A INEQUALITIES FOR GAUSSIAN PROCESSES

In this section, we review some inequalities on the maximum of a Gaussian process. Let $G(x)$ be a separable zero-mean Gaussian process with $x \in \Gamma$. Define the metric on Γ by

$$\mathfrak{d}_g(G(x_1), G(x_2)) = \sqrt{\mathbb{E}(G(x_1) - G(x_2))^2}.$$

The ϵ-covering number of the metric space $(\Gamma, \mathfrak{d}_g)$, denoted as $N(\epsilon, \Gamma, \mathfrak{d}_g)$, is the minimum integer N so that there exist N distinct balls in $(\Gamma, \mathfrak{d}_g)$ with radius ϵ, and the union of these balls covers Γ. Let D be the diameter of Γ with respect to the metric $\mathfrak{d}_g$. The supremum of a Gaussian process is closely tied to a quantity called the *entropy integral*, defined as

$$\int_0^{D/2} \sqrt{\log N(\epsilon, \Gamma, \mathfrak{d}_g)} d\epsilon. \tag{14}$$

For detailed discussion of entropy integral, we refer to Adler & Taylor (2009).

Lemma 1 provides an upper bound on the expectation of the maximum value of a Gaussian process, which is Theorem 1.3.3 of Adler & Taylor (2009).

Lemma 1 *Let $G(x)$ be a separable zero-mean Gaussian process with x lying in a $\mathfrak{d}_g$-compact set Γ, where $\mathfrak{d}_g$ is the metric. Let N be the ϵ-covering number. Then there exists a universal constant η such that*

$$\mathbb{E}\left(\sup_{x\in\Gamma} G(x)\right) \leq \eta \int_0^{D/2} \sqrt{\log N(\epsilon, \Gamma, \mathfrak{d}_g)} d\epsilon. \tag{15}$$

Lemma 2, which is Theorem 2.1.1 of Adler & Taylor (2009), presents a concentration inequality.

Lemma 2 *Let G be a separable Gaussian process on a $\mathfrak{d}_g$-compact Γ with mean zero, then for all $u > 0$,*

$$\mathbb{P}\left(\sup_{x\in\Gamma} G(x) - \mathbb{E}(\sup_{x\in\Gamma} G(x)) > u\right) \leq e^{-u^2/2\sigma_\Gamma^2}, \tag{16}$$

where $\sigma_\Gamma^2 = \sup_{x\in\Gamma} \mathbb{E}G(x)^2$.

Theorem 3 is a slightly strengthened version of Theorem 1 of Wang et al. (2020). Its proof, in Section E, is based on Lemmas 1-2 and some machinery from scattered data approximation Wendland (2004).

Theorem 3 *Suppose Condition 1 holds. Let $\mu(x)$ and $\sigma(x)$ be as in Equation 2 and Equation 3, respectively, and $D_\Omega = diam(\Omega)$ be the Euclidean diameter of Ω. Then for any $u > 0$, and any closed deterministic subset $A \subset \Omega$, with probability at least $1 - \exp\{-u^2/(2\sigma_A^2)\}$, the kriging prediction error has the upper bound*

$$\sup_{x\in A} Z(x) - \mu(x) \leq \eta_1 \sigma_A \sqrt{p(1 \vee \log(A_0 D_\Omega))} \sqrt{\log(e\sigma/\sigma_A))} + u, \tag{17}$$

where A_0 is defined in Condition 1, η_1 is a universal constant, and $\sigma_A = \sup_{x\in A} \sigma(x)$.

B PROOF OF THEOREM 1

We proof Theorem 1 by partitioning Ω into subregions, and applying Theorem 3 on each of them. Let $\Omega_i = \{x \in \Omega | \sigma e^{-i} \leqslant \sigma(x) \leqslant \sigma e^{-i+1}\}$, for $i = 1, \ldots$. Let $\sigma_i = \sup_{x\in\Omega_i} \sigma(x)$.

