Supplementary materials for "Efficient basis selection for smoothing splines via rotated lattices"

April 2, 2023

1 Proof of Proposition 1 in Section 3.1

Proposition 1. *For any two covariates $\boldsymbol{x}, \boldsymbol{x}' \in [0,1]^p$ located in Vor($\boldsymbol{c}_i$), $i = 1, 2, \ldots, n$, we have*

$$\|\boldsymbol{x} - \boldsymbol{x}'\| \leq 2(p+1)^{1/(2p)}\{p(p+2)\}^{1/2}\{12(p+1)\}^{-1/2}n^{-1/p}.$$

Proof. Because Voronoi cells $\{\text{Vor}(\boldsymbol{c}_i)\}_{i=1}^n$ from the lattice L are identical and central symmetric spheres with volume $|det(lM_L)|$, a reasonable scaling parameter $l = n^{-1/p}(p+1)^{1/(2p)}$, and one of the most commonly used generator matrices $M_L = I_p - [\{(1+p)^{1/2} + 1 + p\}/\{p(p+1)\}]\mathbf{1}_p^T\mathbf{1}_p$, then $|det(lM_L)| = n^{-1}$. Moreover, according to He (2017), we have $|det(lM_L)| = \Psi\varrho^p/\Theta$. Where Ψ is the volume of a unit ball in $\mathcal{R}^p$ and $\Theta = \Psi(p+1)^{1/2}[p(p+2)\{12(p+1)\}^{-1}]^{p/2}$ is the volume of a ball in $\mathcal{R}^p$ with radius ϱ divided by the volume of one Voronoi cell. Therefore, for any two covariates $\boldsymbol{x}, \boldsymbol{x}' \in [0,1]^p$ located in Vor($\boldsymbol{c}_i$), $i = 1, 2, \ldots, n$, we have $\|\boldsymbol{x} - \boldsymbol{x}'\| \leq 2\varrho$, that is

$$\|\boldsymbol{x} - \boldsymbol{x}'\| \leq 2(p+1)^{1/(2p)}\{p(p+2)\}^{1/2}\{12(p+1)\}^{-1/2}n^{-1/p}.$$

□

2 Proof of essential lemmas to Theorem 1

Lemma 1. *If* $n^{1+2/p} = O(N)$*, under Conditions* 4–6*, for all* μ *and* γ*,* $\sum_{i=1}^{n}(N_i/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)$ *is an asymptotically unbiased estimate for* $\int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})f_X(\boldsymbol{x})d\boldsymbol{x}$*. when* $\mathcal{X} \subseteq [0,1]^p$*, we have*

$$\left\{\int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})f_X(\boldsymbol{x})d\boldsymbol{x} - \sum_{i=1}^{n}\frac{N_i}{N}\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\right\}^2 = O_P(n^{-1-2/p}).$$

Proof. Let $\{\boldsymbol{x}_i^*\}_{i=1}^n$ represent the collected covariates from the training dataset $\{\boldsymbol{x}_t\}_{t=1}^N$. We start by showing that $\sum_{i=1}^{n}(N_i/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)$ is an asymptotically unbiased estimate for $\int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})f_X(\boldsymbol{x})d\boldsymbol{x}$. Let the partition $\{\mathrm{Vor}(\boldsymbol{c}_i)\}_{i=1}^n$ be the cells of L in $[0,1]^p$ and

$$\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}) = \begin{cases} 1 & \boldsymbol{x} \in \mathrm{Vor}(\boldsymbol{c}_j) \\ 0 & \boldsymbol{x} \notin \mathrm{Vor}(\boldsymbol{c}_j) \end{cases}$$

be the indicator function. Furthermore, denote N_j as the number of covariates located in $\mathrm{Vor}(\boldsymbol{c}_j)$, $j = 1, 2, \ldots, n$. Obviously, $\mathrm{Vor}(\boldsymbol{c}_j)$ does not depend on the training data. Recall a random sampling procedure is performed in the cell $\mathrm{Vor}(\boldsymbol{c}_j)$ of Algorithm 1. Thus, for any $j \in \{1, 2, \ldots, n\}$, we have

$$E\left[\sum_{i=1}^{n}\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*)\middle|\{\boldsymbol{x}_t\}_{t=1}^N\right] = \frac{1}{N_j}\sum_{t=1}^{N}\phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_t). \tag{1}$$

Equation (1) indicates

$$E\left[\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{N_j}{N}\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*)\middle|\{\boldsymbol{x}_t\}_{t=1}^N\right] = \frac{1}{N}\sum_{t=1}^{N}\phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t). \tag{2}$$

