Nonparametric Regression for 3D Point Cloud Learning

Xinyi Li   LIXINYI@CLEMSON.EDU
School of Mathematical and Statistical Sciences, Clemson University
Clemson, SC 29634, USA

Shan Yu   SY5JX@VIRGINIA.EDU
Department of Statistics, University of Virginia
Charlottesville, VA 22904, USA

Yueying Wang[1]   YUEYINGW@AMAZON.COM
Amazon.com, Inc.
Seattle, WA 98121, USA

Guannan Wang   GWANG01@WM.EDU
Department of Mathematics, College of William & Mary
Williamsburg, VA 23185, USA

Li Wang   LWANG41@GMU.EDU
Department of Statistics, George Mason University
Fairfax, VA 22030, USA

Ming-Jun Lai   MJLAI@UGA.EDU
Department of Mathematics, University of Georgia
Athens, GA 30602, USA

for the Alzheimer's Disease Neuroimaging Initiative[2]

Editor: Xiaotong Shen

Abstract

In recent years, there has been an exponentially increased amount of point clouds collected with irregular shapes in various areas. Motivated by the importance of solid modeling for point clouds, we develop a novel and efficient smoothing tool based on multivariate splines over the triangulation to extract the underlying signal and build up a 3D solid model from the point cloud. The proposed method can denoise or deblur the point cloud effectively, provide a multi-resolution reconstruction of the actual signal, and handle sparse and irregularly distributed point clouds to recover the underlying trajectory. In addition, our method provides a natural way of numerosity data reduction. We establish the theoretical guarantees of the proposed method, including the convergence rate and asymptotic normality of the estimator, and show that the convergence rate achieves optimal nonparametric convergence. We also introduce a bootstrap method to quantify the uncertainty of the estimators. Through extensive simulation studies and a real data example, we demonstrate

---

the superiority of the proposed method over traditional smoothing methods in terms of estimation accuracy and efficiency of data reduction.

Keywords: 3D pattern recognition, complex domain, penalized splines, triangulation, trivariate splines

1 Introduction

Recent advances in computer and information technology have dramatically boosted the availability of three-dimensional (3D) point clouds in many fields, such as geography, environmental science, computer graphics, engineering, economics, and medical imaging. These point clouds are usually collections of enormous measurements in space defined by a given coordinates system. Sometimes, these point clouds are rendered and inspected directly; examples can be found in Rusinkiewicz and Levoy (2000). More often, the point clouds are converted to polygon mesh or triangle mesh models; for example, 3D reconstruction is widely used for automated driving vehicles (Lee et al., 2016; Shao, 2021). Many efforts have been devoted to investigating the converting technique, and a survey of surface reconstruction from point clouds can be found in Berger et al. (2017).

These 3D data typically contain more information than the shape of an object, as they represent some meaningful values in real life. For example, healthcare professionals can use brain activity levels contained in 3D neuroimages to access new angles, resolutions, and details to better understand tissue. Another example is using cloud-based 3D scans to explore the location and access the grade and type of ore in the mining industry. Regardless of the potential of these applications, limited research has been developed on identifying underlying signals from 3D solid objects, as there are many challenges in achieving this goal.

First of all, as shown in the point cloud inputs in Figure 1, the shape of the point clouds is usually irregular. To extract the underlying signal over the irregular-shaped 3D solid object, many conventional methods, such as kernel smoothing (Zhu et al., 2014), tensor product smoothing (Reiss and Ogden, 2010), thin plate spline smoothing (Duchon, 1976; Yue and Speckman, 2010), and wavelet smoothing (Morris and Carroll, 2006), suffer from the problem of "leakage", which is referred as the inaccurate signal reconstruction across the complex domains; see the discussions in Ramsay (2002), Wang and Ranalli (2007), and Kim and Wang (2021) for more details. For example, as shown in the pillar of the horseshoe domain in Figure 2, although the Euclidean distance between points A and B is very short, erroneously borrowing information from point B when estimating point A will lead to unsatisfactory results. Several techniques have been developed in the last two decades to tackle the "leakage" problem for two-dimensional (2D) smoothing, such as spline smoothing (Awanou et al., 2006; Wang and Ranalli, 2007; Lai and Wang, 2013; Wilhelm et al., 2016), finite element analysis (Lindgren et al., 2011; Arnone et al., 2022), and kernel smoothing (Guo et al., 2010). For 3D images, Chung et al. (2018) applied the discrete heat kernel smoothing, and for irregular-shaped 3D objects, Huo et al. (2020) considered the finite element method.