Take $\eta_2 = \eta_1\sqrt{2}e$. By Theorem 3, we have

$$\begin{aligned}
&P\left(\sup_{x\in\Omega}\frac{Z(x)-\mu(x)}{\sigma(x)\log^{1/2}(e\sigma/\sigma(x))} > \eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+u\right)\\
\leqslant&\sum_{i=1}^{\infty}P\left(\sup_{x\in\Omega_i}\frac{Z(x)-\mu(x)}{\sigma(x)\log^{1/2}(e\sigma/\sigma(x))} > \eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+u\right)\\
\leqslant&\sum_{i=1}^{\infty}P\left(\sup_{x\in\Omega_i}Z(x)-\mu(x) > (\eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+u)\sigma e^{-i}\sqrt{i}\right)\\
\leqslant&\sum_{i=1}^{\infty}P\left(\sup_{x\in\Omega_i}Z(x)-\mu(x) > (\eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+u)\sigma_i\log^{1/2}(e\sigma/\sigma_i)/(\sqrt{2}e)\right)\\
\leqslant&\sum_{i=1}^{\infty}\exp\left\{-u^2\log(e\sigma/\sigma_i)/(4e^2)\right\}\\
\leqslant&\sum_{i=1}^{\infty}\exp\left\{-iu^2/(4e^2)\right\} = \frac{\exp\left\{-u^2/(4e^2)\right\}}{1-\exp\left\{-u^2/(4e^2)\right\}},
\end{aligned}$$

which, together with the fact that $M \geq 0$, implies the following upper bound of $\mathbb{E}M$

$$\begin{aligned}
\mathbb{E}M &= \int_0^\infty \mathbb{P}(M>x)dx\\
&\leq \left(\int_0^{\eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+1} + \int_{\eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+1}^{\infty}\right)\mathbb{P}(M>x)dx\\
&\leq \eta_2\sqrt{p(1\vee\log(A_0D_\Omega))}+1+\int_1^\infty\frac{2\exp\left\{-x^2/(4e^2)\right\}}{1-\exp\left\{-x^2/(4e^2)\right\}}dx\\
&\leq C_0\sqrt{p(1\vee\log(A_0D_\Omega))}.
\end{aligned}$$

To access the tail probability, we note that $M-\mathbb{E}M$ is also a Gaussian process with mean zero. Thus by Lemma 2, we have

$$\mathbb{P}(M-\mathbb{E}M>t)\leq e^{-t^2/2\sigma_M^2},$$

where

$$\sigma_M^2 = \sup_{x\in\Omega}\mathbb{E}\frac{(Z(x)-\mu(x))^2}{\sigma(x)^2\log(e\sigma/\sigma(x))}\leq 1.$$

Hence, we complete the proof.

C INDEPENDENCE IN SEQUENTIAL GAUSSIAN PROCESS MODELING

The proof of Theorem 2 relies on certain independence properties of sequential Gaussian process modeling shown in Lemmas 3-4. First we introduce some notation. For an arbitrary function f, and $X=(x_1,\ldots,x_n)$, define $f(X)=(f(x_1),\ldots,f(x_n))^T$, and

$$\mathcal{I}_{\Psi,X}f(x) = r^T(x)K^{-1}f(X), \tag{18}$$

where $r=(\Psi(x-x_1),\ldots,\Psi(x-x_n))^T, K=(\Psi(x_j-x_k))_{jk}$. For notational convenience, we define $\mathcal{I}_{\Psi,\varnothing}f=0$.

Lemma 3 *Let Z be a stationary Gaussian process with mean zero and correlation function Ψ. For two sets of scattered points $X'\subset X=(x_1,\ldots,x_n)$, we have*

$$Z-\mathcal{I}_{\Psi,X'}Z = (Z-\mathcal{I}_{\Psi,X}Z)+\mathcal{I}_{\Psi,X}(Z-\mathcal{I}_{\Psi,X'}Z). \tag{19}$$

In addition, if X and X' are deterministic sets, then the residual $Z-\mathcal{I}_{\Psi,X}Z$ and the vector of observed data $(Z(x_1),\ldots,Z(x_n))^T$ are mutually independent Gaussian process and vector, respectively.

