Online Supplemental Materials to "Function-on-Function Gaussian Process with Application in Robust Design"

Jingru Huang[a], Haijie Xu[a], Yifei Gao[a], Chen Zhang[a,*]

[a] Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

jingruhuang@tsinghua.edu.cn, xu-hj22@mails.tsinghua.edu.cn, gao-yf@mail.tsinghua.edu.cn, zhangchen01@tsinghua.edu.cn

[*] Corresponding author

Appendix A: Additional Examples of Operator-Valued Kernels

There are other types of operator-valued kernels that can also be considered as covariance functions for the FFGP, such as multiplication operators and composite operators, which have the following forms:

Multiplication operator: $T_{\mathcal{Y}} y(t) = k_y(t)y(t)$, and $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_x(\boldsymbol{x}_i, \boldsymbol{x}_j)T_{\mathcal{Y}}$,

where k_x a positive kernel and k_y a positive real function.

Composition operator: $C_\varphi : f \longmapsto f \circ \varphi$, and $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = C_{\psi(\boldsymbol{x}_i)} C'_{\psi(\boldsymbol{x}_j)}$,

where C'_ψ is the adjoint operator of C_ψ. $\psi(\boldsymbol{x}_i)$ and $\psi(\boldsymbol{x}_j)$ are maps of Ω_y into itself. More details can be seen in Section 5.4 of Kadri et al. (2016).

Appendix B: Computations of Eigenvalues and Eigenfunctions of $T_{\mathcal{Y}}$

According to Lemma 1, we have

Exponential kernel: $T_{\mathcal{Y}} v(t) = \int_0^t \exp^{-(t-s)/\tau_y} v(s)ds + \int_t^1 \exp^{-(s-t)/\tau_y} v(s)ds = \beta v(t)$.

It is easy to know that v is differentiable because the middle of the above equation is differentiable. Then, the first-order derivative of $\beta v(t)$ to t is $\beta v'(t) = -\frac{1}{\tau_y}\exp^{-t/\tau_y}\int_0^t \exp^{s/\tau_y} v(s)ds + \frac{1}{\tau_y}\exp^{t/\tau_y}\int_t^1 \exp^{-s/\tau_y} v(s)ds$. Note also that $\tau_y v'(0) = v(0)$ and $\tau_y v'(1) = -v(1)$. Similarly, the second-order derivative of $\beta v(t)$ to t is also differentiable. It is given by $\beta\tau_y^2 v''(t) = (\beta - 2\tau_y)v(t)$. According to Keener (2018), we have $v(t) = A\sin(\mu t) + B\cos(\mu t)$ when $\beta < 2\tau_y$. From the above conditions, we obtain $\beta_i = \frac{2\tau_y}{1+\tau_y^2\mu_i^2}$, and $v_i(t) = \tau_y\mu_i\cos(\mu_i t) + \sin(\mu_i t)$. And $\mu_i > 0$ is the i-th solution of the equation $2\cot\mu = \tau_y\mu - 1/(\tau_y\mu)$ for $i \geq 1$.

Wiener kernel: $T_{\mathcal{Y}} v(t) = \int_0^t \frac{s}{\tau_y} v(s)ds + \int_t^1 \frac{t}{\tau_y} v(s)ds = \beta v(t)$.

We have $v(0) = 0$. The derivative of $\beta v(t)$ to t is $\beta v'(t) = \int_t^1 \frac{1}{\tau_y} v(s)ds$. We deduce that $v'(1) = 0$, and v' is differentiable. Then we have $\beta v''(t) = -\frac{1}{\tau_y} v(t)$. The general solution is $v(t) = A\sin(ut) + B\cos(ut)$. From boundary conditions, we have $u = \sqrt{1/(\beta\tau_y)} = \frac{\pi}{2} + (i-1)\pi$. Then, we obtain $\beta_i = \frac{4}{(2i-1)^2\pi^2\tau_y^2}$ and $v_i(t) = \sin\left(\frac{2i-1}{2}\pi t\right)$.

Appendix C: Computations of conditional scalar-valued operator-variance

Under Assumption 2, the conditional covariance operator $\mathrm{Var}\left[f(\boldsymbol{x}_0) \mid \boldsymbol{Y}\right]$ in (4) is positive and a trace class, and its corresponding $s^2(\boldsymbol{x}_0, \boldsymbol{x}_0)$ is finite. Then $f(\boldsymbol{x}_0)$ defines a measure on $\mathcal{Y}$ (i.e. a $\mathcal{Y}$-valued random variable) (Owhadi 2023). From Definition 2 and Lemma 1, we have that $s^2(\boldsymbol{x}_0, \boldsymbol{x}_0) = \sum_{i=1}^\infty \langle K'(\boldsymbol{x}_0, \boldsymbol{x}_0)v_i, v_i\rangle = \sigma^2 \sum_{i=1}^\infty \left\{ \langle K(\boldsymbol{x}_0, \boldsymbol{x}_0)v_i, v_i\rangle - \left\langle \boldsymbol{K}(\boldsymbol{x}_0)^T \left(\mathbf{K} + \lambda/\sigma^2 I_{\mathcal{Y}}\right)^{-1} \boldsymbol{K}(\boldsymbol{x}_0)v_i, v_i \right\rangle \right\} = \sigma^2 \sum_{i=1}^\infty \{I_{1i} - I_{2i}\}$, $\forall \boldsymbol{x}_0 \in \mathcal{X}^p$. For the first item, we have $I_{1i} = \left\langle \sum_{j=1}^\infty k_x(\boldsymbol{x}_0, \boldsymbol{x}_0)\beta_j \langle v_j, v_i\rangle v_j, v_i \right\rangle = k_x(\boldsymbol{x}_0, \boldsymbol{x}_0)\beta_i$. For the second item, we have $\boldsymbol{K}(\boldsymbol{x}_0)v_i = \boldsymbol{k}_x(\boldsymbol{x}_0)\sum_{j=1}^\infty \beta_j \langle v_i, v_j\rangle v_j = \boldsymbol{k}_x(\boldsymbol{x}_0)\beta_i v_i$. Then, there is $\left(\mathbf{K} + \frac{\lambda}{\sigma^2} I_{\mathcal{Y}}\right)^{-1} \boldsymbol{K}(\boldsymbol{x}_0)v_i = \sum_{j=1}^\infty \frac{1}{\alpha_j\beta_i + \lambda/\sigma^2} \langle \boldsymbol{k}_x(\boldsymbol{x}_0)\beta_i v_i, z_j\rangle z_j$.

Since $z = w \otimes v$, we have $\left(\mathbf{K} + \frac{\lambda}{\sigma^2} I_{\mathcal{Y}}\right)^{-1} \boldsymbol{K}(\boldsymbol{x}_0) v_i = \sum_{j=1}^{n} \frac{\beta_i}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x^T(\boldsymbol{x}_0) \boldsymbol{w}_j \boldsymbol{w}_j v_i$. Then we obtain $\boldsymbol{K}(\boldsymbol{x}_0)^T \left(\mathbf{K} + \frac{\lambda}{\sigma^2} I_{\mathcal{Y}}\right)^{-1} \boldsymbol{K}(\boldsymbol{x}_0) v_i = \boldsymbol{k}_x(\boldsymbol{x}_0)^T \sum_{l=1}^{\infty} \beta_l \left\langle \sum_{j=1}^{n} \boldsymbol{a}_j v_i, v_l \right\rangle v_l = \sum_{j=1}^{n} \frac{\beta_i^2}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x^T(\boldsymbol{x}_0) \boldsymbol{w}_j \boldsymbol{w}_j \boldsymbol{k}_x(\boldsymbol{x}_0) v_i$. Thus, $I_{2i} = \sum_{j=1}^{n} \frac{\beta_i^2}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x^T(\boldsymbol{x}_0) \boldsymbol{w}_j \boldsymbol{w}_j \boldsymbol{k}_x(\boldsymbol{x}_0)$. In summary,

$$s^2(\boldsymbol{x}_0, \boldsymbol{x}_0) = \sum_{i=1}^{\infty} \delta_i = \sum_{i=1}^{\infty} \sigma^2 \beta_i \left[k_x(\boldsymbol{x}_0, \boldsymbol{x}_0) - \sum_{j=1}^{n} \frac{\beta_i}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x(\boldsymbol{x}_0)^T \boldsymbol{w}_j \boldsymbol{w}_j^T \boldsymbol{k}_x(\boldsymbol{x}_0) \right]. \tag{A0}$$

Appendix D: The Algorithm and Computational Complexity

Algorithm A1 Parameter estimation for training FFGP

Input: Training data $\mathbf{X}$ and $\boldsymbol{Y}$, pre-specific basis space $\boldsymbol{g}$, initialized $\boldsymbol{\Theta}_0 = \{\boldsymbol{b}_0, \sigma_0^2, \boldsymbol{\tau}_{x0}, \tau_{y0}, \lambda_0\}$.

