E-Companion to "Gaussian Process Based Random Search for Continuous Optimization via Simulation" by Wang, Hong, Jiang and Shen

EC.1. Proof of Critical Lemmas

Section EC.1 contains the proofs of Lemmas 3 and 7-9, which are critical results in this paper.

EC.1.1. Proof of Lemma 3

First notice that $k_n(\boldsymbol{x},\boldsymbol{x}) = \mathrm{Var}\left(g(\boldsymbol{x})|\{\boldsymbol{X}^n,\boldsymbol{G}^n\}\right) \geq 0$. From Equation (3), it is easy to see that

$$k_{n+1}(\boldsymbol{x},\boldsymbol{x}) = k_n(\boldsymbol{x},\boldsymbol{x}) - \frac{[k_n(\boldsymbol{x},\boldsymbol{x}^{n+1})]^2}{k_n(\boldsymbol{x}^{n+1},\boldsymbol{x}^{n+1}) + \lambda^2(\boldsymbol{x}^{n+1})} \leq k_n(\boldsymbol{x},\boldsymbol{x}), \tag{EC.1}$$

which implies that $k_n(\boldsymbol{x},\boldsymbol{x})$ decreases in n. Also note from Equation (3) that reordering the sampling decision-observation pairs $(\boldsymbol{x}_1, G(\boldsymbol{x}_1)),\ldots,(\boldsymbol{x}_n, G(\boldsymbol{x}_n))$ does not alter $k_n(\boldsymbol{x},\boldsymbol{x})$. Fix an $\boldsymbol{x}\in\mathcal{X}$. Then, for any $\epsilon > 0$, $\boldsymbol{x}\in\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},\epsilon)\subset\mathcal{X}$, and

$$k_n(\boldsymbol{x},\boldsymbol{x}) \leq k_{s_n(\boldsymbol{x},\epsilon)}(\boldsymbol{x},\boldsymbol{x}) \leq \tau^2 - \frac{s_n(\boldsymbol{x},\epsilon)\min_{\boldsymbol{x}'\in\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},\epsilon)}[k_0(\boldsymbol{x},\boldsymbol{x}')]^2}{s_n(\boldsymbol{x},\epsilon)\tau^2 + \lambda^2_{max}}, \tag{EC.2}$$

where the second inequality follows from Lemma 2. According to Assumption 3, $[k_0(\boldsymbol{x},\boldsymbol{x}')]^2 = \tau^4\rho^2(|\boldsymbol{x}-\boldsymbol{x}'|)$. Since $\|\boldsymbol{x}-\boldsymbol{x}'\|\leq\epsilon$, Assumption 3 also implies that $\rho(|\boldsymbol{x}-\boldsymbol{x}'|)\geq\rho(\epsilon\mathbf{1})$, where $\mathbf{1}\in\mathbb{R}^d$ is the vector of all ones. Following Equation (EC.2),

$$k_n(\boldsymbol{x},\boldsymbol{x}) \leq \tau^2 - \frac{s_n(\boldsymbol{x},\epsilon)\tau^4\rho^2(\epsilon\mathbf{1})}{s_n(\boldsymbol{x},\epsilon)\tau^2 + \lambda^2_{max}}.$$

By Lemma 1, $s_n(\boldsymbol{x},\epsilon)\to\infty$ almost surely as $n\to\infty$, so $\limsup_{n\to\infty} k_n(\boldsymbol{x},\boldsymbol{x}) \leq \tau^2[1-\rho^2(\epsilon\mathbf{1})]$, with probability one. Sending $\epsilon\to 0$, we have $\rho(\epsilon\mathbf{1})\to 1$, thus $\limsup_{n\to\infty} k_n(\boldsymbol{x},\boldsymbol{x}) \leq 0$, with probability one. Recall that $k_n(\boldsymbol{x},\boldsymbol{x})\geq 0$, then, with probability one, $\limsup_{n\to\infty} k_n(\boldsymbol{x},\boldsymbol{x}) = \liminf_{n\to\infty} k_n(\boldsymbol{x},\boldsymbol{x}) = 0$, which implies $k_n(\boldsymbol{x},\boldsymbol{x})\to 0$ almost surely as $n\to\infty$. □

EC.1.2. Proof of Lemma 7

Let γ_n, p_n and r_n be as defined in Lemma 6 with $\gamma\in(0,1)$ and $b-1>a$. Notice that $k_n(\boldsymbol{x},\boldsymbol{x})$ decreases in n (see Equation (EC.1) in the proof of Lemma 3, and does not depend on the ordering of the sampling decision-observation pairs $(\boldsymbol{x}_1, G(\boldsymbol{x}_1)),\ldots,(\boldsymbol{x}_n, G(\boldsymbol{x}_n))$ (see Equation (3)). Fix an $\boldsymbol{x}\in\mathcal{X}$. Notice that $\boldsymbol{x}\in\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},r_n)\subset\mathcal{X}$. By reordering the decision-observation pairs such that the former $s_n(\boldsymbol{x},r_n)$ points are within $\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},r_n)$, we obtain

$$k_n(\boldsymbol{x},\boldsymbol{x}) \leq k_{s_n(\boldsymbol{x},r_n)}(\boldsymbol{x},\boldsymbol{x}) \leq \tau^2 - \frac{s_n(\boldsymbol{x},r_n)\min_{\boldsymbol{x}'\in\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},r_n)}[k_0(\boldsymbol{x},\boldsymbol{x}')]^2}{s_n(\boldsymbol{x},r_n)\tau^2 + \lambda^2_{max}},$$

where the second inequality follows from Lemma 2. Note that the unconditional covariance function satisfies $k_0(\boldsymbol{x},\boldsymbol{x}')\geq \tau^2(1-C_r\|\boldsymbol{x}-\boldsymbol{x}'\|^{\eta})$ for constants $C_r>0$, $0<\eta\leq 2$ and any close pair of points. By Lemma 6, with probability one, $s_n(\boldsymbol{x},r_n)$ is $\Omega(n^{p_n})$ with $\gamma\in(0,1)$ and $b-a-1>0$, i.e., $s_n(\boldsymbol{x},r_n)\geq cn^{p_n}$ for some $c>0$. Then,

$$\begin{aligned}
k_n(\boldsymbol{x},\boldsymbol{x}) &\leq \tau^2-\frac{s_n(\boldsymbol{x},r_n)\tau^4}{s_n(\boldsymbol{x},r_n)\tau^2+\lambda_{max}^2}\times\min_{\boldsymbol{x}'\in\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},r_n)}(1-C_r\|\boldsymbol{x}-\boldsymbol{x}'\|^{\eta})\\
&\leq \tau^2-\frac{cn^{p_n}\tau^4}{cn^{p_n}\tau^2+\lambda_{max}^2}\times\min_{\boldsymbol{x}'\in\mathcal{X}\cap\mathcal{S}(\boldsymbol{x},r_n)}(1-C_r\|\boldsymbol{x}-\boldsymbol{x}'\|^{\eta})\\
&\leq \tau^2-\frac{cn^{p_n}\tau^4}{cn^{p_n}\tau^2+\lambda_{max}^2}\times(1-C_r r_n^{\eta}),
\end{aligned}$$

where the second inequality holds with probability one, and is due to the fact that $x\tau^2/(x\tau^2+\lambda_{max}^2)$ increases in $x>0$. Hence,

$$\begin{aligned}
k_n(\boldsymbol{x},\boldsymbol{x}) &\leq \tau^2-\frac{cn^{p_n}\tau^4}{cn^{p_n}\tau^2+\lambda_{max}^2}\times\left(1-C_r r_n^{\eta}\right)\\
&=\frac{cn^{p_n}\tau^2}{cn^{p_n}\tau^2+\lambda_{max}^2}\times\left(\frac{\lambda_{max}^2}{c}n^{-p_n}+C_r\tau^2 r_0^{\eta}n^{-\frac{\eta(1-\gamma_n)}{d}}\right)\\
&\leq \text{constant}\times\left(\frac{\lambda_{max}^2}{c}n^{-p_n}+C_r\tau^2 r_0^{\eta}n^{-\frac{\eta(1-\gamma_n)}{d}}\right)\\
&\leq \text{constant}\times n^{-\min\{p_n,\frac{\eta(1-\gamma_n)}{d}\}},
\end{aligned}$$

for sufficiently large n. Thus we have that $k_n(\boldsymbol{x},\boldsymbol{x})$ is $O(n^{-\min\{p_n,\frac{\eta(1-\gamma_n)}{d}\}})$ almost surely.

