Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

Efficient basis selection for smoothing splines via rotated lattices

Huaimin Diao[1] | Mengtong Ai[2] | Yubin Tian[1] | Jun Yu*[1]

[1]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China

[2]Department of Statistics, University of Michigan, Ann Arbor, Michigan, U.S.A.

Correspondence

*Jun Yu. Email: yujunbeta@bit.edu.cn

Summary

Fitting a smoothing spline model on a large-scale dataset is daunting due to the high computational cost. In this study, we develop an efficient basis selection method for smoothing spline calculation. The key idea is to force a nonparametric function in an infinite-dimensional functional space to reside in a relatively small and finite-dimensional model space without the loss of too much prediction accuracy. Such an approximation naturally allows for much faster numerical calculation, especially for large datasets. Among various basis selection methods, space-filling basis selection has been proven to be more efficient since its model space dimension is smaller than that of others. Despite algorithmic benefits, most of the space-filling selection methods only take the overall space-filling property into account. These methods may be less efficient when the underlying response surface is not isomorphic. To overcome this obstacle, we develop an efficient algorithm to improve projective uniformity for space-filling basis selection. It has been proved that the proposed estimator has the same convergence rate as the full bases estimator. Compared with the standard approach, the proposed method significantly reduces the computational cost. Simulation and real data studies demonstrate the efficiency and superiority of the proposed method.

KEYWORDS:
Massive data, Projective uniformity, Rotated lattice, Smoothing spline, Space-filling basis selection

1 | INTRODUCTION

Consider the following nonparametric model:

$$y_i = \eta(\boldsymbol{x}_i) + \epsilon_i, \quad i = 1, 2, \ldots, N, \tag{1}$$

where y_i is the response for the i-th observation, $\eta(\cdot)$ is an unknown function, $\boldsymbol{x}_i \in [0,1]^p$ denotes the covariate, ϵ_i's are independent and identically distributed errors with zero mean and finite variance, and N is the size of training data. In this study, we adopt the smoothing spline to fit $\eta(\cdot)$ which is one of the most pervasive choices (Grace 1990; Gu 2013). To be precise, $\eta(\cdot)$ in Model (1) is assumed to have resided in a reproducing kernel Hilbert space (RKHS), say $\mathcal{H}$. It can be estimated by minimizing the following penalized least squares

$$N^{-1}\sum_{i=1}^{N}\{y_i - \eta(\boldsymbol{x}_i)\}^2 + \lambda J(\eta), \tag{2}$$

where $J(\eta)$ is a roughness penalty, and λ is the tuning parameter to control the smoothness of the estimator.

Despite impressive performance and solid theoretical foundations, solving the optimization problem (2) is a challenging task. Except for the univariate case, which can be solved in $O(N)$, the computational cost for calculating the smoothing spline with $p \geq 2$ using the standard approach is $O(N^3)$. This huge computational cost renders the smoothing spline impractical for a large-scale dataset. Great efforts have been made to alleviate

the computational burden. Typical works include but are not limited to divide-and-conquer kernel ridge regressions (Xu & Wang 2018; Y. Zhang, John, & Martin 2015) and basis selection algorithms (Gu & Kim 2002; Luo & Wahba 1997; Ma, Huang, & Zhang 2015; Meng, Zhang, Zhang, Zhong, & Ma 2020).

Among various techniques designed to reduce the computational cost of the smoothing spline, basis selection methods have gained the attention of data scientists due to the following two reasons. Firstly, it is possible to combine parallel-based strategies with basis selection methods to reap more computational savings. Secondly, basis selection procedures are more flexible than divide-and-conquer kernel ridge regression strategies. The idea of basis selection is to approximate the minimizer of (2) by restricting the estimator $\hat{\eta}$ to a n-dimensional subspace $\mathcal{H}_S \subset \mathcal{H}$. When $n \ll N$, the computational cost can be efficiently reduced from $O(N^3)$ to $O(Nn^2)$. Numerous studies have been developed along this line of thinking. For example, Hastie (1996) suggested fixing n basis functions to approximate splines. Such a method differs from constructing basis functions in smoothing splines. Thus, it is also known as "pseudo splines" or P-splines. Ruppert (2002) considered basis selection for P-splines by adding a penalty to control the number of basis functions. He, Shen, and Shen (2001), Sklar, Wu, Meiring, and Wang (2013), and Yuan, Chen, and Zhou (2013) further considered cases where the regression function has non-homogeneous smoothness across the input space. As for smoothing splines, Luo and Wahba (1997) and H. Zhang Hao et al. (2004) proposed approximating the minimizer of (2) using variable selection techniques. Due to the additional computational cost brought by variable screening, the computational benefits of the above two methods are unclear. Gu and Kim (2002) proposed a uniform basis selection (UBS) method by randomly selecting some basis functions among the N basis functions in $\{G_J(\boldsymbol{x}_i, \cdot)\}_{i=1}^N$, facilitating quick analysis of large datasets. Here, G_J is the reproducing kernel. Ma et al. (2015) developed an adaptive basis selection (ABS) method that uses an adaptive sampling scheme according to the values of the response variable. Meng et al. (2020) proposed a space-filling basis selection (SBS) method that suggests selecting the basis functions $\{G_J(\boldsymbol{x}_i^*, \cdot)\}_{i=1}^n$ with $\{x_i^*\}_{i=1}^n$ being roughly equally-spaced. Meng, Yu, Chen, Zhong, and Ma (2021) extended the basis selection method to cases where the covariates are nonuniformly distributed in $[0,1]^p$. As the space-filling property of the selected sub-data $\{x_i^*\}_{i=1}^n$ is achieved through Hilbert curves, this method is the so-called Hilbert curve basis selection (HBS) method.

A fundamental question regarding basis selection methods is how to choose the basis functions in order to ensure that the restricted estimator converges to the true function η at the identical rate as the full bases estimator $\hat{\eta}$. It has been shown that both the UBS and ABS methods require that n roughly be of the order $O(N^{2/(dr+1)})$, where $d \in [1,2]$ and $r \approx 4$ are constants, depending on the type of spline. In contrast, the SBS only requires that n roughly be of the order $O(N^{1/(dr+1)})$ when $p \leq dr+1$ and the covariates are uniformly distributed in $[0,1]^p$. Thus, the SBS is regarded as a more efficient method. To relax the assumption that all covariates are uniformly distributed in $[0,1]^p$, HBS achieves a similar space-filling property as SBS by increasing the number of bases to $O(N^{2p/\{(p+2)(dr+1)\}})$. The additional basis functions are the price we pay for the irregular distribution of the covariates. Nevertheless, HBS is also more efficient than both UBS and ABS.

To the best of our knowledge, the space-filling type basis selection methods (i.e., SBS and HBS) are only focused on the uniformity of the sub-data $\{x_i^*\}_{i=1}^n$ with respect to the whole input space $[0,1]^p$. Unfortunately, this can result in poor projections of lower-dimensional spaces, which is undesirable when the response curve has non-homogeneous smoothness across the input space. Consider a case where a researcher fits a smoothing spline on a two-dimensional space $[0,1]^2$ while the response curve is $\eta(\boldsymbol{x}) = \sin(15x_1^2)$, which depends solely on the first dimension of the covariate. The true function and the $\{\boldsymbol{x}_i^*\}_{i=1}^n$ generated via UBS, ABS, SBS, HBS, and the proposed method are demonstrated in Figure 1. It is observed that the covariates selected via the UBS and ABS methods are not uniformly distributed in the input space. In contrast, SBS, HBS, and the proposed method enjoy space-filling properties in the whole input space $[0,1]^2$. It is worth mentioning that the three methods have different behaviors when we only consider the projection on x_1. One can conclude that the proposed RBS method has better uniformity when we consider the projection on the subspace of the design region compared with the other methods. Thus, the selected basis functions are more likely to capture the volatility in the first dimension in this case, which will undoubtedly lead to better estimation.