Initialize: $\boldsymbol{b} \leftarrow \boldsymbol{b}_0$, $\sigma^2 \leftarrow \sigma_0^2$, $\boldsymbol{\tau}_x \leftarrow \boldsymbol{\tau}_{x0}$, $\tau_y \leftarrow \tau_{y0}$, $\lambda \leftarrow \lambda_0$.

1: **while** $\boldsymbol{b}$, σ^2, $\boldsymbol{\tau}_x$, τ_y, λ not converge **do**

2: Eigendecomposition $\mathbf{K}_x$. Let $\alpha_j \in \mathbb{R}$ be the j-th eigenvalue, and let $\boldsymbol{w}_j \in \mathbb{R}^n$ be the j-th eigenvector, $j = 1, \cdots, n$.

3: Eigendecomposition $T_{\mathcal{Y}}$. Let $\beta_i \in \mathbb{R}$ be the i-th eigenvalue, and $v_i \in \mathcal{Y}$ be the i-th eigenfunction. Select m by (16).

4: Update $\boldsymbol{b}$ by (20) and $\sigma^2, \boldsymbol{\tau}_x, \tau_y, \lambda$ by (19) with the L-BFGS method.

5: **end while**

REMARK 1. Denote that N_{mc} is the computational complexity of approximating L^2 norm by numerical methods, and $O(n^\omega)$ is the computational complexity of obtaining the eigenvalues and eigenfunctions of $\mathbf{K}_\mathbf{x}$, where $2 < \omega < 2.376$ (Williams 2012). We denote the computational complexity of $T_{\mathcal{Y}}$ feature decomposition as 1 because of the closed forms of eigenvalues and eigenfunctions. Then, the computational complexity of updating $\boldsymbol{b}$ is $O(N_{mc} n^2 m)$. When updating other $p+3$ hyperparameters $\sigma^2, \boldsymbol{\tau}_x, \tau_y, \lambda$, the computational complexity of computing the derivatives of the likelihood function with respect to them is $O(N_{mc} n^2 m + (p+3)mn)$. We use the L-BFGS algorithm to optimize the likelihood, the number of iterations is $O(log(n))$ (Bottou 2010). Then the computational complexity of Algorithm A1 is $O\left(log(n)\left(N_{mc} m n^2 + n^\omega + (p+3)mn\right)\right)$.

REMARK 2. When using the FRGD algorithm to select $\boldsymbol{x}^\star$, for any initial function, the computational complexity of the derivative of $L(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ is $O\left(N_{mc} m^2 n^2 + p n^2\right)$, and the computational complexity of updating $L_m(\boldsymbol{x})$ is $O\left(N_{mc}^2 m^2 n^2\right)$. Therefore, the computational complexity of Algorithm A2 for one time is $O\left(N_{mc}^2 m^2 n^2 + p n^2\right)$.

Appendix E: Computation of Proposition 1

From (22), there is $L_m(\boldsymbol{x}_0) = \|\mu + \boldsymbol{k}_x(\boldsymbol{x}_0)^T \sum_{i=1}^{m} \sum_{j=1}^{n} \eta_{ij} \boldsymbol{w}_j v_i - y^\star\|^2 + \sum_{i=1}^{m} \sigma^2 \beta_i \left[k_x(\boldsymbol{x}_0, \boldsymbol{x}_0) - \sum_{i=1}^{n} \frac{\beta_i}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x(\boldsymbol{x}_0)^T \boldsymbol{w}_j \boldsymbol{w}_j^T \boldsymbol{k}_x(\boldsymbol{x}_0) \right] = L_1(\boldsymbol{x}_0) + L_2(\boldsymbol{x}_0)$. For $L_1(\boldsymbol{x}_0)$, it is easy to derive that $L_1(\boldsymbol{x}_0) = \|\mu - y^\star\|^2 + \boldsymbol{k}_x(\boldsymbol{x}_0)^T \sum_{i=1}^{m} z_i z_i^T \boldsymbol{k}_x(\boldsymbol{x}_0) - 2\boldsymbol{k}_x(\boldsymbol{x}_0)^T \sum_{i=1}^{m} z_i \langle \mu - y^\star, v_i \rangle$, where $z_i = \sum_{j=1}^{n} \eta_{ij} \boldsymbol{w}_j$ is an $n \times 1$ vector. Then, according to the chain rule, we have $\mathrm{D}L_1 |_{\boldsymbol{x}_0} = \left(\mathrm{D}\boldsymbol{k}_x |_{\boldsymbol{x}_0}\right)^T \frac{\mathrm{d}L_1}{\mathrm{d}\boldsymbol{k}_x}$. Here $\frac{\mathrm{d}L_1}{\mathrm{d}\boldsymbol{k}_x} = 2 \sum_{i=1}^{m} \left[z_i z_i^T \boldsymbol{k}_x(\boldsymbol{x}_0) - z_i \langle \mu - y^\star, v_i \rangle \right]$ is an $n \times 1$ vector. Similar to L_1, the Fréchet derivative of L_2 at $\boldsymbol{x}_0$ is $g_2(\boldsymbol{x}_0)$.

Algorithm A2 Functional robust parameter design via FRGD

Input: The FFGP estimator $\hat{f}_m$, the corresponding SVOC $\hat{s}_m^2$, and step size γ, maximum number of iterations I, number of starting points N_s.

Initialize: $\left\{\boldsymbol{x}_i^{\{0\}} \in \mathcal{X}^p,\ i = 1, \cdots, N_s\right\}$ randomly.

1: **for** each $i \in \{1 : N_s\}$ **do**
2: Update $L_i = L_m\left(\hat{\boldsymbol{x}}_i^{\{0\}}\right)$ in (22).
3: **while** L_i has not converged and $1 \le l \le I$ **do**
4: Update $\hat{\boldsymbol{x}}_i^{\{l\}} = \hat{\boldsymbol{x}}_i^{\{l-1\}} - \gamma \mathrm{D}L_m\,|_{\boldsymbol{x}_0}\left(\hat{\boldsymbol{x}}_i^{\{l-1\}}\right)$.
5: **end while**
6: Compute $L_i = L_m\left(\hat{\boldsymbol{x}}_i^{\{l\}}\right)$.
7: **end for**
8: Compute $i^\star = \underset{i\in\{1:N_s\}}{\arg\min}\{L_i\}$.
9: **Output:** the optimal parameter $\hat{\boldsymbol{x}}_m^\star = \hat{\boldsymbol{x}}_{i^\star}^{\{l\}}$ and the corresponding loss value $L_m(\hat{\boldsymbol{x}}_m^\star)$.

Then, we prove Proposition 1. According to the chain rule, we have $\mathrm{D}k_{x_i}\,|_{\boldsymbol{x}_0} = \frac{\mathrm{d}k_x}{\mathrm{d}r_{x_i}}\mathrm{D}r_{x_i}\,|_{\boldsymbol{x}_0}$. For the Gaussian kernel in (7), $\frac{\mathrm{d}k_x}{\mathrm{d}r_{x_i}} = -2r_{x_i}\exp\left\{-r_{x_i}^2\right\}$. For the Matérn kernel in (8), $\frac{\mathrm{d}k_x}{\mathrm{d}r_{x_i}} = \mathrm{D}k_x(\boldsymbol{x}_0, \boldsymbol{x}_i)\,|_{\boldsymbol{x}_0} = \frac{\nu^{\nu/2}}{\Gamma(\nu)2^{\nu/2-1}} m_0(r_{x_i})$, here $m_0(r_{x_i}) = r_{x_i}^{\nu-1}\left[\nu K_\nu(\sqrt{2\nu}r_{x_i}) + r_{x_i}\mathrm{D}K_\nu(\sqrt{2\nu}r_{x_i})\right]$. DK_ν means the derivative operator of the modified Bessel function K_ν at r_{x_i}. Now we compute the Fréchet derivative $\mathrm{D}r_{x_i}\,|_{\boldsymbol{x}_0}$, here $r_{x_i} : \mathcal{X}^p \to \mathbb{R}$ is a functional. Since the dual space of L^2 is equivalent to itself, $\mathrm{D}r_{x_i}\,|_{\boldsymbol{x}_0} \in \mathcal{X}^p$. We first consider the case $p = 1$. Define $\mathrm{D}r_{x_i}\,|_{x_0}(h) = \left\langle h, \frac{x_0 - x_i}{\tau_x \|x_0 - x_i\|}\right\rangle$. Then we prove $\lim\limits_{\|h\|\to 0} \frac{|r_{x_i}(x_0+h) - r_{x_i}(x_0+h) - Dr_{x_i}|_{x_0}(h)|}{\|h\|} = 0$, where $r_{x_i}(x_0) = \frac{\|x_i - x_0\|}{\tau_x}$.