To obtain the maximum rate, we let $p_n=\frac{\eta(1-\gamma_n)}{d}$, i.e., $\gamma-b\varepsilon(n)=\frac{\eta(1-\gamma+a\varepsilon(n))}{d}$, then it can be obtained $\gamma=[\eta+(a\eta+db)\varepsilon(n)]/(d+\eta)$. Take $a=-db/\eta$. It makes $\gamma=\eta/(d+\eta)$, which satisfies $\gamma\in(0,1)$. Let $b>\eta/(d+\eta)$. Then $b-a=b(d+\eta)/\eta>1$, which satisfies $b-1>a$. Finally, $\min\{p_n,\frac{\eta(1-\gamma_n)}{d}\}=p_n=\gamma-b\varepsilon(n)=\eta/(d+\eta)-b\varepsilon(n)$, with $b>\eta/(d+\eta)$. $\square$

EC.1.3. Proof of Lemma 8

For any $n\geq 1$, it can be obtained that

$$\mathbb{P}\{|\mu_n(\boldsymbol{x})-g(\boldsymbol{x})|>\epsilon_n\}\leq\mathbb{P}\{\mu_n(\boldsymbol{x})-g(\boldsymbol{x})>\epsilon_n\}+\mathbb{P}\{\mu_n(\boldsymbol{x})-g(\boldsymbol{x})<-\epsilon_n\}. \tag{EC.3}$$

Applying the Chernoff bound, we have

$$\begin{aligned}
\mathbb{P}\{\mu_n(\boldsymbol{x})-g(\boldsymbol{x})>\epsilon_n\} &=\mathbb{E}\left[\mathbb{P}\{\mu_n(\boldsymbol{x})-g(\boldsymbol{x})>\epsilon_n|\{\boldsymbol{X}^n,\boldsymbol{G}^n\}\}\right]\\
&\leq\mathbb{E}\left[\min_{t>0}e^{-t\epsilon_n}\cdot\mathbb{E}[e^{t(\mu_n(\boldsymbol{x})-g(\boldsymbol{x}))}|\{\boldsymbol{X}^n,\boldsymbol{G}^n\}]\right]\\
&=\mathbb{E}\left[\min_{t>0}e^{-t\epsilon_n}\cdot e^{\frac{t^2}{2}k_n(\boldsymbol{x},\boldsymbol{x})}\right]\\
&=\mathbb{E}\left[\min_{t>0}e^{\frac{t^2}{2}k_n(\boldsymbol{x},\boldsymbol{x})-t\epsilon_n}\right],
\end{aligned}$$

where the second equality is due to $g(\boldsymbol{x})|\{\boldsymbol{X}^n, \boldsymbol{G}^n\} \sim \mathcal{N}(\mu_n(\boldsymbol{x}), k_n(\boldsymbol{x},\boldsymbol{x}))$ and the moment-generating function of normal random variable. Notice that $\frac{t^2}{2}k_n(\boldsymbol{x},\boldsymbol{x}) - t\epsilon_n$ is minimized at $t = \epsilon_n / k_n(\boldsymbol{x},\boldsymbol{x})$ with value $-\epsilon_n^2/(2k_n(\boldsymbol{x},\boldsymbol{x}))$, then

$$\mathbb{P}\{\mu_n(\boldsymbol{x}) - g(\boldsymbol{x}) > \epsilon_n\} \leq \mathbb{E}\left[e^{-\epsilon_n^2/(2k_n(\boldsymbol{x},\boldsymbol{x}))}\right]. \tag{EC.4}$$

In the similar way, we can also get

$$\mathbb{P}\{\mu_n(\boldsymbol{x}) - g(\boldsymbol{x}) < -\epsilon_n\} \leq \mathbb{E}\left[e^{-\epsilon_n^2/(2k_n(\boldsymbol{x},\boldsymbol{x}))}\right]. \tag{EC.5}$$

Combining Equations (EC.3), (EC.4) and (EC.5) finishes the proof. $\square$

EC.1.4. Proof of Lemma 9

Step 1. Arbitrarily choose an optimal solution $\boldsymbol{x}^* \in \mathcal{X}^*$. For $j = 1, 2, \ldots, d$, let $\dot{g}(\boldsymbol{x})_j$ be the sample path of the derivative surface $\dot{f}_{\mathcal{GP}}(\boldsymbol{x})_j$. According to the mean value theorem, it can be obtained that

$$g(\boldsymbol{x}) = g(\boldsymbol{x}^*) + \nabla g(\boldsymbol{\xi})^\intercal(\boldsymbol{x} - \boldsymbol{x}^*), \tag{EC.6}$$

where $\nabla g(\boldsymbol{x}) = (\dot{g}(\boldsymbol{x})_1, \ldots, \dot{g}(\boldsymbol{x})_d)^\intercal$ denotes the gradient of $g(\boldsymbol{x})$ and $\boldsymbol{\xi} \in \mathcal{X}$ (Fitzpatrick 2009, Theorem 15.29). By the Cauchy-Schwarz inequality, it then follows that

$$|g(\boldsymbol{x}^*) - g(\boldsymbol{x})| = \left|\nabla g(\boldsymbol{\xi})^\intercal(\boldsymbol{x} - \boldsymbol{x}^*)\right| \leq \|\nabla g(\boldsymbol{\xi})\| \|\boldsymbol{x}^* - \boldsymbol{x}\|. \tag{EC.7}$$

Recall that $\mathcal{X}$ is compact under Assumption 6. Based on Assumption 5, $\dot{f}_{\mathcal{GP}}(\boldsymbol{x})_j$ has continuous sample paths on $\mathcal{X}$, for all $j = 1, \ldots, d$, and is thus almost surely bounded on $\mathcal{X}$ (Adler and Taylor 2007, Theorem 1.5.4). So,

$$\dot{g}^* = \max_{j=1,\ldots,d}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}} |\dot{g}(\boldsymbol{x})_j|\right\},$$

is a well defined random variable. Since

$$\|\nabla g(\boldsymbol{\xi})\| = \left[\sum_{j=1}^d (\dot{g}(\boldsymbol{\xi})_j)^2\right]^{1/2} \leq \sqrt{d}\dot{g}^*,$$

then

$$|g(\boldsymbol{x}^*) - g(\boldsymbol{x})| \leq \sqrt{d}\dot{g}^* \|\boldsymbol{x}^* - \boldsymbol{x}\|. \tag{EC.8}$$

So, we have

$$\begin{aligned}
&\mathbb{P}\left(\cap_{i=1}^n \{g^* - g(\boldsymbol{x}_i) > \epsilon_n\}\right) \\
=& \mathbb{P}\left(\cap_{i=1}^n \{g(\boldsymbol{x}^*) - g(\boldsymbol{x}_i) > \epsilon_n\}\right) \\
\leq& \mathbb{P}\left(\cap_{i=1}^n \{\sqrt{d}\dot{g}^* \|\boldsymbol{x}^* - \boldsymbol{x}_i\| > \epsilon_n\}\right) \\
\leq& \mathbb{P}\left(\cap_{i=1}^n \left\{\|\boldsymbol{x}^* - \boldsymbol{x}_i\| > \frac{\epsilon_n}{(\log n)^{1/d}}\right\}\right) + \mathbb{P}\left(\sqrt{d}\dot{g}^* \geq (\log n)^{1/d}\right).
\end{aligned} \tag{EC.9}$$

We now establish the bounds for the two terms respectively.