In this study, we propose a rotated lattice-based basis selection (RBS) method to improve the projection uniformity with space-filling properties of the selected sub-data. The contributions of this work are three-fold as follows. Firstly, a novel sub-data selection is proposed for basis selection when the covariates are not uniformly distributed in $[0,1]^p$ and the response curve may have non-homogeneous smoothness across the input space. Secondly, it has been proved that only $O(N^{2p/\{(dr+1)(p+2)\}})$ basis functions are needed to achieve the same convergence rate as the full bases estimator, which is the same as with HBS. This implies that our method increases the projective uniformity without involving additional computational costs. Thirdly, numerical experiments show that the RBS method has comparative performance when compared to the HBS and UBS methods and that it outperforms UBS and ABS when the response curves are isomorphic. Moreover, it is uniformly better than the UBS, ABS, SBS, and HBS methods when the response curve has non-homogeneous smoothness.

The remainder of this paper is organized as follows. In Section 2, the preliminaries of smoothing splines and the sub-sampling of basis functions are reviewed. In Section 3, we present the RBS method and an asymptotic analysis. In Section 4, simulation and real data examples are provided to demonstrate the advantage of the proposed basis selection method. Several conclusions and remarks are given in Section 5. All proofs are provided in the supplementary materials.

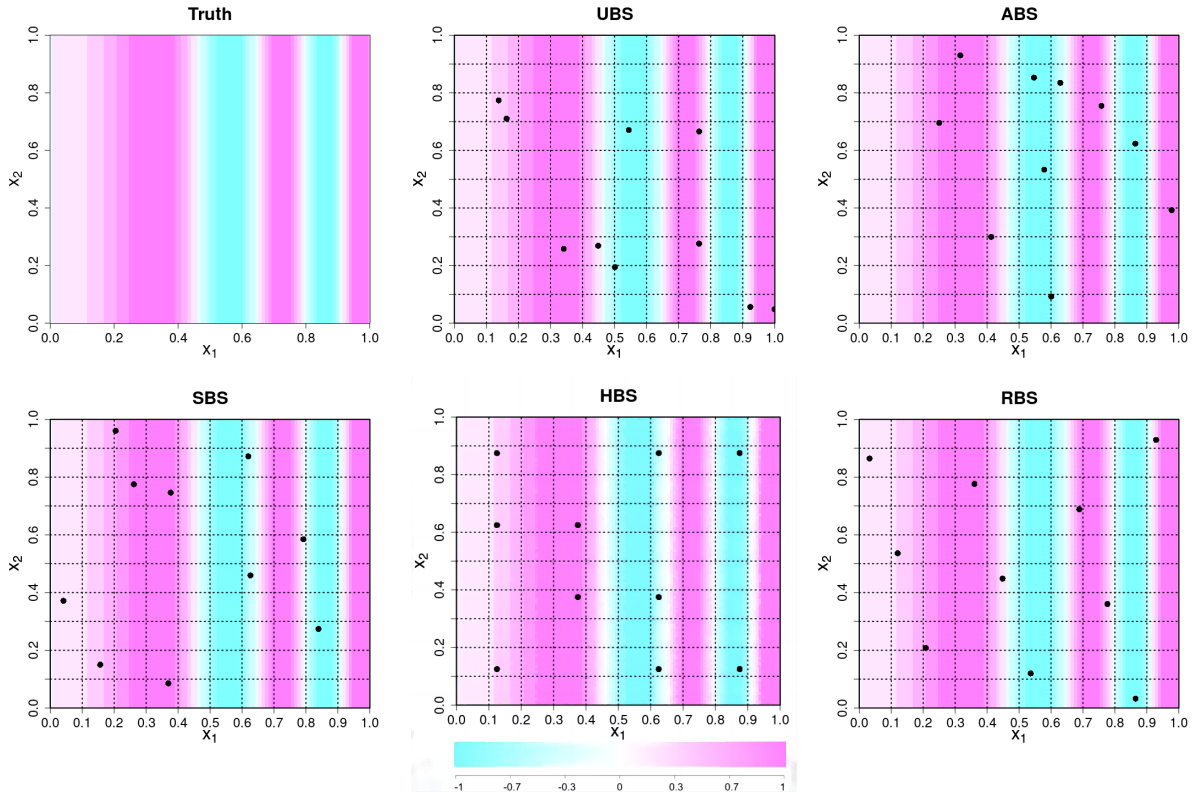

Figure 1 Illustration of different basis selection methods. The leftmost panel at the top shows the heat map for the true function. The heat maps for the spline estimates based on UBS, ABS, SBS, HBS, and the proposed method are presented in the other five panels, respectively. Black dots are the sampled basis functions.

2 | SMOOTHING SPLINES ESTIMATION AND THEIR BASIS SELECTION

In this section, we briefly review the basic background of smoothing splines and introduce the general basis selection methods. For ease of presentation, all vectors are row vectors throughout this study.

Let $\mathcal{H} = \{\eta : J(\eta) < \infty\}$ be an RKHS equipped with a squared semi-norm $J(\cdot)$. Then, $\mathcal{H}$ can be decomposed by $\mathcal{H} = \mathcal{N}_J \oplus \mathcal{H}_J$ with $\mathcal{N}_J = \{\eta : J(\eta) = 0\}$ being the null space of $J(\eta)$ and $\mathcal{H}_J$ being the orthogonal complement of $\mathcal{N}_J$ in $\mathcal{H}$. It is proved that the space $\mathcal{H}_J$ is also a reproducing kernel space associated with a reproducing kernel $G_J(\cdot,\cdot)$ (Gu 2013).

Assume that $\mathcal{N}_J$ is a finite W-dimensional linear subspace of $\mathcal{H}$ with the basis functions $\{\omega_j(\cdot)\}_{j=1}^W$. According to the well-known Wahba's representer theorem (Grace 1990), the unique solution (usually called the smoothing spline estimator) to (2) can be represented by

$$\eta(\boldsymbol{x}) = \sum_{j=1}^{W} \alpha_j \omega_j(\boldsymbol{x}) + \sum_{i=1}^{N} \beta_i G_J(\boldsymbol{x}_i, \boldsymbol{x}) \qquad \boldsymbol{x} \in [0,1]^p.$$

Denote $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_W)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)$. The computation of the smoothing spline estimator is reduced to computing the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Denote Y as the response vector, Ω as the $N \times W$ matrix, with the (i,j)-th element being $\omega_j(\boldsymbol{x}_i)$, and G as the $N \times N$ matrix, with the (i,j)-th element being $G_J(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Then, the original problem (2) becomes that of solving

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \arg \min_{\boldsymbol{\alpha} \in \mathcal{R}^W, \boldsymbol{\beta} \in \mathcal{R}^N} \{N^{-1}(Y - \Omega\boldsymbol{\alpha}^T - G\boldsymbol{\beta}^T)(Y - \Omega\boldsymbol{\alpha}^T - G\boldsymbol{\beta}^T)^T + \lambda \boldsymbol{\beta} G \boldsymbol{\beta}^T\}.$$

The computational cost of the standard approach is $O(N^3)$, which is prohibitive when the observation size N is considerable. To reduce this computational burden, basis selection is adopted. The key idea is to further restrict the unknown function $\eta(\cdot)$ in a smaller functional space $\mathcal{H}_S = \mathcal{N}_J \oplus \text{span}\{G_J(\boldsymbol{x}_i^*, \cdot), i = 1, 2, \ldots, n\}$, where $\{\boldsymbol{x}_i^*\}_{i=1}^n$ is a subset of $\{\boldsymbol{x}_i\}_{i=1}^N$. Consequently, the computation of the smoothing spline estimator is reduced to the computation of coefficients $\boldsymbol{\alpha}_S$ and $\boldsymbol{\beta}_S$ by solving

$$(\hat{\boldsymbol{\alpha}}_S, \hat{\boldsymbol{\beta}}_S) = \arg \min_{\boldsymbol{\alpha}_S \in \mathcal{R}^W, \boldsymbol{\beta}_S \in \mathcal{R}^n} \{N^{-1}(Y - \Omega\boldsymbol{\alpha}_S^T - G_*\boldsymbol{\beta}_S^T)(Y - \Omega\boldsymbol{\alpha}_S^T - G_*\boldsymbol{\beta}_S^T)^T + \lambda \boldsymbol{\beta}_S G_{**} \boldsymbol{\beta}_S^T\}, \tag{3}$$

