small $\epsilon>0$,

$$\log\mathcal{N}(\epsilon,\mathcal{B}_n,\|\cdot\|_{L_\infty(\Omega)})\lesssim\left(\frac{m_n}{\epsilon}\right)^{2p/(2\alpha+p)}+\left(\frac{\epsilon_n}{\epsilon}\right)^{p/\beta},$$

and it follows by simple algebra that

$$\begin{aligned}J_{[\cdot]}(M_n\epsilon_n,\mathcal{B}_n,\|\cdot\|_{L_2(\Omega)})&\lesssim\int_0^{M_n\epsilon_n}\sqrt{\log\mathcal{N}\left(\epsilon,\mathcal{B}_n,\|\cdot\|_{L_\infty(\Omega)}\right)}\mathrm{d}\epsilon\\&\lesssim m_n^{p/(2\alpha+p)}\left(M_n\epsilon_n\right)^{2\alpha/(2\alpha+p)}+\epsilon_n^{p/2\beta}(M_n\epsilon_n)^{(2\beta-p)/(2\beta)}\\&\asymp M_n^{2\alpha/(2\alpha+p)}\sqrt{n}\epsilon_n^2+M_n^{(2\beta-p)/(2\beta)}\epsilon_n\\&\lesssim M_n^{2\alpha/(2\alpha+p)}\sqrt{n}\epsilon_n^2\end{aligned}$$

for sufficiently large n.

D Proof of Lemma 3

Before proceeding, we establish the following fact: if $(\mathcal{W}_n)_{n=1}^{\infty}$ is a sequence of event such that $\Pi(\mathcal{W}_n \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$, then

$$\begin{aligned}\int_{\mathcal{W}_n} \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) &= \Pi(\mathcal{W}_n \mid \mathcal{D}_n) \int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) \\ &= o_{\mathbb{P}_0}(D_n),\end{aligned} \tag{D.1}$$

where

$$D_n := \int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta).$$

Recall that the RKHS $\mathbb{H}_{\Psi_\alpha}(\Omega)$ of the Matérn Gaussian process with a smoothness parameter $\alpha > p/2$ coincides with the Sobolev space $\mathcal{H}_{\alpha+p/2}(\Omega)$ (Wendland, 2004; Tuo and Wu, 2016), and the RKHS norm $\|\cdot\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}$ is equivalent to the Sobolev norm $\|\cdot\|_{\mathcal{H}_{\alpha+p/2}(\Omega)}$. Recall the definition of the isometry U. Then under the prior distribution Π, for any $h \in \mathbb{H}_{\Psi_\alpha}(\Omega)$, $U(h) \sim \mathrm{N}\left(0, \|h\|^2_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right)$. Hence by Lemma 17 in Castillo (2012), for any measurable function $T : \mathfrak{C}(\Omega) \to \mathbb{R}$, any $g, h \in \mathbb{H}_{\Psi_\alpha}(\Omega)$, and any $\rho > 0$,

$$\begin{aligned}&\mathbb{E}_\Pi \left[\mathbb{1}\{|U(g)| \le \rho\} T(\eta - h)\right] \\ &\quad = \mathbb{E}_\Pi \left\{\mathbb{1}\left[\left|U(g) + \langle g, h\rangle_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right| \le \rho\right] T(\eta) \exp\left[U(-h) - \frac{1}{2}\|h\|^2_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right]\right\}.\end{aligned} \tag{D.2}$$

Let $\epsilon_n = n^{-\alpha/(2\alpha+p)}$. Denote $\mathcal{A}_{1n} = \{\|\eta - \eta_0\|_{L_2(\Omega)} \le M_n \epsilon_n\}$, $\mathcal{A}_{2n} = \{\|\eta - \eta_0\|_{L_\infty(\Omega)} \le M\}$, and take

$$g(\mathbf{x}) = 2\sigma^2 \mathbf{t}^{\mathrm{T}} \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_0^*), \quad h(\mathbf{x}) = \frac{2\sigma^2}{\sqrt{n}} \mathbf{t}^{\mathrm{T}} \mathbf{V}_0^{-1} \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}_0^*).$$

Since $U(g/\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)})$ follows the standard normal distribution under the prior, it follows that for sufficiently large L,

$$\Pi(\mathcal{C}_n^c) = \Pi\left\{\left|U\left(\frac{g}{\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}}\right)\right| > L\sqrt{n}\epsilon_n\right\} \le 2\exp\left(-\frac{L}{2} n \epsilon_n^2\right).$$

Then by the proof of Lemma 2, we know that $\Pi(\mathcal{C}_n^c \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$ by taking a sufficiently large L. This completes the proof of the first assertion.

Now we focus on proving the second assertion. Observe that

$$\begin{aligned}\left|\langle g,h\rangle_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right| &= \frac{4\sigma^4}{\sqrt{n}}\left\|\mathbf{t}^{\mathrm{T}}\mathbf{V}_0^{-1}\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\cdot,\boldsymbol{\theta}_0^*)\right\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}^2\\ &\leq \frac{4\sigma^4}{\sqrt{n}}\|\mathbf{V}_0^{-1}\mathbf{t}\|^2\sum_{j=1}^q\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial y^s}{\partial\theta_j}(\cdot,\boldsymbol{\theta})\right\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}^2 = o(\sqrt{n}\epsilon_n),\end{aligned}$$

which implies that for sufficiently large n,

$$\begin{aligned}\left\{|U(g)|\leq (L/2)\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\} &\subset \left\{\left|U(g)+\langle g,h\rangle_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right|\leq L\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}\\ &\subset \left\{|U(g)|\leq 2L\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}. \end{aligned}\tag{D.3}$$

On the other hand,

$$\|h\|_{L_2(\Omega)}\leq \frac{2q\sigma^2}{\sqrt{n}}\|\mathbf{V}_0^{-1}\mathbf{t}\|\max_{j=1,\cdots,q}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial y^s}{\partial\theta_j}(\cdot,\boldsymbol{\theta})\right\|_{L_2(\Omega)} = o(\epsilon_n),$$

implying that

$$\begin{aligned}\mathcal{A}_{1n} &= \left\{\|\eta_{\mathbf{t}}-\eta_0+h\|_{L_2(\Omega)}\leq M_n\epsilon_n\right\}\\ &\subset \left\{\|\eta_{\mathbf{t}}-\eta_0\|_{L_2(\Omega)}\leq M_n\epsilon_n+\|h\|_{L_2(\Omega)}\right\}\\ &\subset \left\{\|\eta_{\mathbf{t}}-\eta_0\|_{L_2(\Omega)}\leq 2M_n\epsilon_n\right\} := \mathcal{A}_{1n}^u(\mathbf{t})\end{aligned}\tag{D.4}$$

for sufficiently large n, where the fact $n^{-1/2}\leq\epsilon_n$ is applied. Similarly, for sufficiently large n it holds that

$$\mathcal{A}_{1n}\supset\{\|\eta_{\mathbf{t}}-\eta_0\|_{L_2(\Omega)}\leq M_n\epsilon_n/2\} := \mathcal{A}_{1n}^l(\mathbf{t}).\tag{D.5}$$

Similarly, by taking $\mathcal{A}_{2n}^l(\mathbf{t}) = \{\|\eta_{\mathbf{t}}-\eta_0\|_{L_\infty(\Omega)}\leq M/2\}$ one can also show that $\mathcal{A}_{2n}^l(\mathbf{t})\subset\mathcal{A}_{2n}$. We break the rest of the proof into two components.