Step 2. For the first term in the right-hand side of Equation (EC.9), let $\delta_n(\epsilon_n)=\epsilon_n/(\log n)^{1/d}$, and construct a ball $\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n))$. Then, we can have

$$\begin{aligned}\mathbb{P}\left(\cap_{i=1}^n\left\{\|\boldsymbol{x}^*-\boldsymbol{x}_i\|>\frac{\epsilon_n}{(\log n)^{1/d}}\right\}\right)&=1-\mathbb{P}\left(\cup_{i=1}^n\{\|\boldsymbol{x}^*-\boldsymbol{x}_i\|\le\epsilon_n/(\log n)^{1/d}\}\right)\\&=1-\mathbb{P}\left(\sum_{i=1}^n\mathbb{1}_{\{\boldsymbol{x}_i\in\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n))\}}>0\right).\end{aligned}\tag{EC.10}$$

By Lemma 5, for small enough $\delta_n(\epsilon_n)>0$, $\nu(\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n)))\cap\mathcal{X})\ge C\nu(\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n)))$. Since each design point $\boldsymbol{x}_i\in\boldsymbol{X}^n$ is generated from density function ψ_i, which satisfies $\psi_i\ge\alpha>0$ on $\mathcal{X}$, for $i=1,\dots,n$,

$$\begin{aligned}\mathbb{P}\{\boldsymbol{x}_i\in\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n))\}&\ge\alpha\nu(\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n))\cap\mathcal{X})\\&\ge\alpha C\nu(\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n)))=\frac{\alpha C\pi^{\frac{d}{2}}\delta_n(\epsilon_n)^d}{\Gamma(\frac{d}{2}+1)},\end{aligned}$$

where the equality is due to the volume formula of a d-dimensional ball. Let B_i, $i=1,\dots,n$, be i.i.d. Bernoulli random variables with parameter $\frac{\alpha C\pi^{d/2}\delta_n(\epsilon_n)^d}{\Gamma(d/2+1)}\in(0,1)$. Then,

$$\mathbb{P}\left(\sum_{i=1}^n\mathbb{1}_{\{\boldsymbol{x}_i\in\mathcal{S}(\boldsymbol{x}^*,\delta_n(\epsilon_n))\}}>0\right)\ge\mathbb{P}\left(\sum_{i=1}^nB_i>0\right)=1-\left(1-\frac{\alpha C\pi^{\frac{d}{2}}\delta_n(\epsilon_n)^d}{\Gamma(\frac{d}{2}+1)}\right)^n.\tag{EC.11}$$

Combing Equations (EC.10) and (EC.11), we have

$$\begin{aligned}\mathbb{P}\left(\cap_{i=1}^n\{\|\boldsymbol{x}^*-\boldsymbol{x}_i\|>\frac{\epsilon_n}{(\log n)^{1/d}}\}\right)&\le\left(1-\frac{\alpha C\pi^{\frac{d}{2}}\delta_n(\epsilon_n)^d}{\Gamma(\frac{d}{2}+1)}\right)^n\le\exp\left\{-\frac{\alpha C\pi^{\frac{d}{2}}\delta_n(\epsilon_n)^dn}{\Gamma(\frac{d}{2}+1)}\right\}\\&=\exp\left\{-\frac{\alpha C\pi^{\frac{d}{2}}\epsilon_n^dn}{\Gamma(\frac{d}{2}+1)\log n}\right\},\end{aligned}\tag{EC.12}$$

where the second inequality is due to $e^x\ge1+x$.

Step 3. For the second term in the right-hand side of Equation (EC.9), we need to bound the tail probability of $\dot{g}^*$. As mentioned before, $\dot{f}_{\mathcal{GP}}(\boldsymbol{x})_j$ is a Gaussian process with continuous and thus bounded sample paths on $\mathcal{X}$ with probability one, for each $j=1,\dots,d$. Then by the Borell-TIS inequality (Adler and Taylor 2007, Section 2.1), for $j=1,\dots,d$, and sufficiently large t,

$$\mathbb{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}}\ (\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j)>t\right\}\le\exp\left\{C_jt-\frac{t^2}{2\sigma_j^2}\right\},$$

where C_j is a constant depending only on $\mathbb{E}[\sup_{\boldsymbol{x}\in\mathcal{X}}(\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j)]$, and $\sigma_j^2=\sup_{\boldsymbol{x}\in\mathcal{X}}\mathbb{E}[(\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j)^2]$. Let $a_j=\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{\mu}_0(\boldsymbol{x})_j|$. Notice that $\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j|+a_j\ge\sup_{\boldsymbol{x}\in\mathcal{X}}(|\dot{g}(\boldsymbol{x})_j|-|\dot{\mu}_0(\boldsymbol{x})_j|)+a_j\ge\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j|$. So, for sufficiently large t,

$$\mathbb{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j|>t\right\}\le\mathbb{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j|+a_j>t\right\}=\mathbb{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j|>t-a_j\right\}$$

$$\begin{aligned}&\leq 2\,\mathbb{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}}\ (\dot{g}(\boldsymbol{x})_j-\dot{\mu}_0(\boldsymbol{x})_j)>t-a_j\right\}\\&\leq 2\exp\left\{C_j(t-a_j)-\frac{(t-a_j)^2}{2\sigma_j^2}\right\}.\end{aligned}\tag{EC.13}$$

Replacing t with $(\log n)^{1/d}/\sqrt{d}$ where n is sufficiently large, we have

$$\begin{aligned}\mathbb{P}\left(\sqrt{d}\dot{g}^*\geq(\log n)^{\frac{1}{d}}\right)&=\mathbb{P}\left(\max_{j=1,\ldots,d}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j|\right\}\geq(\log n)^{\frac{1}{d}}/\sqrt{d}\right)\\&\leq\sum_{j=1}^{d}\mathbb{P}\left(\sup_{\boldsymbol{x}\in\mathcal{X}}|\dot{g}(\boldsymbol{x})_j|\geq(\log n)^{\frac{1}{d}}/\sqrt{d}\right)\\&\leq 2\sum_{j=1}^{d}\exp\left\{C_j\Big(\frac{(\log n)^{1/d}}{\sqrt{d}}-a_j\Big)-\frac{\left(\frac{(\log n)^{1/d}}{\sqrt{d}}-a_j\right)^2}{2\sigma_j^2}\right\},\end{aligned}\tag{EC.14}$$

where the second inequality is by Equation (EC.13). Finally, combining Equations (EC.9), (EC.12) and (EC.14), we have

$$\begin{aligned}&\mathbb{P}\left(\cap_{i=1}^n\{g^*-g(\boldsymbol{x}_i)>\epsilon_n\}\right)\\&\leq\exp\left\{-\frac{\alpha C\pi^{\frac{d}{2}}\epsilon_n^d n}{\Gamma(\frac{d}{2}+1)\log n}\right\}+2\sum_{j=1}^{d}\exp\left\{C_j\Big(\frac{(\log n)^{1/d}}{\sqrt{d}}-a_j\Big)-\frac{\left(\frac{(\log n)^{1/d}}{\sqrt{d}}-a_j\right)^2}{2\sigma_j^2}\right\},\end{aligned}$$

for sufficiently large n. So the proof is completed. □

EC.2. Proof of Other Lemmas

Section EC.2 contains the proofs of Lemmas 6 and 10.