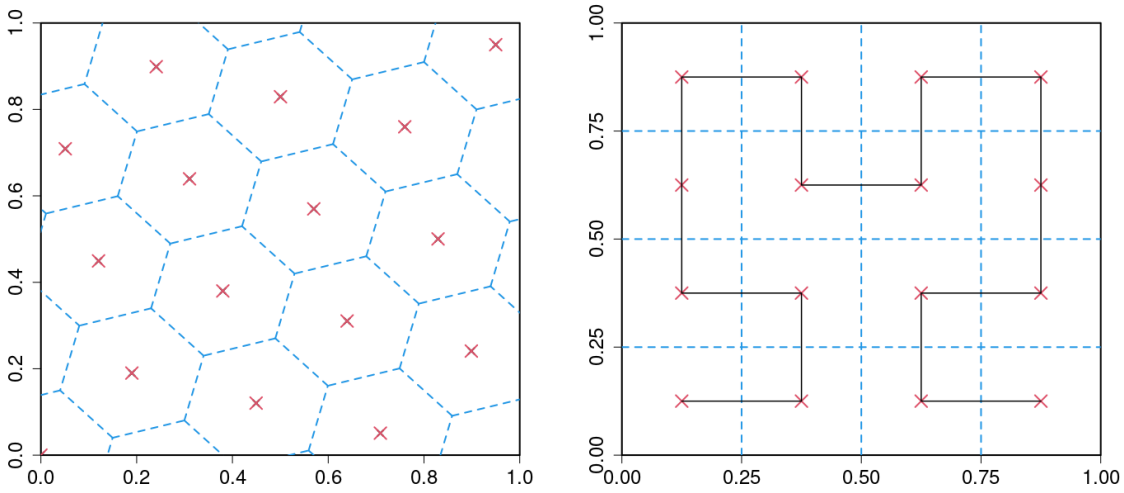

Figure 2 The hexagonal partitioning of $[0,1]^2$ with the rotated thinnest covering lattice on the left and the 2-th Hilbert's space-filling curve on the right shows cells (blue squares or hexagons), centers (red crosses), and the 2-th Hilbert's space-filling curve (black solid line).

where G_* is an $N \times n$ matrix, with the (i,j)-th entry $G_J(\boldsymbol{x}_i, \boldsymbol{x}_j^*)$ and G_{**} is an $n \times n$ matrix, with the (i,j)-th entry $G_J(\boldsymbol{x}_i^*, \boldsymbol{x}_j^*)$. Evaluation of the restricted estimator $\tilde{\eta}$ based on sample observations satisfies $\hat{H}_S^T = \Omega\hat{\boldsymbol{\alpha}}_S^T + G_*\hat{\boldsymbol{\beta}}_S^T$, where $\hat{H}_S = \{\tilde{\eta}(\boldsymbol{x}_1), \ldots, \tilde{\eta}(\boldsymbol{x}_n)\}$. Obviously, when $n \ll N$, the computational cost is only $O(N^2 n)$, which is a significant reduction compared with $O(N^3)$.

3 | BASIS SELECTION USING ROTATED LATTICES

3.1 | Rotated thinnest covering lattice

The thinnest covering lattice entails finding the best placement of the identical p-dimensional spheres that jointly cover the input space. For $2 \le p \le 22$, one can systematically construct the rotated thinnest covering lattice (He 2017) via

$$L = \{l\boldsymbol{a}M_L R : \boldsymbol{a} \in \mathcal{Z}^p\}, \tag{4}$$

where l is a positive constant giving the scale of the lattice cells, M_L is a generator matrix, and R is a $p \times p$ rotated matrix. Let V be the volume of the input space, i.e., the covariate space. A reasonable scaling parameter is $l = n^{-1/p}(p+1)^{1/(2p)}V^{1/p}$, which yields n lattice cells in the covariate space $[0,1]^p$ (He 2017 2020). One of the most commonly used generator matrices is $M_L = I_p - [\{(1+p)^{1/2}+1+p\}/\{p(p+1)\}]\mathbf{1}_p^T\mathbf{1}_p$, where I_p is the $p \times p$ identity matrix, and $\mathbf{1}_p$ is the p-dimensional row vector with all elements as one. The rotated matrix R can be implemented by any orthogonal matrix. More details can be refered to He (2017). Denote $\{\boldsymbol{c}_i\}_{i=1}^n$ as the elements of the lattice L in $[0,1]^p$; then, the Voronoi cell of point $\boldsymbol{c}_i$ is defined as the region $\text{Vor}(\boldsymbol{c}_i) = \{\boldsymbol{z} | \|\boldsymbol{z} - \boldsymbol{c}_i\|_2 = \min_{\boldsymbol{c}\in L}\|\boldsymbol{z} - \boldsymbol{c}\|_2\}$, where $\boldsymbol{c}_i$ is called the center of the Voronoi cell $\text{Vor}(\boldsymbol{c}_i)$, and $\|\cdot\|$ denotes the Euclidean norm. Here, we focus on the best known thinnest coverings, other lattices such as the densest packings listed in Conway and Sloane (1998) can also be applied. One may expect to use some other types of lattices to achieve the desired results when p is beyond 22.

A two-dimensional case for a rotated thinnest covering lattice in $[0,1]^2$ is demonstrated in the left panel of Figure 2. The hexagons are the Voronoi cells with their centers marked by the crosses, which are the elements of the lattice L. Clearly, the hexagon partitioned the space $[0,1]^2$. For comparison, we also show the space partitioned by Hilbert's space-filling curve, which is the main ingredient in the HBS method. This is presented in the right panel of Figure 2. From Figure 2, one can easily see that both the 2-dimensional rotated thinnest covering lattice and the 2-th Hilbert's space-filling curve yield reasonable partitions of $[0,1]^2$. Clearly, the 16 crosses in the two panels are uniform over the whole covariate space; however, the uniformity in the left panel is better than that in the right panel. Moreover, all projections of the 16 crosses in the left panel onto each axis have a space-filling character. However, some projections of the 16 crosses in the right panel onto any axis are coincident, and only four different projection points are obtained. Judging from the projection uniformity of each subspace, the rotated thinnest covering lattice is far

superior to Hilbert's space-filling curve. Thus, the $\{\boldsymbol{x}_i^*\}_{i=1}^n$ selected by the rotated thinnest covering lattice will benefit cases where the response curve has non-homogeneous smoothness. Proposition 1 gives the space-filling property for the general rotated thinnest covering lattice.

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

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

3.2 | The rotated lattice-based basis selection algorithm

The proposed RBS algorithm includes the following three main steps:

Step 1. Generate a rotated thinnest covering lattice L with n Voronoi cells $\{\mathrm{Vor}(\boldsymbol{c}_i)\}_{i=1}^n$ in $[0,1]^p$. Divide the covariates of N observations into n cells according to the distance information of p-dimensional covariates to Voronoi cells $\{\mathrm{Vor}(\boldsymbol{c}_i)\}_{i=1}^n$;

Step 2. Adjust the estimated value of the volume of non-empty cells, so that the size of the non-empty cells $\hat{n}$ is roughly equal to n;

Step 3. Randomly select one covariate from each non-empty cell to form $\{\boldsymbol{x}_i^*\}_{i=1}^n$, and then, estimate the smoothing spline on the space $\mathcal{H}_S$ as in (3).

A rotated thinnest covering lattice L is generated in the first step. After achieving the lattice L, we will map the covariates to the Voronoi cells in L. Clearly, performing an exhaustive search of all the Voronoi cells in L and then deciding to which cell a particular covariate belongs requires $O(np^2)$ time.

In the second step, denote $\hat{n}$ as the size of non-empty cells. Then, we should adjust the estimated value of the volume of non-empty cells so that $\hat{n}$ is closer to n. For clarity, we denote the estimated value by $\hat{V}$. Initially, $\hat{V} = 1$, i.e., the volume of the covariate space $[0,1]^p$. Then, l is estimated by $n^{-1/p}(p+1)^{1/(2p)}\hat{V}^{1/p}$. When the difference between $\hat{n}$ and the target n is too large, we should adjust l according to a new estimated $\hat{V}$, so that the difference between $\hat{n}$ and the target n does not exceed $0.1n$.