Proof It is easily seen that $\mathcal{I}_{\Psi,X}$ and $\mathcal{I}_{\Psi,X'}$ are linear operators and $\mathcal{I}_{\Psi,X'}\mathcal{I}_{\Psi,X} = \mathcal{I}_{\Psi,X}$, which implies Equation 19.

The residual $Z - \mathcal{I}_{\Psi,X}Z$ is a Gaussian process because $\mathcal{I}_{\Psi,X}$ is linear. The independence between the Gaussian process and the vector can be proven by calculation the covariance

$$\begin{aligned} & \operatorname{Cov}(Z(x') - \mathcal{I}_{\Psi,X'}Z(x'), Z(X)) \\ = & \operatorname{Cov}(Z(x') - r^T(x')K^{-1}Z(X), Z(X)) \\ = & r(x') - r(x') = 0, \end{aligned}$$

which completes the proof. ■

Lemma 4 *For any instance algorithm of Bayesian optimization, the following statements are true.*

1. *Conditional on $\mathcal{F}_{n-1}$ and X_n, the residual process $Z(\cdot) - \mu_n(\cdot)$ is independent of $\mathcal{F}_n$.*
2. *Conditional on $\mathcal{F}_n$, the residual process $Z(\cdot) - \mu_n(\cdot)$ is a Gaussian process with same law as $Z'(\cdot) - \mathcal{I}_{\Psi,X_{1:n}}Z'(\cdot)$, where Z' is an independent copy of Z.*

Proof We use induction on n. For $n = 1$, the desired results are direct consequences of Lemma 3, because the design set is suppressed conditional on $\mathcal{F}_0$.

Now suppose that we have proven already the assertion for n and want to conclude it for $n + 1$. First, we invoke the decomposition given by Lemma 3 to have

$$Z' - \mathcal{I}_{\Psi,X_{1:n}}Z' = (Z' - \mathcal{I}_{\Psi,X_{1:(n+1)}}Z') + \mathcal{I}_{\Psi,X_{1:(n+1)}}(Z' - \mathcal{I}_{\Psi,X_{1:n}}Z'). \tag{20}$$

Because $\mu_n = \mathcal{I}_{\Psi,X_{1:n}}Z$, we also have

$$Z - \mu_n = (Z - \mu_{n+1}) + \mathcal{I}_{\Psi,X_{1:(n+1)}}(Z - \mu_n). \tag{21}$$

By the inductive hypothesis, $Z - \mu_n$ has the same law as $Z' - \mathcal{I}_{\Psi,X_{1:n}}Z'$ conditional on $\mathcal{F}_n$, denoted by $Z - \mu_n \overset{\mathrm{d}}{=} Z' - \mathcal{I}_{\Psi,X_{1:n}}Z'|\mathcal{F}_n$. Our assumption that X_{n+1} is independent of (Z, Z') conditional on $\mathcal{F}_n$ implies that X_{n+1} is independent of $(Z - \mu_n, Z' - \mathcal{I}_{\Psi,X_{1:n}}Z')$ as well. Thus,

$$Z - \mu_n \overset{\mathrm{d}}{=} Z' - \mathcal{I}_{\Psi,X_{1:n}}Z'|\mathcal{F}_n, X_{n+1}.$$

Clearly, this equality in distribution is preserved by acting $\mathcal{I}_{\Psi,X_{1:(n+1)}}$ on both sides, which implies

$$\left(Z - \mu_n, \mathcal{I}_{\Psi,X_{1:(n+1)}}(Z - \mu_n)\right) \overset{\mathrm{d}}{=} \left(Z' - \mathcal{I}_{\Psi,X_{1:n}}Z', \mathcal{I}_{\Psi,X_{1:(n+1)}}(Z' - \mathcal{I}_{\Psi,X_{1:n}}Z')\right)|\mathcal{F}_n, X_{n+1}.$$

