

Appendices

A Proof of Theorem 3.1

Consider a location of interest $x \in \Omega$ and the nearest *design point* $x_i \in \bar{X}$. Let $\sigma_i^2 = \sigma_\tau^2(x_i)$. The uppermost terms in (4) can be expressed as

$$\begin{aligned}
& \Psi_\theta(x, x) - \Psi_\theta(x, \bar{X})[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}\Psi_\theta(\bar{X}, x) \\
&= \Psi_\theta(x, x) \\
&\quad - [(\Psi_\theta(x, \bar{X}) - \Psi_\theta(x_i, \bar{X}) - \sigma_i^2 e_i^T)[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}(\Psi_\theta(\bar{X}, x) - \Psi_\theta(\bar{X}, x_i) - \sigma_i^2 e_i) \\
&\quad + 2(\Psi_\theta(x_i, \bar{X}) + \sigma_i^2 e_i^T)[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}\Psi_\theta(\bar{X}, x) \\
&\quad - (\Psi_\theta(x_i, \bar{X}) + \sigma_i^2 e_i^T)[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}(\Psi_\theta(\bar{X}, x_i) + \sigma_i^2 e_i)] \\
&= \Psi_\theta(x, x) - 2e_i^T \Psi_\theta(\bar{X}, x) + e_i^T (\Psi_\theta(\bar{X}, x_i) + \sigma_i^2 e_i) \\
&\quad - (\Psi_\theta(x, \bar{X}) - \Psi_\theta(x_i, \bar{X}) - \sigma_i^2 e_i^T)[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}(\Psi_\theta(\bar{X}, x) - \Psi_\theta(\bar{X}, x_i) - \sigma_i^2 e_i) \\
&= \Psi_\theta(x, x) + \Psi_\theta(x_i, x_i) + \sigma_i^2 - 2\Psi_\theta(x_i, x) \\
&\quad - (\Psi_\theta(x, \bar{X}) - \Psi_\theta(x_i, \bar{X}) - \sigma_i^2 e_i^T)[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}(\Psi_\theta(\bar{X}, x) - \Psi_\theta(\bar{X}, x_i) - \sigma_i^2 e_i),
\end{aligned} \tag{A.1}$$

where e_i denotes the i^{th} column of an $n \times n$ identity matrix. The last term on the right-hand side of (A.1) can be bounded as

$$\begin{aligned}
& - (\Psi_\theta(x, \bar{X}) - \Psi_\theta(x_i, \bar{X}) - \sigma_i^2 e_i^T)[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1}(\Psi_\theta(\bar{X}, x) - \Psi_\theta(\bar{X}, x_i) - \sigma_i^2 e_i) \\
&\leq - \frac{\|\Psi_\theta(\bar{X}, x) - \Psi_\theta(\bar{X}, x_i) - \sigma_i^2 e_i\|_2^2}{\lambda_{\max}[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]} \\
&\leq - \frac{(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i) - \sigma_i^2)^2}{\lambda_{\max}[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]} \\
&\leq - \frac{(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i))^2 - 2\sigma_i^2(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)) + \sigma_i^4}{\lambda_{\max}[\Psi_\theta(\bar{X}, \bar{X})] + \lambda_{\max}(\bar{\Sigma}_\epsilon)} \\
&\leq - \frac{(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i))^2 - 2\sigma_i^2(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)) + \sigma_i^4}{n \sup_{u, v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)}
\end{aligned}$$

$$= -\frac{(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i))^2}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)} + \frac{2\sigma_i^2(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)) - \sigma_i^4}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)}, \quad (\text{A.2})$$

where the first inequality is true because for any vector d and matrix G , $d^T G^{-1} d \geq \lambda_{\min}(G^{-1}) \|d\|_2^2$ and $\lambda_{\min}(G^{-1}) = 1/\lambda_{\max}(G)$, the second inequality is true because the sum of squares $\|\cdot\|_2^2$ is larger than any one of its elements squared, the third inequality is true because the maximum eigenvalue of a sum is at most the sum of the maximum eigenvalues, and the final inequality is true because Gershgorin's theorem (Varga, 2010) implies

$$\lambda_{\max}(\Psi_\theta(\bar{X}, \bar{X})) \leq \max_j \sum_{i=1}^n \Psi_\theta(x_i, x_j) \leq n \sup_{u,v \in \Omega} \Psi_\theta(u, v). \quad (\text{A.3})$$

Combining (A.1) and (A.2) gives

$$\begin{aligned} & \Psi_\theta(x, x) - \Psi_\theta(x, \bar{X})[\Psi_\theta(\bar{X}, \bar{X}) + \bar{\Sigma}_\epsilon]^{-1} \Psi_\theta(\bar{X}, x) \\ & \leq \Psi_\theta(x, x) + \Psi_\theta(x_i, x_i) - 2\Psi_\theta(x_i, x) \\ & \quad - \frac{(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i))^2}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)} + \sigma_i^2 + \frac{2\sigma_i^2(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)) - \sigma_i^4}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)}. \end{aligned} \quad (\text{A.4})$$

Consider the concave, quadratic function

$$f_1(t) = t + \frac{2t(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)) - t^2}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)},$$

where $t \in [0, \lambda_{\max}(\bar{\Sigma}_\epsilon)]$. $f_1(\cdot)$ has axis of symmetry

$$\begin{aligned} t &= \frac{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon) + 2(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i))}{2} \\ &\geq \frac{(n-2) \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)}{2}, \end{aligned}$$

where the last inequality is true because $\Psi_\theta(x_i, x) \geq 0$ and $\Psi_\theta(x_i, x_i) < \sup_{u,v \in \Omega} \Psi_\theta(u, v)$.

If $(n-2) \sup_{u,v \in \Omega} \Psi_\theta(u, v) > \lambda_{\max}(\bar{\Sigma}_\epsilon)$, then the axis of symmetry lies to the right of the

interval $[0, \lambda_{\max}(\bar{\Sigma}_\epsilon)]$ and $f_1(t)$ is increasing in $[0, \lambda_{\max}(\bar{\Sigma}_\epsilon)]$. This indicates

$$\begin{aligned} f_1(t) &\leq \lambda_{\max}(\bar{\Sigma}_\epsilon) + \frac{2\lambda_{\max}(\bar{\Sigma}_\epsilon)(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)) - \lambda_{\max}(\bar{\Sigma}_\epsilon)^2}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)} \\ &= \frac{\lambda_{\max}(\bar{\Sigma}_\epsilon)(n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + 2(\Psi_\theta(x_i, x) - \Psi_\theta(x_i, x_i)))}{n \sup_{u,v \in \Omega} \Psi_\theta(u, v) + \lambda_{\max}(\bar{\Sigma}_\epsilon)}. \end{aligned} \quad (\text{A.5})$$

Plugging (A.5) into (A.4), gives the result.