Step 1: We provide an upper bound for $\int_{\mathcal{A}_n\cap\mathcal{C}_n}\exp(\ell_n(\eta_{\mathbf{t}})-\ell_n(\eta_0))\Pi(\mathrm{d}\eta)$.

Write

$$\begin{aligned}&\int_{\mathcal{A}_n\cap\mathcal{C}_n}\exp(\ell_n(\eta_{\mathbf{t}})-\ell_n(\eta_0))\Pi(\mathrm{d}\eta)\\ &\quad\leq\int\mathbb{1}\left\{|U(g)|\leq L\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}\mathbb{1}(\mathcal{A}_{1n}^u(\mathbf{t}))\exp(\ell_n(\eta_{\mathbf{t}})-\ell_n(\eta_0))\Pi(\mathrm{d}\eta).\end{aligned}$$

We obtain the upper bound of the right-hand side of the last display using the change of measure formulas (D.2), (D.3), and (D.4):

$$
\begin{aligned}
&\int \mathbb{1}\left\{|U(g)| \le L\sqrt{n}\epsilon_n \|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\} \mathbb{1}(\mathcal{A}_{1n}^u(\mathbf{t}))\exp(\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)\\
&\quad\le \int \mathbb{1}\left\{\left|U(g) + \langle g, h\rangle_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right| \le L\sqrt{n}\epsilon_n \|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\} \mathbb{1}(\|\eta - \eta_0\|_{L_2(\Omega)} \le 2M_n\epsilon_n)\\
&\qquad\quad \times \exp(\ell_n(\eta) - \ell_n(\eta_0))\exp\left[U(-h) - \frac{2\sigma^4}{n}\left\|\mathbf{t}^{\mathrm{T}}\mathbf{V}_0^{-1}\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\cdot, \boldsymbol{\theta}_0^*)\right\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}^2\right]\Pi(\mathrm{d}\eta)\\
&\quad\le \int \mathbb{1}\left\{|U(g)| \le 2L\sqrt{n}\epsilon_n \|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\} \mathbb{1}(\|\eta - \eta_0\|_{L_2(\Omega)} \le 2M_n\epsilon_n)\\
&\qquad\quad \times \exp(\ell_n(\eta) - \ell_n(\eta_0))\exp\left[U\left(-\frac{g}{\sqrt{n}}\right)\right]\Pi(\mathrm{d}\eta)\\
&\quad\le \int_{\{\|\eta-\eta_0\|_{L_2(\Omega)}\le 2M_n\epsilon_n\}} \exp(\ell_n(\eta) - \ell_n(\eta_0))\exp\left(2L\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right)\Pi(\mathrm{d}\eta)\\
&\quad\le [1 + o(1)]\int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta).
\end{aligned}
$$

Therefore we conclude that

$$
\int_{\mathcal{A}_n} \exp(\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) \le [1 + o_{\mathbb{P}_0}(1)]\int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta). \tag{D.6}
$$

Step 2: We provide a lower bound for $\int_{\mathcal{A}_n\cap\mathcal{C}_n} \exp(\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)$.

Recall the construction of $\mathcal{B}_n$ in the proof of Lemma 2: $\mathcal{B}_n = \epsilon_n\mathfrak{C}_\beta^1(\Omega) + m_n\mathbb{H}_{\Psi_\alpha}(\Omega)$, where

$$
\mathfrak{C}_\beta^1(\Omega) = \left\{f \in \mathfrak{C}_\beta(\Omega) : \|f\|_{\mathfrak{C}_\beta(\Omega)} \le 1\right\}, \quad \mathbb{H}_{\Psi_\alpha}^1(\Omega) = \left\{f \in \mathbb{H}_{\Psi_\alpha}(\Omega) : \|f\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)} \le 1\right\},
$$

and $m_n = -2\Phi^{-1}(\exp[-(2C + 1/\sigma^2)n\epsilon_n^2])$. Now take $\widetilde{\mathcal{B}}_n = \epsilon_n\mathfrak{C}_\beta^1(\Omega) + (3m_n/4)\mathbb{H}_{\Psi_\alpha}(\Omega)$. Then again by the Borell's inequality (van der Vaart and van Zanten, 2008) we have

$$
\begin{aligned}
\Pi(\widetilde{\mathcal{B}}_n) &\ge \Phi\left(\Phi^{-1}\left(\exp\left(-Cn\epsilon_n^2\right)\right) + \frac{3m_n}{4}\right)\\
&= \Phi\left(\Phi^{-1}\left(\exp\left(-Cn\epsilon_n^2\right)\right) - \frac{3}{2}\Phi^{-1}\left(\exp\left[-\left(2C + \frac{1}{\sigma^2}\right)n\epsilon_n^2\right]\right)\right)\\
&\ge \Phi\left(-\frac{1}{2}\Phi^{-1}\left(\exp\left[-\left(2C + \frac{1}{\sigma^2}\right)n\epsilon_n^2\right]\right)\right).
\end{aligned}
$$

Using the facts that $\Phi^{-1}(u) \le (-1/2)\sqrt{\log(1/u)}$ for $u \in (0, 1/2)$, $1 - \Phi(x) \le (1/2)\mathrm{e}^{-x^2/2}$ for sufficiently large x (see, for example, Lemma K.6 in Ghosal and van der Vaart, 2017), and

$n\epsilon_n^2 \to \infty$, we further lower bound the last display as follows:

$$\Pi(\widetilde{\mathcal{B}}_n) \geq \Phi\left(-\frac{1}{2}\Phi^{-1}\left(\exp\left[-\left(2C+\frac{1}{\sigma^2}\right)n\epsilon_n^2\right]\right)\right)$$
$$\geq \Phi\left(\frac{1}{4}\sqrt{\left(2C+\frac{1}{\sigma^2}\right)n\epsilon_n^2}\right) \geq 1-\frac{1}{2}\exp\left[-\frac{1}{32}\left(2C+\frac{1}{2\sigma^2}\right)n\epsilon_n^2\right].$$

Then we conclude that $\Pi(\widetilde{\mathcal{B}}_n \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$ by following an argument that is similar to that for proving $\Pi(\mathcal{B}_n \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$. Furthermore, for any $\eta \in \widetilde{\mathcal{B}}_n$, there exists $\eta_1 \in \mathfrak{C}_\beta^1(\Omega)$ and $\eta_2 \in \mathbb{H}_{\Psi_\alpha}(\Omega)$ such that $\eta = \epsilon_n\eta_1 + (3m_n/4)\eta_2$. Consequently, if $\eta_{\mathbf{t}} \in \widetilde{\mathcal{B}}_n$, then

$$\eta = \eta_{\mathbf{t}} + h = \epsilon_n(\eta_{\mathbf{t}})_1 + (3m_n/4)(\eta_{\mathbf{t}})_2 + h = \epsilon_n(\eta_{\mathbf{t}})_1 + m_n\left(\frac{3(\eta_{\mathbf{t}})_2}{4} + \frac{h}{m_n}\right).$$