In the last step, we randomly select one covariate allocated in each non-empty cell. Then, the total $\hat{n}$ covariates $\{\boldsymbol{x}_i^*\}_{i=1}^{\hat{n}}$ are collected. If $n > \hat{n}$, we randomly add $n - \hat{n}$ covariates from the remaining $N - \hat{n}$ covariates $\{\boldsymbol{x}_i^*\}_{i=\hat{n}+1}^N$. If $n < \hat{n}$, we randomly delete $\hat{n} - n$ covariates from the collected $\hat{n}$ covariates $\{\boldsymbol{x}_i^*\}_{i=1}^{\hat{n}}$. Suppose that, without loss of generality, the selected covariates are denoted as $\{\boldsymbol{x}_i^*\}_{i=1}^n$.

Finally, the selected covariates $\{\boldsymbol{x}_i^*\}_{i=1}^n$ are used to construct the effective subspace $\mathcal{H}_S$. Then, the smoothing spline estimator $\tilde{\eta}$ of η is achieved through (3). We summarize the proposed method in Algorithm 1.

Algorithm 1 Basis selection method using a rotated lattice

1: Compute $l = n^{-1/p}(p+1)^{1/(2p)}\hat{V}^{1/p}$ with the initial $\hat{V} = 1$.
2: **for** $1 \leq i \leq N$ **do**
3: Determine to which cell $\boldsymbol{x}_i$ belongs.
4: **end for**
5: Obtain the number of non-empty cells, denoted as $\hat{n}$.
6: Adjust l according to a new estimated $\hat{V}$ and repeat the above process until the difference between $\hat{n}$ and n does not exceed $0.1n$.
7: Randomly select one covariate from each non-empty cell; then, collect $\hat{n}$ covariates $\{\boldsymbol{x}_i^*\}_{i=1}^{\hat{n}}$.
8: **if** $\hat{n} > n$ **then**
9: Randomly delete $\hat{n} - n$ covariates from the collected $\hat{n}$ covariates $\{\boldsymbol{x}_i^*\}_{i=1}^{\hat{n}}$.
10: **else**
11: Randomly add $n - \hat{n}$ covariates from the remaining $N - \hat{n}$ covariates $\{\boldsymbol{x}_i^*\}_{i=\hat{n}+1}^N$.
12: **end if**
13: Denote the selected sub-data as $\{\boldsymbol{x}_i^*\}_{i=1}^n$.
14: Minimize the penalized least squares criterion (2) over the effective subspace $\mathcal{H}_S = \mathcal{N}_J \oplus \mathrm{span}\{G_J(\boldsymbol{x}_i^*, \cdot), i = 1, 2, \ldots, n\}$.

The computational cost for Algorithm 1 mainly depends on dividing N covariates into n cells and estimating the smoothing spline estimator in the last step. Note that Algorithm 1 requires $O(np)$ operations to obtain the cell closest to a given covariate. Therefore, the computational complexity of finding the cells closest to N covariates is of the order $O(Nnp)$. As discussed in Meng et al. (2021 2020), estimating the smoothing spline estimator $\tilde{\eta}$ of η over the effective subspace $\mathcal{H}_S$ is of the order $O(Nn^2)$. Since n is much larger than p, $O(Nnp)$ is negligible compared with $O(Nn^2)$. In summary, the overall computational cost of the RBS method in Algorithm 1 is in the order of $O(Nn^2)$.

3.3 | Theoretical analysis

In this section, the asymptotic properties of the smoothing spline estimator using the proposed RBS method are presented. Let f_X denote the probability density function of the covariate. Define the mean square error of the estimator $\hat{\eta}$ in estimating η as the quadratic functional

$$\Lambda(\hat{\eta}-\eta)=\int_{\mathcal{X}}(\hat{\eta}(\boldsymbol{x})-\eta(\boldsymbol{x}))^2 f_X(\boldsymbol{x})d\boldsymbol{x}.$$

Conventionally, the rate of asymptotic convergence of $\hat{\eta}$ is described by eigenvalue analysis of $J(\eta)$ in (2) with respect to $\Lambda(\eta)$. In the following, we shall use $\Lambda(\cdot,\cdot)$ and $J(\cdot,\cdot)$ to represent the (semi) inner products associated with the square (semi) norms Λ and J.

Below, we introduce six primary regularity conditions, which are the same as in the HBS method. A more detailed discussion can be found in Meng et al. (2021).

Condition 1. The functional Λ is completely continuous with respect to J, where $\Lambda(\phi_\gamma,\phi_\mu)=\delta_{\gamma\mu}$, $J(\phi_\gamma,\phi_\mu)=\rho_\gamma\delta_{\gamma\mu}$, ϕ_γ and ϕ_μ are the eigenfunctions associated with Λ and J in $\mathcal{H}$, ρ_γ is the nonnegative eigenvalue associated with ϕ_γ, and $\delta_{\gamma\mu}$ denotes the Kronecker delta;

Condition 2. For some $\beta>0$ and $r>1$, $\rho_\gamma>\beta\gamma^r$ for sufficiently large γ;

Condition 3. For all μ, γ and a positive constant C, $\text{var}\{\phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})\}\le C$;

Condition 4. For all μ and γ, $\phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})$ is Lipschitz continuous; that is, for any $\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}$, there exists a positive constant B such that $|\phi_\gamma(\boldsymbol{x})\phi_\mu(\boldsymbol{x})-\phi_\gamma(\boldsymbol{x}')\phi_\mu(\boldsymbol{x}')|\le B\|\boldsymbol{x}-\boldsymbol{x}'\|$;

Condition 5. Suppose $\max\{(nN_i)/N\}_{i=1}^n=O_P(1)$, where N_i is the number of covariates in the i-th cell;

Condition 6. The number of bases n satisfy that $n^{1+2/p}=O(N)$.

The first four conditions are the regularity conditions for smoothing spline estimators even if the full bases are used. Condition 5 prevents the extreme case of the probability density function f_X, i.e., when there is only one non-empty cell; then, we have $\max\{(nN_j)/N\}_{j=1}^n=nN/N=n$, which conflicts with Condition 5. Lastly, Condition 6 naturally holds when n is a manageable size. For example, when $p=2$, Condition 6 holds for $n=O(N^{1/2})$.

Denote $\tilde{\eta}_R$ as the estimator of η obtained by the basis functions $\{G_J(\boldsymbol{x}_i^*,\cdot)\}_{i=1}^n$ selected using the RBS method. The following theorem shows that $\tilde{\eta}_R$ converges to the true function η_0 at the same rate as $\hat{\eta}$, whose proof is provided in the supplementary materials.

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 $\lambda\to0$ 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 $(\Lambda+\lambda J)(\tilde{\eta}_R-\eta_0)=O_P(N^{-dr/(dr+1)})$.

Denote $\tilde{\eta}_U$, $\tilde{\eta}_A$, $\tilde{\eta}_S$, and $\tilde{\eta}_H$ as the estimators of η obtained based on observations selected using the UBS, ABS, SBS, and HBS methods, respectively. Obviously, the convergence rates of $\tilde{\eta}_U$, $\tilde{\eta}_A$, $\tilde{\eta}_S$, and $\tilde{\eta}_H$ are identical to that of $\hat{\eta}$ with different essential choices of n. The convergence rates of $\tilde{\eta}_U$ and $\tilde{\eta}_A$ are the same as that of $\hat{\eta}$ when the size of the basis functions n is of the order $O(N^{2/(dr+1)})$. For the estimator $\tilde{\eta}_R$, n is of the order $O(N^{2p/(dr+1)(p+2)})$, which is smaller than those of $\tilde{\eta}_U$ and $\tilde{\eta}_A$. The order of n in estimating $\tilde{\eta}_S$ is $O(N^{1/(dr+1)})$, which is smaller than that for the proposed $\tilde{\eta}_R$. However, the estimator $\tilde{\eta}_S$ requires that the covariates be uniformly distributed in the whole covariate space. Once the covariates are not uniformly distributed, $\tilde{\eta}_S$ may not perform well, compared with the proposed $\tilde{\eta}_R$. The order of n in $\tilde{\eta}_H$ is the same as $\tilde{\eta}_R$ for the response curve in general. When the response curve has homogeneous smoothness across the covariate space, our method performs similarly to the HBS method. Additionally, the selected sub-data enjoy a pretty projection property in lower-dimensional subspaces through a well-designed rotation, instead of increasing the sub-data size. Thus, it has some advantages in cases where the response curves have non-homogeneous smoothness.