B Proof of Theorem 5.1

Here, the derivatives of the log-likelihood and emulator are expressed in terms of the equivalent parameters $\vartheta = (\beta^T, \sigma^2, \rho, \gamma)^T$, where $\gamma_i = \text{Var}(\epsilon(x_i))/\sigma^2$. The vector of derivatives of the emulator with respect to the parameters $\frac{\partial \hat{f}(x)}{\partial \vartheta}$ has block components

$$\begin{aligned} c_1 &= \frac{\partial \hat{f}(x)}{\partial \beta} = h(x) - H(X)^T(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1}\Phi_\rho(X, x), \\ c_2 &= \frac{\partial \hat{f}(x)}{\partial \sigma^2} = 0, \\ (c_3)_j &= \frac{\partial \hat{f}(x)}{\partial \rho_j} = \frac{\partial \Phi_\rho(x, X)}{\partial \rho_j}(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1}(f(X) - H(X)\beta) \\ &\quad - \Phi_\rho(x, X)(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_j}(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1}(f(X) - H(X)\beta), \\ (c_4)_t &= \frac{\partial \hat{f}(x)}{\partial \tau_t} \Phi_\rho(x, X)(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} \text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n} \right\} \\ &\quad \times (\Phi_\rho(X, X) + \Sigma_\gamma)^{-1}(y(X) - H(X)\beta), \end{aligned}$$

where $\Sigma_\gamma = \text{diag}(\gamma_1 I_{k_1}, \dots, \gamma_n I_{k_n})$. The vector of derivatives of the log-likelihood with respect to the parameters $\frac{\partial l}{\partial \vartheta}$ has block components

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \frac{1}{\sigma^2} (X)^T [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} (f(X) - H(X)\beta), \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} (f(X) - H(X)\beta)^T (\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} (f(X) - H(X)\beta), \\ \frac{\partial l}{\partial \rho_j} &= -\frac{1}{2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_j} \right) \\ &\quad + \frac{1}{2\sigma^2} (f(X) - H(X)\beta)^T [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_j} [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} (f(X) - H(X)\beta), \end{aligned}$$

$$\begin{aligned}
\frac{\partial l}{\partial \tau_t} = & -\frac{1}{2\sigma^2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n} \right\} \right) \\
& + \frac{1}{2\sigma^4} (f(X) - H(X)\beta)^T [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \text{diag} \left(\frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n} \right) \\
& \times [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} (f(X) - H(X)\beta).
\end{aligned}$$

So, the information matrix has block components

$$\begin{aligned}
\mathbb{E} - \frac{\partial^2 l}{\partial \beta^2} &= \frac{1}{\sigma^2} (X)^T [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} H(X), \\
\mathbb{E} - \frac{\partial^2 l}{\partial \beta \partial \sigma^2} &= 0, \\
\mathbb{E} - \frac{\partial^2 l}{\partial \beta \partial \tau_t} &= 0, \\
\mathbb{E} - \frac{\partial^2 l}{\partial \beta \partial \rho_j} &= 0, \\
\mathbb{E} - \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} &= \frac{m}{2\sigma^4} \\
\mathbb{E} - \frac{\partial^2 l}{\partial \sigma^2 \partial \rho_j} &= \frac{1}{2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_j} \right), \\
\mathbb{E} - \frac{\partial^2 l}{\partial \rho_{j_1} \partial \rho_{j_2}} &= \frac{1}{2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_{j_2}} [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_{j_1}} \right), \\
\mathbb{E} - \frac{\partial^2 l}{\partial \tau_t \partial \sigma^2} &= \frac{1}{2\sigma^2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n} \right\} \right), \\
\mathbb{E} - \frac{\partial^2 l}{\partial \tau_t \partial \rho_j} &= \frac{1}{2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \frac{\partial \Phi_\rho(X, X)}{\partial \rho_j} [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \right. \\
& \quad \left. \times \text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n} \right\} \right), \\
\mathbb{E} - \frac{\partial^2 l}{\partial \tau_{t_1} \partial \tau_{t_2}} &= \frac{1}{2} \text{trace} \left([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_{t_1}} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_{t_1}} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_{t_1}} I_{k_n} \right\} \right. \\
& \quad \left. \times [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} \text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_{t_2}} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_{t_2}} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_{t_2}} I_{k_n} \right\} \right).
\end{aligned}$$

Building from (14) and the block representations above gives

$$\begin{aligned}
\mathbb{E}\{\hat{f}_{\vartheta_*}(x) - \hat{f}_{\hat{\vartheta}}(x)\}^2 &\approx (c_1^T, c_2^T, c_3^T, c_4^T) \begin{pmatrix} a_{11}^{-1} & 0 \\ 0 & \mathcal{I}^{-1} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \\
&= c_1^T a_{11}^{-1} c_1 + (c_2^T, c_3^T, c_4^T) \mathcal{I}^{-1} \begin{pmatrix} c_2 \\ c_3 \\ c_4 \end{pmatrix} \\
&= \text{Part(I)} + \text{Part(II)},
\end{aligned}$$

where

$$a_{11} = \frac{\partial^2 l}{\partial \beta^2} \quad \text{and} \quad \mathcal{I} = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}$$

with

$$\begin{aligned}
\mathcal{I}_{11} &= -\mathbb{E} \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2}, \quad \mathcal{I}_{12} = -(\mathbb{E} \frac{\partial^2 l}{\partial \sigma^2 \partial \rho^T}, \mathbb{E} \frac{\partial^2 l}{\partial \sigma^2 \partial \tau^T}), \quad \mathcal{I}_{21} = \mathcal{I}_{12}^T, \quad \mathcal{I}_{22} = \begin{pmatrix} D_1 & D_2^T \\ D_2 & D_3 \end{pmatrix}, \\
D_1 &= -\mathbb{E} \frac{\partial^2 l}{\partial \rho \partial \rho^T} \quad D_2 = -\mathbb{E} \frac{\partial^2 l}{\partial \tau \partial \rho^T}, \quad \text{and} \quad D_3 = -\mathbb{E} \frac{\partial^2 l}{\partial \tau \partial \tau^T}.
\end{aligned} \tag{B.6}$$