Second, the information comprised in the 3D point clouds usually contains unwanted noise that can obscure the features of the underlying signal of interest. Consequently, denoising is a crucial step in the modeling process. Various smoothing methods have been
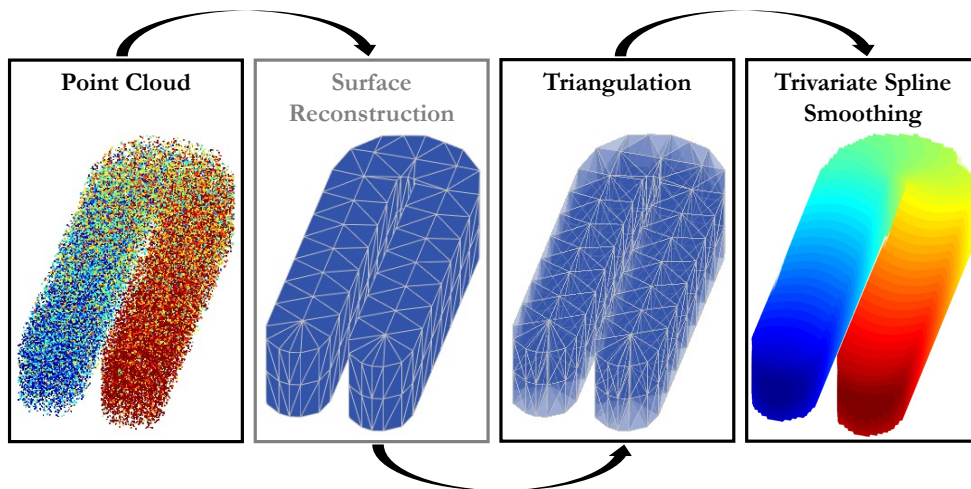
Figure 1: A flowchart demonstrating the overall procedure of TPST method analyzing a 3D point cloud.

adopted for denoising in many 2D studies (Goldsmith et al., 2014; Kang, 2020). However, the unique characteristics of 3D point clouds present significant challenges in denoising along with analysis (Wong et al., 2016).

Third, due to the nature of the devices for generating point cloud data, raw point cloud data is often sparse, uneven, or even partially missing. Therefore, accurately extracting the underlying signal from these point clouds becomes incredibly challenging. Consequently, it is necessary to preprocess the raw data to generate complete, dense, and uniform point cloud data.

Last but not least, as aforementioned, the size of the point clouds is usually enormous, and thus, makes it challenging in both storage and the corresponding data analysis. For example, a 3D image produced from a standard positron emission tomography (PET) brain scan contains over half a million voxels, which leaves data processing and analysis computing intensive tasks.
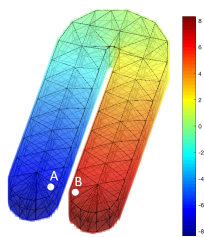

Figure 2: A illustration of "leakage" problem.

To address the prevalent challenges in converting a 3D irregular-shaped point cloud to a solid model and accurately estimating the underlying function, we propose a novel and efficient smoothing method based on trivariate penalized spline over triangulations (TPST) with a directional derivative-based penalty function. Figure 1 demonstrates the flowchart of the overall procedure of the TPST method for analyzing a 3D point cloud. The proposed TPST smoothing method has several appealing properties for handling data collected from a point cloud.

One of the main advantages is its ability to effectively solve the problem of "leakage" across complex domains. By borrowing the information from neighboring tetrahedra, the TPST method is able to accurately denoise or deblur the collected data while preserving inherent geometric features or spatial structures. This makes it particularly useful for analyzing irregularly shaped 3D point clouds. Similar to P-splines of Eilers and Marx (1996), TPST uses compactly supported basis functions and a sparse penalty, which is computationally efficient for smoothing unevenly distributed data, such as sparse and irregular 3D point cloud data. By using an appropriate partition, the method can reconstruct or recover the underlying function from both global and local missing data, and provide a multi-resolution reconstruction of the actual signal.