Moreover, because N_i is the number of covariates allocated in the cell that $\boldsymbol{x}_i^*$ lies in and only one covariate is randomly selected from each non-empty cell,

$$\sum_{j=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) = \begin{cases} (N_i/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) & \boldsymbol{x}_i^* \in \mathrm{Vor}(\boldsymbol{c}_j); \\ 0 & \boldsymbol{x}_i^* \notin \mathrm{Vor}(\boldsymbol{c}_j). \end{cases}$$

Consequently, we have

$$\begin{aligned} \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) &= \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) \right\} \\ &= \sum_{i=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*). \end{aligned} \quad (3)$$

Combining Equation (2) and Equation (3), we have

$$E\left\{ \sum_{i=1}^{n} \frac{N_i}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \right\} = \frac{1}{N}\sum_{t=1}^{N} \phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t). \quad (4)$$

Therefore, $\sum_{i=1}^{n}(N_i/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)$ is an asymptotically unbiased estimate for $\int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})f_X(\boldsymbol{x})d\boldsymbol{x}$. We now derive the convergence rate for $\sum_{i=1}^{n}(N_i/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)$ in the following. Recall the Proposition 1 in Section 3.1. For any two covariates $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{Vor}(\boldsymbol{c}_j)$, $j = 1, 2, \ldots, n$, we have

$$\|\boldsymbol{x} - \boldsymbol{x}'\| \leq 2(p+1)^{1/(2p)}\{p(p+2)\}^{1/2}\{12(p+1)\}^{-1/2}n^{-1/p}.$$

Consequently, under Condition 4, we get

$$\begin{aligned} &\left|\phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) - \phi_\gamma(\boldsymbol{x}')\phi_\mu(\boldsymbol{x}')\right| \\ &\leq B\|\boldsymbol{x} - \boldsymbol{x}'\| \\ &\leq 2B(p+1)^{1/(2p)}\{p(p+2)\}^{1/2}\{12(p+1)\}^{-1/2}n^{-1/p}. \end{aligned} \quad (5)$$

Because N_j represents the number of covariates located in $\mathrm{Vor}(\boldsymbol{c}_j)$, Inequality (5) indicates that for any $\boldsymbol{x}_i^* \in \mathrm{Vor}(\boldsymbol{c}_j)$, $j = 1, 2, \ldots, n$,

$$
\begin{aligned}
&\left|\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) - \frac{1}{N_j}\sum_{t=1}^{N}\phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_t)\right| \\
&= \left|\frac{1}{N_j}N_j\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) - \frac{1}{N_j}\sum_{t=1}^{N}\phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_t)\right| \\
&\le \frac{1}{N_j}\sum_{\boldsymbol{x}_i^*, \boldsymbol{x}_t \in \mathrm{Vor}(\boldsymbol{c}_j)}\left|\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) - \phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_t)\right| \\
&\le 2B(p+1)^{1/(2p)}\{p(p+2)\}^{1/2}\{12(p+1)\}^{-1/2}n^{-1/p}.
\end{aligned}
$$

In other words, the conditional variance of $(N_j/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*)$ is bounded by

$$
\begin{aligned}
&\mathrm{var}\left[\frac{N_j}{N}\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*)\middle|\{\boldsymbol{x}_t\}_{t=1}^N\right] \\
&= \frac{N_j^2}{N^2}\left\{\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*) - \frac{1}{N_j}\sum_{t=1}^{N}\phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_t)\right\}^2 \\
&\le \frac{N_j^2}{N^2}4B^2(p+1)^{1/p}\{p(p+2)\}\{12(p+1)\}^{-1}n^{-2/p}. \qquad (6)
\end{aligned}
$$

Consequently, the conditional variance of $\sum_{j=1}^n (N_j/N)\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*)$ is

$$
\begin{aligned}
&\mathrm{var}\left[\sum_{j=1}^n\frac{N_j}{N}\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\mathbb{I}_{\mathrm{Vor}(\boldsymbol{c}_j)}(\boldsymbol{x}_i^*)\middle|\{\boldsymbol{x}_t\}_{t=1}^N\right] \\
&= \mathrm{var}\left[\sum_{i=1}^n\frac{N_i}{N}\phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*)\middle|\{\boldsymbol{x}_t\}_{t=1}^N\right] \\
&\le \sum_{i=1}^n\frac{N_j^2}{N^2}4B^2(p+1)^{1/p}\{p(p+2)\}\{12(p+1)\}^{-1}n^{-2/p} \\
&\le \max_{1\le i\le n}\frac{N_j^2}{N^2}\times n\times 4B^2(p+1)^{1/p}\{p(p+2)\}\{12(p+1)\}^{-1}n^{-2/p} \\
&= \{O_P(1)n^{-1}\}\times 4B^2(p+1)^{1/p}\{p(p+2)\}\{12(p+1)\}^{-1}n^{-2/p} \\
&= O_P(n^{-1-2/p}) \qquad (7)
\end{aligned}
$$