Applying block matrix inverse results (Harville, 1997) and noticing that $c_2 = 0$ gives

$$\text{Part(II)} = c^T B_1^{-1} c,$$

where $B_1 = \mathcal{I}_{22} - \mathcal{I}_{21} \mathcal{I}_{11}^{-1} \mathcal{I}_{12}$, and $c = (c_3^T, c_4^T)^T$. With the aim of bounding Part(II), the following notation is introduced. Let

$$\begin{aligned}
a_j &= \text{vec} \left(\sigma^2 \frac{\partial \Phi_\rho(X, X)}{\partial \rho_j} \right), \\
b_t &= \text{vec} \left(\text{diag} \left\{ \frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n} \right\} \right),
\end{aligned}$$

$$A_1 = (a_1, \dots, a_{|\rho|}, b_1, \dots, b_{|\tau|}).$$

Then,

$$B_1 = \frac{1}{\sigma^4} A_1^T ((\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} \otimes (\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} - \frac{1}{n} \text{vec}([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1}) \text{vec}([\Phi_\rho(X, X) + \Sigma_\gamma]^{-1})^T) A_1.$$

For simplicity, let

$$W_1 = \Phi_\rho(X, X) + \Sigma_\gamma \quad \text{and} \quad w = \frac{\text{vec}(W_1)}{\|\text{vec}(W_1)\|_2}.$$

The matrix inside the quadratic form has eigenvector w with corresponding eigenvalue 0. Following the approach in Haaland et al. (2018), the minimum eigenvalue of B_1 can be bounded below by

$$\frac{1}{\sigma^4} \lambda_{\min}(A_1^T (I - ww^T) A_1) \times \lambda_2 \left((W_1^{-1} \otimes W_1^{-1} - \frac{1}{m} \text{vec}(W_1^{-1}) \text{vec}(W_1^{-1})^T) \right),$$

where λ_2 denotes the second smallest eigenvalue of its argument. Weyl's theorem (Ipsen and Nadler, 2009) implies that the second smallest eigenvalue can be bounded below by

$$\lambda_{\min}(W_1^{-1} \otimes W_1^{-1}) = \frac{1}{\lambda_{\max}(\Psi_\theta(X, X) + \Sigma_\epsilon)^2} \geq \frac{1}{(m \sup_{u,v \in \Omega} \Phi_\rho(u, v) + \lambda_{\max}(\Sigma_\gamma))^2}.$$

For $\lambda_{\min}(A_1^T (I - ww^T) A_1)$, an approximate lower bound is given. Let $\xi = (\rho, \tau)$. Notice that

$$\begin{aligned} & A_1^T (I - ww^T) A_1 \\ &= \left[\sum_{i,j} \frac{\partial W_1(x_i, x_j)}{\partial \xi} \frac{\partial W_1(x_i, x_j)}{\partial \xi^T} - \frac{1}{\|\text{vec}(W_1)\|_2^2} \left(\sum_{i,j} \frac{\partial W_1(x_i, x_j)}{\partial \xi} W_1(x_i, x_j) \right) \left(\sum_{i,j} \frac{\partial W_1(x_i, x_j)}{\partial \xi^T} W_1(x_i, x_j) \right) \right] \\ &\approx m^2 \left[\int \frac{\partial W_1(x, y)}{\partial \xi} \frac{\partial W_1(x, y)}{\partial \xi^T} dF^2(x, y) \right] \end{aligned}$$

$$\begin{aligned}
& - \frac{1}{\|W_1\|_{L_2(F^2)}^2} \left(\int \frac{\partial W_1(x, y)}{\partial \xi} W_1(x, y) dF^2(x, y) \right) \left(\int \frac{\partial W_1(y)}{\partial \xi^T} W_1(x, y) dF^2(x, y) \right) \Big] \\
& \succeq m^2 s_1,
\end{aligned} \tag{B.7}$$

where $W_1(x, y) = \Phi_\rho(x, y) + \frac{\sigma_\tau^2(x)}{\sigma^2} \mathbb{I}_{\{x=y\}}$ and F^2 denotes the large sample distribution of point pairs. Applying a version of the Cauchy-Schwarz inequality for random vectors (Tripathi, 1999), gives $s_1 \geq 0$ with $s_1 > 0$ unless

$$\frac{\partial W_1(x, y)}{\partial \xi} a = W_1(x, y) b$$

with probability 1 with respect to large sample distribution of point pairs F^2 for some vectors a and b . So, Part(II) has approximate upper bound

$$\frac{\sigma^4 (m \sup_{u, v \in \Omega} \Phi_\rho(u, v) + \lambda_{\max}(\Sigma_\gamma))^2}{m^2 s_1} \|c\|_2^2. \tag{B.8}$$

Also,

$$\begin{aligned}
\text{Part(I)} & \leq \frac{\sigma^2 \|c_1\|_2^2}{\lambda_{\min}(H(X)^T [\Phi_\rho(X, X) + \Sigma_\gamma]^{-1} H(X))} \\
& \leq \frac{\sigma^2 \|c_1\|_2^2 \lambda_{\max}(\Phi_\rho(X, X) + \Sigma_\gamma)}{\lambda_{\min}(H(X)^T H(X))}.
\end{aligned} \tag{B.9}$$

Following development similar to above, $\lambda_{\min}(H(X)^T H(X))$ admits approximation

$$\lambda_{\min}(H(X)^T H(X)) = \lambda_{\min}\left(\sum_{i=1}^n h(x_i) h(x_i)^T\right) \approx m \lambda_{\min}\left(\int h(y) h(y)^T dF(y)\right) = m s_2, \tag{B.10}$$

with respect to the large sample distribution of the input locations, F . Further, $s_2 \geq 0$ with equality if and only if there exists $a \neq 0$ such that $h(y)^T a = 0$ with probability 1. Combining (B.8) and (B.9) gives approximate upper bound for Part(I) + Part(II)

$$\begin{aligned}
& \frac{\sigma^2 \|c_1\|_2^2 \lambda_{\max}(\Phi_\rho(X, X) + \Sigma_\gamma)}{m s_2} + \frac{\sigma^4 \|c\|_2^2 (m \sup_{u, v \in \Omega} \Phi_\rho(u, v) + \lambda_{\max}(\Sigma_\gamma))^2}{m^2 s_1}, \\
& \leq \frac{\sigma^2 \|c_1\|_2^2 (m \sup_{u, v \in \Omega} \Phi_\rho(u, v) + \lambda_{\max}(\Sigma_\gamma))}{m s_2} + \frac{\sigma^4 \|c\|_2^2 (m \sup_{u, v \in \Omega} \Phi_\rho(u, v) + \lambda_{\max}(\Sigma_\gamma))^2}{m^2 s_1},
\end{aligned} \tag{B.11}$$

finishing the proof of Theorem 5.1.