Then we directly conclude that $\eta \in \mathcal{B}_n$, namely, $\mathbb{1}(\eta_{\mathbf{t}} \in \widetilde{\mathcal{B}}_n) \leq \mathbb{1}(\eta \in \mathcal{B}_n)$, by noting that

$$\left\|\frac{3(\eta_{\mathbf{t}})_2}{4} + \frac{h}{m_n}\right\|_{\Psi_\alpha(\Omega)} \leq \frac{3}{4}\|\eta_{\mathbf{t}}\|_{\Psi_\alpha(\Omega)} + \frac{1}{m_n}\|h\|_{\Psi_\alpha(\Omega)} \leq 1.$$

Now we turn to the computation of the desired lower bound. Write

$$\int_{\mathcal{A}_n\cap\mathcal{C}_n} \exp(\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)$$
$$\geq \int \mathbb{1}\left\{|U(g)| \leq L\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_\Psi(\Omega)}\right\}\mathbb{1}(\mathcal{A}_{1n}^l(\mathbf{t}))\mathbb{1}(\mathcal{A}_{2n}^l(\mathbf{t}))\mathbb{1}(\eta_{\mathbf{t}} \in \widetilde{\mathcal{B}}_n)\exp(\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta).$$

We lower bound the preceeding display using (D.2), (D.3), and (D.5):

$$\int \mathbb{1}\left\{|U(g)| \leq L\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}\mathbb{1}(\mathcal{A}_{1n}^l(\mathbf{t}))\mathbb{1}(\mathcal{A}_{2n}^l(\mathbf{t}))\mathbb{1}(\eta_{\mathbf{t}} \in \widetilde{\mathcal{B}}_n)\exp(\ell_n(\eta_{\mathbf{t}}) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)$$
$$= \int \mathbb{1}\left\{\left|U(g) + \langle g, h\rangle_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right| \leq L\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}\mathbb{1}\left\{\|\eta-\eta_0\|_{L_2(\Omega)} \leq M_n\epsilon_n/2\right\}$$
$$\times \exp(\ell_n(\eta) - \ell_n(\eta_0))\exp\left[U\left(-\frac{g}{\sqrt{n}}\right) - \frac{2\sigma^2}{n}\left\|\mathbf{t}^{\mathrm{T}}\mathbf{V}_0^{-1}\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\cdot,\boldsymbol{\theta}_0^*)\right\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}^2\right]\Pi(\mathrm{d}\eta)$$
$$\geq \int \mathbb{1}\left\{|U(g)| \leq (L/2)\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}\mathbb{1}\left\{\|\eta-\eta_0\|_{L_2(\Omega)} \leq M_n\epsilon_n/2\right\}$$
$$\times \mathbb{1}\left\{\|\eta-\eta_0\|_{L_\infty(\Omega)} \leq M/2\right\}\mathbb{1}\left(\eta \in \widetilde{\mathcal{B}}_n\right)\exp(\ell_n(\eta) - \ell_n(\eta_0))$$
$$\times \exp\left(-\frac{1}{\sqrt{n}}|U(g)|\right)[1-o(1)]\Pi(\mathrm{d}\eta)$$
$$\geq [1-o(1)]\int \mathbb{1}\left\{|U(g)| \leq (L/2)\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}\mathbb{1}\left\{\|\eta-\eta_0\|_{L_2(\Omega)} \leq M_n\epsilon_n/2\right\}$$
$$\times \mathbb{1}\left\{\|\eta-\eta_0\|_{L_\infty(\Omega)} \leq M/2\right\}\mathbb{1}(\widetilde{\mathcal{B}}_n)\exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta).$$

Since $\Pi(\|\eta - \eta_0\|_{L_2(\Omega)} > M_n\epsilon_n/2 \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$, $\Pi(\widetilde{\mathcal{B}}_n^c) = o_{\mathbb{P}_0}(1)$, and for sufficiently large L and M, $\Pi(|U(g)| > (L/2)\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_\Psi(\Omega)} \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$, $\Pi(\|\eta - \eta_0\|_{L_\infty(\Omega)} > M/2 \mid \mathcal{D}_n) = o_{\mathbb{P}_0}(1)$, the last display can be further computed

$$
\begin{aligned}
&\int \mathbb{1}\left\{|U(g)| \leq (L/2)\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\} \mathbb{1}\left\{\|\eta - \eta_0\|_{L_2(\Omega)} \leq M_n\epsilon_n/2\right\}\\
&\quad\times\left\{\|\eta - \eta_0\|_{L_\infty(\Omega)} \leq M/2\right\} \mathbb{1}(\widetilde{\mathcal{B}}_n)\exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)\\
&\quad\geq \int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) - \int_{\left\{|U(g)| > (L/2)\sqrt{n}\epsilon_n\|g\|_{\mathbb{H}_{\Psi_\alpha}(\Omega)}\right\}} \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)\\
&\quad\quad - \int_{\left\{\|\eta-\eta_0\|_{L_2(\Omega)} > M_n\epsilon_n/2\right\}} \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) - \int_{\widetilde{\mathcal{B}}_n} \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)\\
&\quad\quad - \int_{\left\{\|\eta-\eta_0\|_{L_\infty(\Omega)} > M/2\right\}} \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta)\\
&\quad= \int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) - o_{\mathbb{P}_0}(D_n).
\end{aligned}
$$

Hence we conclude that

$$
\begin{aligned}
\int_{\mathcal{A}_n} \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) &\geq [1 - o(1)]\int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta) - o_{\mathbb{P}_0}(D_n)\\
&= [1 - o_{\mathbb{P}_0}(1)]\int \exp(\ell_n(\eta) - \ell_n(\eta_0))\Pi(\mathrm{d}\eta). \qquad \text{(D.7)}
\end{aligned}
$$

The proof is completed by combining (D.6) and (D.7).

E Proof of Corollary 2

The proof is similar to that of Corollary of Yang et al. (2015) and is included here for completeness. For each $k = 1, \ldots, q$, let the event $A = \mathbb{R} \times \ldots \times A_s \times \ldots \times \mathbb{R}$ in Theorem 2, where the sth component is A_s and the rest are $\mathbb{R}$. Then it follows directly from Theorem 2 that

$$
\sup_{A_s \subset \mathbb{R}}\left|\Pi\left([\boldsymbol{\theta}_\eta^*]_k \in A_s \;\middle|\; \mathcal{D}_n\right) - \mathrm{N}\left([\widehat{\boldsymbol{\theta}}_{L_2}]_k, \frac{4\sigma^2}{n}[\mathbf{V}_0^{-1}\mathbf{W}\mathbf{V}_0^{-1}]_{kk}\right)(A_s)\right| = o_{\mathbb{P}_0}(1),
$$

where $[\cdot]_k$ is the kth component of the argument vector and $[\cdot]_{kk}$ is the (k,k)th element of the argument matrix. Now set $A_s = (-\infty, [\widehat{\boldsymbol{\theta}}^*]_k]$. It follows that

$$\left|\Phi\left(\sqrt{\frac{n}{4\sigma^2[\mathbf{V}_0^{-1}\mathbf{W}\mathbf{V}_0^{-1}]_{kk}}}\left([\widehat{\boldsymbol{\theta}}^*]_k - [\widehat{\boldsymbol{\theta}}_{L_2}]_k\right)\right) - \frac{1}{2}\right| = o_{\mathbb{P}_0}(1),$$

where Φ is the cumulative distribution function (CDF) of the standard normal distribution. By the continuity of Φ^{-1}, we have $[\widehat{\boldsymbol{\theta}}^*]_k - [\widehat{\boldsymbol{\theta}}_{L_2}]_k = o_{\mathbb{P}_0}(1/\sqrt{n})$. Invoking the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{L_2}$ completes the proof.