In addition, the proposed TPST has the potential to be used as a numerosity reduction method (Han et al., 2011). With a smooth underlying signal, the smaller set of spline coefficients could well represent and thus significantly reduce the amount of original data. Once the spline basis functions are determined, we can use the coefficients to recover the underlying signal of the point clouds. For example, as demonstrated in Section 6, a point cloud of size 50,000 can be represented by a vector of 323 spline coefficients. Let's take the biomedical image for another example, where the proposed method can reduce an image point cloud of 510,340 voxels to a vector of 4,856 coefficients, with a peak signal-to-noise ratio reaching 30.76. See Table 2 in a later section for details. This can greatly alleviate the storage and computational challenges associated with large point clouds.

Furthermore, the proposed TPST is computationally efficient in dealing with point clouds of large size. As shown in Section 3, it is very easy to set up the spline basis functions and penalties of TPST. Similar to many other types of spline smoothing (Wang and Yang, 2009; Ma, 2012; Wang et al., 2020a,b; Liu and Zhao, 2021), TPST is a global estimation method with an explicit model expression, and consequently, it only requires solving a single linear system to obtain the estimate. Thus, TPST could reduce the computational complexity and make it suitable for handling large point clouds.

In this article, we also investigate the theoretical properties of the proposed smoothing method. To be more specific, we establish the convergence rate of the proposed TPST estimator, which is determined by the fineness of the triangulation, the trivariate spline degree and smoothness, penalty parameter, and smoothness of the unknown function. When the penalty parameter is zero, the rate achieves the optimal nonparametric convergence rate. We further derive the asymptotic normality for the TPST estimator.

When working with point clouds, uncertainty quantification is essential after the estimate, as it provides an assessment of the reliability and robustness of the estimators, based on which results one could make more accurate conclusions. The aforementioned asymptotic normality of the proposed TPST estimator could thus be used to establish pointwise confidence intervals in theory. In practice, however, it can be difficult to derive the exact

form of the standard error due to the characteristics of the trivariate spline basis functions. As an alternative, we propose using a wild bootstrap method to estimate the standard error. This method is particularly useful for 3D point clouds that may have varying levels of noise or uncertainty as it can handle the heterogeneity effectively in data. Our simulation results show that the bootstrap standard error is very close to the true standard errors, providing a reliable method for quantifying the uncertainty of the estimators.

The rest of the paper is structured as follows. In Section 2, we give an overview of the triangulations and the methods for constructing the triangulations from a 3D point cloud. Once a triangulation is obtained, we can construct the spline space over this triangulation. We introduce the penalized spline estimators in Section 3. In Section 4, we present the theoretical results. In Section 5, we discuss implementation details for the proposed TPST method, including the creation of the penalty matrix, selection of the penalty parameter, choice of triangulation, and evaluations of uncertainties. We then demonstrate the performance of the proposed smoothers on several simulation examples and a real data example in Section 6. Finally, this paper is concluded, and future work is outlined in Section 7. We put all the theoretical details and more detailed introductions in the supplementary material.

2 Trivariate Splines over Triangulations and Basic Properties

In this section, we provide a basic framework for triangulations and trivariate splines on those triangulations. Please refer to Appendix A for more detailed introductions.

2.1 Triangulations

We assume there exists a tight polyhedral domain, Ω, that contains all the data locations of a point cloud. In the theoretical development below, we consider the Ω to be fixed and given. Additionally, we take into account that the domain may have one or multiple holes that do not contain any observations, as seen in the 3D domain of Figure 3 (b). Triangulations are very popular in approximating the domain Ω. In the following, we use T to denote a tetrahedron, that is, a convex hull of four noncoplanar points in $\mathbb{R}^3$. Then, a collection $\triangle = \{T_1, \ldots, T_N\}$ of N tetrahedra is called a *triangulation* of $\Omega = \cup_{h=1}^{N} T_h$, provided that any pair of tetrahedra in $\triangle$ intersect at most at a common vertex, along a common edge, or along a common triangular face. An example of triangulation $\triangle = \{T_1 = \langle \boldsymbol{v}_5, \boldsymbol{v}_1, \boldsymbol{v}_4, \boldsymbol{v}_3 \rangle, T_2 = \langle \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle\}$ of the domain $\Omega = \langle \boldsymbol{v}_2, \boldsymbol{v}_5, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$ is illustrated in Figure 4 (a), where $\boldsymbol{v}_\imath$, $\imath = 1, \ldots, 5$, are the vertices of the tetrahedra. In contrast, as illustrated in Figure 4 (b) for tetrahedra T_1, $T_3 = \langle \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_6 \rangle$ and $T_4 = \langle \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{v}_6, \boldsymbol{v}_4 \rangle$,