$$\begin{aligned}
\frac{|r_{x_i}(x_0+h) - r_{x_i}(x_0+h) - Dr_{x_i}\,|_{x_0}(h)|}{\|h\|} &= \frac{\left|\|x_0 - x_i\|\|h + x_0 - x_i\| - \|x_0 - x_i\|^2 - \langle h, x_0 - x_i\rangle\right|}{\tau_x\|h\|} \\
&= \frac{|\|x_0 - x_i\|\|h + x_0 - x_i\| - \langle h + x_0 - x_i, x_0 - x_i\rangle|}{\tau_x\|h\|\|x_0 - x_i\|} \\
&= \frac{\left|\|x_0 - x_i\|^2\|h + x_0 - x_i\|^2 - \langle h + x_0 - x_i, x_0 - x_i\rangle^2\right|}{\tau_x\|h\| \cdot \|x_0 - x_i\|\left(\|x_0 - x_i\|\,\|h + x_0 - x_i\|\,\langle h + x_0 - x_i, x_0 - x_i\rangle\right)} \\
&= \frac{\langle x_0 - x_i, x_0 - x_i\rangle\langle h, h\rangle - \langle h, x_0 - x_i\rangle^2}{\tau_x\|h\| \cdot \|x_0 - x_i\|\left(\|x_0 - x_i\|\,\|h + x_0 - x_i\|\,\langle h + x_0 - x_i, x_0 - x_i\rangle\right)}.
\end{aligned}$$

Then we have

$$\begin{aligned}
\lim_{\|h\|\to 0} \frac{|r_{x_i}(x_0+h) - r_{x_i}(x_0+h) - Dr_{x_i}\,|_{x_0}(h)|}{\|h\|} &= \lim_{\|h\|\to 0} \frac{\langle x_0 - x_i, x_0 - x_i\rangle\langle h, h\rangle - \langle h, x_0 - x_i\rangle^2}{2\tau_x\|h\| \cdot \|x_0 - x_i\|^3} \\
&= \lim_{\|h\|\to 0} \frac{\|x_0 - x_i\|^2\|h\| - \langle h, x_0 - x_i\rangle\left\langle \frac{h}{\|h\|}, x_0 - x_i\right\rangle}{2\tau_x\|x_0 - x_i\|^3}.
\end{aligned}$$

Assume $\|x_0 - x_i\| > 0$ is bounded, the denominator of the above equation is bounded. The first term of the numerator of the above equation converges to 0 when $\|h\| \to 0$. According to the Cauchy-Schwarz inequality, $\left\langle \frac{h}{\|h\|}, x_0 - x_i\right\rangle \le \|x_0 - x_i\|$ is bounded. And $\langle h, x_0 - x_i\rangle$ converges to 0 when $\|h\| \to 0$. Then we have $\lim\limits_{\|h\|\to 0} \frac{|r_{x_i}(x_0+h) - r_{x_i}(x_0+h) - Dr_{x_i}|_{x_0}(h)|}{\|h\|} = 0$. Since $Dr_{x_i}\,|_{x_0}$ at x_0 is a function in $\mathcal{X}$, there follows $Dr_{x_i}\,|_{x_0}(s) = \frac{x_0(s) - x_i(s)}{\tau_x\|x_i - x_0\|}$, where $s \in \Omega_x$. By extending it to $p > 1$, we have $Dr_{x_i}\,|_{\boldsymbol{x}_0}(s) = \frac{\boldsymbol{x}_0(s) - \boldsymbol{x}_i(s)}{\tau_x\|\boldsymbol{x}_i - \boldsymbol{x}_0\|}$. By taking it into $\mathrm{D}k_{x_i}\,|_{\boldsymbol{x}_0}$ proves that (25) and (26).

Appendix F: Proof of Theorem 1

Proof. According to Definition 4, for $\forall \boldsymbol{x}_0 \in \mathcal{X}^p$ we have

$$\begin{aligned}
\mathrm{MSPE}(\widetilde{f}) &= \mathrm{E}\left[\left\|\widetilde{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right\|^2\right] = \mathrm{E}\left[\left\|\widetilde{f}(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0) + \hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right\|^2\right] \\
&= \mathrm{E}\left[\left\|\hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right\|^2\right] + \mathrm{E}\left[\left\|\widetilde{f}(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0)\right\|^2\right] + 2\mathrm{E}\left[\left\|\left(\widetilde{f}(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0)\right)\left(\hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right)\right\|\right] \\
&\geq \mathrm{MSPE}(\hat{f}) + 2\mathrm{E}\left[\left\|\left(\widetilde{f}(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0)\right)\left(\hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\right)\right\|\right] \geq \mathrm{MSPE}(\hat{f}).
\end{aligned}$$

Appendix G: Proof of Corollary 1

Without loss of generality, we consider $\boldsymbol{g} = \boldsymbol{0}$ and $\boldsymbol{\tau}_x = \boldsymbol{1}_p$ in the following proofs.

Proof. Since $T_{\boldsymbol{Y}}$ is a trace class, then $\sum_{i=1}^{\infty} \beta_i \leq \infty$. According to (15), the trace of $\mathrm{Var}\left[f(\boldsymbol{x}) \mid \boldsymbol{Y}\right]$ is

$$\begin{aligned}
\sum_{i=1}^{\infty} \delta_i &= \sigma^2 \sum_{i=1}^{\infty} \beta_i \left[k_x(\boldsymbol{x}_0, \boldsymbol{x}_0) - \sum_{j=1}^{n} \frac{\beta_i}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x(\boldsymbol{x}_0)^T \boldsymbol{w}_j \boldsymbol{w}_j^T \boldsymbol{k}_x(\boldsymbol{x}_0)\right] \\
&= \sigma^2 \sum_{i=1}^{\infty} \beta_i \left[k_x(\boldsymbol{x}_0, \boldsymbol{x}_0) - \sum_{j=1}^{n} \boldsymbol{k}_x(\boldsymbol{x}_0)^T \frac{\boldsymbol{w}_j \boldsymbol{w}_j^T}{\alpha_j} \boldsymbol{k}_x(\boldsymbol{x}_0) + \sum_{j=1}^{n} \frac{\lambda/\sigma^2}{\alpha_j \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x(\boldsymbol{x}_0)^T \frac{\boldsymbol{w}_j \boldsymbol{w}_j^T}{\alpha_j} \boldsymbol{k}_x(\boldsymbol{x}_0)\right].
\end{aligned}$$

Denote $\boldsymbol{\Lambda} = \mathrm{diag}(\alpha_1, \cdots, \alpha_n)$, here $\mathrm{diag}(\boldsymbol{a})$ is the diagonal matrix consisting of elements of $\boldsymbol{a}$. Then we have

$$\begin{aligned}
\sum_{i=1}^{\infty} \delta_i &\leq \sigma^2 \sum_{i=1}^{\infty} \beta_i \left[k_x(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}_x(x)^T \mathbf{W} \boldsymbol{\Lambda}^{-1} \mathbf{W}^T \boldsymbol{k}_x(\boldsymbol{x}) + \frac{\lambda/\sigma^2}{\alpha_{min} \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x(\boldsymbol{x})^T \mathbf{W} \boldsymbol{\Lambda}^{-1} \mathbf{W}^T \boldsymbol{k}_x(\boldsymbol{x})\right] \\
&\leq \sigma^2 \sum_{i=1}^{\infty} \beta_i \left[k^{\perp}(\boldsymbol{x}_0, \boldsymbol{x}_0) + \frac{\lambda/\sigma^2}{\alpha_{min} \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x^T(\boldsymbol{x}_0) \mathbf{K}_x(\boldsymbol{x}_0)^{-1} \boldsymbol{k}_x(\boldsymbol{x}_0)\right],
\end{aligned}$$

here $\alpha_{min} > 0$ is the minimum eigenvalue of $\mathbf{K}_x$. For k_x, it can define the unique functional input GP $g^{\star}(\cdot)$ (Sung et al. 2024). Given inputs data $\mathbf{X}$, the conditional covariance at new input $\boldsymbol{x}_0$ is $k^{\perp}(\boldsymbol{x}_0) = k_x(\boldsymbol{x}_0, \boldsymbol{x}_0) - \boldsymbol{k}_x(\boldsymbol{x}_0)^T \mathbf{K}_x^{-1} \boldsymbol{k}_x(\boldsymbol{x}_0)$, satisfying $0 \leq k^{\perp}(\boldsymbol{x}_0) \leq 1$. Then we have $\sum_{i=1}^{\infty} \delta_i \leq 3\sigma^2 \sum_{i=1}^{\infty} \beta_i \leq 3\sigma^2 C_0$, and each $\delta_i > 0$.