C Proof of Proposition 5.1

Recall that

$$(c_4)_t = \frac{\partial \hat{f}(x)}{\partial \tau_t} = \Phi_\rho(x, X)(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} \text{diag}\left(\frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n}\right) \\ \times (\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} (y(X) - H(X)\beta).$$

In this section we would give an upper bound of $(c_4)_t$. Without loss of generality, we can suppose $\Phi_\rho(x, x) = 1$. Let

$$\Phi_\rho(X, X) + \Sigma_\gamma = \begin{bmatrix} B_1 + \Sigma_{\gamma_1} & R^T \\ R & B_2 + \Sigma_{\gamma_2} \end{bmatrix},$$

where

$$B_1 = \mathbf{1}\mathbf{1}^T,$$

$$\Sigma_{\gamma_1} = \sigma_1^2 I_{k_1},$$

$$R = \Phi_\rho(X_2, x_1)\mathbf{1}^T,$$

$$B_2 = \Phi_\rho(X_2, X_2).$$

Thus, we have

$$(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} = \begin{bmatrix} B_1 + \Sigma_{\gamma_1} & R^T \\ R & B_2 + \Sigma_{\gamma_2} \end{bmatrix}^{-1} \\ = \begin{bmatrix} B_{22}^{-1} & -B_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} \\ -(B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1} & (B_2 + \Sigma_{\gamma_2})^{-1} + (B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} \end{bmatrix},$$

where $B_{22} = B_1 + \Sigma_{\gamma_1} - R^T(B_2 + \Sigma_{\gamma_2})^{-1}R$. Notice that

$$(B_1 + \Sigma_{\gamma_1})^{-1}\mathbf{1} = \frac{1}{k_1 + \sigma_1^2}\mathbf{1},$$

we have

$$\begin{aligned}
B_{22}^{-1} &= (B_1 + \Sigma_{\gamma_1} - R^T(B_2 + \Sigma_{\gamma_2})^{-1}R)^{-1} \\
&= (B_1 + \Sigma_{\gamma_1})^{-1}((B_1 + \Sigma_{\gamma_1})^{-1} - (B_1 + \Sigma_{\gamma_1})^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1}R(B_1 + \Sigma_{\gamma_1})^{-1})^{-1}(B_1 + \Sigma_{\gamma_1})^{-1} \\
&= (B_1 + \Sigma_{\gamma_1})^{-1}\left((B_1 + \Sigma_{\gamma_1})^{-1} - \frac{1}{(k_1 + \sigma_1^2)^2}\mathbf{1}\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)\mathbf{1}^T\right)^{-1}(B_1 + \Sigma_{\gamma_1})^{-1} \\
&= (B_1 + \Sigma_{\gamma_1})^{-1}\left((B_1 + \Sigma_{\gamma_1})^{-1} - \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{(k_1 + \sigma_1^2)^2}\mathbf{1}\mathbf{1}^T\right)^{-1}(B_1 + \Sigma_{\gamma_1})^{-1}.
\end{aligned}$$

By binomial inverse theorem,

$$\begin{aligned}
&\left((B_1 + \Sigma_{\gamma_1})^{-1} - \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{(k_1 + \sigma_1^2)^2}\mathbf{1}\mathbf{1}^T\right)^{-1} \\
&= B_1 + \Sigma_{\gamma_1} + \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{(k_1 + \sigma_1^2)^2} \frac{(B_1 + \Sigma_{\gamma_1})\mathbf{1}\mathbf{1}^T(B_1 + \Sigma_{\gamma_1})}{1 - \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{(k_1 + \sigma_1^2)^2}\mathbf{1}^T(B_1 + \Sigma_{\gamma_1})\mathbf{1}} \\
&= B_1 + \Sigma_{\gamma_1} + \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{(k_1 + \sigma_1^2)^2} \frac{(k_1 + \sigma_1^2)^2\mathbf{1}\mathbf{1}^T}{1 - \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{(k_1 + \sigma_1^2)^2}(k_1 + \sigma_1^2)k_1} \\
&= B_1 + \Sigma_{\gamma_1} + \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1) \frac{\mathbf{1}\mathbf{1}^T}{1 - \frac{\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)}{k_1 + \sigma_1^2}k_1}.
\end{aligned}$$

Let $d = \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)$. Thus,

$$\begin{aligned}
B_{22}^{-1} &= (B_1 + \Sigma_{\gamma_1})^{-1}\left(B_1 + \Sigma_{\gamma_1} + d\frac{\mathbf{1}\mathbf{1}^T}{1 - \frac{d}{k_1 + \sigma_1^2}k_1}\right)(B_1 + \Sigma_{\gamma_1})^{-1} \\
&= (B_1 + \Sigma_{\gamma_1})^{-1} + d\frac{(B_1 + \Sigma_{\gamma_1})^{-1}\mathbf{1}\mathbf{1}^T(B_1 + \Sigma_{\gamma_1})^{-1}}{1 - \frac{d}{k_1 + \sigma_1^2}k_1} \\
&= (B_1 + \Sigma_{\gamma_1})^{-1} + d\frac{\frac{1}{(k_1 + \sigma_1^2)^2}\mathbf{1}\mathbf{1}^T}{1 - \frac{d}{k_1 + \sigma_1^2}k_1}.
\end{aligned}$$

Since

$$\Phi_\rho(x, X) = (\Phi_\rho(x, x_1)\mathbf{1}^T, \Phi_\rho(x, X_2)),$$

we have

$$\begin{aligned}
& \Phi_\rho(x, X)(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} \\
&= (\Phi_\rho(x, x_1)\mathbf{1}^T, \Phi_\rho(x, X_2)) \\
& \quad \times \begin{bmatrix} B_{22}^{-1} & -B_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} \\ -(B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1} & (B_2 + \Sigma_{\gamma_2})^{-1} + (B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} \end{bmatrix}, \\
&= \left(\Phi_\rho(x, x_1)\mathbf{1}^T B_{22}^{-1} - \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1}, \right. \\
& \quad - \Phi_\rho(x, x_1)\mathbf{1}^T B_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} + \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \\
& \quad \left. + (B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} \right).
\end{aligned}$$