| (a) | (b) | (c) | (d) | (e) |
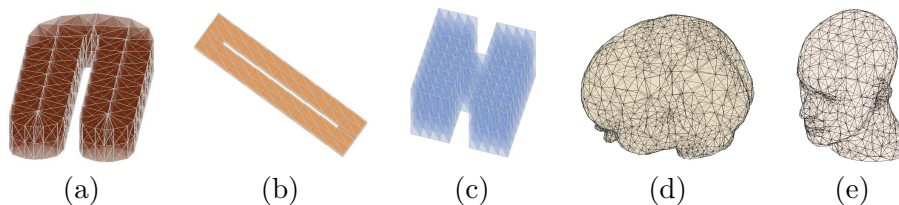
Figure 3: Examples of 3D point clouds and corresponding triangulation.

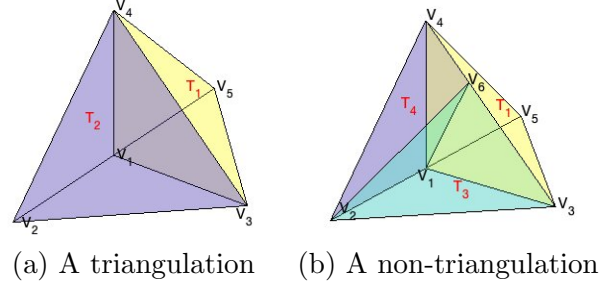(a) A triangulation   (b) A non-triangulation

Figure 4: (a) and (b) provide an example and a counterexample of triangulation.

$\{T_1, T_3, T_4\}$ does not form a triangulation of Ω, because of the intersections between the pairs of tetrahedra (T_1, T_3) and (T_1, T_4).

Furthermore, let $|T|$ be the length of the longest edge of T, and ϱ_T be the radius of the largest ball that can be inscribed in T, then the ratio $\beta_T := |T|/\varrho_T$ is the *shape parameter* of T; see an illustration example in Figure 5 (a). In general, the shape parameter β_T describes the shape of T, and the larger β_T is, the flatter the tetrahedron T becomes. If T is a regular tetrahedron whose six edges are all of the same lengths, then $\beta_T = 2\sqrt{6}$, and for any other tetrahedron, $\beta_T > 2\sqrt{6}$. Based on this shape parameter, we can define that a triangulation $\triangle$ is β-*quasi-uniform* if there is a positive value β such that $\triangle$ satisfies

$$|\triangle|/\varrho_T \leq \beta < \infty, \quad \text{for all } T \in \triangle, \tag{1}$$

where $|\triangle| := max\{|T|, T \in \triangle\}$ is referred as the *size of* $\triangle$, that is, the length of the longest edge of $\triangle$. Let N be the number of the tetrahedra in the polygonal domain Ω. From Equation 1, we have for a β-quasi-uniform partition, $N \leq (4\pi|\triangle|^3)^{-1}3V_\Omega\beta^3$, where V_Ω denotes the volume of Ω.

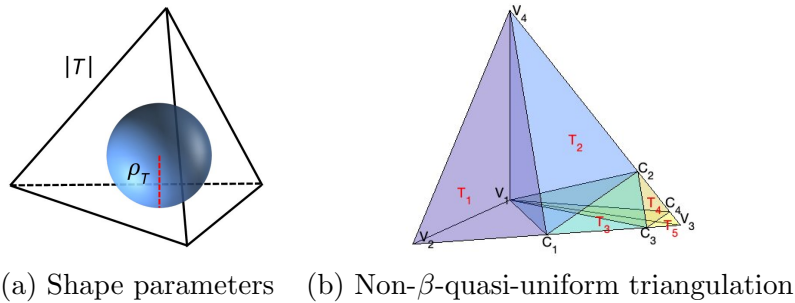(a) Shape parameters   (b) Non-β-quasi-uniform triangulation

Figure 5: (a) shows an illustration of the shape parameters of a tetrahedron; (b) shows an example of a non-β-quasi-uniform triangulation.