Appendix H: Proof of Theorem 2

Proof. From the proof of Corollary 1, we have

$$\begin{aligned}
\mathrm{MSPE}\big(\hat{f}\big) = s^2(\boldsymbol{x}_0, \boldsymbol{x}_0) &\leq \sigma^2 \sum_{i=1}^{\infty} \beta_i \left[\frac{\lambda/\sigma^2}{\alpha_{min} \beta_i + \lambda/\sigma^2} \boldsymbol{k}_x(\boldsymbol{x}_0)^T \mathbf{K}_x(\boldsymbol{x}_0)^{-1} \boldsymbol{k}_x(\boldsymbol{x}_0) + k^{\perp}(\boldsymbol{x}_0, \boldsymbol{x}_0)\right] \\
&\leq \sigma^2 \sum_{i=1}^{\infty} \beta_i \left[2 \frac{\lambda/\sigma^2}{\alpha_{min} \beta_i + \lambda/\sigma^2} + k^{\perp}(\boldsymbol{x}_0, \boldsymbol{x}_0)\right]. \qquad \text{(A1)}
\end{aligned}$$

First, we prove the first term of (4). Under Assumption 4, we have $\zeta = \frac{\lambda}{\sigma^2} < bn^{-c}$. Then we have $\sum_{i=1}^{\infty} \frac{\zeta \beta_i}{\alpha_{min} \beta_i + \zeta} = \sum_{i=1}^{\infty} \frac{1}{\alpha_{min}/\zeta + 1/\beta_i} \leq \sum_{i=1}^{\infty} \frac{1}{(\alpha_{min}/b) n^c + 1/\beta_i}$. According Assumption 1, when k_y is the exponential kernel, we have $\beta_i = \frac{2\tau_y}{1 + \tau_y^2 \mu_i^2}$, and $h(\mu) = 2\cos\mu/\sin\mu - \tau_y \mu + 1/(\tau_y \mu) = 0$ for $i \geq 1$ and $\mu > 0$. Since the cos/sin function is periodic with

period π and $h(\mu) \in (+\infty, -\infty)$ is monotonically decreasing in each period, we have $\mu_i \in ((i-1)\pi, i\pi)$ and $\beta_i \asymp i^{-2}$. Therefore, when k_y is the exponential kernel or the Wiener kernel, there exists $\tau_1, \tau_2 > 0$, s.t. $\tau_1 < \beta_i / i^{-2} < \tau_2$. Then,

$$\sum_{i=1}^{\infty} \frac{\zeta\beta_i}{\alpha_{min}\beta_i + \zeta} \leq \sum_{i=1}^{\infty} \frac{1}{(\alpha_{min}/b)n^c + i^2/\tau_1} = \int_1^{\infty} \frac{1}{(\alpha_{min}/b)n^c + i^2/\tau_1} di = \frac{b}{\alpha_{min}n^c} \int_1^{\infty} \frac{1}{1 + bi^2/(\tau_1\alpha_{min}n^c)} di$$
$$= \sqrt{\frac{b\tau_1}{\alpha_{min}n^c}} \int_{\sqrt{\frac{b}{\alpha_{min}\tau_1 n^c}}}^{\infty} \frac{1}{1+u^2} du = \sqrt{\frac{b\tau_1\pi^2}{4\alpha_{min}}} n^{-c/2} - \sqrt{\frac{b\tau_1}{\alpha_{min}}} \frac{\arctan(\sqrt{\frac{b\tau_1}{\alpha_{\min}} \mathrm{n}^{-\mathrm{c}/2}})}{n^{c/2}} \asymp n^{-c/2}.$$

That is, there exists a constant c_1, s.t. $2\sigma^2 \sum_{i=1}^{\infty} \beta_i \frac{\lambda/\sigma^2}{\alpha_{min}\beta_i + \lambda/\sigma^2} \leq c_1 n^{-c/2}$.

Second, we prove the second term of (A1). From Corollary 1, $g^\star$ is functional input GP (FIGP) with mean 0 and covariance function $k_x(\cdot,\cdot)$ (Sung et al. 2024). Then $k^\perp(\boldsymbol{x})$ represents the conditional covariance of $g^\star(\boldsymbol{x})$ given data $\boldsymbol{g}^\star = g^\star(\boldsymbol{x}_1), \cdots, g^\star(\boldsymbol{x}_n)$. And we have $k^\perp(\boldsymbol{x}) = \mathrm{E}\left[g^\star(\boldsymbol{x}) - \hat{g}^\star(\boldsymbol{x})\right]^2$. Here $\hat{g}^\star(\boldsymbol{x}) = \boldsymbol{k}_x(\boldsymbol{x})^T \mathbf{K}_x^{-1} \boldsymbol{g}^\star$ is the posterior mean function under FIGP with covariance k_x. Next we prove $\mathrm{E}\left[g^\star(\boldsymbol{x}) - \hat{g}^\star(\boldsymbol{x})\right]^2 \to 0$ as $n \to \infty$.

Sung et al. (2024) proves the result when $\boldsymbol{x}$ is a function, i.e., $p = 1$. Now we refer to the proof process in Sung et al. (2024) and expand the case when $p > 1$. We introduce the following Lemmas, which can be found in Lemma 2 and Definition 1 of Sung et al. (2024), Theorem 11.8, and Theorem 11.9 of Wendland (2004).

LEMMA A1. Let Ψ be a Matérn kernel function, and $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ be its eigenvalues. Then, $\lambda_k \asymp k^{-2\nu/d}$.

LEMMA A2. A set $\Omega \subset \mathbb{R}^d$ is said to satisfy an interior cone condition if there exists an angle $\omega \in (0, \pi/2)$ and a radius $\mathfrak{r} > 0$ such that for every $\boldsymbol{t} \in \Omega$, a unit vector $\xi(\boldsymbol{t})$ exists such that the cone $C(\boldsymbol{t}, \xi(\boldsymbol{t}), \omega, \mathfrak{r}) := \left\{\boldsymbol{t} + \eta\tilde{\boldsymbol{t}} : \tilde{\boldsymbol{t}} \in \mathbb{R}^d, \left\|\tilde{\boldsymbol{t}}\right\| = 1, \tilde{\boldsymbol{t}}^T\xi(\boldsymbol{t}) \geq \cos(\omega), \eta \in [0, \mathfrak{r}]\right\}$ is contained in Ω.

LEMMA A3. Let $l \in \mathbb{N}_0$ and $\pi_l\left(\mathbb{R}^d\right)$ be the set of d-variate polynomials with absolute degree no more than l. Suppose $\Omega \subset \mathbb{R}^d$ is bounded and satisfies an interior cone condition. Define $C_7 = 2, \quad C_8 = \frac{16(1+\sin(\omega))^2 l^2}{3\sin^2(\omega)}, \quad c_3 = \frac{\mathfrak{r}}{C_8}$, where ω and $\mathfrak{r}$ are defined in Lemma A2. Then for all $\mathbf{T}_n = \{\boldsymbol{t}_1, \cdots, \boldsymbol{t}_n\} \subset \Omega$ with $h_{\mathbf{T}_n,\Omega} \leq c_3$ and every $\boldsymbol{t} \in \Omega$ there exist numbers $\tilde{u}_1(\boldsymbol{t}), \cdots, \tilde{u}_n(\boldsymbol{t})$ with (1) $\sum_{j=1}^n \tilde{u}_j(\boldsymbol{t}) q\left(\boldsymbol{t}_j\right) = q(\boldsymbol{t})$ for all $q \in \pi_l\left(\mathbb{R}^d\right)$, (2) $\sum_{j=1}^n \left|\tilde{u}_j(\boldsymbol{t})\right| \leq C_7$, (3) $\tilde{u}_j(\boldsymbol{t}) = 0$, if $\left\|\boldsymbol{t} - \boldsymbol{t}_j\right\| > C_8 h_{\mathbf{T}_n,\Omega}$.