The values of d and r are determined by the roughness of η and the order of the fitted spline, respectively. More details can be found in Gu and Kim (2002), Ma et al. (2015), Meng et al. (2020), and Meng et al. (2021). In our numerical study, we set r to 4, as Ma et al. (2015) and Gu and Kim (2002) have suggested. Hence, we can obtain that n is roughly located in the interval $(O(N^{2p/\{9(p+2)\}}),O(N^{2p/\{5(p+2)\}}))$. In the following real data analysis, we take the dimension n of the effective subspace $\mathcal{H}_S$ to be between $15N^{2/9}$ and $30N^{2/9}$.

4 | SIMULATION AND REAL DATA STUDIES

In this section, we compare our proposed basis selection method to several other methods using five numerical examples and two real data examples in terms of prediction accuracy with the log mean absolute error (MAE) given by

$$\log(\text{MAE})=\log\left(N_{test}^{-1}\sum_{i=1}^{N_{test}}|\tilde{\eta}(\boldsymbol{x}_i)-\eta_0(\boldsymbol{x}_i)|\right),$$

where N_{test} is the size of the test data, $\{\boldsymbol{x}_i\}_{i=1}^{N_{test}}$ is an independent test dataset generated using the same probability density function as in the training data, and $\eta_0(\boldsymbol{x}_i)$ and $\tilde{\eta}(\boldsymbol{x}_i)$ denote the real and fitted response values from the i-th test data point, respectively. The methods in the comparison are the UBS method in Gu and Kim (2002), the ABS method in Ma et al. (2015), the SBS method in Meng et al. (2020), and the HBS method in Meng et al. (2021).

4.1 | Simulation studies

We generate a synthetic training dataset with $N = 2000$ from each of the following four probability density functions, and then, the covariate is scaled to $[0, 1]^p$. The signal-to-noise ratio, defined as $\text{var}(\eta(\boldsymbol{x}))/\sigma^2$, is set to two. We then use the five basis selection methods to obtain a subset of $n = \{40, 60, 80, 100, 120\}$ observations and fit the smoothing spline model.

We consider four distributions of the covariate variable $\boldsymbol{x}$ over the domain of interest, identical to those considered in Meng et al. (2021). The p in the following dataset are adapt to the functions f_1 to f_5.

d$_1$: The p-dimension covariate variable $\boldsymbol{x}$ is uniformly distributed in $[0, 1]^p$;

d$_2$: The p-dimension covariate variable $\boldsymbol{x}$ follows a mixture $t-$distribution $(T_1, \ldots, T_p)$, where $\{T_i\}_{i=1}^p$ are independently generated from $t(10, -5)/2 + t(10, 5)/2$;

d$_3$: The p-dimension covariate variable $\boldsymbol{x}$ obeys a banana-shaped distribution, which is generated by $(Z_1, Z_2 + Z_1^2/1.2, \ldots, Z_p + Z_1^2/1.2)$, where $(Z_1, Z_2, \ldots, Z_p)$ is generated from the standard multivariate Gaussian distribution;

d$_4$: The p-dimension covariate variable $\boldsymbol{x}$ obeys a multivariate Gaussian distribution $GP(0, \Sigma)$, where the (i, j)-th element of Σ is $\Sigma_{i,j} = 0.9^{|i-j|}$, $i, j = 1, \ldots, p$.

We adopt five different regression function settings. The first four are the same as those considered in Meng et al. (2021), and the last one is similar to that in Ma et al. (2015):

f$_1$: A two-dimension function given by

$$f_1(\boldsymbol{x}) = \sin\{10/(x_1 + x_2 + 0.15)\};$$

f$_2$: A two-dimension function given by

$$f_2(\boldsymbol{x}) = h_1(x_1, x_2) + h_2(x_1, x_2),$$

where

$$h_1(x_1, x_2) = \{0.75/(\pi\sigma_1\sigma_2)\} \times \exp\{-(x_1 - 0.2)^2/\sigma_1^2 - (x_2 - 0.3)^2/\sigma_2^2\},$$
$$h_2(x_1, x_2) = \{0.75/(\pi\sigma_1\sigma_2)\} \times \exp\{-(x_1 - 0.7)^2/\sigma_1^2 - (x_2 - 0.5)^2/\sigma_2^2\},$$

$\sigma_1 = 0.1$, and $\sigma_2 = 0.2$;

f$_3$: A three-dimension function given by

$$f_3(\boldsymbol{x}) = \sin\{\pi(x_1 + x_2 + x_3)/3\} - x_1 - x_2^2;$$

f$_4$: A four-dimension function given by

$$\begin{aligned} f_4(\boldsymbol{x}) =& x_1 + (2x_2 - 1)^2/2 + [\sin(10\pi x_3)/\{2 - \sin(10\pi x_3)\}]/3 + \{0.1\sin(2\pi x_4) \\ &+ 0.2\cos(4\pi x_4) + 0.3\sin(6\pi x_4)^2 + 0.4\cos(8\pi x_4)^3 + 0.5\sin(10\pi x_4)^3\}/4; \end{aligned}$$

f$_5$: A copula function with only two important covariates given by

$$f_5(\tilde{\boldsymbol{x}}) = (2\pi)^{-1}|\Sigma|^{-1/2}\exp\{-1/2(g_1(\tilde{x}_1), g_2(\tilde{x}_2))^T\Sigma^{-1}(g_1(\tilde{x}_1), g_2(\tilde{x}_2))\}|g_1'(\tilde{x}_1)g_2'(\tilde{x}_2)|,$$

where $g_1(\tilde{x}_1) = 2\text{sign}(\tilde{x}_1)|\tilde{x}_1|^2$, $g_2(\tilde{x}_2) = 3\tilde{x}_2^3$, $\tilde{x}_j = 2.3(x_j - 0.5)$ for $j = 1, 2$, and Σ is a 2×2 matrix, with the (i, j)-th entry being $0.5^{|i-j|}$, $i, j = 1, 2$.

Figure 3 predicts $N_{test} = 2000$ testing data, records the $\log(\text{MAE})$, and calculates the standard errors shown in vertical bars, which are obtained from a hundred replicates. The full bases estimator is omitted here due to the high computation cost. Figure 3 illustrates that all the methods in the first column have similar performances, while the proposed RBS method performs slightly better and the UBS method performs slightly worse, in which cases the observed values are uniformly distributed in a hypercube. Meanwhile, in the right three columns, we observe that the