Incorporating the above equation with Equation 20 and Equation 21 yields

$$(Z - \mu_{n+1}, Z - \mu_n)) \overset{\mathrm{d}}{=} \left(Z' - \mathcal{I}_{\Psi,X_{1:(n+1)}}Z', Z' - \mathcal{I}_{\Psi,X_{1:n}}Z'\right)|\mathcal{F}_n, X_{n+1}. \tag{22}$$

Now we consider the vectors $V := Z(X_{n+1}) - \mu_n(X_{n+1})$ and $V' = Z'(X_{n+1}) - \mathcal{I}_{\Psi,X_{1:n}}Z'(X_{n+1})$. Then Equation 22 implies

$$(Z - \mu_{n+1}, V)) \overset{\mathrm{d}}{=} \left(Z' - \mathcal{I}_{\Psi,X_{1:(n+1)}}Z', V'\right)|\mathcal{F}_n, X_{n+1}. \tag{23}$$

Because V' consists of observed data, we can apply Lemma 3 to obtain that, conditional on $\mathcal{F}_n$ and X_{n+1}, $Z' - \mathcal{I}_{\Psi,X_{1:(n+1)}}Z'$ is independent of V', which, together with Equation 23, implies that $Z - \mu_{n+1}$ and V are independent conditional on $\mathcal{F}_n$ and X_{n+1}. Because $\mu_n(X_{n+1})$ is measurable with respect to the σ-algebra generated by $\mathcal{F}_n$ and X_{n+1}, we obtain that $Z - \mu_{n+1}$ is independent of $Z(X_{n+1})$ conditional on $\mathcal{F}_n$ and X_{n+1}, which proves Statement 1. Combining Statement 1 and Equation 22 yields Statement 2. ■

D PROOF OF THEOREM 2

The law of total probability implies

$$\begin{aligned}
& \mathbb{P}(M_T - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t) \\
= & \sum_{i=n}^{\infty} \mathbb{P}(M_T - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t | T = n)\mathbb{P}(T = n) \\
= & \sum_{n=1}^{\infty} \mathbb{P}(M_n - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t | T = n)\mathbb{P}(T = n) \\
= & \sum_{n=1}^{\infty} \mathbb{E}\left\{\mathbb{P}(M_n - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t | \mathcal{F}_n)\Big| T = n\right\}\mathbb{P}(T = n),
\end{aligned}$$

where the last equality follows from the fact that $\{T = n\} \in \mathcal{F}_n$, namely, T is a stopping time. Clearly, the desired results are proven if we can show $\mathbb{P}(M_n - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t|\mathcal{F}_n) < e^{-t^2/2}$. Now we resort to part 2 of Lemma 4, which states that conditional on $\mathcal{F}_n$, $Z(\cdot) - \mu_n(\cdot)$ is identical in law to its independent copy $Z'(\cdot) - \mathcal{I}_{\Psi, X_{1:n}} Z'(\cdot)$. Although the event $\{M_n - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t\}$ looks complicated, it is measurable with respect to $Z(\cdot) - \mu_n(\cdot)$. Thus, we arrive at

$$\begin{aligned}
& \mathbb{P}(M_n - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t | \mathcal{F}_n) \\
= & \mathbb{P}\left(\sup_{x \in \Omega} \frac{Z'(x) - \mathcal{I}_{\Phi, X_{1:n}} Z'(x)}{\sigma_n(x) \log^{1/2}(e\sigma/\sigma_n(x))} - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t | \mathcal{F}_n\right).
\end{aligned} \tag{24}$$

Because Z' is independent of Z, the part of conditioning with respect to $Z(X_{1:n})$ in Equation 24 has no effect on Z'. The only thing that matters is the effect of the conditioning on the design points $X_{1:n}$. Hence, Equation 24 is reduced to

$$\mathbb{P}\left(\sup_{x \in \Omega} \frac{Z'(x) - \mathcal{I}_{\Phi, X_{1:n}} Z'(x)}{\sigma_n(x) \log^{1/2}(e\sigma/\sigma_n(x))} - C\sqrt{p(1 \vee \log(A_0 D_\Omega))} > t | X_{1:n}\right). \tag{25}$$

Clearly, we can regard the points $X_{1:n}$ in the formula above as a fixed design. Then the probability Equation 25 is bounded above by $e^{-t^2/2}$ as asserted by Corollary 1.