LEMMA A4. Suppose $\Phi = r(\|\cdot\|_2) \in C^k\left(\mathbb{R}^d\right)$ is positive definite. Let Ω be a compact region satisfying an interior cone condition. Then for $h_{\mathbf{T}_n,\Omega} \leq h_0$, $\Phi(\boldsymbol{t},\boldsymbol{t}) - 2\sum_{j=1}^n \tilde{u}_j \Phi\left(\boldsymbol{t}, \boldsymbol{t}_j\right) + \sum_{j=1}^n \sum_{k=1}^n \tilde{u}_j \tilde{u}_k \Phi\left(\boldsymbol{t}_j, \boldsymbol{t}_k\right) \leq (1 + C_7)^2 \max_{0 \leq s \leq 2C_8 h_{\boldsymbol{T}_n,\Omega}} \left|\phi(s) - p\left(s^2\right)\right|$, where $p \in \pi_{\lfloor l/2 \rfloor}(\mathbb{R})$, $h_{\mathbf{T}_n,\Omega} = \sup_{\boldsymbol{t} \in \Omega} \min_{1 \leq j \leq n} \left\|\boldsymbol{t} - \boldsymbol{t}_j\right\|_2$.

We first provide a characterization of the function class $\mathcal{X}^p$. Since Φ is a positive definite function, we can apply Mercer's theorem to Φ and obtain $\Phi(\boldsymbol{s}, \boldsymbol{s}') = \sum_{j=1}^{\infty} \lambda_{\Phi,j} \phi_j(\boldsymbol{s}) \phi_j(\boldsymbol{s}'), \quad \boldsymbol{s}, \boldsymbol{s}' \in \Omega_x$, where $\lambda_{\Phi,1} \geq \lambda_{\Phi,2} \geq \cdots > 0$ are the eigenvalues and $\{\phi_k\}_{k \in \mathbb{N}}$ are the eigenfunctions of $\mathcal{N}_\Phi(\Omega_x)$, and the summation is uniformly and absolutely convergent. Because $\boldsymbol{x} = (x_1, \cdots, x_p)^T \in \mathcal{X}^p \subset [\mathcal{N}_\Phi(\Omega_x)]^p$, by Theorem 10.29 of Wendland (2004), for each element $x_i \in \mathcal{X}$ of $\boldsymbol{x}$, the summation is $\|\boldsymbol{x}\|^2_{[\mathcal{N}_\Phi(\Omega)]^p} = \sum_{i=1}^p \|x_i\|^2_{\mathcal{N}_\Phi(\Omega)} = \sum_{i=1}^p \sum_{j=1}^{\infty} \lambda^{-1}_{\Phi,j} \left\langle x_i, \phi_j \right\rangle^2 \leq p$.

Define a map $h : \mathcal{X}^p \to W$ between the function class $\mathcal{X}^p$ and the set $W = \left\{\mathbf{a} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n, \cdots)^T : \sum_{j=1}^{\infty} \lambda^{-1}_{\Phi,j} a^2_{ij} \leq 1, i = 1, \cdots, p\right\} \subset l_2\left(\mathbb{R}^{\infty \times p}\right)$ as $h(\boldsymbol{x}) = \left(h_1(x_1), \cdots, h_p(x_p)\right)$. where $h_i(x_i) = \left(\langle x_i, \phi_1 \rangle, \cdots, \langle x_i, \phi_n \rangle, \cdots\right)^T$.

It can be verified that $\|x_i\| = \|h_i(x_i)\|_2$ and $\|\boldsymbol{x}\| = \|h(\boldsymbol{x})\|_2$, here $\|\cdot\|_2$ is the Euclidean norm. Therefore, we can define a new positive definite function K_1 on W which satisfies $\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^p$, $K_1(\mathbf{a}, \mathbf{a}') = \psi(\|\mathbf{a} - \mathbf{a}'\|_2) = \psi(\|\boldsymbol{x} - \boldsymbol{x}'\|) = k_x(\boldsymbol{x}, \boldsymbol{x}')$. Here $\mathbf{a} = h(\boldsymbol{x})$ and $\mathbf{a}' = h(\boldsymbol{x}')$. Define $\mathbf{a}^{(j)} = h(\boldsymbol{x}_j)$. For any $\mathbf{u}_n = (u_1, \cdots, u_n)^T \in \mathbb{R}^n$, it follows that

$$\begin{aligned}\mathbb{E}\left(g^\star(\boldsymbol{x}) - \sum_{j=1}^n u_j g^\star(\boldsymbol{x}_j)\right)^2 =& k_x(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_{j=1}^n u_j k_x(\boldsymbol{x}, \boldsymbol{x}_j) + \sum_{j=1}^n \sum_{k=1}^n u_j u_k k_x(\boldsymbol{x}_j, \boldsymbol{x}_k) \\ =& \psi(0) - 2\sum_{j=1}^n u_j \psi\left(\left\|\mathbf{a} - \mathbf{a}^{(j)}\right\|_2\right) + \sum_{j=1}^n \sum_{k=1}^n u_j u_k \psi\left(\left\|\mathbf{a}^{(j)} - \mathbf{a}^{(k)}\right\|_2\right).\end{aligned}$$

Let $\mathbf{b}_j = \left(\boldsymbol{a}_1^{(j)}, \cdots, \boldsymbol{a}_{n_0}^{(j)}\right)^T$, $\mathbf{b} = \left(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{n_0}\right)^T$, $\mathbf{a}_c^{(j)} = \left(\boldsymbol{a}_{n_0+1}^{(j)}, \boldsymbol{a}_{n_0+2}^{(j)}, \cdots\right)$, and $\mathbf{a}_c = \left(\boldsymbol{a}_{n_0+1}, \boldsymbol{a}_{n_0+2}, \cdots\right)$, where n_0 will be determined later. Then $\mathbf{a}^{(j)} = \left(\mathbf{b}_j^T, \left(\mathbf{a}_c^{(j)}\right)^T\right)^T$ and $\mathbf{a} = \left(\mathbf{b}^T, \mathbf{a}_c^T\right)^T$. Denote $\boldsymbol{d}$ and $\boldsymbol{d}_j$ are the $n_0 p$ vector of the matrix vec operator on the $\mathbf{b}$ and $\mathbf{b}_j$. Applying Lemma A3, we obtain that for some $\tilde{u}_j$, $\sum_{j=1}^n \tilde{u}_j q(\boldsymbol{d}_j) = q(\boldsymbol{d})$, for all $q \in \pi_l(\mathbb{R}^{n_0 p})$, $\sum_{j=1}^n |\tilde{u}_j(\boldsymbol{d})| \le C_6$, and $\tilde{u}_j(\boldsymbol{d}) = 0$, if $\|\boldsymbol{d} - \boldsymbol{d}_j\|_2 > C_7 h_{\mathbf{B}_n, \Omega_x}$, when $h_{\mathbf{B}_n, \Omega_x} \le c_3$, where $\mathbf{B}_n = \{\boldsymbol{d}_1, \cdots, \boldsymbol{d}_n\}$. Note that C_8 and c_3 depend on the interior cone condition and l. In particular, they change as the dimension of $\boldsymbol{d}$ and the degree l of polynomials p in change.

Since $\mathbf{a}, \mathbf{a}^{(j)} \in W$, it follows that $\mathbf{a} \le \sqrt{\lambda_{\Phi,j}}$ and $\mathbf{a}^{(j)} \le \sqrt{\lambda_{\Phi,j}}$. Define a set $\mathcal{X}_0 = \prod_{k=1}^{n_0 p}\left[0, \sqrt{\lambda_{\Phi,k}}\right]$. It can be verified that $\boldsymbol{d}, \boldsymbol{d}_j \in \mathcal{X}_0$. Set $\omega = \pi/6$ and $\mathfrak{r} = \sqrt{\lambda_{\Phi,n_0 p}}/2$. It can be verified that the interior cone condition is satisfied. Then $C_8 = 48l^2$ and $c_3 = \frac{\sqrt{\lambda_{\Phi,n_0 p}}}{96l^2}$.