F Proof of Theorem 3

Before presenting the proof, we need several auxiliary Lemmas from Mairal (2013) and Li and Orabona (2018).

Lemma F.1 (Mairal (2013), Lemma A.5)**.** *Let $(a_t)_{t\geq 1}, (b_t)_{t\geq 1}$ be two non-negative real sequences such that b_t's are bounded, $\sum_{t=1}^{\infty} a_t b_t$ converges and $\sum_{t=1}^{\infty} a_t$ diverges, and $|b_{t+1}-b_t| \lesssim a_t$. Then $b_t \to 0$ as $t \to \infty$.*

Lemma F.2 (Lemma 4, Li and Orabona (2018))**.** *Let $(a_t)_{t=1}^N$ be a non-negative real sequences such that $a_0 > 0$, and $\beta > 1$. Then $\sum_{t=1}^N a_t/(a_0 + \sum_{j=1}^t a_t)^\beta \leq (\beta-1)^{-1} a_0^{1-\beta}$.*

Lemma F.3 (Lemma 5, Li and Orabona (2018))**.** *Assume conditions A2 and A4 hold, and the sample path η is squared-integrable. Then the iterates of Algorithm 1 satisfy the following inequality*

$$\begin{aligned}
&\mathbb{E}_{\mathbf{w}}\left[\sum_{t=1}^N \left\langle \frac{\partial f_\eta(\boldsymbol{\theta}^{(t)})}{\partial \boldsymbol{\theta}}, \sum_{k=1}^q \alpha_{tk} \frac{\partial f_\eta(\boldsymbol{\theta}^{(t)})}{\partial \theta_k}\right\rangle\right] \\
&\quad \leq f_\eta(\boldsymbol{\theta}^{(1)}) - f_\eta(\boldsymbol{\theta}^*_\eta) + \frac{1}{2}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\int_\Omega \frac{\partial}{\partial \boldsymbol{\theta}}[y^s(\mathbf{x},\boldsymbol{\theta}^{(t)}) - \eta(\mathbf{x})]^2 \mathrm{d}\mathbf{x}\right\|_{L_2(\Omega)} \\
&\quad\quad \times \mathbb{E}_{\mathbf{w}}\left\{\sum_{t=1}^N \sum_{k=1}^q \alpha_{tk}^2 \left[\frac{\partial}{\partial \theta_k}(y^s(\mathbf{w}_t, \boldsymbol{\theta}^{(t)}) - \eta(\mathbf{w}_t))^2\right]^2\right\}
\end{aligned}$$

The proof is based on a modification of the Theorem 2 in Li and Orabona (2018), which is provided here for completeness. Observe that by Lemma F.2, conditions A2 and A4, and the fact that η is continuous over Ω, we have,

$$\begin{aligned}
&\sum_{t=1}^{\infty}\sum_{k=1}^{q}\alpha_{tk}^2\left[\frac{\partial}{\partial\theta_k}(y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t))^2\right]^2\\
&\quad=\sum_{t=1}^{\infty}\sum_{k=1}^{q}\alpha_{(t+1)k}^2\left[\frac{\partial}{\partial\theta_k}(y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t))^2\right]^2\\
&\qquad+\sum_{t=1}^{\infty}\sum_{k=1}^{q}(\alpha_{tk}^2-\alpha_{(t+1)k}^2)\left[\frac{\partial}{\partial\theta_k}(y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t))^2\right]^2\\
&\quad\le\frac{a_0^2}{2\epsilon b_0^{2\epsilon}}+\sup_{(\mathbf{w},\boldsymbol{\theta})\in\Omega\times\Theta}\max_{1\le k\le q}\left|\frac{\partial}{\partial\theta_k}\left[y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t)\right]^2\right|^2\sum_{t=1}^{\infty}\sum_{k=1}^{q}(\alpha_{tk}^2-\alpha_{(t+1)k}^2)\\
&\quad\le\frac{a_0^2}{2\epsilon b_0^{2\epsilon}}+\sup_{(\mathbf{w},\boldsymbol{\theta})\in\Omega\times\Theta}\max_{1\le k\le q}\left|\frac{\partial}{\partial\theta_k}\left[y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t)\right]^2\right|^2\alpha_{1k}^2<\infty.
\end{aligned}$$

Therefore, for any $m\in\mathbb{N}_+$, we obtain by Cauchy-Schwarz inequality that

$$\begin{aligned}
\|\boldsymbol{\theta}^{(N+m)}-\boldsymbol{\theta}^{(N)}\|^2&=\left\|\sum_{t=N}^{N+m-1}(\boldsymbol{\theta}^{(t+1)}-\boldsymbol{\theta}^{(t)})\right\|^2\le m\sum_{t=N}^{N+m-1}\|\boldsymbol{\theta}^{(t+1)}-\boldsymbol{\theta}^{(t)}\|^2\\
&\le m\sum_{t=N}^{N+m-1}\left\|2[y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t)]\mathrm{diag}(\alpha_{t1},\dots,\alpha_{tq})\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})\right\|^2\\
&\le m\sum_{t=N}^{N+m-1}\sum_{k=1}^{q}\alpha_{tk}^2\left|\frac{\partial}{\partial\theta_k}[y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t)]^2\right|^2,
\end{aligned}$$

and the previous infinite sum being finite implies that $\lim_{N\to\infty}\|\boldsymbol{\theta}^{(N+m)}-\boldsymbol{\theta}^{(N)}\|=0$ a.s., *i.e.*, $(\boldsymbol{\theta}^{(N)})_N$ forms a Cauchy sequence, and thus must converges to some point $\boldsymbol{\theta}^*\in\Theta$ a.s.. Note that $\boldsymbol{\theta}^*$ is still a random variable.