Under Condition 4, we have

$$\left\{ \frac{1}{N} \sum_{t=1}^{N} \phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t) - \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) f_X(\boldsymbol{x}) d\boldsymbol{x} \right\}^2 = O_P(N^{-1}). \tag{8}$$

Then, combining Condition 6, Equation (4), Inequality (7), and Equation (8), we obtain

$$\begin{aligned}
&\operatorname{var}\left[\sum_{i=1}^{n} \frac{N_i}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \right] \\
&= E\left[\operatorname{var}\left\{ \sum_{j=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \middle| \{\boldsymbol{x}_t\}_{t=1}^{N} \right\} \right] + \operatorname{var}\left[E\left\{ \sum_{j=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \middle| \{\boldsymbol{x}_t\}_{t=1}^{N} \right\} \right] \\
&= O_P(n^{-1-2/p}) + O_P(N^{-1}) \\
&= O_P(n^{-1-2/p}).
\end{aligned} \tag{9}$$

Therefore, because of Equation (8), Equation (9), and Holder's inequality, we obtain

$$\begin{aligned}
&\left\{ \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) f_X(\boldsymbol{x}) d\boldsymbol{x} - \sum_{i=1}^{n} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \right\}^2 \\
&\leq 2\left\{ \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) f_X(\boldsymbol{x}) d\boldsymbol{x} - \frac{1}{N} \sum_{t=1}^{N} \phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t) \right\}^2 \\
&+ 2\left\{ \frac{1}{N} \sum_{t=1}^{N} \phi_\gamma(\boldsymbol{x}_t)\phi_\mu(\boldsymbol{x}_t) - \sum_{i=1}^{n} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \right\}^2 \\
&= O_P(n^{-1-2/p}) + O_P(N^{-1}) \\
&= O_P(n^{-1-2/p}).
\end{aligned} \tag{10}$$

□

Lemma 2. *Under Condition 2, when $\lambda \to 0$, we have*

$$\sum_{\gamma}(1+\lambda\rho_\lambda)^{-1} = O(\lambda^{-1/r}),$$
$$\sum_{\gamma}(1+\lambda\rho_\lambda)^{-2} = O(\lambda^{-1/r}),$$
$$\sum_{\gamma}\lambda\rho_\lambda(1+\lambda\rho_\lambda)^{-2} = O(\lambda^{-1/r}). \tag{11}$$

Proof. We prove the first equation, and the proof of the other two equations can be obtained similarly.

$$\begin{aligned}\sum_{\gamma}(1+\lambda\rho_\lambda)^{-1} &= \left(\sum_{\gamma<\lambda^{-1/r}} + \sum_{\gamma\geq\lambda^{-1/r}}\right)(1+\lambda\rho_\lambda)^{-1} \\ &= O(\lambda^{-1/r}) + O\left(\int_{\lambda^{-1/r}}^{+\infty}(1+\lambda\boldsymbol{x}^r)^{-1}d\boldsymbol{x}\right) \\ &= O(\lambda^{-1/r}) + \lambda^{-1/r}O\left(\int_{1}^{+\infty}(1+\boldsymbol{x}^r)^{-1}d\boldsymbol{x}\right) \\ &= O(\lambda^{-1/r}).\end{aligned}$$

□

Lemma 3. *Under Conditions 1-5, when $\lambda \to 0$ and $n^{1+2/p}\lambda^{2/r} \to \infty$, for any $h \in \mathcal{H}\ominus\mathcal{H}_S$, we obtain $\Lambda(h) = o_P\{\lambda J(h)\}$, where $\mathcal{H}\ominus\mathcal{H}_S$ is the orthogonal complement of $\mathcal{H}_S$ in the reproducing kernel Hilbert space $\mathcal{H}$.*

Proof. For any $h \in \mathcal{H}\ominus\mathcal{H}_S$ and $i \in \{1, 2, \ldots, N\}$, we have $h(\boldsymbol{x}_i) = J(G_J(\boldsymbol{x}_i, \cdot), h) = 0$.