And then we have

$$\begin{aligned}&\mathbb{E}(g^\star(\boldsymbol{x}) - \hat{g^\star}(\boldsymbol{x}))^2 \le \psi(0) - 2\sum_{j=1}^n \tilde{u}_j \psi\left(\left\|\mathbf{a} - \mathbf{a}^{(j)}\right\|_2\right) + \sum_{j=1}^n \sum_{l=1}^n \tilde{u}_j \tilde{u}_l \psi\left(\left\|\mathbf{a}^{(j)} - \mathbf{a}^{(l)}\right\|_2\right) \\ =&\left(\psi(0) - 2\sum_{j=1}^n \tilde{u}_j \psi\left(\|\mathbf{b} - \mathbf{b}_j\|_2\right) + \sum_{j=1}^n \sum_{k=1}^n \tilde{u}_j \tilde{u}_k \psi\left(\|\mathbf{b}_j - \mathbf{b}_k\|_2\right)\right) \\ &+ \left[-2\sum_{j=1}^n \tilde{u}_j\left(\psi\left(\left\|\mathbf{a} - \mathbf{a}^{(j)}\right\|_2\right) - \psi\left(\|\mathbf{b} - \mathbf{b}_j\|_2\right)\right) + \sum_{j=1}^n \sum_{k=1}^n \tilde{u}_j \tilde{u}_k\left(\psi\left(\left\|\mathbf{a}^{(j)} - \mathbf{a}^{(l)}\right\|_2\right) - \psi\left(\|\mathbf{b}_j - \mathbf{b}_k\|_2\right)\right)\right] \\ =& I_1 + I_2.\end{aligned}$$

The first term can be bounded by Lemma A4, which gives $I_1 \le 9\max_{0 \le s \le 2C_8 h_{\mathbf{B}_n,\Omega}} \left|\psi(s) - q\left(s^2\right)\right| = 9\max_{0 \le s \le \sqrt{\lambda_{\Phi,n_0 p}}} \left|\psi(s) - q\left(s^2\right)\right|$, for some $q \in \pi_{\lfloor l/2 \rfloor}(\mathbb{R})$, provided $h_{\mathbf{B}_n,\Omega_x} \le c_3$. Since $\left\|\boldsymbol{d}_j - \boldsymbol{d}\right\|_2 \le \left\|\mathbf{b}_j - \mathbf{b}\right\|_2 \le \left\|\mathbf{a} - \mathbf{a}^{(j)}\right\|_2 = \left\|\boldsymbol{x} - \boldsymbol{x}_j\right\|$, we have $h_{\mathbf{B}_n,\Omega_x} \le h_{\mathbf{X},\mathcal{X}^p}$ so $h_{\mathbf{B}_n,\Omega} \le c_3$ holds.

Next, we consider bounding I_1 with a Matérn kernel function ψ. Lemma A1 implies that $\lambda_{\Phi,j} \asymp j^{-2\nu_1/d}$. By the expansion of modified Bessel function (Abramowitz and Stegun 1948), ψ can be written as $\psi(r) = \sum_{k=0}^{\lfloor \nu \rfloor} e_k r^{2k} + e_\psi(r)$, where $e_\psi(r) = e r^{2\nu} \log r + O\left(r^{2\nu}\right)$ for $\nu = 1, 2, \cdots$, or $e_\psi(r) = e r^{2\nu} + O\left(r^{2(\lfloor \nu \rfloor + 1)}\right)$ for otherwise.

Therefore, we can take $q\left(s^2\right) = -\sum_{k=0}^{\lfloor \nu \rfloor} e_k s^{2k}$. For $\max_{0 \le s \le \sqrt{\lambda_{\Phi,n_0 p}}} \left|\psi(s) - q\left(s^2\right)\right|$, we can obtain that it is less than $C_9 (n_0 p)^{-2\nu\nu_1/d} \log(n_0 p)$ when $\nu = 2, \cdots$, or $C_{10}(n_0 p)^{-2\nu\nu_1/d}$ for otherwise.

Then we have

$$I_1 \le \begin{cases} C_{11} n_0^{-2\nu\nu_1/d} \log n_0, & \nu = 2, \cdots \\ C_{12} n_0^{-2\nu\nu_1/d}, & \text{otherwise.} \end{cases}$$

It remains to bound I_2. For a Matérn kernel function ψ, it can be verified that for all $s_1, s_2 \in [0, s]$, $|\psi(s_1) - \psi(s_2)| \le C_{13} |s_1 - s_2|^{2\nu_0}$, where $\nu_0 = \min(\nu, 1)$. From S7.9 of Sung et al. (2024), we have $|I_2| \le C_{13}\left(4C_7 + 2C_7^2\right)\left(p\lambda_{\Phi,n_0+1}\right)^{\nu_0} \le$

$C_{13}\left(4C_7+2C_7^2\right)p^{\nu_0}n_0^{-2\nu_0\nu_1/d}$. Because $n_0^{-2\nu\nu_1/d}\log n_0 \le n_0^{-2\nu_1\nu_0/d}\log n_0$ and $\log n_0 > 1$, then we have $\mathbb{E}(g^\star(\boldsymbol{x}) - \hat{g^\star}(\boldsymbol{x}))^2 \le C_{14}n_0^{-2\nu_1\nu_0/d}\log n_0$.

Since it is known that a unit ball of $\mathcal{N}_\Phi(\Omega_x)$ has a covering number $N\left(\delta,\mathcal{X},\|\cdot\|_{L_\infty}\right) \le C_{15}\exp\left(C_{16}\delta^{-d/(\nu_1+d/2)}\right)$. Consider p-dimensional independent $\mathcal{N}_\Phi(\Omega_x)$, the covering number is $C_{15}\exp\left(C_{16}(\delta/p)^{-pd/(\nu_1+d/2)}\right)$, here we set $\delta = C_2n_0^{-2\nu_1/d}$. Therefore, according to Assumption 4, we have $h_{\mathbf{X},\mathcal{X}^p} \le C_2n_0^{-2\nu_1/d}$ and $n_0 = \lfloor C_3(\log n)^{(\nu_1+d/2)/2\nu_1 p}\rfloor > e$. That implies $n \ge \exp\left(((e+1)/C_3)^{2\nu_1 p/(\nu_1+d/2)}\right) =: N_0$. As $n\to\infty$, we have $\mathrm{E}\left(g^\star(\boldsymbol{x}_0)-\hat{g}^\star(\boldsymbol{x}_0)\right)^2 \le C_4(\log n)^{-\frac{(\nu_1+d/2)\nu_0}{pd}}\mathrm{loglog}n \to 0$. Therefore, as $n\to\infty$, it follows that

$$\mathrm{MSPE}\big(\hat{f}(\boldsymbol{x}_0)\big) \le \sigma^2\sum_{i=1}^{\infty}\beta_i\left[2\frac{\lambda/\sigma^2}{\alpha_{min}\beta_i+\lambda/\sigma^2}+k^\perp(\boldsymbol{x}_0,\boldsymbol{x}_0)\right] \le c_1n^{-c/2}+C_4(\log n)^{-\frac{(\nu_1+d/2)\nu_0}{pd}}\mathrm{loglog}n \to 0. \tag{A1}$$

The convergence rate is $O\left(n^{-c/2}+(\log n)^{-\frac{(\nu_1+d/2)\nu_0}{pd}}\mathrm{loglog}n\right)$.

Appendix I: Prove of Theorem 3

Proof. From (17), we can obtain that $\|\hat{f}_m(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0)\| = \left\|\sum_{i=m+1}^{\infty}\sum_{j=1}^{n}\eta_{ij}\boldsymbol{k}_x(\boldsymbol{x}_0)^T\boldsymbol{w}_j v_i\right\| \le \sum_{i=m+1}^{\infty}\left|\sum_{j=1}^{n}\eta_{ij}\boldsymbol{k}_x(\boldsymbol{x}_0)^T\boldsymbol{w}_j\right|$, here $\eta_{ij} = \frac{\beta_i}{\alpha_j\beta_i+\lambda/\sigma^2}\left\langle \boldsymbol{Y}-\mathbf{1}_n\mu, \boldsymbol{w}_j v_i\right\rangle$. According to Appendix B, when k_y is the exponential kernel, we have $\beta_i = \frac{2\tau_y}{1+\tau_y^2\mu_i^2}$, and $h(\mu) = 2\cos\mu/\sin\mu - \tau_y\mu + 1/(\tau_y\mu) = 0$ for $i \ge 1$ and $\mu > 0$. Since the cos/sin function is periodic with period π and $h(\mu)\in(+\infty,-\infty)$ is monotonically decreasing in each period, we have $\mu_i \in ((i-1)\pi, i\pi)$ and $\beta_i < \frac{2\tau_y}{1+\tau_y^2(i-1)^2}$. Therefore, when k_y is the Exponential kernel or the Wiener Kernel, there exists a constant C_{17}, such that $\beta_1 < C_{17}$ and $\beta_i < C_{17}(i-1)^{-2}$, $\quad i > 1$. And $|\eta_{ij}| < C_{18}(i-1)^{-2}|\left\langle \boldsymbol{Y}-\mathbf{1}_n\mu, \boldsymbol{w}_j v_i\right\rangle|$.