To illustrate the definition of quasi-uniform partition, we show an example of non-β-quasi-uniform triangulation in Figure 5 (b),

$$\triangle = \{T_1 = \langle \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{c}_1, \boldsymbol{v}_4 \rangle, T_2 = \langle \boldsymbol{v}_1, \boldsymbol{c}_2, \boldsymbol{c}_1, \boldsymbol{v}_4 \rangle, T_3 = \langle \boldsymbol{c}_1, \boldsymbol{v}_1, \boldsymbol{c}_3, \boldsymbol{c}_2 \rangle,$$
$$T_4 = \langle \boldsymbol{c}_3, \boldsymbol{v}_1, \boldsymbol{c}_4, \boldsymbol{c}_2 \rangle, T_5 = \langle \boldsymbol{c}_3, \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{c}_4 \rangle, \ldots \},$$

where the edges $\langle \boldsymbol{v}_4, \boldsymbol{c}_1 \rangle$ and $\langle \boldsymbol{c}_2, \boldsymbol{c}_3 \rangle$ are perpendicular to the edge $\langle \boldsymbol{v}_2, \boldsymbol{v}_3 \rangle$, and $\langle \boldsymbol{c}_1, \boldsymbol{c}_2 \rangle$ and $\langle \boldsymbol{c}_3, \boldsymbol{c}_4 \rangle$ are perpendicular to $\langle \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$.

2.2 Point Clouds to Triangulations

We now describe how to construct a triangulation from a 3D point cloud. The first step is to determine the domain of the point cloud, which is often given as a polygonal domain Ω, a geometric region defined by a set of polygons. This can be used to represent a wide range of objects, such as buildings, terrain, and objects in medical imaging. If the domain of the point cloud is not known, we can use surface reconstruction techniques, such as triangular meshes, to find a tight polygonal domain that encases the point cloud. As Ω has piecewise linear boundary faces, we have a normal direction perpendicular to each planar face of Ω, which can be used to determine whether a point is inside or outside of Ω. Then, we use this normal direction information of all boundary faces and the observed data locations to partition Ω into a collection of tetrahedra. Many mesh generation algorithms for 3D domains available in various software packages and toolboxes, such as the MATLAB functions "delaunay" and "distmesh/distmeshnd" (Persson and Strang, 2004), the C++ library "CGAL" (Jamin et al., 2015; Project, 2020), the "TetGen" (Si, 2015), and "iso2mesh" (Fang and Boas, 2009) built on "CGAL". In this article, we consider a triangulation, $\triangle$, with vertices containing the partial or whole set of observed data locations to achieve interpolation at the given data locations. This type of triangulation can be constructed using the constrained Delaunay triangulation method, as described in Shewchuk (1998). Based on a similar method proposed in Xu (2019), we obtain all the triangulations using MATLAB.
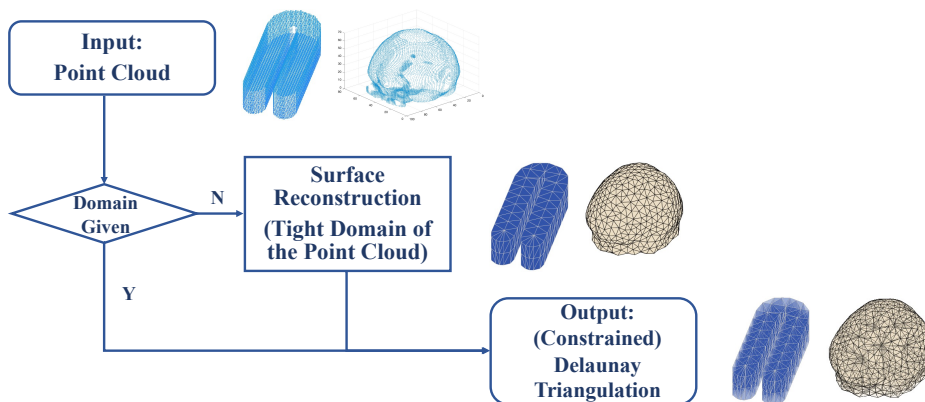

Figure 6: Flowchart demonstrating the process of constructing a triangulation from a 3D point cloud.