E PROOF OF THEOREM 3

This proof is similar to Theorem 1 of Wang et al. (2020) but with a few technical improvements.

Because $\mu(x)$ is a linear combination of $Z(x_i)$'s, $\mu(x)$ is also a Gaussian process. The main idea of the proof is to invoke a maximum inequality for Gaussian processes, which states that the supremum of a Gaussian process is no more than a multiple of the integral of the covering number with respect to its natural distance $\mathfrak{d}$. See Adler & Taylor (2009); van der Vaart & Wellner (1996) for related discussions.

Let $g(x) = Z(x) - \mu(x)$. For any $x, x' \in A$, because A is deterministic, we have

$$\begin{aligned}
\mathfrak{d}(x, x')^2 = & \mathbb{E}(g(x) - g(x'))^2 \\
= & \mathbb{E}(Z(x) - \mu(x) - (Z(x') - \mu(x')))^2 \\
= & \sigma^2(\Psi(x - x) - r^T(x)K^{-1}r(x) + \Psi(x' - x') - r^T(x')K^{-1}r(x') \\
& - 2[\Psi(x - x') - r^T(x')K^{-1}r(x)]),
\end{aligned}$$

where $r(\cdot) = (\Psi(\cdot - x_1), \ldots, \Psi(\cdot - x_n))^T$, $K = (\Psi(x_j - x_k))_{jk}$.

The rest of our proof consists of the following steps. In step 1, we bound the covering number $N(\epsilon, A, \mathfrak{d})$. Next we bound the diameter D. In step 3, we obtain a bound for the entropy integral. In the last step, we invoke Lemmas 1 and 2 to obtain the desired results.

Step 1: Bounding the covering number

Let $h(\cdot) = \Psi(x - \cdot) - \Psi(x' - \cdot)$. It can verified that

$$\mathfrak{d}(x, x')^2 = -\sigma^2[h(x') - \mathcal{I}_{\Psi,X}h(x')] + \sigma^2[h(x) - \mathcal{I}_{\Psi,X}h(x)].$$

By Theorem 11.4 of Wendland (2004),

$$\mathfrak{d}(x, x')^2 \leq 2\sigma^2(\sigma_A/\sigma\|h\|_{\mathcal{N}_\Psi(\mathbf{R}^d)}) = 2\sigma\sigma_A\|h\|_{\mathcal{N}_\Psi(\mathbf{R}^d)}, \tag{26}$$

where

$$\sigma_A^2 = \sup_{x\in A}\sigma(x)^2 = \sigma^2\sup_{x\in A}(\Psi(x-x) - r^T(x)K^{-1}r(x)).$$

Denote the Euclidean norm by $\|\cdot\|$. Then, by the definition of the spectral density and the mean value theorem, we have

$$\begin{aligned}
\|h\|^2_{\mathcal{N}_\Psi(\mathbf{R}^d)} &= \Psi(x-x) - 2\Psi(x'-x) + \Psi(x'-x') \\
&= 2\int_{\mathbf{R}^d}(1-\cos((x-x')^T\omega))\tilde{\Psi}(\boldsymbol{\omega})d\boldsymbol{\omega} \\
&\leq \left(2\int_{\mathbf{R}^d}\|\omega\|\tilde{\Psi}(\boldsymbol{\omega})d\boldsymbol{\omega}\right)\|x-x'\| \\
&\leq 2A_0\|x-x'\|,
\end{aligned} \tag{27}$$

where the last inequality follows from the fact that $\|\omega\| \leq \|\omega\|_1$. Combining Equation 26 and Equation 27 yields

$$\mathfrak{d}(x, x')^2 \leq 2A_0^{1/2}\sigma\sigma_A\|x-x'\|^{1/2}. \tag{28}$$

Therefore, the covering number is bounded above by

$$\log N(\epsilon, A, \mathfrak{d}) \leq \log N\left(\frac{\epsilon^4}{4A_0\sigma^2\sigma_A^2}, A, \|\cdot\|\right). \tag{29}$$