The error of $\hat{f}_m$ can be written as

$$\|\hat{f}_m(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0)\| \le \sum_{i=m+1}^{\infty}\left|\sum_{j=1}^{n}\eta_{ij}\boldsymbol{k}_x(\boldsymbol{x}_0)^T\boldsymbol{w}_j\right| < C_{18}\sum_{i=m+1}^{\infty}(i-1)^{-2}\left|\boldsymbol{k}_x(\boldsymbol{x}_0)^T\sum_{j=1}^{n}\boldsymbol{w}_j\boldsymbol{w}_j^T\|\boldsymbol{Y}-\mathbf{1}_n\mu\|\right|$$

$$< C_{18}\alpha_1\sum_{i=m+1}^{\infty}(i-1)^{-2}\left|\boldsymbol{k}_x(\boldsymbol{x}_0)^T\sum_{j=1}^{n}\boldsymbol{w}_j\boldsymbol{\Lambda}^{-1}\boldsymbol{w}_j^T\|\boldsymbol{Y}-\mathbf{1}_n\mu\|\right| < C_{18}\alpha_1\sum_{i=m+1}^{\infty}(i-1)^{-2}\left|\boldsymbol{k}_x(\boldsymbol{x}_0)^T\mathbf{W}\boldsymbol{\Lambda}^{-1}\mathbf{W}^T\|\boldsymbol{Y}-\mathbf{1}_n\mu\|\right|.$$

Note that $\boldsymbol{k}_x(\boldsymbol{x}_0)^T\mathbf{W}\boldsymbol{\Lambda}^{-1}\mathbf{W}^T\|(\boldsymbol{Y}-\mathbf{1}_n\mu)\| = \|\boldsymbol{k}_x(\boldsymbol{x}_0)^T\mathbf{W}\boldsymbol{\Lambda}^{-1}\mathbf{W}^T(\boldsymbol{Y}-\mathbf{1}_n)\mu\|$. And $\boldsymbol{k}_x(\boldsymbol{x}_0)^T\mathbf{W}\boldsymbol{\Lambda}^{-1}\mathbf{W}^T(\boldsymbol{Y}-\mathbf{1}_n\mu)\in\mathcal{Y}$, then $\left|\boldsymbol{k}_x(\boldsymbol{x}_0)^T\mathbf{W}\boldsymbol{\Lambda}^{-1}\mathbf{W}^T\|\boldsymbol{Y}-\mathbf{1}_n\mu\|\right| \le C_1$. Therefore, $\|\hat{f}_m(\boldsymbol{x}_0) - \hat{f}(\boldsymbol{x}_0)\| < C_{18}C_1\alpha_1\sum_{i=m+1}^{\infty}(i-1)^{-2} < C_5(m-1)^{-1}$. The bound of the error of truncated mean is $C_5(m-1)^{-1}$ with the convergence rate $O(m^{-1})$. Similarly, the error of $\hat{s}_m^2(\boldsymbol{x}_0,\boldsymbol{x}_0)$ follows that $|\hat{s}_m^2(\boldsymbol{x}_0,\boldsymbol{x}_0) - s^2(\boldsymbol{x}_0,\boldsymbol{x}_0)| = \sum_{i=m+1}^{\infty}\delta_i \le 3\sigma^2\sum_{i=m+1}^{\infty}\beta_i \le c_2(m-1)^{-1}$. The bound of the error of truncated SVOC is $c_2(m-1)^{-1}$ with the convergence rate is $O\left(m^{-1}\right)$.

Appendix J: Prove of Theorem 4

Proof. Denote $L_0(\cdot) = \|f(\cdot) - y^\star\|^2$, we will prove the result through three steps. First, we prove that $L_m(\boldsymbol{x}) \to L_0(\boldsymbol{x}),\ n,m\to\infty, \forall\boldsymbol{x}\in\mathcal{X}^p$. From Theorem 2 and Theorem 3, it is easy to prove the result. Then, we prove that $\min_{\boldsymbol{x}\in\mathcal{X}^p}L_m(\boldsymbol{x}) \to \min_{\boldsymbol{x}\in\mathcal{X}^p}L_0(\boldsymbol{x}),\ n,m\to\infty$. Referring to the proof idea in Han and Ouyang (2021), we replace minimum with infimum for proof. Since $L_m(\boldsymbol{x})$ converges to $L_0(\boldsymbol{x})$, there exists $\epsilon > 0$, $N\in\mathbb{N}_+$ such that $\forall n > N$, $|L_m(\boldsymbol{x}) - L_0(\boldsymbol{x})| < \epsilon$. This gives $L_0(\boldsymbol{x}) - \epsilon < L_m(\boldsymbol{x}) < L_0(\boldsymbol{x}) + \epsilon$. Choose $\delta_0 = 1/n$, pick up $\boldsymbol{x}_0$ such that $L_0(\boldsymbol{x}_0) \le \inf L_0(\boldsymbol{x}) + 1/n$. Then, we have $L_m(\boldsymbol{x}_0) < L_0(\boldsymbol{x}_0) + \epsilon \le \inf L_0(\boldsymbol{x}) + 1/n + \epsilon$. Since $\inf L_m(x) \le L_m(\boldsymbol{x}_0)$, we have $\inf L_m(\boldsymbol{x}) < L_0(\boldsymbol{x}_0) \le \inf L_0(\boldsymbol{x}) + 1/n + \epsilon$. By taking limits, we have $\lim_{n,m\to\infty}\inf L_m(\boldsymbol{x}) \le \inf L_0(\boldsymbol{x})$ if both of them are finite. On the other hand, pick up $\boldsymbol{x}_0$ such that $L_m(\boldsymbol{x}_0) \le \inf L_m(\boldsymbol{x}) + 1/n$. Since $\inf L_0(\boldsymbol{x}_0) - \epsilon \le L_m(\boldsymbol{x}_0)$ and

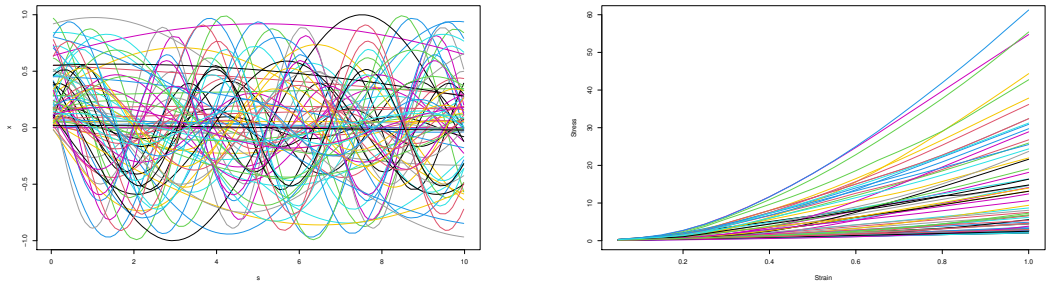
$\inf L_0(\boldsymbol{x}) \le L_0(\boldsymbol{x}_0)$, we have $\inf L_0(\boldsymbol{x}) - \epsilon < L_m(\boldsymbol{x}_0) \le \inf L_m(\boldsymbol{x}) + 1/n$. Therefore, we have $\lim_{m\to\infty} \inf L_m(\boldsymbol{x}) = \inf L_0(\boldsymbol{x})$. Finally, we prove that $\lim_{m\to\infty} L_0(\hat{\boldsymbol{x}}_m^\star) = L_0(\boldsymbol{x}^\star)$. We have obtained that $\lim_{m\to\infty} L_m(\hat{\boldsymbol{x}}_m^\star) = L_0(\hat{\boldsymbol{x}}_m^\star)$ and $\lim_{m\to\infty} L_m(\hat{\boldsymbol{x}}_m^\star) = L_0(\boldsymbol{x}^\star)$. Then there exists $\epsilon_1 > 0$, $\epsilon_2 > 0$, $N_1 \in \mathbb{N}_+$, and $N_2 \in \mathbb{N}_+$ such that $\forall n > N_1$, $\forall m > N_2$, $|L_m(\hat{\boldsymbol{x}}_m^\star) - L_0(\hat{\boldsymbol{x}}_m^\star)| < \epsilon_1$, $|L_m(\hat{\boldsymbol{x}}_m^\star) - L_0(\boldsymbol{x}^\star)| < \epsilon_2$. Then, $|L_0(\hat{\boldsymbol{x}}_m^\star) - L_0(\boldsymbol{x}^\star)| \le |L_0(\hat{\boldsymbol{x}}_m^\star) - L_m(\hat{\boldsymbol{x}}_m^\star)| + |L_m(\hat{\boldsymbol{x}}_m^\star) - L_0(\boldsymbol{x}^\star)| < \epsilon_1 + \epsilon_2$.

In summary, we have $\lim_{m\to\infty} L_0(\hat{\boldsymbol{x}}_m^\star) = L_0(\boldsymbol{x}^\star)$.