Notice that

$$\begin{aligned}
B_{22}^{-1}\mathbf{1} &= (B_1 + \Sigma_{\gamma_1})^{-1}\mathbf{1} + d \frac{\frac{1}{(k_1 + \sigma_1^2)^2}\mathbf{1}\mathbf{1}^T\mathbf{1}}{1 - \frac{d}{k_1 + \sigma_1^2}k_1} \\
&= (B_1 + \Sigma_{\gamma_1})^{-1}\mathbf{1} + d \frac{\frac{1}{(k_1 + \sigma_1^2)^2}\mathbf{1}\mathbf{1}^T\mathbf{1}}{1 - \frac{d}{k_1 + \sigma_1^2}k_1} \\
&= \left(\frac{1}{k_1 + \sigma_1^2} + \frac{\frac{dk_1}{(k_1 + \sigma_1^2)^2}}{1 - \frac{dk_1}{k_1 + \sigma_1^2}} \right) \mathbf{1},
\end{aligned}$$

we have (let $d_1 = (\frac{1}{k_1 + \sigma_1^2} + \frac{\frac{dk_1}{(k_1 + \sigma_1^2)^2}}{1 - \frac{dk_1}{k_1 + \sigma_1^2}}) = \frac{1}{k_1 + \sigma_1^2 - dk_1}$)

$$\begin{aligned}
& \Phi_\rho(x, x_1)\mathbf{1}^T B_{22}^{-1} - \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1} \\
&= \Phi_\rho(x, x_1)d_1\mathbf{1}^T - d_1\Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)\mathbf{1}^T,
\end{aligned}$$

and

$$\begin{aligned}
& - \Phi_\rho(x, x_1)\mathbf{1}^T B_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} + \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} + (B_2 + \Sigma_{\gamma_2})^{-1}RB_{22}^{-1}R^T(B_2 + \Sigma_{\gamma_2})^{-1} \\
&= - \Phi_\rho(x, x_1)k_1d_1\Psi_\theta(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} + \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \\
& \quad + \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1)k_1d_1\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \\
&= - \Phi_\rho(x, x_1)k_1d_1\Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} + \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}
\end{aligned}$$

$$+ dk_1 d_1 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}.$$

With the same procedure, we have

$$\begin{aligned} & (\Phi_\rho(X, X) + \Sigma_\gamma)^{-1}(y(X) - H(X)\beta) \\ = & \left((y(x_1) - H(x_1)\beta)d_1 \mathbf{1} - d_1 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}(y(X_2) - H(X_2)\beta)\mathbf{1}, \right. \\ & - (y(x_1) - H(x_1)\beta)k_1 d_1 (B_2 + \Sigma_{\gamma_2})^{-1} \Phi_\rho(X_2, x_1) + (B_2 + \Sigma_{\gamma_2})^{-1}(y(X_2) - H(X_2)\beta) \\ & \left. + dk_1 d_1 (B_2 + \Sigma_{\gamma_2})^{-1} \Phi_\rho(X_2, x_1) \right). \end{aligned}$$

Thus,

$$\begin{aligned} (c_4)_t &= \frac{\partial \hat{f}(x)}{\partial \tau_t} = \Phi_\rho(x, X)(\Phi_\rho(X, X) + \Sigma_\gamma)^{-1} \text{diag}\left(\frac{\partial \gamma_1}{\partial \tau_t} I_{k_1}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n}\right) \\ & \quad \times (\Phi_\rho(X, X) + \Sigma_\gamma)^{-1}(y(X) - H(X)\beta) \\ &= k_1 \frac{\partial \gamma_1}{\partial \tau_t} (\Phi_\rho(x, x_1)d_1 - d_1 \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \Phi_\rho(X_2, x_1))((y(x_1) - H(x_1)\beta)d_1 \\ & \quad - d_1 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}(y(X_2) - H(X_2)\beta)) \\ & \quad + (-\Phi_\rho(x, x_1)k_1 d_1 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} + \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \\ & \quad + dk_1 d_1 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}) \text{diag}\left(\frac{\partial \gamma_2}{\partial \tau_t} I_{k_2}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n}\right) \\ & \quad \times (-(y(x_1) - H(x_1)\beta)k_1 d_1 (B_2 + \Sigma_{\gamma_2})^{-1} \Phi_\rho(X_2, x_1) + (B_2 + \Sigma_{\gamma_2})^{-1}(y(X_2) - H(X_2)\beta) \\ & \quad + dk_1 d_1 (B_2 + \Sigma_{\gamma_2})^{-1} \Phi_\rho(X_2, x_1)) \end{aligned}$$

Let $d_2 = \frac{1}{1 + \sigma_1^2/k_1 - d}$, we have

$$\begin{aligned} (c_4)_t &= \frac{1}{k_1} \frac{\partial \gamma_1}{\partial \tau_t} (\Phi_\rho(x, x_1)d_2 - d_2 \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \Phi_\rho(X_2, x_1))((y(x_1) - H(x_1)\beta)d_2 \\ & \quad - d_2 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}(y(X_2) - H(X_2)\beta)) \\ & \quad + (-\Phi_\rho(x, x_1)d_2 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} + \Phi_\rho(x, X_2)(B_2 + \Sigma_{\gamma_2})^{-1} \\ & \quad + dd_2 \Phi_\rho(x_1, X_2)(B_2 + \Sigma_{\gamma_2})^{-1}) \text{diag}\left(\frac{\partial \gamma_2}{\partial \tau_t} I_{k_2}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n}\right) \end{aligned}$$

$$\begin{aligned}
& \times (-(y(x_1) - H(x_1)\beta)d_2(B_2 + \Sigma_{\gamma_2})^{-1}\Phi_\rho(X_2, x_1) + (B_2 + \Sigma_{\gamma_2})^{-1}(y(X_2) - H(X_2)\beta) \\
& + dd_2(B_2 + \Sigma_{\epsilon_2})^{-1}\Phi_\rho(X_2, x_1)) \\
& = \Phi_\rho(x, X)(\Phi_\rho(X', X') + \Sigma_\gamma)^{-1} \text{diag}\left(\frac{1}{k_1} \frac{\partial \gamma_1}{\partial \tau_t}, \dots, \frac{\partial \gamma_i}{\partial \tau_t} I_{k_i}, \dots, \frac{\partial \gamma_n}{\partial \tau_t} I_{k_n}\right) \\
& \quad \times (\Phi_\rho(X', X') + \Sigma_\gamma)^{-1}(y(X') - H(X')\beta),
\end{aligned}$$

where $X' = (x_1, X_2)$. Thus, by continuing this procedure, we have

$$|(c_4)_t| \leq \frac{\|\Phi_\rho(x, \bar{X})\|_2 \|\bar{Y} - H(\bar{X})\beta\|_2}{(\lambda_{\min}(\Phi_\rho(\bar{X}, \bar{X}) + \bar{\Sigma}_\gamma))^2} \max_{i: x_i \in \bar{X}} \left| \frac{1}{k_i} \frac{\partial \gamma_i}{\partial \tau_t} \right|.$$