The right side of Equation 29 involves the covering number of a Euclidean ball, which is well understood in the literature. See Lemma 4.1 of Pollard (1990). This result leads to the bound

$$\log N(\epsilon, A, \mathfrak{d}) \leq p\log\left(\frac{48A_0D_A\sigma^2\sigma_A^2}{\epsilon^4}+1\right) \leq p\log\left(\frac{48A_0D_\Omega\sigma^2\sigma_A^2}{\epsilon^4}+1\right), \tag{30}$$

where $D_A = \text{diam}(A)$ and $D_\Omega = \text{diam}(\Omega)$ are the Euclidean diameter of A and Ω, respectively.

Step 2: Bounding the diameter D

Define the diameter under metric $\mathfrak{d}$ by $D = \sup_{x,x'\in A}\mathfrak{d}(x, x')$. For any $x, x' \in A$,

$$\begin{aligned}
\mathfrak{d}(x, x')^2 =& \mathbb{E}(g(x) - g(x'))^2 \leq 4\sup_{x\in A}\mathbb{E}(g(x))^2 \\
=& 4\sup_{x\in A}\mathbb{E}(Z(x) - \mathcal{I}_{\Psi,\mathbf{X}}Z(x))^2 \\
=& 4\sigma^2\sup_{x\in A}(\Psi(x-x) - r^T(x)K^{-1}r(x)) = 4\sigma_A^2.
\end{aligned} \tag{31}$$

Thus we conclude that

$$D \leq 2\sigma_A. \tag{32}$$

Step 3: Bounding the entropy integral

By Equation 30 and Equation 32,

$$\begin{aligned}\int_0^{D/2}\sqrt{\log N(\epsilon,A,\mathfrak{d})}d\epsilon &\leq \int_0^{\sigma_A}\sqrt{p\log\left(\frac{48A_0D_\Omega\sigma^2\sigma_A^2}{\epsilon^4}+1\right)}d\epsilon\\ &\leq\left(\int_0^{\sigma_A}d\epsilon\right)^{1/2}\left(\int_0^{\sigma_A}p\log\left(\frac{48A_0D_\Omega\sigma^2\sigma_A^2}{\epsilon^4}+1\right)d\epsilon\right)^{1/2}\\ &=\left(\int_0^{\sigma_A}d\epsilon\right)^{1/2}\left(\sigma\int_0^{\sigma_A/\sigma}p\log\left(\frac{48A_0D_\Omega\sigma_A^2}{u^4\sigma^2}+1\right)du\right)^{1/2}\\ &\leq\sigma_A^{1/2}\left(\sigma\int_0^{\sigma_A/\sigma}p\log\left(\frac{48A_0D_\Omega\sigma_A^2}{u^4\sigma^2}+\frac{\sigma_A^2}{u^4\sigma^2}\right)du\right)^{1/2}\\ &\leq\sqrt{2p}\sigma_A\sqrt{\log(e^2\sqrt{1+48A_0D_\Omega}\sigma/\sigma_A))}\\ &\leq\sqrt{4p}\sigma_A\sqrt{\log(e\sqrt{1+48A_0D_\Omega})}\sqrt{\log(e\sigma/\sigma_A))}\\ &\leq c\sqrt{p(1\vee\log(A_0D_\Omega))}\sigma_A\sqrt{\log(e\sigma/\sigma_A))},\end{aligned}\tag{33}$$

where $c=\sqrt{6\log(7e)}$.