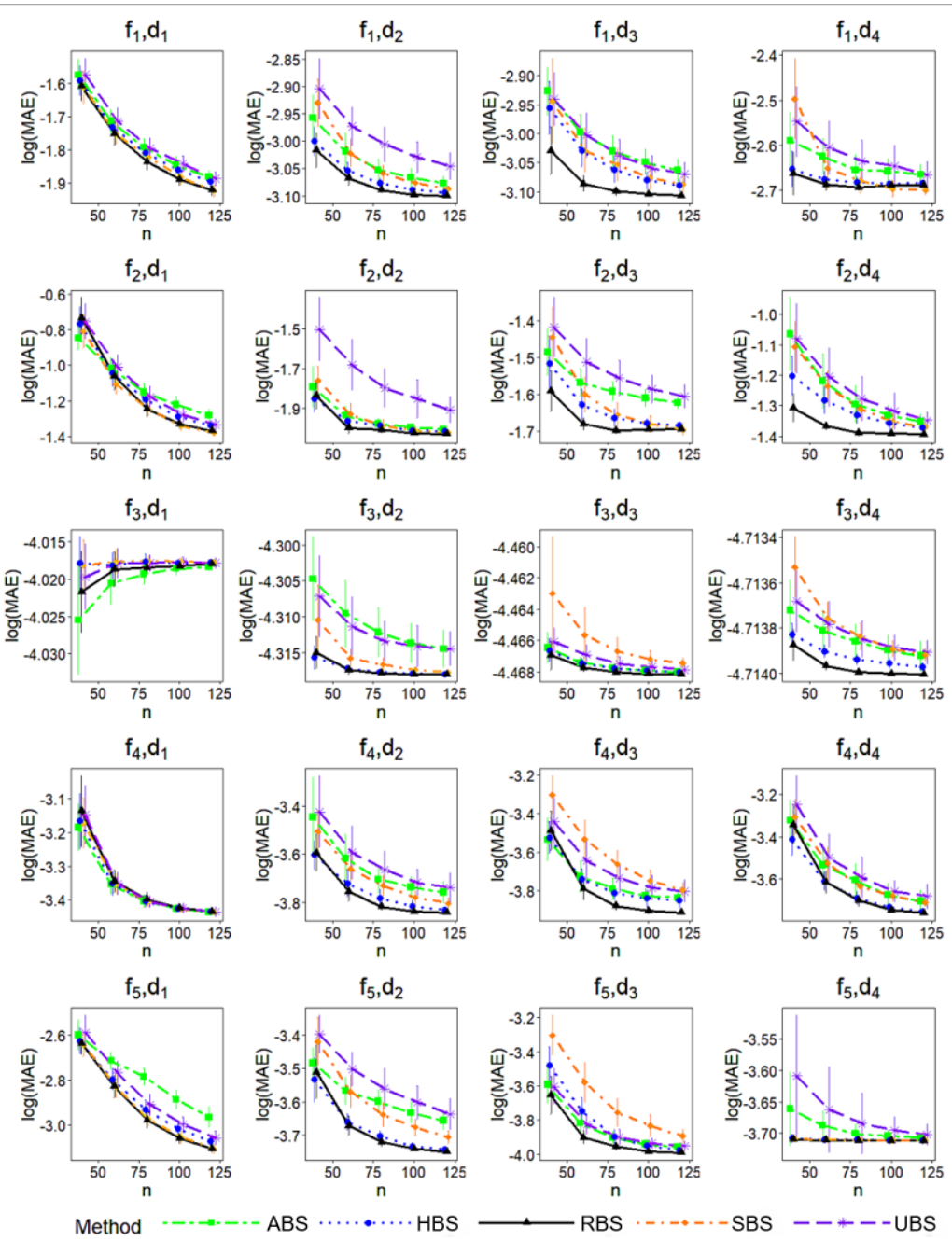

Figure 3 $\log(\text{MAE})$ values from the smoothing spline model for simulation under five different regression functions (from upper to lower) and four different probability density functions (from left to right) are plotted versus different n.

proposed RBS method yields perfect performance compared with the UBS, ABS, and SBS methods, followed by the HBS method, in cases where the covariates are not uniformly distributed in a hypercube. Moreover, the UBS, ABS, and SBS methods perform unstably in the nonuniformly distributed case. Overall, our RBS method has the upper hand in all settings, yielding lower $\log(\text{MAE})$ and smaller standard errors, which implies that the RBS method is feasible for a broad range of covariate distributions, surface functions, and dimensions and thus emerges as the best estimator of the true function, compared with other methods.

In the following, we consider a scenario in which the underlying response surface is not isomorphic to further illustrate the advantages of the proposed method. To be precise, we consider a smoothing spline on d_5 with p varying from three to eight. The underlying response surface is taken as f_5 which only relies on the first two dimensions of d_5. Note that there is at least one redundant covariate in this example which makes

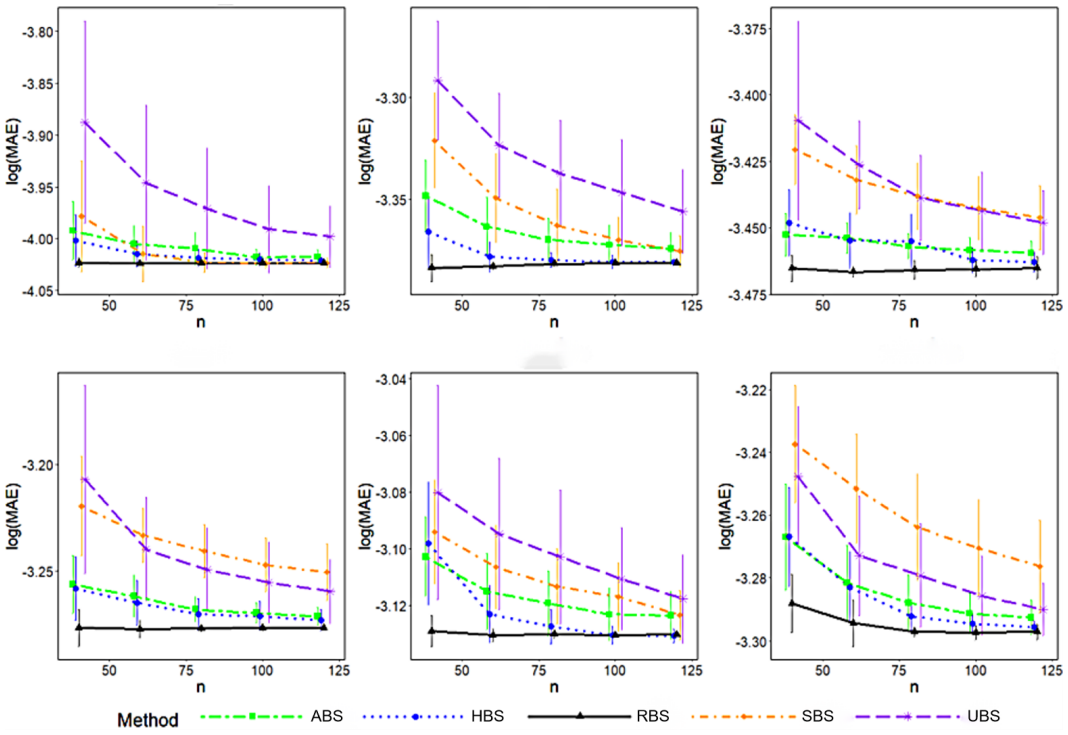

Figure 4 $\log(\text{MAE})$ values from the smoothing spline model for simulation under regression function f_5 and probability density function d_4 with 1, 2, 3, 4, 5, and 6 redundant covariates (from upper to lower and left to right) are plotted versus different n. The response surface is not isomorphic due to the existence of redundant covariates.

the response surface not isomorphic. Figure 4 reports the $\log(\text{MAE})$s of all five basis selection methods when redundant covariates exist in the settings of regression functions and the dataset is not uniformly generated in a hypercube.

The results show that the performance of the RBS method is remarkably better when the function has non-homogeneous smoothness and that the performance of the RBS method is comparable with that of the HBS method when the function is isomorphic. This is because when the response surface is not isomorphic, the uniformity on the subspace with relatively large volatility is more important than other subspaces. Compared with the uniformity in full data space, the projective uniformity is much more appreciated in this scenario since it achieves a better uniformity on all $k(< p)$ dimensional subspace than the uniform design in full space. Consequently, the proposed RBS method is more advantageous than the other four methods since it achieves better projective uniformity.

To clearly see the projective uniformity of the proposed method, we evaluate the proposed method together with the other four methods by the log maximum projection criterion proposed in Joseph, Gul, and Ba (2015). To be precise, the log maximum projection is calculated by

$$\log(\psi(D)) = \log\left(\frac{1}{n(n-1)}\sum_{1\le i<j\le n}\prod_{k=1}^{p}(x^*_{i,k}-x^*_{j,k})^{-2}\right)^{1/p},$$

with $D = \{\boldsymbol{x}^*_i\}_{i=1}^n$, which is the data set we use to construct the basis functions in $\mathcal{H}_S$. Clearly, a smaller value implies better projection uniformity since the data points in D will not be close to each other in any dimensions. For cases where $(x^*_{i,k} - x^*_{j,k}) = 0$ for some i, j, and k, we simply denote the results as NA, implying that projective uniformity is not available for the sub-data D. The projection uniformity as measured by $\log(\psi(D))$ for the five methods is shown in Table 1 under f_5 and d_4 with redundant covariates. Based on the result presented in Table 1, one can clearly see that the $\log\{\psi(D)\}$ value of RBS is the smallest among the other four methods, which implies the data points selected via our methods have a better projective uniformity on the subspace spanned by the first two covariates.