D Proof of Theorem 6.1

The following lemma, which describes the accuracy of solving linear systems (Golub and Van Loan, 1996), will be used to develop a bound on the numeric error.

Lemma D.1. *Suppose $Ax = b$ and $\tilde{A}\tilde{x} = \tilde{b}$ with $\|\tilde{A} - A\|_2 \leq \delta\|A\|_2$, $\|\tilde{b} - b\|_2 \leq \delta\|b\|_2$, and $\kappa(A) = r/\delta < 1/\delta$ for some $\delta > 0$. Then, \tilde{A} is non-singular,*

$$\begin{aligned}
\frac{\|\tilde{x}\|_2}{\|x\|_2} & \leq \frac{1+r}{1-r}, \\
\frac{\|\tilde{x} - x\|_2}{\|x\|_2} & \leq \frac{2\delta}{1-r} \kappa(A),
\end{aligned} \tag{D.12}$$

where $\kappa(A) = \|A\|_2 \|A^{-1}\|_2$.

Further, for conformable A , b , \tilde{A} , and \tilde{b} , we have

$$\begin{aligned}
\|Ab - \tilde{A}\tilde{b}\|_2 & = \|A(b - \tilde{b}) - (\tilde{A} - A)\tilde{b}\|_2 \\
& \leq \|A(b - \tilde{b})\|_2 + \|(\tilde{A} - A)\tilde{b}\|_2 \leq \|A\|_2 \|b - \tilde{b}\|_2 + \|(\tilde{A} - A)\|_2 \|\tilde{b}\|_2.
\end{aligned} \tag{D.13}$$

In order to satisfy the conditions of Lemma D.1, we make a few assumptions *in addition to Assumption 4.1*, in particular, with regard to the accuracy of numeric optimization.

Assumption D.1. *Assume $\kappa(\hat{A}) = r/\delta$ with $r < 1$ and*

$$\|\hat{A} - \tilde{A}\|_2 \leq \delta \|\hat{A}\|_2, \|\Psi_{\hat{\theta}}(\bar{X}, x) - \Psi_{\tilde{\theta}}(\bar{X}, x)\|_2 \leq \delta \|\Psi_{\hat{\theta}}(\bar{X}, x)\|_2.$$

Note that this assumption does not concern the parameter estimates themselves, but instead the accuracy of the solution to the optimization problem. If the optimization problem is solved with sufficient accuracy, then this assumption will be satisfied. However, as we will see in the following, the regression function coefficients β have great potential to cause problems. Briefly, in order to control parameter estimation numeric error, we need that numeric properties are even more tightly controlled, in particular, an even smaller condition number of $\Psi_\theta(\bar{X}, \bar{X}) + \Sigma_\epsilon$, which is stated in the following assumption.

Assumption D.2.

$$\delta\kappa(\hat{A})\kappa(H(\bar{X})^T H(\bar{X})) \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) < 1.$$

Assumption D.2 is a strong assumption, since it requires $\delta\kappa(\hat{A})^2$ to be relatively small, at least smaller than 1. However, since our goal is to make $\kappa(\hat{A})$ small, in practice this condition is not too difficult to achieve, since we can control the condition number of \hat{A} .

The following lemma states that if Assumption D.2 holds, combining Assumption D.1, the conditions of Lemma D.1 holds.

Lemma D.2. *Let*

$$\begin{aligned} r_1 &= \delta\kappa(\hat{A})\kappa(H(\bar{X})^T H(\bar{X})) \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) \\ &\quad + \frac{1}{2} \min\{\delta, 1 - \delta\kappa(\hat{A})\kappa(H(\bar{X})^T H(\bar{X})) \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right)\} \\ \delta_1 &= \frac{r_1}{\kappa(H(\bar{X})^T \hat{A}^{-1} H(\bar{X}))}. \end{aligned} \tag{D.14}$$

Suppose Assumptions 4.1, D.1, and D.2 hold, we have $r_1 < 1$ and

$$\|H(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 < \delta_1 \|H(\bar{X})^T \hat{A}^{-1} H(\bar{X})\|_2. \tag{D.15}$$

Thus, we have all tools to give an upper bound of $|\hat{f}_{\hat{\theta}} - \hat{f}_{\tilde{\theta}}|$. By triangle inequality,

$$\begin{aligned} &|\hat{f}_{\hat{\theta}} - \hat{f}_{\tilde{\theta}}| \\ &= |h(x)^T \hat{\beta} + \Psi_{\hat{\theta}}(x, \bar{X}) \hat{A}^{-1} (\bar{Y} - H(\bar{X}) \hat{\beta}) - (h(x)^T \tilde{\beta} + \Psi_{\tilde{\theta}}(x, \bar{X}) \tilde{A}^{-1} (\bar{Y} - H(\bar{X}) \tilde{\beta}))| \end{aligned}$$

$$\begin{aligned}
&= |h(x)^T(\hat{\beta} - \tilde{\beta}) + \bar{Y}^T(\hat{A}^{-1}\Psi_{\hat{\theta}}(\bar{X}, x) - \tilde{A}^{-1}\Psi_{\tilde{\theta}}(\bar{X}, x)) \\
&\quad - [\Psi_{\hat{\theta}}(x, \bar{X})\hat{A}^{-1}H(\bar{X})\hat{\beta} - \Psi_{\tilde{\theta}}(x, \bar{X})\tilde{A}^{-1}H(\bar{X})\tilde{\beta}]| \\
&\leq \|h(x)\|_2\|\hat{\beta} - \tilde{\beta}\|_2 + \|\bar{Y}\|_2\|\hat{A}^{-1}\Psi_{\hat{\theta}}(\bar{X}, x) - \tilde{A}^{-1}\Psi_{\tilde{\theta}}(\bar{X}, x)\|_2 \\
&\quad + \|\Psi_{\hat{\theta}}(x, \bar{X})\hat{A}^{-1}H(\bar{X})\hat{\beta} - \Psi_{\tilde{\theta}}(x, \bar{X})\tilde{A}^{-1}H(\bar{X})\tilde{\beta}\|_2 \\
&= \text{Part}(i) + \text{Part}(ii) + \text{Part}(iii).
\end{aligned} \tag{D.16}$$