Step 4: Bounding $\mathbb{P}(\sup_{x\in A}Z(x)-\mu(x)>\eta\int_0^{D/2}\sqrt{\log N(\epsilon,A,\mathfrak{d})}d\epsilon+u)$

By Lemmas 1 and 2, we have

$$P\left(\sup_{x\in A}Z(x)-\mu(x)>\eta\int_0^{D/2}\sqrt{\log N(\epsilon,A,\mathfrak{d})}d\epsilon+t\right)\leq e^{-t^2/(2\sigma_A^2)}.\tag{34}$$

By plugging Equation 33 into Equation 34, we obtain the desired inequality with $\eta_1=c\eta$, which completes the proof.

F Calibrating C via simulation

An upper bound of the constant C in Theorem 1 can be obtained by examine the proof of Lemma 1 and Theorem 3. However, this theoretical upper bound can be too large for practical use. In this section, we consider estimating C via numerical simulation.

According to Part 1 of Theorem 1,

$$C_0\geq\mathbb{E}M/\sqrt{p(1\vee\log(A_0D_\Omega))},$$

where $M=\sup_{x\in\Omega}\frac{Z(x)-\mu(x)}{\sigma(x)\log^{1/2}(e\sigma/\sigma(x))}$, A_0 is as in Equation 1, and D_Ω is the Euclidean diameter of Ω. For a specific Gaussian process, $\mathbb{E}M/\sqrt{p(1\vee\log(A_0D_\Omega))}$ is a constant and can be obtained by Monte Carlo. Let $\mathcal{M}$ be the collection of Gaussian processes satisfying the conditions of Theorem 1. Then

$$C_0=\sup_{M\in\mathcal{M}}\mathbb{E}M/\sqrt{p(1\vee\log(A_0D_\Omega))}=:\sup_{M\in\mathcal{M}}H(M).$$

The idea is to consider various Gaussian processes and find the maximum value of $\mathbb{E}M/\sqrt{p(1\vee\log(A_0D_\Omega))}$. This value can be close to C when we cover a broad range of Gaussian processes.

In the numerical studies, we consider $\Omega=[0,1]^p$ for $p=1,2,3$. We consider different A_0 values to get different A_0D_Ω's. In each Monte Carlo sampling, we approximate M using

$$M_1=\sup_{x\in\Omega_1}\frac{Z(x)-\mu(x)}{\sigma(x)\log^{1/2}(e\sigma/\sigma(x))},$$

where Ω_1 is the first $100, 1000, 2000$ points of the Halton sequence (Niederreiter, 1992) for $p=1,2,3$, respectively. We calculate the average of $M_1/\sqrt{p(1\vee\log(A_0D_\Omega))}$ over all the simulated realizations of each Gaussian process.

Specifically, We simulate 1000 realizations of the Gaussian processes for $p = 1$, 100 realizations for $p = 2, 3$ and consider the following four cases. In Cases 1-3, we use maximin Latin hypercube designs (Santner et al., 2003), and use independent samples from the uniform distribution in Case 4.

Case 1: We consider $p = 1$ with 20 and 50 design points. We consider the Gaussian correlation functions and Matérn correlation functions with $\nu = 1.5, 2.5, 3.5$. The results are presented in Table 2.

Case 2: We consider $p = 2$ with 20, 50, and 100 design points. We consider the Gaussian correlation functions and product Matérn correlation functions with $\nu = 1.5, 2.5, 3.5$. The results are presented in Table 3.

Case 3: We consider $p = 3$ with 20, 50, 100 and 500 design points. We consider the product Matérn correlation functions with $\nu = 1.5, 2.5, 3.5$. The results are shown in Table 4.

Case 4: We consider $p = 2$ with 20, 50, and 100 design points. We consider the product Matérn correlation functions with $\nu = 1.5, 2.5, 3.5$. The results are shown in Table 5.