Next we show that $\boldsymbol{\theta}^*$ is a stationary point of f_η. We obtain, by Lemma F.3 and taking $N\to\infty$ that

$$\begin{aligned}
\mathbb{E}_{\mathbf{w}}\left[\sum_{t=1}^{\infty}\sum_{k=1}^{q}\alpha_{tk}\left(\frac{\partial f_\eta(\boldsymbol{\theta}^{(t)})}{\partial\theta_k}\right)^2\right]&\le f_\eta(\boldsymbol{\theta}^{(1)})-f_\eta(\boldsymbol{\theta}_\eta^*)+\frac{1}{2}\sup_{\boldsymbol{\theta}\in\Theta}\left\|\int_\Omega\frac{\partial}{\partial\boldsymbol{\theta}}[y^s(\mathbf{x},\boldsymbol{\theta}^{(t)})-\eta(\mathbf{x})]^2\mathrm{d}\mathbf{x}\right\|_{L_2(\Omega)}\\
&\quad\times\mathbb{E}_{\mathbf{w}}\left\{\sum_{t=1}^{\infty}\sum_{k=1}^{q}\alpha_{tk}^2\left[\frac{\partial}{\partial\theta_k}(y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t))^2\right]^2\right\}<\infty.
\end{aligned}$$

Therefore, $\sum_{t=1}^{\infty}\alpha_{tk}[\partial f_\eta(\boldsymbol{\theta}^{(t)})/\partial\theta_k]^2<\infty$ a.s., for all $k=1,\dots,q$. In addition, observe that

$$\sup_{\mathbf{w}_t,\boldsymbol{\theta}^{(t)}}\left\|2[y^s(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})-\eta(\mathbf{w}_t)]\mathrm{diag}(\alpha_{t1},\dots,\alpha_{tq})\frac{\partial y^s}{\partial\boldsymbol{\theta}}(\mathbf{w}_t,\boldsymbol{\theta}^{(t)})\right\|$$

$$\leq \max_{t,k} \alpha_{tk} \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta} \left\| \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\theta}) \right\| \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta} |2[y^s(\mathbf{x},\boldsymbol{\theta}) - \eta(\mathbf{x})]| < \infty.$$

Since by the construction of Algorithm 1, $\boldsymbol{\theta}^{(t)} \in \Theta\backslash\partial\Theta$, we see that there exists an integer m^*, such that for all $t \in \mathbb{N}_+$, the number of times that line 11 of Algorithm 1 is called is no greater than m^*. This implies that

$$\frac{a_0}{2^{m^*}} \left\{ b_0 + \sum_{j=1}^{t-1} \left[\frac{\partial g(\mathbf{w}_j, \boldsymbol{\theta}^{(j)})}{\partial \theta_k} \right]^2 \right\}^{-(1/2+\epsilon)} \leq \alpha_{tk} \leq a_0 \left\{ b_0 + \sum_{j=1}^{t-1} \left[\frac{\partial g(\mathbf{w}_j, \boldsymbol{\theta}^{(j)})}{\partial \theta_k} \right]^2 \right\}^{-(1/2+\epsilon)}$$

for all $t \in \mathbb{N}_+$ and all $k = 1, \dots, q$, where $g(\mathbf{x},\boldsymbol{\theta}) = [y^s(\mathbf{x},\boldsymbol{\theta}) - \eta(\mathbf{w}_t)]^2$. This further implies that

$$\sum_{t=1}^{\infty} \alpha_{tk} \geq \frac{a_0}{2^{m^*}} \sum_{t=1}^{\infty} \left\{ b_0 + (t-1) \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta} \left[\frac{\partial g(\mathbf{w}_j, \boldsymbol{\theta}^{(j)})}{\partial \theta_k} \right]^2 \right\}^{-(1/2+\epsilon)} = \infty.$$

Since condition A4 implies that almost surely,

$$\begin{aligned}
\left| \frac{\partial f_\eta(\boldsymbol{\theta}^{(t+1)})}{\partial \theta_k} - \frac{\partial f_\eta(\boldsymbol{\theta}^{(t)})}{\partial \theta_k} \right| &\leq |\theta_k^{(t+1)} - \theta_k^{(t)}| \int_\Omega \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta} \frac{\partial^2}{\partial \theta_k^2} [y^s(\mathbf{x},\boldsymbol{\theta}) - \eta(\mathbf{x})]^2 \mathrm{d}\mathbf{x} \\
&\leq \alpha_{tk} \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta} \left| \frac{\partial g}{\partial \theta_k} g(\mathbf{x},\boldsymbol{\theta}) \right| \int_\Omega \sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta} \frac{\partial^2}{\partial \theta_k^2} [y^s(\mathbf{x},\boldsymbol{\theta}) - \eta(\mathbf{x})]^2 \mathrm{d}\mathbf{x} \\
&\lesssim \alpha_{tk},
\end{aligned}$$

then by the facts that $\sum_{t=1}^\infty \alpha_{tk} [\partial f_\eta(\boldsymbol{\theta}^{(t)})/\partial\theta_k]^2 < \infty$ a.s., and $\sum_{t=1}^\infty \alpha_{tk} = \infty$, we invoke Lemma F.1 to conclude that $\lim_{N\to\infty} \partial f_\eta(\boldsymbol{\theta}^{(N)})/\partial\theta_k = 0$ a.s., for all $k = 1, \dots, q$. The continuity of $\nabla f_\eta(\boldsymbol{\theta})$ and the almost sure convergence of $\boldsymbol{\theta}^{(N)} \to \boldsymbol{\theta}^*$ directly yield that $\nabla f_\eta(\boldsymbol{\theta}^*) = \mathbf{0}$ a.s., completing the proof.

G Proof of Theorem 4

The idea of the proof is based on the proof of Theorem 2 and a fine control between $\boldsymbol{\theta}_\eta^*$ and $\widetilde{\boldsymbol{\theta}}_\eta$. By the proof of Lemma 1, there exists some $\epsilon > 0$ such that over $\{\eta \in \mathcal{F} : \|\eta - \eta_0\|_{L_2(\Omega)} < \epsilon\}$, the functional $\boldsymbol{\theta}_\eta^* : \eta \mapsto \arg\min_{\boldsymbol{\theta}\in\Theta} \|\eta(\cdot) - y^s(\cdot,\boldsymbol{\theta})\|_{L_2(\Omega)}^2$ is continuous, the Fréchet derivative $\dot{\boldsymbol{\theta}}_\eta^* : \mathcal{F} \to \mathbb{R}^q$ for $\boldsymbol{\theta}_\eta^*$ exists, and can be computed by

$$\dot{\boldsymbol{\theta}}_\eta^*(h) = -\left[\mathbf{F}_\theta(\eta, \boldsymbol{\theta}_\eta^*)\right]^{-1} \left[\mathbf{F}_\eta(\eta, \boldsymbol{\theta}_\eta^*)\right](h) = 2\mathbf{V}_\eta^{-1} \int_\Omega h(\mathbf{x}) \frac{\partial y^s}{\partial \boldsymbol{\theta}}(x, \boldsymbol{\theta}_\eta^*) \mathrm{d}\mathbf{x}.$$