Therefore, we obtain $\sum_{j=1}^{n}(N_j/N)h^2(\boldsymbol{x}_j^*) = 0$. Write $h = \sum_\gamma h_\gamma \phi_\gamma$, then, it satisfies that

$$\begin{aligned}
\Lambda(h) &= \int_{\mathcal{X}} h^2(\boldsymbol{x}) f_X(\boldsymbol{x}) d\boldsymbol{x} \\
&= \sum_\gamma \sum_\mu h_\gamma h_\mu \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) d\boldsymbol{x} \\
&= \sum_\gamma \sum_\mu h_\gamma h_\mu \left\{ \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) d\boldsymbol{x} - \sum_{j=1}^{n} (N_j/N)\phi_\gamma(\boldsymbol{x}_j^*)\phi_\mu(\boldsymbol{x}_j^*) \right\}. \qquad (12)
\end{aligned}$$

Following the Cauchy inequality, we have

$$\begin{aligned}
\Lambda(h) \le & \left[\sum_\gamma \sum_\mu (1+\lambda\rho_\gamma)^{-1}(1+\lambda\rho_\mu)^{-1} \left\{ \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) d\boldsymbol{x} - \sum_{j=1}^{n} \frac{N_j}{N} \phi_\gamma(\boldsymbol{x}_j^*)\phi_\mu(\boldsymbol{x}_j^*) \right\}^2 \right]^{1/2} \\
& \times \left\{ \sum_\gamma \sum_\mu (1+\lambda\rho_\gamma)(1+\lambda\rho_\mu) h_\gamma^2 h_\mu^2 \right\}^{1/2}. \qquad (13)
\end{aligned}$$

According to Lemma 2, we get

$$\sum_\gamma \sum_\mu (1+\lambda\rho_\gamma)^{-1}(1+\lambda\rho_\mu)^{-1} = O(\lambda^{-2/r}). \qquad (14)$$

Meanwhile, form Lemma 1, we have

$$\left\{ \int_{\mathcal{X}} \phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x}) f_X(\boldsymbol{x}) d\boldsymbol{x} - \sum_{i=1}^{n} \frac{N_i}{N} \phi_\gamma(\boldsymbol{x}_i^*)\phi_\mu(\boldsymbol{x}_i^*) \right\}^2 = O_P(n^{-1-2/p}). \qquad (15)$$

Because ϕ_γ's simultaneously diagonalize Λ and J,

$$\sum_\gamma (1+\lambda\rho_\gamma) h_\gamma^2 = (\Lambda + \lambda J)(h). \qquad (16)$$

Moreover, together with equations (14), (15), and (16), the desired result follows by the Equation (13), i.e.,

$$\Lambda(h) \le \left\{ O_P(n^{-1-2/p})\lambda^{-2/r} \right\} (\Lambda + \lambda J)(h). \qquad (17)$$

Finally, when $n^{1+2/p}\lambda^{2/r} \to \infty$, from Inequality (17), we obtain

$$\Lambda(h) = o_P\left\{\lambda J(h)\right\}. \tag{18}$$

□

3 Proof of Theorem 1 in Section 3.3

Theorem 1. *Assume that* $\sum_{\gamma} \rho_{\gamma}^{d}\Lambda(\eta_0, \phi_\gamma)^2 < \infty$ *for some* $d \in [1, 2]$. *Under Conditions 1-6, as* $\lambda \to 0$ *and* $n^{1+2/p}\lambda^{2/r} \to \infty$, *we have* $(\Lambda + \lambda J)(\tilde{\eta}_R - \eta_0) = O_P(N^{-1}\lambda^{-1/r} + \lambda^{d})$. *Particularly, when* $\lambda \asymp N^{-r/(dr+1)}$, $\tilde{\eta}_R$ *achieves the optimal convergence rate* $(\Lambda+\lambda J)(\tilde{\eta}_R - \eta_0) = O_P(N^{-dr/(dr+1)})$.

The condition in Theorem 9.17 in Gu (2013) is $n\lambda^{2/r} \to \infty$, compared with that in Theorem 1, the condition is $n^{1+2/p}\lambda^{2/r} \to \infty$. Therefore, $\tilde{\eta}_R$ has the same convergence rate as the full bases estimator with a smaller n. Under this condition and conditions 1-5, as $\lambda \to 0$, $\forall h \in \mathcal{H} \ominus \mathcal{H}_S$, $\Lambda(h)$ is dominated by $\lambda J(h)$, which is guaranteed by Lemma 3. Theorem 1 can therefore be proved directly according to the proof of Theorem 9.17 in Gu (2013).

4 Additional Simulation Results

Aside from MAE, log mean squared error (MSE) given by $\log\left(N_{test}^{-1}\sum_{i=1}^{N_{test}}(\tilde{\eta}(\boldsymbol{x}_i) - \eta_0(\boldsymbol{x}_i))^2\right)$ is another commonly used measurement. In this section, we will further evaluate the performance of all five methods in Section 4 via log(MSE). Except the performance measurement, all the settings are the same as Section 4.