EC.2.1. Proof of Lemma 6

Since $\gamma_n=\gamma-a\varepsilon(n)\to\gamma$ as $n\to\infty$, it can be obtained that $r_n\to 0$. Notice that $\gamma\in(0,1)$. Then, there exists some $N\in\mathbb{N}$ such that $\gamma_n\in(0,1)$ and $r_n/(3\sqrt{d})\leq\epsilon$ for all $n\geq N$, where ϵ is a sufficiently small positive constant which satisfies Equation (7) in Lemma 5. Suppose that for each $n>N$, we partition each coordinate of $\mathbb{R}^d$ into segments of length $r_n/(3\sqrt{d})$, and by doing so we obtain closed subsets, which are referred as grid boxes that together cover $\mathbb{R}^d$. For each $\boldsymbol{x}\in\mathcal{X}$, let $T_{\boldsymbol{x}}$ be the grid box containing $\boldsymbol{x}$. Define $H_{\boldsymbol{x}}$ as the union of $T_{\boldsymbol{x}}$ and all the other grid box adjacent to $T_{\boldsymbol{x}}$ (i.e., with common vertex, edge or surface). Under the Euclidean distance used in this paper, $H_{\boldsymbol{x}}$ covers all points which are at most $r_n/(3\sqrt{d})$ from $T_{\boldsymbol{x}}$, and hence it can be obtained that $\mathcal{S}(\boldsymbol{x},r_n/(3\sqrt{d}))\subset H_{\boldsymbol{x}}$. Thus, for $n\geq N$,

$$\nu(H_{\boldsymbol{x}}\cap\mathcal{X})\geq\nu(\mathcal{S}(\boldsymbol{x},r_n/(3\sqrt{d}))\cap\mathcal{X})\geq C_1\cdot\nu(\mathcal{S}(\boldsymbol{x},r_n/(3\sqrt{d})))=C_2\cdot(r_n)^d=\Omega\left(n^{-(1-\gamma_n)}\right),\tag{EC.15}$$

where C_1 and C_2 are positive constants, the second inequality follows from Lemma 5, and the last equality follows from the choice that $r_n = r_0 n^{-\frac{1-\gamma_n}{d}}$.

Because $\mathcal{X}$ is bounded by Assumption 6 and r_n is of order $\Omega(n^{-\frac{1-\gamma_n}{d}})$, each dimension needs to be partitioned into $O(n^{\frac{1-\gamma_n}{d}})$ segments, and thus the total number of grid boxes $T_{\boldsymbol{x}}$ necessary to cover $\mathcal{X}$ is $O(n^{1-\gamma_n})$. Because $T_{\boldsymbol{x}} \subset H_{\boldsymbol{x}}$ for each $\boldsymbol{x} \in \mathcal{X}$, obviously $\mathcal{X}$ can be covered with a set of $H_{\boldsymbol{x}}$, whose cardinality is $l(n) = O(n^{1-\gamma_n})$. Denote such set as $\mathcal{H}(n) = \{H_k(n)\}_{k=1}^{l(n)}$. By Equation (EC.15), it can be obtained that for each $n \geq N$, $\nu(H_k(n) \cap \mathcal{X})$ is $\Omega(n^{-(1-\gamma_n)})$ for all k. Note that for each $n \geq N$ and $\boldsymbol{x} \in \mathcal{X}$, we can find $1 \leq k \leq l(n)$ such that $\boldsymbol{x} \in H_k(n)$. Also note that the maximum distance between any two points in $H_k(n)$ is r_n. We can obtain that $H_k(n) \cap \mathcal{X} \subset \mathcal{S}(\boldsymbol{x}, r_n)$ for each $\boldsymbol{x} \in H_k(n)$. To summarize, for sufficiently large n, (i.e., $n \geq N$), we obtain the three properties about $\mathcal{H}(n)$:

(i) $l(n)$ is $O(n^{1-\gamma_n})$.

(ii) $\nu(H_k(n) \cap \mathcal{X})$ is $\Omega(n^{-(1-\gamma_n)})$ for all k, where $\nu(\cdot)$ is the d-dimensional volume.

(iii) For $\boldsymbol{x} \in \mathcal{X}$, if $H_k(n)$ is the box in $\mathcal{H}(n)$ such that $\boldsymbol{x} \in H_k(n)$, then $H_k(n) \cap \mathcal{X} \subset \mathcal{S}(\boldsymbol{x}, r_n) \subseteq \mathcal{S}(\boldsymbol{x}, r_i),\ i = 1, \ldots, n$.

Recall that $p_n = \gamma - b\varepsilon(n)$. Let $s(n)$ be an integer-valued function of n with order $\Theta(n^{p_n})$, and $N_k(n) = \sum_{i=1}^n \mathbb{1}_{\boldsymbol{x}_i \in H_k(n) \cap \mathcal{X}}$ be the number of sampled points that fall into $H_k(n) \cap \mathcal{X}$ among all n points. By the property (iii), if $\boldsymbol{x} \in H_k(n)$,

$$\{N_k(n) \geq s(n)\} \subseteq \{s_n(\boldsymbol{x}, r_n) \geq s(n)\}.$$

Because $\mathcal{H}(n)$ covers $\mathcal{X}$ and property (iii) holds for all $\boldsymbol{x} \in \mathcal{X}$, we can then have

$$D(n) = \bigcap_{k=1}^{l(n)} \{N_k(n) \geq s(n)\} \subseteq \bigcap_{\boldsymbol{x} \in \mathcal{X}} \{s_n(\boldsymbol{x}, r_n) \geq s(n)\}.$$

Taking complement set on both sides yields

$$D(n)^{\mathsf{c}} = \bigcup_{k=1}^{l(n)} \{N_k(n) < s(n)\} \supseteq \bigcup_{\boldsymbol{x} \in \mathcal{X}} \{s_n(\boldsymbol{x}, r_n) < s(n)\}. \tag{EC.16}$$

Let us first look at the probability $\mathbb{P}\{N_k(n) < s(n)\}$. For a fixed n, consider $H_k(n) \in \mathcal{H}(n)$. Since each design point $\boldsymbol{x}_i \in \boldsymbol{X}^n$ is generated from density $\psi_i \geq \alpha > 0$ on $\mathcal{X}$, for $i = 1, \ldots, n$, then

$$\mathbb{P}\{\boldsymbol{x}_i \in H_k(n)\} \geq \alpha \nu(\mathcal{X} \cap H_k(n)).$$

Notice that $\nu(H_k(n) \cap \mathcal{X}) \geq \frac{c_0}{n^{1-\gamma_n}}$ for some constant $c_0 > 0$, from the property (ii). Let B_i, $i = 1, 2, \ldots$, be independent Bernoulli random variables with parameter $\frac{\alpha c_0}{n^{1-\gamma_n}}$. So, by letting $c = \frac{1}{\alpha c_0}$,

$$\mathbb{P}\{N_k(n) < s(n)\} = \mathbb{P}\left\{\sum_{i=1}^n \mathbb{1}_{\boldsymbol{x}_i \in H_k(n) \cap \mathcal{X}} < s(n)\right\}$$

$$
\begin{aligned}
&\leq \mathbb{P}\left\{\sum_{i=1}^{n} B_i < s(n)\right\} \\
&\leq \sum_{j=0}^{s(n)-1} \binom{n}{j} \left(\frac{1}{cn^{1-\gamma_n}}\right)^{j} \left(1-\frac{1}{cn^{1-\gamma_n}}\right)^{n-j}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathbb{P}\{D(n)^{\mathrm{c}}\} = \mathbb{P}\left(\bigcup_{k=1}^{l(n)} \{N_k(n) < s(n)\}\right) &\leq \sum_{k=1}^{l(n)} \mathbb{P}\{N_k(n) < s(n)\} \\
&\leq l(n) \sum_{j=0}^{s(n)-1} \binom{n}{j} \left(\frac{1}{cn^{1-\gamma_n}}\right)^{j} \left(1-\frac{1}{cn^{1-\gamma_n}}\right)^{n-j} \\
&= l(n) \left(1-\frac{1}{cn^{1-\gamma_n}}\right)^{n} \sum_{j=0}^{s(n)-1} \binom{n}{j} \left(cn^{1-\gamma_n}-1\right)^{-j}.
\end{aligned}
$$