Table 2: Simulation results of Case 1

	design points	$A_0D_\Omega = 1$	$A_0D_\Omega = 3$	$A_0D_\Omega = 5$	$A_0D_\Omega = 10$	$A_0D_\Omega = 25$
Gaussian	20	0.11640290	0.1978563	0.2450737	0.4542654	0.859318
	50	0.08102775	0.0916648	0.1206034	0.1683377	0.422786
$\nu = 1.5$	20	0.9640650	1.065597	0.9537634	0.9429957	1.0197966
	50	0.9442937	1.009187	0.8981430	0.8331926	0.8372607
$\nu = 2.5$	20	0.7432965	0.8554707	0.7804686	0.8371662	1.0074204
	50	0.7304104	0.8218710	0.7346077	0.6987832	0.7563067
$\nu = 3.5$	20	0.6054239	0.7248086	0.6833789	0.7711124	0.9608837
	50	0.3367513	0.6941391	0.6244660	0.6278185	0.6928741

Table 3: Simulation results of Case 2

	design points	$A_0D_\Omega = 1$	$A_0D_\Omega = 3$	$A_0D_\Omega = 5$	$A_0D_\Omega = 10$	$A_0D_\Omega = 25$
Gaussian	20	0.2801128	0.4767259	0.5644628	0.7408401	1.0554507
	50	0.1465512	0.2927036	0.3789438	0.5683807	0.9309326
	100	0.1156139	0.1961319	0.2436626	0.4189444	0.7641615
$\nu = 1.5$	20	0.8106718	0.9528429	0.8748865	0.9365989	1.0894451
	50	0.8114071	0.9299506	0.8568070	0.8576984	0.9964256
	100	0.8137517	0.9108342	0.8224467	0.7951887	0.9168643
$\nu = 2.5$	20	0.6072854	0.7709362	0.7411921	0.8540687	1.0933120
	50	0.6316136	0.7218077	0.7218077	0.7690956	0.9703693
	100	0.5651732	0.6677120	0.6677120	0.7090934	0.8791792
$\nu = 3.5$	20	0.5243251	0.6881401	0.6915576	0.8290974	1.0876019
	50	0.3947094	0.6420423	0.6434791	0.7030224	0.9494486
	100	0.2898865	0.6279639	0.6036111	0.6420049	0.8373886

Table 4: Simulation results of Case 3

Cases	$H(M)$
20 design points, $\nu = 1.5$, $A_0D_\Omega = 1$	0.6977030
500 design points, $\nu = 3.5$, $A_0D_\Omega = 5$	0.4961581
100 design points, $\nu = 2.5$, $A_0D_\Omega = 3$	0.6628567
50 design points, $\nu = 1.5$, $A_0D_\Omega = 10$	0.7632713

From Tables 2-5, we find the following patterns:

Table 5: Simulation results of Case 4

Cases	$H(M)$
100 design points, $\nu = 3$, $A_0 D_\Omega = 3$	0.6778535
50 design points, $\nu = 1.5$, $A_0 D_\Omega = 1$	0.8144700
20 design points, $\nu = 2.5$, $A_0 D_\Omega = 5$	0.7735112
100 design points, $\nu = 1.5$, $A_0 D_\Omega = 10$	0.8164859

- All numerical values ($H(M)$) in Tables 2-5 are less than 1.10. Only eight entries are greater than one.
- In most scenarios, the obtained values are decreasing in ν. This implies that $H(M)$ is smaller when M is smoother.
- $H(M)$ is not monotonic in $A_0 D_\Omega$, which implies a more complicated function relationship between $H(M)$ and $A_0 D_\Omega$.
- In most scenarios, $H(M)$ decreases as the dimension p increases.
- The obtained values are decreasing in the number of design points.

In summary, the largest $H(M)$ values are observed when the sample size is small, the smoothness is low and the dimension is low. Therefore, we believe that our simulation study covers the largest possible $H(M)$ values and our suggestion of choosing $C_0 = 1$ can be used in most practical situations.