Part(ii) can be bounded using Lemma D.1 as

$$\text{Part}(ii) \leq \|\bar{Y}\|_2 \frac{2\delta}{1-r} \kappa(\hat{A}) \|\hat{A}^{-1}\Psi_{\hat{\theta}}(\bar{X}, x)\|_2. \tag{D.17}$$

Similarly, Part(iii) can be bounded using (D.13) and Lemma D.1 as

$$\begin{aligned}
\text{Part}(iii) &\leq \|H(\bar{X})\|_2\|\hat{\beta}\|_2 \frac{2\delta}{1-r} \kappa(\hat{A}) \|\hat{A}^{-1}\Psi_{\hat{\theta}}(\bar{X}, x)\|_2 + \|H(\bar{X})\|_2\|\hat{\beta} - \tilde{\beta}\|_2 \|\tilde{A}^{-1}\Psi_{\tilde{\theta}}(\bar{X}, x)\|_2 \\
&\leq \|H(\bar{X})\|_2\|\hat{\beta}\|_2 \frac{2\delta}{1-r} \kappa(\hat{A}) \|\hat{A}^{-1}\Psi_{\hat{\theta}}(\bar{X}, x)\|_2 + \|H(\bar{X})\|_2\|\hat{\beta} - \tilde{\beta}\|_2 \frac{1+r}{1-r} \|\hat{A}^{-1}\Psi_{\hat{\theta}}(\bar{X}, x)\|_2.
\end{aligned} \tag{D.18}$$

Combining (D.16), (D.17) and (D.18) gives

$$\begin{aligned}
|\hat{f}_{\hat{\theta}} - \hat{f}_{\tilde{\theta}}| &\leq \frac{2\delta\kappa(\hat{A})}{(1-r)\lambda_{\min}(\hat{A})} \|\Psi_{\hat{\theta}}(\bar{X}, x)\|_2 (\|\bar{Y}\|_2 + \|H(\bar{X})\|_2\|\hat{\beta}\|_2) \\
&\quad + \|\hat{\beta} - \tilde{\beta}\|_2 (\|h(x)\|_2 + \frac{1+r}{(1-r)\lambda_{\min}(\hat{A})} \|H(\bar{X})\|_2 \|\Psi_{\hat{\theta}}(\bar{X}, x)\|_2).
\end{aligned} \tag{D.19}$$

Notice that the first term in (D.19) can be controlled by restraining $g(\Sigma_M, \Sigma_{\epsilon})$, as defined in (13). The second part can be controlled by, in addition, restraining $\|\hat{\beta} - \tilde{\beta}\|_2$. Recall that

$$\begin{aligned}
\hat{\beta} &= (H(\bar{X})^T \hat{A}^{-1} H(\bar{X}))^{-1} H(\bar{X})^T \hat{A}^{-1} \bar{Y}, \\
\tilde{\beta} &= (\tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X}))^{-1} \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{Y}.
\end{aligned}$$

Since by Lemma D.2, the condition of Lemma D.1 holds. Thus, by Lemma D.1, we have

$$\|\hat{\beta} - \tilde{\beta}\|_2 \leq \frac{2\delta_1}{1-r_1} \kappa(H(\bar{X})^T \hat{A}^{-1} H(\bar{X})) \|\hat{\beta}\|_2 = \frac{2r_1}{1-r_1} \|\hat{\beta}\|_2.$$

By plugging in (D.14), we have

$$\|\hat{\beta} - \tilde{\beta}\|_2 \leq 2\delta \left(\kappa(\hat{A}) \kappa(H(\bar{X})^T H(\bar{X})) \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) + 1 \right) \|\hat{\beta}\|_2. \quad (\text{D.20})$$

Combining (D.19) and (D.20), we finish the proof.

E Proof of Lemma D.2

Notice that if Assumption D.2 holds, we have $r_1 < 1$. We only need to prove (D.15). Notice that

$$\begin{aligned} & \|H(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 \\ & \leq \|H(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \hat{A}^{-1} H(\bar{X})\|_2 + \|\tilde{H}(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 \\ & \leq \delta \|H(\bar{X})\|_2 \|\hat{A}^{-1} H(\bar{X})\|_2 + \|\tilde{H}(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 \\ & \leq \delta \|H(\bar{X})\|_2^2 \|\hat{A}^{-1}\|_2 + \|\tilde{H}(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2, \end{aligned} \quad (\text{E.21})$$

where the first inequality is true because of the triangle inequality, the second inequality is true because of Assumption D.1, and the third inequality is true because $\|G^{-1}d\|_2 \leq \|G^{-1}\|_2 \|d\|$ for any vector d and non-singular matrix G . The second term in (E.21) has

$$\begin{aligned} & \|\tilde{H}(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 \\ & \leq \|\tilde{H}(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \hat{A}^{-1} \tilde{H}(\bar{X})\|_2 + \|\tilde{H}(\bar{X})^T \hat{A}^{-1} \tilde{H}(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 \\ & \leq \delta \|\tilde{H}(\bar{X})\|_2 \|\hat{A}^{-1} \tilde{H}(\bar{X})\|_2 + \|\tilde{H}(\bar{X})^T (\hat{A}^{-1} - \tilde{A}^{-1}) \tilde{H}(\bar{X})\|_2 \\ & \leq \delta \|\tilde{H}(\bar{X})\|_2^2 \|\hat{A}^{-1}\|_2 + \|\hat{A}^{-1} - \tilde{A}^{-1}\|_2 \|\tilde{H}(\bar{X})\|_2^2 \\ & \leq \delta (1 + \delta)^2 \|H(\bar{X})\|_2^2 \|\hat{A}^{-1}\|_2 + (1 + \delta)^2 \|H(\bar{X})\|_2^2 \|\hat{A}^{-1} - \tilde{A}^{-1}\|_2, \end{aligned} \quad (\text{E.22})$$

where the first inequality is true is because of the triangle inequality, the second inequality is true because of Assumption D.1, the third inequality is true because $\|G^{-1}d\|_2 \leq \|G^{-1}\|_2 \|d\|$, and the last inequality is true because by Assumption 4.1, $\|\tilde{H}(\bar{X})\|_2 \leq (1 + \delta) \|H(\bar{X})\|_2$. Next, $\|\hat{A}^{-1} - \tilde{A}^{-1}\|_2$ is bounded.