By the property (i), $l(n) \leq Cn^{1-\gamma_n}$ for some constant $C>0$. Recall that $s(n)$ is $\Theta(n^{p_n})$, so $s(n) \leq \bar{a} n^{p_n}$ for some constant $\bar{a} > 0$. We then follow the similar steps as in the proof of Lemma 4 in Andradóttir and Prudius (2010) to show

$$
\sum_{n=n_0}^{\infty} \mathbb{P}\{D(n)^{\mathrm{c}}\} < \infty, \tag{EC.17}
$$

for sufficiently large n_0. Note that $p_n = \gamma - b\varepsilon(n)$, so $p_n \to \gamma \in (0,1)$ as $n \to \infty$. Recall that $\gamma_n \to \gamma \in (0,1)$ as $n \to \infty$, too. Hence, let n be sufficiently large so that $s(n) \leq n/2$ and $cn^{1-\gamma_n} > 2$. Then, for $j < s(n) \leq n/2$, $\binom{n}{j} \leq \binom{n}{s(n)} \leq n^{s(n)} \leq n^{\bar{a} n^{p_n}}$. So,

$$
\mathbb{P}\{D(n)^{\mathrm{c}}\} \leq Cn^{1-\gamma_n+\bar{a}n^{p_n}} \left(1-\frac{1}{cn^{1-\gamma_n}}\right)^{n} \sum_{j=0}^{s(n)-1} \left(cn^{1-\gamma_n}-1\right)^{-j}.
$$

Since $cn^{1-\gamma_n} > 2$, then

$$
\sum_{j=0}^{s(n)-1} \left(cn^{1-\gamma_n}-1\right)^{-j} \leq \sum_{j=0}^{\infty} \left(\frac{1}{cn^{1-\gamma_n}-1}\right)^{j} = \frac{cn^{1-\gamma_n}-1}{cn^{1-\gamma_n}-2}.
$$

Thus,

$$
\mathbb{P}\{D(n)^{\mathrm{c}}\} \leq Cn^{1-\gamma_n+\bar{a}n^{p_n}} \left(1-\frac{1}{cn^{1-\gamma_n}}\right)^{n} \times \frac{cn^{1-\gamma_n}-1}{cn^{1-\gamma_n}-2} \leq \text{constant} \times n^{1+\bar{a}n^{p_n}} \left(1-\frac{1}{cn^{1-\gamma_n}}\right)^{n},
$$

for sufficiently large n. Observe that $n^{1-\gamma_n} \to \infty$ as $n \to \infty$ since $\gamma_n \to \gamma \in (0,1)$. Then, $\left(1 + 1/(-cn^{1-\gamma_n})\right)^{-cn^{1-\gamma_n}} \to e$ as $n \to \infty$, which implies that $\left(1+1/(-cn^{1-\gamma_n})\right)^{cn^{1-\gamma_n}} \to 1/e < 1$. So, there exists $0 < \beta < 1$ such that for sufficiently large n, $\left(1+1/(-cn^{1-\gamma_n})\right)^{n^{1-\gamma_n}} \leq \beta$, which further implies that $\left(1-1/(cn^{1-\gamma_n})\right)^{n} \leq \beta^{n^{\gamma_n}}$. Note that $b-1>a$. We can find $\delta > 0$ such that $b-1-\delta > a$. Observe that $\log\left(n^{1+\bar{a}n^{p_n}}\right) = \log n + \bar{a}n^{\gamma-(b-1)\varepsilon(n)}$. It can be obtained that $\log\left(n^{1+\bar{a}n^{p_n}}\right)/n^{\gamma-(b-1-\delta)\varepsilon(n)} \to 0$ as $n \to \infty$, which implies that $n^{1+\bar{a}n^{p_n}} \le e^{n^{\gamma-(b-1-\delta)\varepsilon(n)}}$ for sufficiently large n. Then,

$$\begin{aligned}\mathbb{P}\{D(n)^{\mathrm{c}}\} \le \text{constant} \times e^{n^{\gamma-(b-1-\delta)\varepsilon(n)}} \times \beta^{n^{\gamma_n}} &= \text{constant} \times \exp\left(n^{\gamma-(b-1-\delta)\varepsilon(n)} + n^{\gamma-a\varepsilon(n)}\log\beta\right)\\ &\le \text{constant} \times \exp\left(-n^{\gamma-(b-1-\delta)\varepsilon(n)}\right),\end{aligned}$$

for sufficiently large n, where the second inequality comes from the facts that $\log\beta < 0$ and $b-1-\delta > a$. Notice that $\gamma-(b-1-\delta)\varepsilon(n) \to \gamma \in (0,1)$ as $n \to \infty$. So we can find $t \in (0,1)$ such that $t \le \gamma-(b-1-\delta)\varepsilon(n)$ for sufficiently large n. Thus,

$$\mathbb{P}\{D(n)^{\mathrm{c}}\} \le \text{constant} \times \exp\left(-n^{\gamma-(b-1-\delta)\varepsilon(n)}\right) \le \text{constant} \times \exp(-n^t).$$

So, to prove Equation (EC.17), by the integral test for convergence, it suffices to prove $\int_{n_0}^{\infty} e^{-x^t}\,\mathrm{d}x < \int_0^{\infty} e^{-x^t}\,\mathrm{d}x < \infty$, for $t \in (0,1)$. Note that by the change of variable in integral,

$$\int_0^{\infty} e^{-x^t}\,\mathrm{d}x = \int_0^{\infty} \frac{1}{t} y^{\frac{1}{t}-1} e^{-y}\,\mathrm{d}y = \frac{1}{t}\Gamma(\frac{1}{t}) < \infty,$$

where $\Gamma(\cdot)$ is the gamma function. Thus, Equation (EC.17) is proved.

According to Equations (EC.16) and (EC.17), for any $\boldsymbol{x} \in \mathcal{X}$,

$$\sum_{n=n_0}^{\infty} \mathbb{P}\{s_n(\boldsymbol{x}, r_n) < s(n)\} \le \sum_{n=n_0}^{\infty} \mathbb{P}\left(\bigcup_{\boldsymbol{x}\in\mathcal{X}} \{s_n(\boldsymbol{x}, r_n) < s(n)\}\right) \le \sum_{n=n_0}^{\infty} \mathbb{P}\{D(n)^{\mathrm{c}}\} < \infty,$$

for sufficiently large n_0. Then, by Borel-Cantelli lemma, $\mathbb{P}\{s_n(\boldsymbol{x}, r_n) < s(n) \text{ infinitely often}\} = 0$. Since $s(n)$ is $\Theta(n^{p_n})$, we then immediately have $\mathbb{P}\{s_n(\boldsymbol{x}, r_n) \text{ is } \Omega(n^{p_n})\} = 1$. $\square$

EC.2.2. Proof of Lemma 10

Let $Z \sim \mathcal{N}(0,1)$, then

$$\begin{aligned}\mathbb{P}\{Z(\boldsymbol{x}) > c\} &= \mathbb{P}\{\mu_n^{\text{cap}}(\boldsymbol{x}) + [k_n^{\text{cap}}(\boldsymbol{x}, \boldsymbol{x})]^{1/2} Z > c\}\\ &= \mathbb{P}\left\{Z > \frac{c - \mu_n^{\text{cap}}(\boldsymbol{x})}{[k_n^{\text{cap}}(\boldsymbol{x}, \boldsymbol{x})]^{1/2}}\right\}\\ &\ge \mathbb{P}\left\{Z > \frac{\overline{M} - \underline{M}}{\underline{\tau}}\right\}\\ &= 1 - \Phi((\overline{M} - \underline{M})/\underline{\tau}),\end{aligned}$$

where the inequality is due to the facts that $c \le \overline{M}$, $\mu_n^{\text{cap}}(\boldsymbol{x}) \ge \underline{M}$, and $k_n^{\text{cap}}(\boldsymbol{x}, \boldsymbol{x}) \ge \underline{\tau}^2$. Besides, since $\mu_n^{\text{cap}}(\boldsymbol{x}) \le c$, $\mathbb{P}\{Z(\boldsymbol{x}) > c\} \le 0.5$. Hence,

$$f_n(\boldsymbol{x}) = \frac{\mathbb{P}\{Z(\boldsymbol{x}) > c\}}{\int_{\mathcal{X}} \mathbb{P}\{Z(\boldsymbol{x}) > c\}\mathrm{d}\boldsymbol{z}} \ge \frac{1 - \Phi((\overline{M} - \underline{M})/\underline{\tau})}{\int_{\mathcal{X}} 0.5\,\mathrm{d}\boldsymbol{z}} = \alpha.$$

$\square$

EC.3. Analyses of Some Correlation Functions

Section EC.3 analyzes the sample path differentiability and lower bound of several correlation functions.