In the following, we will evaluate the performance of all five methods in computing time. Since all the cases have similar performance, we only illustrate the results that the covariates are generated as d_4 and the regression function is f_5 with n varying from 40 to 120. We record the computing time for our desktop PC with 8 G memory and a 2.4 GHz Celeron processor. Each method has been repeated 100 times and the corresponding results are reported in Table 2.

Table 1 $\log\{\psi(D)\}$ values of the five basis selection methods for simulation under regression function f_5 and probability density function d_4 with $n = 80$ basis functions. NA represents that $\log\{\psi(D)\}$ is not applicable to the sub-data D.

p	UBS	ABS	SBS	HBS	RBS
3	9.551428	9.501454	NA	NA	**8.357445**
4	9.108542	8.869277	NA	NA	**7.794891**
5	8.710636	8.749605	NA	NA	**7.684428**
6	8.642750	8.738538	NA	8.865160	**7.744047**
7	8.358219	8.379835	NA	8.198035	**7.439442**
8	8.105443	7.993638	NA	7.989379	**7.290941**

Table 2 Computing time (in seconds) of the five basis selection methods and full data for simulation under regression function f_5 and probability density function d_4 with $n = \{40, 60, 80, 100, 120\}$ basis functions.

n	UBS	ABS	SBS	HBS	RBS	full data
40	1.2402	1.2221	1.2063	1.2610	1.5928	
60	1.5383	1.5937	1.3282	1.6333	1.9558	
80	1.7584	1.7986	1.5685	1.8367	2.2261	37.7657
100	1.9162	1.9991	1.7520	2.0945	2.5139	
120	2.0824	2.1675	1.9644	2.3453	2.7677	

From Table 2, it can be clearly seen that all methods significantly reduce the computing time compared with the full data approach since the computational results are reduced from $O(N^3)$ to $O(Nn^2)$. As discussed in Section 3.3, all the basis selection methods have comparable performance since fitting the smoothing spline ANOVA model takes the most computing time.

4.2 | White river pollutant data

This dataset contains the pollutant concentrations at different locations in China's White River. The main task is to predict the pollutant concentration of unmonitored locations. The variable pollutant concentration is treated as the response, and the corresponding longitude and latitude are treated as two covariates. The dataset contains $N = 3,965$ training data and $N_{test} = 396$ test data. We select $n = N^{2/9}\{15, 20, 25, 30\}$ basis functions from the $N = 3,965$ training data using the UBS, ABS, SBS, HBS, and RBS methods. We fit the cubic tensor product smoothing spline analysis of the variance model to the dataset and consider the following model settings:

M_1: An additive smoothing spline regression model with all two main effects and the two-way interaction function of corresponding covariates,

$$y_i = \eta_\emptyset + \eta_1(x_{i1}) + \eta_2(x_{i2}) + \eta_{1,2}(x_{i1}, x_{i2}) + \epsilon_i,\ i = 1, 2, \ldots, N;$$

M_2: A redundant covariate x_{i3}, for example, the depth of the river, independently generated through a uniform distribution on $[0, 1]$, is added to the respected model. To be precise, an additive smoothing spline regression model with all three main effects and the two-way interaction of x_{i1} and x_{i2},

$$y_i = \eta_\emptyset + \eta_1(x_{i1}) + \eta_2(x_{i2}) + \eta_3(x_{i3}) + \eta_{1,2}(x_{i1}, x_{i2}) + \epsilon_i,\ i = 1, 2, \ldots, N,$$

where y_i is the contaminant concentration measurement at the i-th observation, x_{ij} is the value of the j-th dimension at the i-th covariate, $\eta_\emptyset$ is a constant function, $\{\eta_j\}_{j=1}^3$ are main effect functions, $\eta_{1,2}$ is the two-way interaction function of the corresponding covariate, and $\{\epsilon_i\}_{i=1}^N$ represent the independent and identically distributed random errors with zero mean and unknown variance σ^2. It is worth mentioning that the $R^2 = 0.7543$ which implies the fitness of M_1. By applying model checking tools in (Gu 2013, Chapter 3.7), one can find a strong correlation between all three terms with the response. To show the effectiveness of the proposed estimator, we compare it with four mainstream competitors, as mentioned in the previous section, in terms of the prediction $\log(\text{MAE})$, calculated on a holdout testing set, and summarize the results in Figure 5.

As shown in the left panel of Figure 5, the $\log(\text{MAE})$s of all five basis selection methods come closer with the increasing basis function size n. As expected, in models M_1 and M_2, RBS is superior to the other four basis selection methods, which echoes the results in Section 4.1.

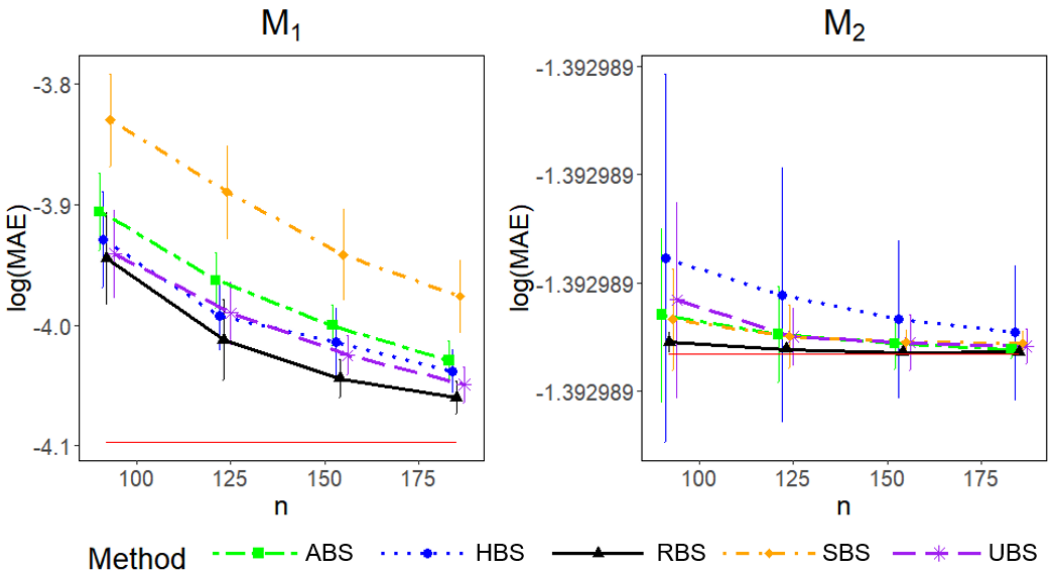

Figure 5 $\log(\text{MAE})$ values versus different n under model setting M_1 and M_2 from the smoothing spline model for the pollutant data. The horizontal line represents the performance for all the training data.

4.3 | Debutanizer column data

This dataset comprises debutanizer distillation column data from the process of separating butane from gasoline. The task is to predict the butane content by using the conditions of the debutanizer columns and other relevant information measured using soft sensors in the process of petroleum refining. Seven relevant features are provided in the dataset, such as temperature, pressure, flow, and so forth. More details about the dataset can be found in Fortuna, Graziani, Rizzo, and Maria (2007).

We first scale the dataset to $[0,1]^7$, then fit the cubic tensor product smoothing spline analysis of the variance model to the scaled dataset. We consider the following model setting:

M_3: An additive smoothing spline regression model with all main effects,

$$y_i = \eta_\emptyset + \sum_{j=1}^{7} \eta_j(x_{ij}) + \epsilon_i,\ i = 1, 2, \ldots, N,$$

where y_i is the butane concentration of the i-th observation, x_{ij} is the value of the j-th dimension of the i-th covariate, $\eta_\emptyset$ is a constant function, $\{\eta_j\}_{j=1}^7$ are main effect functions, and $\{\epsilon_i\}_{i=1}^N$ represent the independent and identically distributed random errors with zero mean and unknown variance. It is worth mentioning that the cosine values of all possible interaction terms are no more than 0.29. Thus, according to Gu (2013), we opt to consider this smoothing spline ANOVA model with main effects only for model brevity.