Appendix K: Additional information in the Case Study

Background: Three-dimensional (3D) printing prototypes have been widely used to simulate organ shapes in medical applications due to their ability to provide accurate simulations at acceptable prices and time costs. Currently, the state-of-the-art approach involves embedding metamaterial structures to simulate the mechanical properties of biological tissues (Wang et al. 2016). Therefore, optimizing the structure of these metamaterials to ensure that the mechanical properties of the printed tissues closely match those of real biological tissues is an urgent challenge in the medical engineering field, i.e., robust parameter design. However, due to the complex functional nature of metamaterial structures, this adjusted optimization process can take days or even weeks to complete, significantly limiting the medical applicability of tissue simulation prototypes (Chen et al. 2021). To address this challenge, it is crucial to develop a cost-effective and accurate surrogate model that simulates the relationship between metamaterial structure and mechanical properties based on true observations. In this context, we focus on 3D printed human aortic valve tissue in this section. As shown in Figure A1, the various sine wave curves represent the shapes of sinusoidal metamaterials and can be regarded as functional inputs to be designed. Their corresponding stress-strain curves describe the mechanical characteristics of the printed aortic valves and can be regarded as functional output. The goal is to learn how the wave curves influence the mechanical characteristics. Consequently, it can select the wave curve parameters such that the mechanical characteristics can fit a desired curve and be insensitive to variations in the wave curves caused by printing noises.

Figure A1 **Left: Sine wave curves of input samples. Right: Stress-strain curves of output samples.**



Data and setting: We have a total of 77 data samples, with 71 samples allocated to the training set, a collection of 5 samples are utilized as the testing set, and the remaining data function used as the target input and output for RPD. Datasets of training and testing observations can be referred from `https://www.tandfonline.com/doi/suppl/10.1080/00401706.2020.1801255?scroll=top` (Chen et al. 2021). As shown in Figure A1, the various sinusoidal wave curves represent the shapes of metamaterials and can be regarded as functional inputs

to be designed, which can be described and collected by three parameters: the amplitude $A \in [0, 1]$ mm, frequency $\omega \in [0, 0.8]$ mm^{-1} and initial phase $\phi \in [0, 2\pi]$. It is expressed as

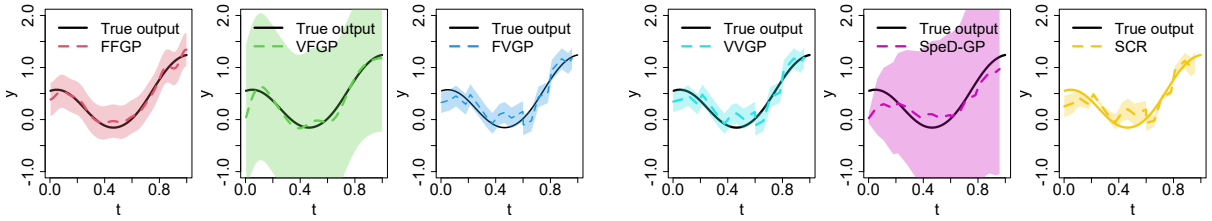
$$x(t) = A\sin(2\pi\omega t + \phi).$$

The corresponding stress-strain curve of a given $x(s)$ is achieved via finite element analysis (FEA). Here we consider the strain from 0% to 15% with each curve collected as a 101-dimensional vector. To simplify the calculation, we scale it to a 20-dimensional vector as the curve exhibits clear exponential properties. In line with Chen et al. (2021), we preprocess the data by dividing each observation by 10^4. The random error can be set to extra small values in this case since the physical data is generated via FEA. To exclude the influence of magnitude on the prediction effect, we employ mean relative error (MRE) as the criteria to evaluate the performance of FFGP and other baselines, i.e., MRE $= \frac{1}{N}\sum_{i=1}^{N} \frac{\|\hat{f}(x_i) - f(x_i)\|}{\|f(x_i)\|}$, where $\{x_i\}_{i=1}^{N}$ represents the testing inputs. In addition, we use the mean standard deviation (MSD) to measure the uncertainty of $\hat{f}(\boldsymbol{x})$. In particular, for FFGP and VFGP, we use $\text{MSD}_1 = \frac{1}{N}\sum_{i=1}^{N} \hat{s}_m^2(\boldsymbol{x}_i, \boldsymbol{x}_i)$, where $\{\boldsymbol{x}_i\}_{i=1}^{N}$ represents the testing inputs. For FVGP, VVGP, and SpeD-GP, we use $\text{MSD}_2 = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} \text{Var}\left[\hat{f}(\boldsymbol{x}_i, t_j) \mid \boldsymbol{Y}\right]$. For SCR, we use the bootstrap to qualify the uncertainty. According to domain knowledge, $(A, \omega, \phi) = (0.1236, 1.0618, 0.3029)$ gives the optimal input $x^\star$, and its corresponding output $y^\star$ can be emulated via FEA.

Appendix L: Additional Predicted Results in Numerical Simulations and Case Study

Based on the kernel selection results in Table 2, Figure A2 gives the predicted curve for one sample of the testing set with the corresponding 95% confidence interval when the sample size is 50. It can be seen that FFGP shows

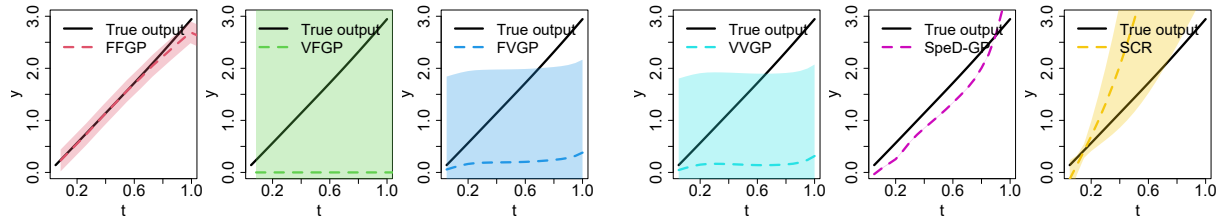
Figure A2 **The prediction using different estimators, together with their corresponding point-wise 95%-credible intervals (in the shaded area) when $n = 50$.**



the best prediction performance. Its predicted output function almost approximates the true output everywhere. The corresponding confidence interval covers the true curve. For other methods, VFGP is poorly predicted at the boundary and has the largest variance. FVGP and VVGP have a good prediction performance. However, the predicted curve is non-smooth because of the discretized output. As to SpeD-GP, its fitting performance is much more poor. This is, on the one hand, because SpeD-GP also discretizes the output function into a vector. On the other hand, it indicates the proposed spectral distance may not be appropriate for describing the input correlations.

Figure A3 shows the emulated stress-strain curve for a test metamaterial structure and corresponding 95% confidence intervals based on the results in Table 7 in the case study. It can be seen that FFGP has the best prediction results with the narrowest confidence intervals. As to other baselines, their predicted outputs are quite far from the true ones.

Figure A3 **The prediction using different estimators, respectively, together with their corresponding point-wise 95%-credible intervals (in the shaded area) in the case study.**



References

Abramowitz M, Stegun IA (1948) *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 (US Government printing office).

Bottou L (2010) Large-scale machine learning with stochastic gradient descent. *Proc. of COMPSTAT* .

Chen J, Mak S, Joseph VR, Zhang C (2021) Function-on-function kriging, with applications to three-dimensional printing of aortic tissues. *Technometrics* 63(3):384–395.

Han M, Ouyang L (2021) Robust functional response-based metamodel optimization considering both location and dispersion effects for aeronautical airfoil designs. *Structural and Multidisciplinary Optimization* 64(3):1545–1565.

Kadri H, Duflos E, Preux P, Canu S, Rakotomamonjy A, Audiffren J (2016) Operator-valued kernels for learning from functional response data. *Journal of Machine Learning Research* 17(20):1–54.

Keener JP (2018) *Principles of applied mathematics: transformation and approximation* (CRC Press).

Owhadi H (2023) Do ideas have shape? idea registration as the continuous limit of artificial neural networks. *Physica D: Nonlinear Phenomena* 444:133592.

Sung CL, Wang W, Cakoni F, Harris I, Hung Y (2024) Functional-input gaussian processes with applications to inverse scattering problems. *Statistica Sinica* .

Wang K, Zhao Y, Chang YH, Qian Z, Zhang C, Wang B, Vannan MA, Wang MJ (2016) Controlling the mechanical behavior of dual-material 3d printed meta-materials for patient-specific tissue-mimicking phantoms. *Materials & Design* 90:704–712.

Wendland H (2004) *Scattered Data Approximation* (Cambridge University Press).

Williams VV (2012) Multiplying matrices faster than coppersmith-winograd. *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, 887–898.