The sample path differentiability

The Gaussian correlation function

$$\rho(\boldsymbol{x},\boldsymbol{x}') = \rho(\boldsymbol{\delta}) = \exp\{-\sum_{j=1}^{d}\theta_j\delta_j^2\}$$

with $\boldsymbol{\delta}=\boldsymbol{x}-\boldsymbol{x}'$. It is well known that the sample paths from a Gaussian process having this correlation function are infinitely differentiable almost surely. Moreover, the correlation function of the first-order derivative surface $\ddot{\rho}_j(\boldsymbol{\delta})/\ddot{\rho}_j(\mathbf{0}) = (1-2\theta_j\delta_j^2)\rho(\boldsymbol{\delta})$ (a function of $\boldsymbol{\delta}$) is still stationary. Therefore, the Gaussian process with Gaussian correlation function satisfies Assumption 5.

The Matérn correlation function

$$\rho_v(\boldsymbol{x},\boldsymbol{x}') = \rho(\boldsymbol{\delta}) = \prod_{j=1}^{d}\exp\left(-\frac{\sqrt{2v}|\delta_j|}{l_j}\right)\frac{\Gamma(t+1)}{\Gamma(2t+1)}\sum_{i=0}^{t}\frac{(t+i)!}{i!(t-i)!}\left(\frac{\sqrt{8v}|\delta_j|}{l_j}\right)^{t-i}.$$

with $\boldsymbol{\delta}=\boldsymbol{x}-\boldsymbol{x}'$, $l_j>0$ and $v=t+1/2$ (t is a positive integer). It is well known that the sample paths from a one-dimensional Gaussian process having Matérn correlation function are continuously differentiable almost surely of order $\lceil v\rceil-1$ (Santner et al. 2003, p. 43-44). In addition, it can be checked that the one-dimensional Matérn correlation function (with v being a half integer and greater than 1) and the correlation function of its first-order derivative surface all satisfy Assumption 3 (iii). The d-dimensional Matérn correlation function is the product of the one-dimensional correlation functions. It can be verified that the Gaussian process with this d-dimensional Matérn correlation function still has continuously differentiable sample paths. Therefore, the Gaussian process having Matérn correlation function with $v=t+1/2$ (t is a positive integer) satisfies Assumption 5.

The rational quadratic function

$$\rho(\boldsymbol{x},\boldsymbol{x}') = \rho(\boldsymbol{\delta}) = \left(1+\frac{\sum_{j=1}^{d}\delta_j^2}{2\alpha l^2}\right)^{-\alpha}$$

with $\boldsymbol{\delta}=\boldsymbol{x}-\boldsymbol{x}'$ and $\alpha, l>0$. By taking twice differentiation, we can obtain that $\ddot{\rho}_j(\boldsymbol{\delta}) = -\frac{1}{l^2}\left(1+\frac{\sum_{j=1}^{d}\delta_j^2}{2\alpha l^2}\right)^{-\alpha-1} + \frac{(\alpha+1)\delta_j^2}{\alpha l^4}\left(1+\frac{\sum_{j=1}^{d}\delta_j^2}{2\alpha l^2}\right)^{-\alpha-2}$ exists and is continuous with $\ddot{\rho}_j(0) = -1/l^2$. Then, it can be checked that $\ddot{\rho}_j(\boldsymbol{\delta})/\ddot{\rho}_j(0)$ satisfies Assumption 3 (iii). So it can be concluded that the sample paths of the Gaussian process having this correlation function are continuously differentiable almost surely (Abrahamsen 1997). Moreover, the correlation function $\ddot{\rho}_j(\boldsymbol{\delta})/\ddot{\rho}_j(0)$ is still stationary. Therefore, the Gaussian process having rational quadratic function satisfies Assumption 5.

The Lower Bound of Correlation Functions

The power exponential correlation function

$$\rho(\boldsymbol{x},\boldsymbol{x}') = \exp\{-\sum_{j=1}^{d}\theta_j|x_j - x_j'|^{\eta}\}$$

with $\theta_j > 0$ and $0 < \eta \le 2$. Define θ_{max} be the maximum among θ_j, $j = 1, 2, \ldots, d$, and $\|\cdot\|_\eta$ is a norm on $\mathbb{R}^d$ such that $\|\boldsymbol{x}\|_\eta = (\sum_{j=1}^{d}|x_j|^{\eta})^{1/\eta}$. Since the space $\mathbb{R}^d$ is of finite dimension, any norm $\|\cdot\|_{\eta_1}$ is equivalent to any other norm $\|\cdot\|_{\eta_2}$, i.e., $\|\boldsymbol{x}\|_{\eta_1} \le C_{\eta_1\eta_2}\|\boldsymbol{x}\|_{\eta_2}$ for a positive number $C_{\eta_1\eta_2}$ and all $\boldsymbol{x}$ in $\mathcal{X}$ (Kreyszig 1978). Then, we can have $\sum_{j=1}^{d}\theta_j|x_j - x_j'|^{\eta} \le \theta_{max}\|\boldsymbol{x}-\boldsymbol{x}'\|_\eta^{\eta} \le \theta_{max}C_{\eta 2}\|\boldsymbol{x}-\boldsymbol{x}'\|^{\eta} = C_r\|\boldsymbol{x}-\boldsymbol{x}'\|^{\eta}$ for $\boldsymbol{x}'$ in $S(x, r_n)$. According to $e^x \ge 1 + x$, we can obtain $\rho(\boldsymbol{x},\boldsymbol{x}') \ge 1 - C_r\|\boldsymbol{x}-\boldsymbol{x}'\|^{\eta}$. We note that the power exponential correlation function is the Gaussian correlation function with $\eta = 2$.

When v is half integer, i.e., $v = t + 1/2$, where t is a positive integer, the Matérn correlation function is

$$\rho_v(\boldsymbol{x},\boldsymbol{x}') = \prod_{j=1}^{d}\exp\left(-\frac{\sqrt{2v}\left|x_j - x_j'\right|}{l_j}\right)\frac{\Gamma(t+1)}{\Gamma(2t+1)}\sum_{i=0}^{t}\frac{(t+i)!}{i!(t-i)!}\left(\frac{\sqrt{8v}\left|x_j - x_j'\right|}{l_j}\right)^{t-i}.$$

It can be verified that

$$\begin{aligned}
\rho_v(\boldsymbol{x},\boldsymbol{x}') &= \prod_{j=1}^{d}\exp\left(-\frac{\sqrt{2v}\left|x_j - x_j'\right|}{l_j}\right)\frac{\Gamma(t+1)}{\Gamma(2t+1)}\sum_{i=0}^{t}\frac{(t+i)!}{i!(t-i)!}\left(\frac{\sqrt{8v}\left|x_j - x_j'\right|}{l_j}\right)^{t-i} \\
&= \prod_{j=1}^{d}\left(1 + \frac{\sqrt{2v}\left|x_j - x_j'\right|}{l_j} + \frac{\Gamma(t+1)}{\Gamma(2t+1)}\sum_{i=0}^{t-2}\frac{(t+i)!}{i!(t-i)!}\left(\frac{\sqrt{8v}\left|x_j - x_j'\right|}{l_j}\right)^{t-i}\right)\cdot\exp\left(-\sqrt{2v}\sum_{j=1}^{d}\frac{\left|x_j - x_j'\right|}{l_j}\right) \\
&\ge \prod_{j=1}^{d}\left(1 + \frac{\sqrt{2v}\left|x_j - x_j'\right|}{l_j}\right)\cdot\exp\left(-\sqrt{2v}\sum_{j=1}^{d}\frac{\left|x_j - x_j'\right|}{l_j}\right) \\
&\ge \left(1 + \sum_{j=1}^{d}\frac{\sqrt{2v}\left|x_j - x_j'\right|}{l_j}\right)\left(1 - \sum_{j=1}^{d}\frac{\sqrt{2v}\left|x_j - x_j'\right|}{l_j}\right) \\
&\ge 1 - 2v\left(\sum_{j=1}^{d}\frac{\left|x_j - x_j'\right|}{l_j}\right)^2 \\
&\ge 1 - \frac{2v}{l_{min}}\left(\sum_{j=1}^{d}\left|x_j - x_j'\right|\right)^2 \\
&\ge 1 - C_{12}\frac{2v}{l_{min}}\|\boldsymbol{x}-\boldsymbol{x}'\|^2,
\end{aligned}$$

where $l_{min} = \min_{j=1,\cdots,d} l_j$ and the last inequality is due to the equivalence of the norm $\|\cdot\|_\eta$ on finite dimensional space $\mathbb{R}^d$.