For any $x \in \mathbb{R}^n$ such that $\|x\|_2 = 1$, let $y_1, y_2 \in \mathbb{R}^n$ such that $\hat{A}y_1 = x$ and $\tilde{A}y_2 = x$. Let

$\delta_A = \tilde{A} - \hat{A}$. Thus, $(\hat{A} + \delta_A)y_2 = x$. Notice that by assumption,

$$\|\hat{A}^{-1}\delta_A\|_2 \leq \delta\|\hat{A}^{-1}\|_2\|\hat{A}\|_2 = r < 1 \quad \text{and} \quad (I + \hat{A}^{-1}\delta_A)y_2 = y_1.$$

The following Lemma from Golub and Van Loan (1996) will be used.

Lemma E.1. *Suppose $F \in \mathbb{R}^{n \times n}$, $\|F\|_2 < 1$. Then $I - F$ is invertible and*

$$\|(I - F)^{-1}\|_2 \leq \frac{1}{1 - \|F\|_2},$$

where I is identity matrix in $\mathbb{R}^{n \times n}$.

By Lemma E.1, we have

$$\|y_2\|_2 \leq \|(I + \hat{A}^{-1}\delta_A)^{-1}\|_2\|y_2\|_2 \leq \frac{1}{1 - r}\|y_1\|_2 \quad \text{and} \quad y_1 - y_2 = \hat{A}^{-1}\delta_A y_2.$$

So,

$$\|y_1 - y_2\|_2 \leq \|\hat{A}^{-1}\delta_A\|_2\|y_2\|_2 \leq \delta\|\hat{A}^{-1}\|_2\|\hat{A}\|_2\|y_2\|_2 = \frac{\delta}{1 - r}\kappa(\hat{A})\|y_1\|_2.$$

Plugging in y_1 and y_2 gives

$$\|(\hat{A}^{-1} - \tilde{A}^{-1})x\|_2 \leq \frac{\delta}{1 - r}\kappa(\hat{A})\|\hat{A}^{-1}x\|_2 \leq \frac{\delta}{1 - r}\kappa(\hat{A})\|\hat{A}^{-1}\|_2 = \frac{\delta}{1 - r}\kappa(\hat{A})\frac{1}{\lambda_{\min}(\hat{A})}, \quad (\text{E.23})$$

indicating

$$\|\hat{A}^{-1} - \tilde{A}^{-1}\|_2 \leq \frac{\delta}{1 - r}\kappa(\hat{A})\frac{1}{\lambda_{\min}(\hat{A})}, \quad (\text{E.24})$$

since (E.23) is true for any x with $\|x\|_2 = 1$. Combining (E.21), (E.22), and (E.24) gives

$$\begin{aligned} & \|H(\bar{X})^T \hat{A}^{-1} H(\bar{X}) - \tilde{H}(\bar{X})^T \tilde{A}^{-1} \tilde{H}(\bar{X})\|_2 \\ & \leq \delta \|H(\bar{X})\|_2^2 \|\hat{A}^{-1}\|_2 + \delta(1 + \delta)^2 \|H(\bar{X})\|_2^2 \|\hat{A}^{-1}\|_2 + (1 + \delta)^2 \|H(\bar{X})\|_2^2 \|\hat{A}^{-1} - \tilde{A}^{-1}\|_2 \\ & \leq \delta \frac{\|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A})} + \delta(1 + \delta)^2 \frac{\|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A})} + \frac{\delta(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \frac{\|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A})} \end{aligned}$$

$$= \frac{\delta \|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A})} \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right). \quad (\text{E.25})$$

Thus, (D.15) holds if

$$\frac{\delta \|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A})} \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) < \delta_1 \|H(\bar{X})^T \hat{A}^{-1} H(\bar{X})\|_2,$$

or equivalently

$$\frac{\delta \|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A}) \lambda_{\min}(H(\bar{X})^T \hat{A}^{-1} H(\bar{X}))} \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) < \delta_1 \kappa(H(\bar{X})^T \hat{A}^{-1} H(\bar{X})). \quad (\text{E.26})$$

Next, we simplify (E.26). Notice that the left-hand side of (E.26) has

$$\begin{aligned} & \frac{\delta \|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A}) \lambda_{\min}(H(\bar{X})^T \hat{A}^{-1} H(\bar{X}))} \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) \\ & \leq \frac{\lambda_{\max}(\hat{A})}{\lambda_{\min}(H(\bar{X})^T H(\bar{X}))} \frac{\delta \|H(\bar{X})\|_2^2}{\lambda_{\min}(\hat{A})} \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) \\ & = \delta \kappa(\hat{A}) \kappa(H(\bar{X})^T H(\bar{X})) \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right), \end{aligned}$$

so, if

$$\delta \kappa(\hat{A}) \kappa(H(\bar{X})^T H(\bar{X})) \left(1 + (1 + \delta)^2 + \frac{(1 + \delta)^2}{1 - r} \kappa(\hat{A}) \right) < \delta_1 \kappa(H(\bar{X})^T \hat{A}^{-1} H(\bar{X})), \quad (\text{E.27})$$

(D.15) holds. By plugging in (D.14), we have (E.27) holds, which finishes the proof.

References

- Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computations (3rd Ed.)*. Johns Hopkins University Press, Baltimore, MD, USA.
- Haaland, B., Wang, W., and Maheshwari, V. (2018). A framework for controlling sources of inaccuracy in gaussian process emulation of deterministic computer experiments. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):497–521.

- Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*, volume 1. Springer.
- Ipsen, I. C. and Nadler, B. (2009). Refined perturbation bounds for eigenvalues of hermitian and non-hermitian matrices. *SIAM Journal on Matrix Analysis and Applications*, 31(1):40–53.
- Tripathi, G. (1999). A matrix extension of the cauchy-schwarz inequality. *Economics Letters*, 63(1):1–3.
- Varga, R. S. (2010). *Gershgorin and his circles*, volume 36. Springer Science & Business Media.