By Proposition 1 in Tuo and Wu (2015), $\|\widehat{\eta}-\eta_0\|_{L_2(\Omega)} = O_{\mathbb{P}_0}(n^{-\alpha/(2\alpha+p)})$, since the RKHS $\mathbb{H}_{\Psi_\nu}$ coincides with $\mathcal{H}_\alpha(\Omega)$ for $\nu = \alpha - p/2$. Therefore, with probability tending to one, $\|\widehat{\eta}-\eta_0\|_{L_2(\Omega)} < \epsilon$. We now assume this event occurs and denote it by $\mathcal{E}_n$. Then similar to the proof of Lemma 1, for any η in the $L_2(\Omega)$-neighborhood of η_0 with radius ϵ, we apply the fundamental theorem of calculus and mean-value theorem to obtain

$$\boldsymbol{\theta}^*_\eta - \widehat{\boldsymbol{\theta}}_{L_2} = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}u}\boldsymbol{\theta}^*_{\eta[u]}\mathrm{d}u = 2\int_\Omega [\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})]\mathbf{V}^{-1}_{\eta[u']}\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u']})\mathrm{d}\mathbf{x},$$

where $\eta[u] = \widehat{\eta} + (\eta - \widehat{\eta})u$ for $0 \le u \le 1$ and $u' \in [0,1]$. Then following the same argument in the proof of Lemma 1, we have, $\|\boldsymbol{\theta}^*_\eta - \widehat{\boldsymbol{\theta}}_{L_2}\| \le L^{(1)}_{\eta_0}\|\eta - \widehat{\eta}\|_{L_2(\Omega)}$ for some constant $L^{(1)}_{\eta_0} > 0$ depending on η_0 only. Furthermore, $\|\mathbf{V}^{-1}_\eta - \mathbf{V}^{-1}_0\| \le 2qC_{\eta_0}\|\mathbf{V}^{-1}_0\|\|\eta - \eta_0\|_{L_2(\Omega)}$ for some constant $C_{\eta_0} > 0$ whenever $\|\eta - \eta_0\|_{L_2(\Omega)} < \epsilon$. Therefore, using a technique similar to that in the proof of Lemma 1,

$$\begin{aligned}
\mathbf{r}(\eta, \widehat{\eta}) = \boldsymbol{\theta}^*_\eta - \widetilde{\boldsymbol{\theta}}_\eta &= \boldsymbol{\theta}^*_\eta - \widehat{\boldsymbol{\theta}}_{L_2} - 2\int_\Omega [\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})]\,\mathbf{V}^{-1}_{\widehat{\eta}}\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{L_2})\mathrm{d}\mathbf{x} \\
&= 2\int_\Omega [\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})]\left[\mathbf{V}^{-1}_{\eta[u']}\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u']}) - \mathbf{V}^{-1}_{\widehat{\eta}}\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{L_2})\right]\mathrm{d}\mathbf{x} \\
&= 2\int_\Omega [\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})]\left[(\mathbf{V}^{-1}_{\eta[u']} - \mathbf{V}^{-1}_0 + \mathbf{V}^{-1}_0 - \mathbf{V}^{-1}_{\widehat{\eta}})\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u']})\right]\mathrm{d}\mathbf{x} \\
&\quad + 2\int_\Omega [\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})](\mathbf{V}^{-1}_0 + \mathbf{V}^{-1}_{\widehat{\eta}} - \mathbf{V}^{-1}_0)\left[\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u']}) - \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{L_2})\right]\mathrm{d}\mathbf{x},
\end{aligned}$$

and hence,

$$\begin{aligned}
\|\mathbf{r}(\eta, \widehat{\eta})\| &\le 2\int_\Omega |\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})|\left[\left(\left\|\mathbf{V}^{-1}_{\eta[u']} - \mathbf{V}^{-1}_0\right\| + \left\|\mathbf{V}^{-1}_{\widehat{\eta}} - \mathbf{V}^{-1}_0\right\|\right)\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})\right\|\right]\mathrm{d}\mathbf{x} \\
&\quad + 2\int_\Omega |\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})|\left[\left(\left\|\mathbf{V}^{-1}_0\right\| + \left\|\mathbf{V}^{-1}_{\widehat{\eta}} - \mathbf{V}^{-1}_0\right\|\right)\left\|\frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}^*_{\eta[u']}) - \frac{\partial y^s}{\partial \boldsymbol{\theta}}(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{L_2})\right\|\right]\mathrm{d}\mathbf{x} \\
&\lesssim \left(\|\eta[u'] - \eta_0\|_{L_2(\Omega)} + \|\widehat{\eta} - \eta_0\|_{L_2(\Omega)}\right)\int_\Omega |\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})|\mathrm{d}\mathbf{x} \\
&\quad + \left(\left\|\mathbf{V}^{-1}_0\right\| + 2qC_{\eta_0}\epsilon\right)\int_\Omega |\eta(\mathbf{x}) - \widehat{\eta}(\mathbf{x})|\mathrm{d}\mathbf{x}\sup_{(\mathbf{x},\boldsymbol{\theta})\in\Omega\times\Theta}\left\|\frac{\partial^2 y^s}{\partial \boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}(\mathbf{x}, \boldsymbol{\theta})\right\|\|\boldsymbol{\theta}^*_{\eta[u']} - \widehat{\boldsymbol{\theta}}_{L_2}\| \\
&\lesssim \left(\|\eta - \widehat{\eta}\|_{L_2(\Omega)} + \|\widehat{\eta} - \eta_0\|_{L_2(\Omega)}\right)\|\eta - \widehat{\eta}\|_{L_2(\Omega)} + \|\eta - \widehat{\eta}\|^2_{L_2(\Omega)} \\
&\lesssim \|\eta - \eta_0\|^2_{L_2(\Omega)} + \|\widehat{\eta} - \eta_0\|^2_{L_2(\Omega)}.
\end{aligned}$$

Recall that we use $\mathcal{A}_n = \{\|\eta - \eta_0\|_{L_2(\Omega)} \leq M_n\epsilon_n\} \cap \{\|\eta - \eta_0\|_{L_\infty(\Omega)}\} \cap \mathcal{B}_n$ in the proof of Theorem 2 for $M_n = \log n$ and $\epsilon_n = n^{-\alpha/(2\alpha+p)}$. Let $\mathcal{J}_n = \{\mathcal{D}_n : \|\widehat{\eta} - \eta_0\|_{L_2(\Omega)} \leq M_n\epsilon_n\}$. Clearly, By the argument of the proof of Theorem 2, it suffices to show that

$$\int_{\mathcal{A}_n\cap\mathcal{C}_n} \exp\left[\mathbf{t}^{\mathrm{T}}\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_\eta - \widehat{\boldsymbol{\theta}}_{L_2}\right)\right]\Pi(\mathrm{d}\eta \mid \mathcal{D}_n) \to \exp\left[\frac{1}{2}\mathbf{t}^{\mathrm{T}}\left(4\sigma^2\mathbf{V}_0^{-1}\mathbf{W}\mathbf{V}_0^{-1}\right)\mathbf{t}\right]$$

in $\mathbb{P}_0$-probability for any fixed vector $\mathbf{t} \in \mathbb{R}^q$. First observe that by the previous derivation, for any $\mathcal{D}_n \in \mathcal{J}_n$,

$$\begin{aligned}\sup_{\eta\in\mathcal{A}_n\cap\mathcal{C}_n}\left|\mathbf{t}^{\mathrm{T}}\sqrt{n}(\widetilde{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}^*_\eta)\right| &\leq \sqrt{n}\|\mathbf{t}\|\sup_{\eta\in\mathcal{A}_n}\left(\left\|\widetilde{\boldsymbol{\theta}}_\eta - \widehat{\boldsymbol{\theta}}_{L_2} - \boldsymbol{\theta}^*_\eta + \widehat{\boldsymbol{\theta}}_{L_2}\right\|\right) = \sqrt{n}\|\mathbf{t}\|\sup_{\eta\in\mathcal{A}_n}\|\mathbf{r}(\eta,\widehat{\eta})\| \\ &\lesssim \sqrt{n}M_n^2\epsilon_n^2 = (\log n)^2 n^{-(\alpha-p/2)/(2\alpha+p)} \to 0.\end{aligned}$$