The quadratic covariance function

$$\rho(\boldsymbol{x},\boldsymbol{x}') = \left(1 + \frac{\sum_{j=1}^{d}\left|x_j - x_j'\right|^2}{2\alpha l^2}\right)^{-\alpha},$$

with $\alpha, l > 0$. It can also be verified that

$$\begin{aligned}\rho(r) &= \frac{1}{\left(1+\frac{\sum_{j=1}^d (x_j - x_j')^2}{2\alpha l^2}\right)^\alpha} \\ &\geq \frac{1}{(e^{\sum_{j=1}^d (x_j - x_j')^2/2\alpha l^2})^\alpha} \\ &= e^{-\frac{\sum_{j=1}^d (x_j - x_j')^2}{2l^2}} \\ &\geq 1 - \frac{\sum_{j=1}^d (x_j - x_j')^2}{2l^2} \\ &= 1 - \frac{1}{2l^2}\|\boldsymbol{x} - \boldsymbol{x}'\|^2.\end{aligned}$$

EC.4. Sampling Scheme

In the following, we will describe the acceptance-rejection sampling (ARS) scheme and the Markov chain coordinate sampling (MCCS) scheme in details. According to Equation (15), we have that $\mathbb{P}\{Z(\boldsymbol{x}) > c\} \leq \mathbb{P}\{Z(\boldsymbol{x}) > \mu_n^{\text{cap}}(\boldsymbol{x})\} = 1/2$, and

$$f_n(\boldsymbol{x}) = \frac{\mathbb{P}\{Z(\boldsymbol{x}) > c\}}{\int_{\mathcal{X}} \mathbb{P}\{Z(\boldsymbol{z}) > c\} \, \mathrm{d}\boldsymbol{z}} \leq \frac{(1/2)\nu(\mathcal{X})}{\int_{\mathcal{X}} \mathbb{P}\{Z(\boldsymbol{z}) > c\} \, \mathrm{d}\boldsymbol{z}} \cdot \frac{1}{\nu(\mathcal{X})}.$$

Since $\nu(\mathcal{X})/\int_{\mathcal{X}} \mathbb{P}\{Z(\boldsymbol{z}) > c\} \, \mathrm{d}\boldsymbol{z}$ is a constant and $\frac{1}{\nu(\mathcal{X})}$ is the probability density of the uniform distribution on $\mathcal{X}$, it is easy to see that the following ARS scheme can output a sample following the density $f_n(\boldsymbol{x})$.

ARS Scheme

Step 1. Generate a sample $\boldsymbol{y}$ uniformly in $\mathcal{X}$ and u uniformly in $(0, 1)$.

Step 2. If $u \leq 2\,\mathbb{P}\{Z(\boldsymbol{y}) > c\}$, return $\boldsymbol{y}$; otherwise, go to Step 1.

The ARS algorithm avoids the calculation of the integration in the denominator of Equation (15). However, when the sampling probabilities are concentrated on small regions, the acceptance rate may be very low, and thus impacts the efficiency of the ARS scheme. Therefore, a MCCS scheme is proposed.

MCCS Scheme

Step 0. Let $t = 0$, $\boldsymbol{y} = \boldsymbol{y}_0$. Specify the iteration number T.

Step 1. Let $t = t + 1$. Sample an integer j from 1 to d uniformly. Let $l(\boldsymbol{y}, j)$ be the line that passes through $\boldsymbol{y}$ and parallel to the y_j coordinate axis. Then $l(\boldsymbol{y}, j) \cap \mathcal{X}$ is the line segment that is contained in $\mathcal{X}$. Sample a point on $l(\boldsymbol{y}, j) \cap \mathcal{X}$ uniformly, whose j-th coordinate is denoted as b. Set $\boldsymbol{z} = \boldsymbol{y}$ and $z_j = b$.

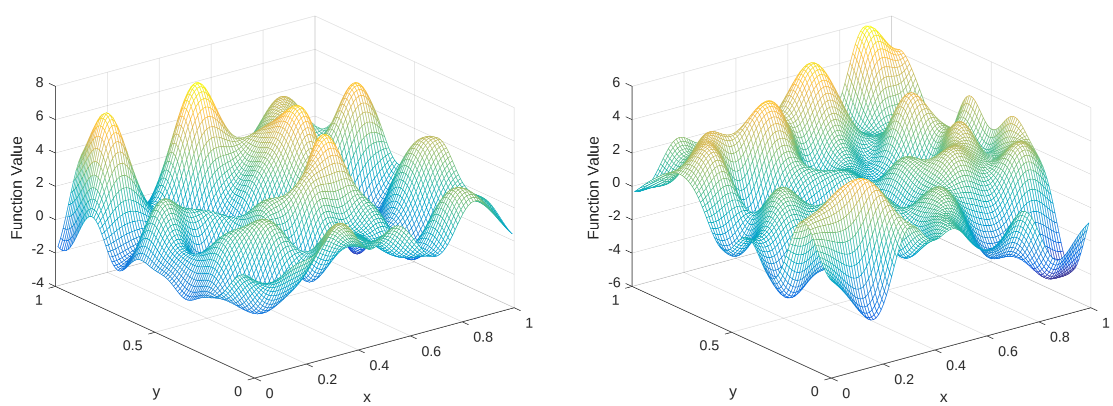
Step 2. Sample an u uniformly in $(0,1)$. If $u \leq f_n(\boldsymbol{z})/f_n(\boldsymbol{y}) = \mathbb{P}\{Z(\boldsymbol{z}) > c\}/\mathbb{P}\{Z(\boldsymbol{y}) > c\}$, set $\boldsymbol{y} = \boldsymbol{z}$.

Step 3. If $t = T$, return $\boldsymbol{y}$; otherwise go to Step 1.

Similar to the proof in Baumert et al. (2009), it can be show that, as $T \to \infty$, the probability density function of the random output $\boldsymbol{y}$ converges to $f_n(\boldsymbol{x})$, regardless of $\boldsymbol{y}_0$. In practice, the starting point $\boldsymbol{y}_0$ is usually the current optimal solution. The MCCS scheme guarantees to sample (approximately) a point every T steps. Therefore, it may become more efficient than the ARS scheme when the acceptance rate in the ARS scheme becomes very low (i.e., lower than $1/T$). We note that the sampling scheme, which can be used in the GPS-C algorithm, is not restricted to the ARS or the MCCS scheme. Many other sampling schemes, e.g., the Gaussian mixture model of Sun et al. (2018), can also be considered.

EC.5. Numerical Experiments

Section EC.5 contains the parameter settings of the algorithms used in Section 6, and some supplementary figures in the numerical experiments.