The simulation is performed under the setting that the size of the full data is $2,395$, 10% of the full data is randomly selected without repetition as test data, the remaining data is used as training data, and set $n = N^{2/9}\{15, 20, 25, 30\}$.

As shown in Figure 6, with the increase of the size n of the basis functions, the $\log(\text{MAE})$s of all the five basis selection methods come closer to one another. Furthermore, as expected, in Figure 6, RBS is superior to the other four basis selection methods, which is identical to the results in Section 4.1 and Section 4.2.

5 | CONCLUSION AND DISCUSSION

This study introduces a new basis selection method called RBS for large N datasets. The RBS method is an efficient stratified basis selection method in which the strata are Voronoi cells of rotated thinnest covering lattices, which does not require the assumption that the covariates are uniformly distributed on a hypercube and that there is homogeneous smoothness across the covariate space.

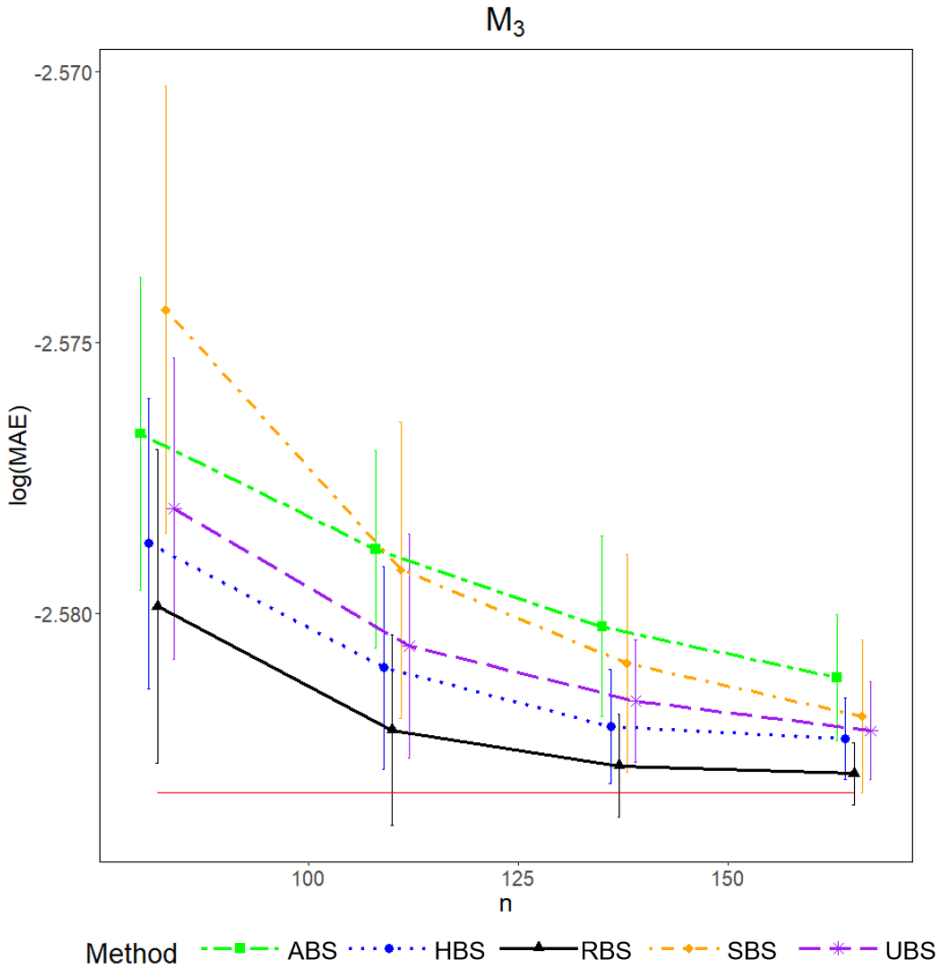

Figure 6 $\log(\text{MAE})$ values versus different n under model setting M_3 from the smoothing spline model for the debutanizer distillation column data. The horizontal line represents the performance for all the training data.

Theoretically, we prove that for the proposed RBS method, when the size of the basis functions n is roughly in the order $O\{N^{2p/(dr+1)(p+2)}\}$, the convergence rate of $\tilde{\eta}_R$ is identical to that of $\hat{\eta}$, which is the same as that of the HBS method, and is also faster than that of the UBS and ABS methods. Furthermore, comprehensive simulation and real data studies were conducted to illustrate the performance of the RBS method, and the numerical results are consistent with our theoretical results. Compared with the other four existing methods, the RBS method is more effective, especially when the covariates are not uniformly distributed on a hypercube and the response is not isotropic since the RBS method improve the projective uniformity substantially via rotating the lattice.

The newly proposed basis selection method may be suitable for other nonparametric models, including but not limited to the Gaussian process model and the nearest neighbor model. However, more work needs to be done to verify its performance. It would also be interesting to explore whether the new method can be modified by considering the response information.

ACKNOWLEDGMENTS

The authors sincerely thank the co-editor, associate editor, and two referees for their valuable comments and insightful suggestions, which led to significant improvement of this article. Tian's work was supported by NSFC grants 12131001; Yu's work was supported by Beijing Municipal Natural Science Foundation No. 1232019, NSFC grants 12001042, and Beijing Institute of Technology Research Fund Program for Young Scholars.

SUPPORTING INFORMATION

The online supplementary materials include detailed proofs of all results, additional simulation results, and codes for the numerical experiments.

References

Conway, J. H., & Sloane, N. J. A. (1998). *Sphere packings, lattices and groups*. New York: Springer.

Fortuna, L., Graziani, S., Rizzo, A., & Maria, X., Gabriella. (2007). *Soft sensors for monitoring and control of industrial processes*. Springer Science and Business Media.

Grace, W. (1990). *Spline models for observational data*. SIAM.

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

Gu, C., & Kim, Y.-J. (2002). Penalized likelihood regression: general formulation and efficient approximation. *Canadian Journal of Statistics*, *30*(4), 619-628.

Hastie, T. (1996). Pseudosplines. *Journal of the Royal Statistical Society. Series B (Methodological)*, 379-396.

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

He, X. (2020). Lattice-based designs with quasi-optimal separation distance on all projections. *Biometrika*, *108*(2), 443-454.

He, X., Shen, L., & Shen, Z. (2001). A data-adaptive knot selection scheme for fitting splines. *IEEE Signal Processing Letters*, *8*(5), 137-139.

Joseph, V. R., Gul, E., & Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, *102*(2), 371-380.

Luo, Z., & Wahba, G. (1997). Hybrid adaptive splines. *Journal of the American Statistical Association*, *92*(437), 107-116.

Ma, P., Huang, J. Z., & Zhang, N. (2015). Efficient computation of smoothing splines via adaptive basis sampling. *Biometrika*, *102*(3), 631–645.

Meng, C., Yu, J., Chen, Y., Zhong, W., & Ma, P. (2021). Smoothing splines approximation using hilbert curve basis selection. *arXiv preprint arXiv:2109.11727*.

Meng, C., Zhang, X., Zhang, J., Zhong, W., & Ma, P. (2020). More efficient approximation of smoothing splines via space-filling basis selection. *Biometrika*, *107*(3), 723-735.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics*, *11*(4), 735-757.

Sklar, C., Jeffrey, Wu, J., Meiring, W., & Wang, Y. (2013). Nonparametric regression with basis selection from multiple libraries. *Technometrics*, *55*(2), 189-201.

Xu, D., & Wang, Y. (2018). Divide and recombine approaches for fitting smoothing spline models with large datasets. *Journal of computational and graphical statistics*, *27*(3), 677–683.

Yuan, Y., Chen, N., & Zhou, S. (2013). Adaptive B-spline knot selection using multi-resolution basis set. *Iie Transactions*, *45*(12), 1263–1277.

Zhang, H., Hao, Wahba, G., Lin, Y., Voelker, M., Ferris, M., Klein, R., & Klein, B. (2004). Variable selection and model building via likelihood basis pursuit. *Journal of the American Statistical Association*, *99*(467), 659–672.

Zhang, Y., John, D., & Martin, W. (2015). Divide and conquer kernel ridge regression: A distributed algorithm with minimax optimal rates. *The Journal of Machine Learning Research*, *16*(1), 3299–3340.