As can be seen from Figure 1, as expected, the performance under MSE is quite similar compared with MAE. The RBS method is still superior to the other four basis selection methods, which is identical to the results in Section 4.

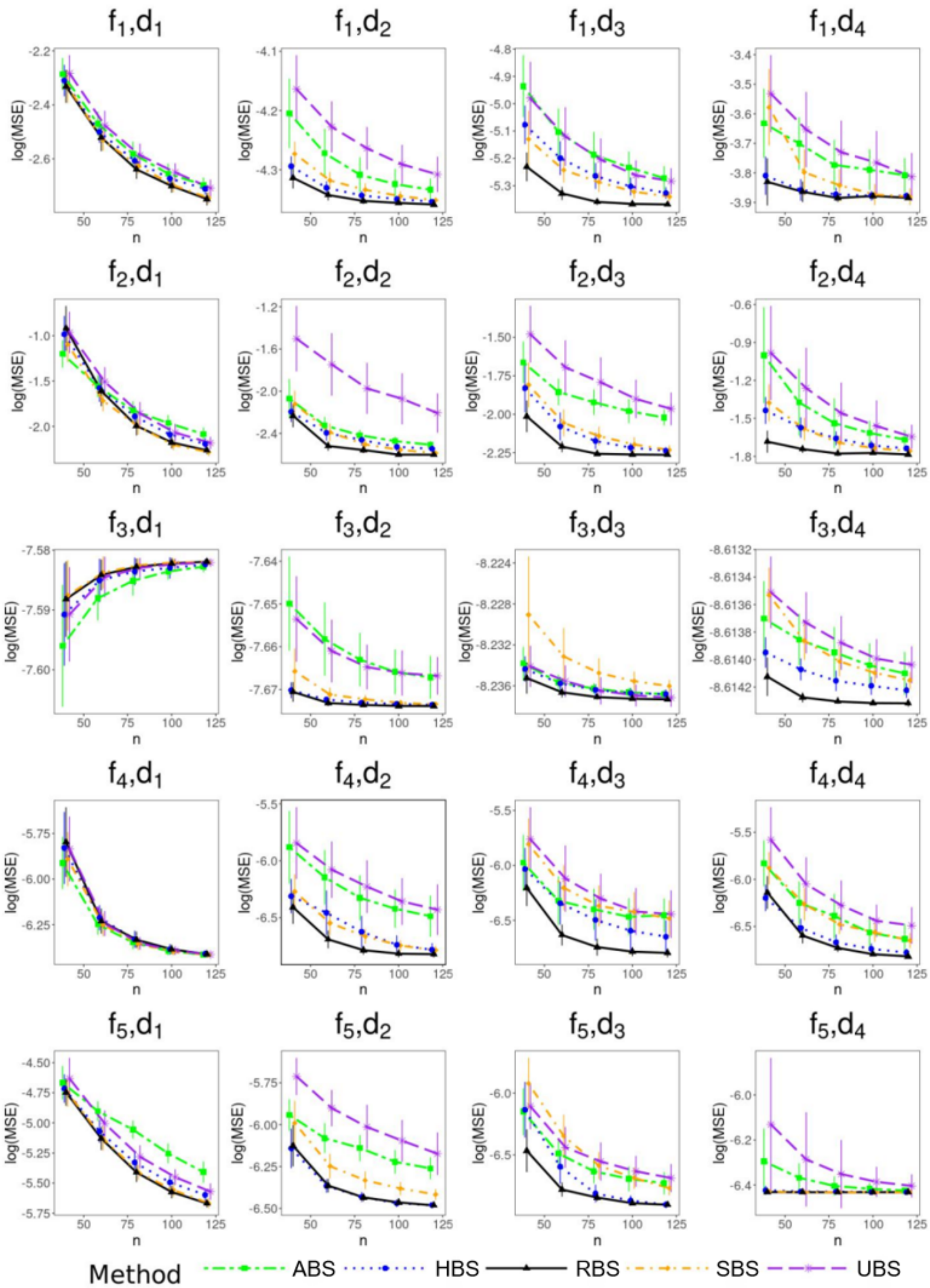

Figure 1: log(MSE) values from the smoothing spline model for simulation under five different regression functions (from upper to lower) and four different probability density functions (from left to right) are plotted versus different n.

References

Gu, C. (2013). *Smoothing spline ANOVA models*. Springer Science & Business Media.

He, X. (2017). Rotated sphere packing designs. *Journal of the American Statistical Association 112*(520), 1612–1622.