(a) Example 1

(b) Example 2

Figure EC.1 Two Examples of the Generated Sample Paths with $\mu_0 = 1$, $\sigma^2 = 4$ and $\theta = (80, 80)$

Table EC.1 The Parameter Settings of the GPS-C Algorithm in Section 6.1

Problem	Parameters
Two-dimensional	$\mu_0 = 1$, $\theta = (80, 80)$, $\tau^2 = 4$, $r = 10$, $\lambda^2 = 0.25$, $\underline{\tau}^2 = 1$, $\underline{M} = 1$, $\overline{M} = 20$, $T = 100$
Three-dimensional	$\mu_0 = 1$, $\theta = (40, 40)$, $\tau^2 = 9$, $r = 10$, $\lambda^2 = 0.25$, $\underline{\tau}^2 = 3$, $\underline{M} = 1$, $\overline{M} = 40$, $T = 100$

Branin problem

$$\max_{-5\leq x_1\leq 10, 0\leq x_2\leq 15} g(x_1,x_2) = -\left[\left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1-\frac{1}{8\pi}\right)\cos x_1 + 10\right].$$

Six-hump problem

$$\max_{-2\leq x_1,x_2\leq 2} g(x_1,x_2) = -\left[4x_1^2 - 2.1x_1^4 + 1/3x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4\right].$$

Hills Problem

$$\max_{0\leq x_1,x_2\leq 100} g(x_1,x_2) = 10\cdot\frac{\sin^6(0.05\pi x_1)}{2^{2((x_1-90)/50)^2}} + 10\cdot\frac{\sin^6(0.05\pi x_2)}{2^{2((x_2-90)/50)^2}}.$$

10-Dimensional Rosenbrock Problem

$$\max_{-10\leq x_1,\dots,x_{10}\leq 10} g(x_1,\dots,x_{10}) := -10^{-4}\times\sum_{i=1}^{9}((1-x_i)^2 + 100(x_{i+1}-x_i^2)^2),$$

d-Dimensional Weighted Sphere Problem

$$\max_{-5.12\leq x_1,\dots,x_d\leq 5.12} g(x_1,\dots,x_d) := -\sum_{i=1}^{d}(i\times x_i^2),$$

Table EC.2 The Basic Information of the Test Problems

Problem	Dimension	Optimal value	Optimal solution
Branin	2	-0.398	(9.425,2.475) or (-3.142,12.275) or (3.142,2.275)
Six-hump	2	1.032	(0.090, -0.713) or (-0.090,0.713)
Hills	2	20.0	(90.0,90.0)
Rosenbrock	10	0	(1,1,...,1)
Weighted Sphere	d	0	(0,0,...,0)

Table EC.3 The Parameter Settings of the GPS-C Algorithm in Section 6.2.1

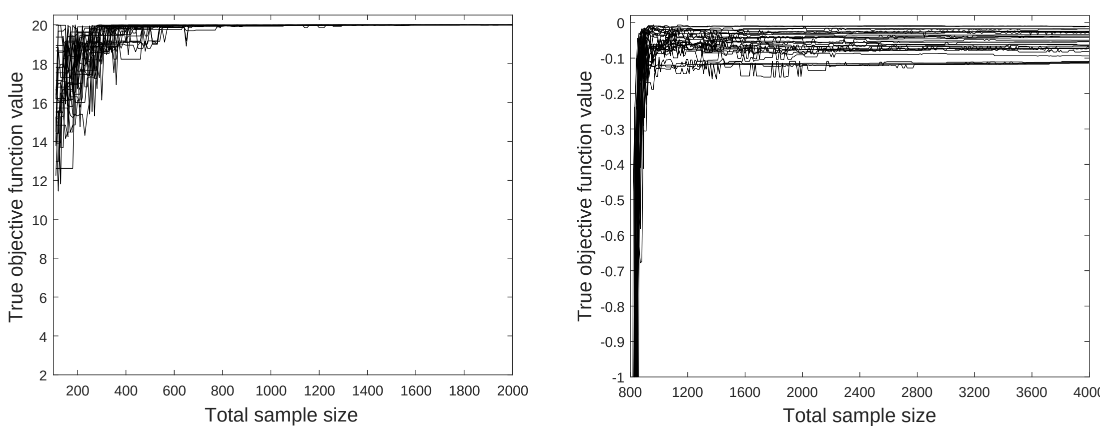
Algorithm	Problem	Variance	Parameters
GPS-C	Branin	equal	$r=10$, $\underline{\tau}^2=0.01$, $\underline{M}=-10$, $\overline{M}=20$, $T=100$
GPS-C	Six-hump	equal	$r=10$, $\underline{\tau}^2=0.01$, $\underline{M}=-10$, $\overline{M}=20$, $T=100$
GPS-C	Hills	equal	$r=10$, $\underline{\tau}^2=0.25$, $\underline{M}=0$, $\overline{M}=40$, $T=100$

Table EC.4 The Parameter Settings of the Algorithms in Section 6.2.2

Algorithm	Problem	Variance	Parameters
GPS-C	Hills	unequal	$h_0 = 2,\ \beta = 0.009,\ r = 10,\ \underline{\tau}^2 = 4,\ \underline{M} = 0,\ \overline{M} = 40,\ T = 100$
ASR	Hills	unequal	$b = 1.1,\ C = 1,\ g = 0.5,\ \delta = 1,\ K = 10,\ T = 1$
IHR-SO	Hills	unequal	$\gamma = 0.91,\ \beta = 0.009,\ s = 0.9,\ \kappa_r = 1$
AP-SO	Hills	unequal	$\gamma = 0.91,\ \beta = 0.009,\ s = 0.9,\ \kappa_r = 1,\ R = 1$
GPS-C	Rosenbrock	unequal	$h_0 = 5,\ \beta = 0.009,\ r = 10,\ \underline{\tau}^2 = 0.01,\ \underline{M} = -100,\ \overline{M} = 10,\ T = 100$
ASR	Rosenbrock	unequal	$b = 1.1,\ C = 1,\ g = 0.5,\ \delta = 0.01,\ K = 10,\ T = 1$
IHR-SO	Rosenbrock	unequal	$\gamma = 0.91,\ \beta = 0.009,\ s = 0.9,\ \kappa_r = 1$
AP-SO	Rosenbrock	unequal	$\gamma = 0.91,\ \beta = 0.009,\ s = 0.9,\ \kappa_r = 1,\ R = 0.4$

Table EC.5 The Performance of ASR, SOSA and GPS-C Algorithms up to Different Sample Size on the Hills and Rosenbrock Problems with Heteroscedastic Simulation Noise

Problem	**ASR**			**AP-SO**			**IHR-SO**			**GPS-C**		
	600	**1000**	**2000**	**600**	**1000**	**2000**	**600**	**1000**	**2000**	**600**	**1000**	**2000**
Hills	1.382	0.836	0.548	1.654	1.354	0.926	1.726	1.007	0.669	0.063	0.018	0.007
Problem	**ASR**			**AP-SO**			**IHR-SO**			**GPS-C**		
	1000	**2000**	**4000**	**1000**	**2000**	**4000**	**1000**	**2000**	**4000**	**1000**	**2000**	**4000**
Rosenbrock	0.451	0.201	0.062	0.869	0.138	0.051	0.379	0.273	0.141	0.043	0.034	0.026



(a) Hills problem (b) Rosenbrock problem

Figure EC.2 Performance of the Revised GPS-C Algorithm on the Hills Problem and Rosenbrock Problem

Table EC.6 The Parameter Settings of the GPS-C Algorithm in Section 6.3

Weighted Sphere problem	Parameters
d-dimensional	$r = 10,\ \underline{\tau}^2 = 0.2,\ \underline{M} = -1000,\ \overline{M} = 10,\ T = 100$

References

Abrahamsen P (1997) A review of Gaussian random fields and correlation functions. Technical Report, Norwegian Computing Center, Oslo, 1997.

Adler RJ, Taylor JE (2007) *Random Fields and Geometry* (Springer Science & Business Media).

Andradóttir S, Prudius AA (2010) Adaptive random search for continuous simulation optimization. *Naval Research Logistics (NRL)* 57(6):583–604.

Baumert S, Ghate A, Kiatsupaibul S, Shen Y, Smith RL, Zabinsky ZB (2009) Discrete hit-and-run for sampling points from arbitrary distributions over subsets of integer hyperrectangles. *Operations Research* 57(3):727–739.

Fitzpatrick P (2009) *Advanced Calculus*, volume 5 (American Mathematical Soc.).

Kreyszig E (1978) *Introductory Functional Analysis with Applications*, volume 1 (New York: Wiley).

Santner TJ, Williams BJ, Notz WI (2003) *The Design and Analysis of Computer Experiments* (New York: Springer).

Sun W, Hu Z, Hong, LJ (2018) Gaussian mixture model-based random search for continuous optimization via simulation. *2018 Winter Simulation Conference (WSC)*, 2003–2014 (IEEE).