Therefore, for any $\epsilon > 0$,

$$\mathbb{P}_0\left(\sup_{\eta\in\mathcal{A}_n\cap\mathcal{C}_n}\left|\mathbf{t}^{\mathrm{T}}\sqrt{n}(\widetilde{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}^*_\eta)\right| > \epsilon\right) \leq \mathbb{P}_0(\mathcal{J}_n^c) + \mathbb{P}_0\left(\sup_{\eta\in\mathcal{A}_n}\left|\mathbf{t}^{\mathrm{T}}\sqrt{n}(\widetilde{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}^*_\eta)\right| > \epsilon, \mathcal{D}_n \in \mathcal{J}_n\right) \to 0.$$

Since

$$\int_{\mathcal{A}_n\cap\mathcal{C}_n}\exp\left[\mathbf{t}^{\mathrm{T}}\sqrt{n}(\boldsymbol{\theta}^*_\eta - \widehat{\boldsymbol{\theta}})\right]\Pi(\mathrm{d}\eta \mid \mathcal{D}_n) = \exp\left[\frac{1}{2}\mathbf{t}^{\mathrm{T}}(4\sigma^2\mathbf{V}_0^{-1}\mathbf{W}\mathbf{V}_0^{-1})\mathbf{t}\right] + o_{\mathbb{P}_0}(1)$$

by the proof of Theorem 2, it follows that

$$\begin{aligned}&\int_{\mathcal{A}_n\cap\mathcal{C}_n}\exp\left[\mathbf{t}^{\mathrm{T}}\sqrt{n}(\widetilde{\boldsymbol{\theta}}_\eta - \widehat{\boldsymbol{\theta}})\right]\Pi(\mathrm{d}\eta \mid \mathcal{D}_n) \\ &\quad= \int_{\mathcal{A}_n\cap\mathcal{C}_n}\exp\left\{\mathbf{t}^{\mathrm{T}}\sqrt{n}\left[(\widetilde{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}^*_\eta) + (\boldsymbol{\theta}^*_\eta - \widehat{\boldsymbol{\theta}}_{L_2})\right]\right\}\Pi(\mathrm{d}\eta \mid \mathcal{D}_n) \\ &\quad= (1 + o_{\mathbb{P}_0}(1))\left\{\exp\left[\frac{1}{2}\mathbf{t}^{\mathrm{T}}(4\sigma^2\mathbf{V}_0^{-1}\mathbf{W}\mathbf{V}_0^{-1})\mathbf{t}\right] + o_{\mathbb{P}_0}(1)\right\} \to \exp\left[\frac{1}{2}\mathbf{t}^{\mathrm{T}}(4\sigma^2\mathbf{V}_0^{-1}\mathbf{W}\mathbf{V}_0^{-1})\mathbf{t}\right]\end{aligned}$$

in $\mathbb{P}_0$-probability. This completes the proof.

H Additional Numerical Results on KO Calibration

In this section we provide additional results regarding the computation issue of the classical KO approach for calibrating computer models. Recall that Kennedy and O'Hagan (2001) formulate the computer model calibration problem as the following statistical model:

$$\eta(\mathbf{x}) = y^s(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}),$$

where η is the physical system, y^s is the computer model involving the calibration parameter $\boldsymbol{\theta}$, and δ is the discrepancy between them. Classical KO approach and the variations thereof are built on the assumption that δ follows a Gaussian process prior $\delta \sim \mathrm{GP}(\mu, \Psi_\psi)$ for some mean function $\mu : \Omega \to \mathbb{R}$ and some covariance function $\Psi(\cdot, \cdot \mid \psi) : \Omega \times \Omega \to \mathbb{R}_+$ that is typically governed by a range parameter ψ, and $\boldsymbol{\theta}$ follows some prior $\pi(\boldsymbol{\theta})$ based on certain expert knowledge. It is routine in the Bayes literature to further impose a hyperprior distribution $\pi(\psi)$ on the range parameter ψ. For example, in Section 5 of the manuscript we take $\pi(\psi)$ to be the inverse-Gamma distribution. For simplicity we assume that μ is zero. After collecting noisy physical observations $y_i = \eta(\mathbf{x}_i) + e_i$, $e_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \sigma^2)$, the joint posterior density of $\boldsymbol{\theta}$ and ψ is

$$\begin{aligned}\pi(\boldsymbol{\theta}, \psi \mid \mathcal{D}_n) \propto\; & \frac{\pi(\boldsymbol{\theta})\pi(\psi)}{\det(\boldsymbol{\Psi}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n} \mid \psi) + \sigma^2 \mathbf{I}_n)^{1/2}} \\ & \times \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{y}^s_{\boldsymbol{\theta}})^{\mathrm{T}}(\boldsymbol{\Psi}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n} \mid \psi) + \sigma^2\mathbf{I}_n)^{-1}(\mathbf{y} - \mathbf{y}^s_{\boldsymbol{\theta}})\right], \end{aligned} \tag{H.1}$$

where $\mathbf{y}^s_{\boldsymbol{\theta}} = [y^s(\mathbf{x}_1, \boldsymbol{\theta}), \ldots, y^s(\mathbf{x}_n, \boldsymbol{\theta})]^{\mathrm{T}}$ and $\boldsymbol{\Psi}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n} \mid \psi) = [\Psi_\psi(\mathbf{x}_i, \mathbf{x}_j \mid \psi)]_{n\times n}$.

In principle, posterior computation can be directly carried out by routinely drawing samples of $\boldsymbol{\theta}$ and ψ using Metropolis-Hastings algorithm. This could be cumbersome when n is large, since each iteration of the algorithm requires inverting an $n \times n$ matrix. Here we present an alternative strategy to reduce the computational complexity. Rather than drawing ψ from the Markov chain, we propose to directly estimate ψ by maximizing the posterior density (H.1), *i.e.*, we seek to find the maximum *a posteriori* (MAP) estimate of $\boldsymbol{\theta}$ and ψ. In order for the MAP estimation to be valid, the hyperprior $\pi(\psi)$ for the range parameter needs to be carefully selected. We follow the suggestion of Gu (2018) and take $\pi(\psi)$ to be of the form

$$\pi(\psi) \propto \left(\psi + \sigma^2\right)^{a_\psi} \exp\left[-b_\psi(\psi + \sigma^2)\right] \tag{H.2}$$

for some $a_\psi > -(p+1)$ and $b_\psi > 0$. Eq. (H.2) is the one-dimensional version of the jointly robust prior proposed in Gu (2018), and has been shown to yield valid MAP estimate of ψ for Matérn covariance function.

Having an estimate $\widehat{\psi}$ of ψ by maximizing (H.1) with $\pi(\psi)$ in (H.2), the posterior inference regarding $\boldsymbol{\theta}$ can be conveniently carried out by Metropolis-Hastings scheme, and the precision matrix $(\boldsymbol{\Psi}(\mathbf{x}_{1:n}, \mathbf{x}_{1:n} \mid \widehat{\psi}) + \sigma^2 \mathbf{I}_n)^{-1}$ can be computed before the MCMC. As pointed out by one of the reviewers, besides MCMC sampling, the normalizing constant in $\pi(\boldsymbol{\theta} \mid \mathcal{D}_n)$ can also be computed by numerical integration method when Θ is low-dimensional. Namely, one first computes

$$Z(\widehat{\psi}) = \int_{\Theta} \pi(\theta, \widehat{\psi} \mid \mathcal{D}_n) \mathrm{d}\boldsymbol{\theta}$$

using numerical integration methods (*e.g.*, quadrature method), and then obtain the exact posterior density of $\boldsymbol{\theta}$ via $\pi(\boldsymbol{\theta} \mid \mathcal{D}_n) = \pi(\boldsymbol{\theta}, \widehat{\psi} \mid \mathcal{D}_n)/Z(\widehat{\psi})$. The posterior density of $\boldsymbol{\theta}$ obtained via numerical integration can serve as an auxiliary result to check the accuracy of MCMC samples. In what follows we provide an illustrative numerical example.

We adopt the same simulation setup as that of configuration 1 in Section 5.1, and is included here for readers' convenience. The computer model is

$$y^s(x, \boldsymbol{\theta}) = 7[\sin(2\pi\theta_1 - \pi)]^2 + 2[(2\pi\theta_2 - \pi)^2 \sin(2\pi x - \pi)],$$

and the physical system coincides with the computer model when $\boldsymbol{\theta}_0^* = [0.2, 0.3]^{\mathrm{T}}$, *i.e.*, $\eta_0(x) = y^s(x, \boldsymbol{\theta}_0^*)$. The design space Ω is $[0, 1]$, and the parameter space Θ for $\boldsymbol{\theta}$ is $[0, 0.25] \times [0, 0.5]$. We simulate $n = 50$ observations from the randomly perturbed physical system $y_i = \eta_0(x_i) + e_i$, where $(x_i)_{i=1}^n$ are uniformly sampled from Ω, and the variance for the noises $(e_i)_{i=1}^n$ is set to 0.2^2. We follow the aforementioned strategy to compute $\widehat{\psi}$ and draw 1000 posterior samples from the MCMC after 1000 burn-in iterations. These post-burn-in samples are collected every 10 iterations during the Markov chain. The comparison between the posterior samples and the exact posterior density obtained via numerical integration is visualized in Figure 1. It can be seen that the distribution of these MCMC samples are in high accordance with the exact posterior density.

Furthermore, we compute the means, standard deviations, and covariance matrices of $\boldsymbol{\theta}$ using the drawn MCMC samples and the exact posterior density, respectively, and tabulate them in Table 1. It can be seen that the results computed using MCMC samples are close to their exact values, and there is no sign of non-accuracy occurring in these MCMC samples.

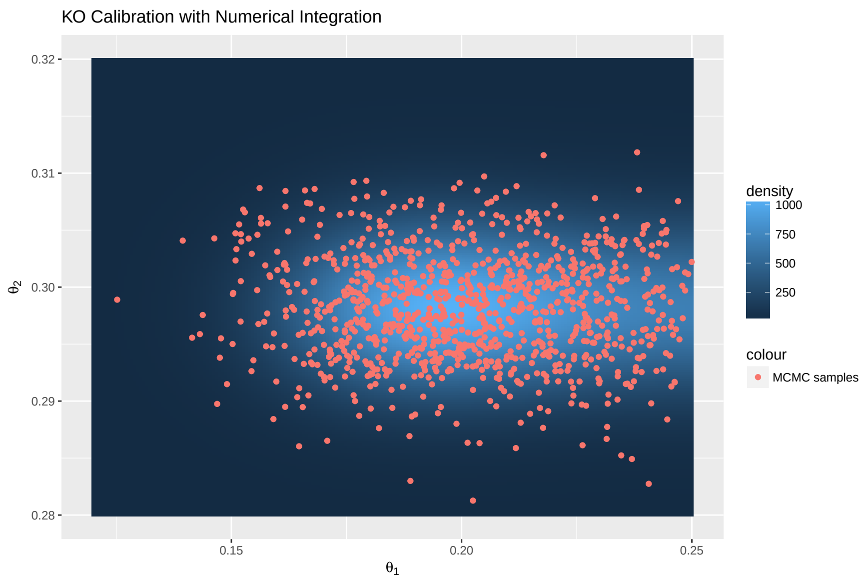

Figure 1: Visualization of the comparison of MCMC sampling and numerical integration for posterior inference in KO method for configuration 1 in Section 5.1. The heatmap is the posterior density of $\boldsymbol{\theta}$ in KO method, the normalizing constant of which is computed using the cubature numerical integration method; The orange scatter points are the samples drawn from MCMC.

Table 1: Summary statistics comparison of MCMC sampling and numerical integration for posterior inference in KO method for configuration 1 in Section 5.1.

	MCMC Sampling		Numerical Integration	
$\boldsymbol{\theta}$	θ_1	θ_2	θ_1	θ_2
Mean	0.2013	0.2982	0.2037	0.2984
Standard Deviation	0.0244	0.0048	0.0255	0.0052
Covariance Matrix	$10^{-4} \times \begin{bmatrix} 5.91 & -0.0354 \\ -0.0354 & 0.23 \end{bmatrix}$		$10^{-4} \times \begin{bmatrix} 6.48 & -0.0024 \\ -0.0024 & 0.27 \end{bmatrix}$	

References

Castillo, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 152(1):53–99.

Edmunds, D. E. and Triebel, H. (2008). *Function spaces, entropy numbers, differential operators*, volume 120. Cambridge University Press.

Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–223.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press.

Gu, M. (2018). Jointly robust prior for gaussian stochastic process in emulation, calibration and variable selection. *Bayesian Anal.* Advance publication.

Hoffman, A. J. and Wielandt, H. W. (2003). The variation of the spectrum of a normal matrix. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 118–120. World Scientific.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Li, X. and Orabona, F. (2018). On the convergence of stochastic gradient descent with adaptive stepsizes. *arXiv preprint arXiv:1805.08114*.

Mairal, J. (2013). Stochastic majorization-minimization algorithms for large-scale optimization. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 2283–2291. Curran Associates, Inc.

Tuo, R. and Wu, C. F. J. (2015). Efficient calibration for imperfect computer models. *The Annals of Statistics*, 43(6):2331–2352.

Tuo, R. and Wu, C. F. J. (2016). A theoretical framework for calibration in computer models: parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):767–795.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.

van der Vaart, A. and Zanten, H. v. (2011). Information rates of nonparametric gaussian process methods. *Journal of Machine Learning Research*, 12(Jun):2095–2119.

van der Vaart, A. W. and van Zanten, J. H. (2008). Rates of contraction of posterior distributions based on gaussian process priors. *The Annals of Statistics*, pages 1435–1463.

Wendland, H. (2004). *Scattered data approximation*, volume 17. Cambridge university press.

Xie, F., Jin, W., and Xu, Y. (2017). A theoretical framework for Bayesian nonparametric regression: Orthonormal random series and rates of contraction. *arXiv preprint arXiv:1712.05731*.

Yang, Y., Cheng, G., and Dunson, D. B. (2015). Semiparametric bernstein-von mises theorem: Second order studies. *arXiv preprint arXiv:1503.04493*.