Figure 6 illustrates our flowchart for obtaining a triangulation. This generation procedure can be applied to many other 3D point clouds collected in various fields, as shown in the triangulation examples in Figure 3.

2.3 Trivariate Splines on a Triangulation

Suppose we have obtained a triangulation $\triangle$ for a tight domain Ω of a given point cloud. For any tetrahedron $T = \langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle \in \triangle$, any point $\boldsymbol{p} = (x, y, z) \in \mathbb{R}^3$ has a unique representation in terms of $\langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$,

$$\boldsymbol{p} = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + b_3\boldsymbol{v}_3 + b_4\boldsymbol{v}_4, \text{ with } b_1 + b_2 + b_3 + b_4 = 1,$$

where (b_1, b_2, b_3, b_4) are called the *barycentric coordinates* of $\boldsymbol{p}$ *relative to the tetrahedron* T, and they are nonnegative if $\boldsymbol{p}$ is inside or on the faces of T. Accordingly, for some nonnegative integers i, j, k, l with $i + j + k + l = d$, define *trivariate Bernstein basis polynomial of degree d relative to T* as

$$B_{ijkl}^{d,T}(\boldsymbol{p}) := \frac{d!}{i!j!k!l!}b_1^i b_2^j b_3^k b_4^l, \text{ with } i + j + k + l = d. \tag{2}$$

For any positive integer d, let $\mathcal{P}_d$ be the space of all trivariate polynomials with total degrees less than or equal to d. Note that the dimension of $\mathcal{P}_d$ is $\binom{d+3}{3}$. According to Theorem 15.8 in Lai and Schumaker (2007) and Lemma A.6 in the Appendix A, the set of Bernstein basis polynomials in Equation 2 forms a basis for the space of polynomials $\mathcal{P}_d$. Thus, any polynomial $\phi(\boldsymbol{p}) \in \mathcal{P}_d$ on with domain T can be written uniquely as the *B-form*,

$$\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{T;ijkl} B_{ijkl}^{d,T}(\boldsymbol{p}) = \mathbf{B}_T^d(\boldsymbol{p})^\top \boldsymbol{\gamma}_T, \tag{3}$$

where the coefficients $\boldsymbol{\gamma}_T = \{\gamma_{T;ijkl}\}_{i+j+k+l=d}$ are called *B-coefficients* of $\phi(\cdot)$.

For any nonnegative integer r, we use $\mathcal{C}^r(\Omega)$ to denote the collection of all r-th continuously differentiable functions over Ω. For triangulation $\triangle = \{T_1, \ldots, T_N\}$, let $\mathcal{S}_d^r(\triangle) = \{s \in \mathcal{C}^r(\Omega) : s|_T \in \mathcal{P}_d(T), T \in \triangle\}$ be a spline space of degree d and smoothness r over $\triangle$, where $s|_T$ is the polynomial piece of spline s restricted on tetrahedron T. According to Equation 3, for any $s \in \mathcal{S}_d^r(\triangle)$, there exists a coefficient vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{T_1}^\top, \ldots, \boldsymbol{\gamma}_{T_N}^\top)^\top$ with $\boldsymbol{\gamma}_{T_J} = \{\gamma_{T_J;ijkl}\}_{i+j+k+l=d}$ such that

$$s|_{T_J}(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{T_J;ijkl} B_{ijkl}^{d,T_J}(\boldsymbol{p}) = \mathbf{B}_{T_J}^d(\boldsymbol{p})^\top \boldsymbol{\gamma}_{T_J}, \ J = 1, \ldots, N. \tag{4}$$

That is, $s(\boldsymbol{p}) = \mathbf{B}_d(\boldsymbol{p})^\top \boldsymbol{\gamma}$, where $\mathbf{B}_d = \{(\mathbf{B}_{T_1}^d)^\top, \ldots, (\mathbf{B}_{T_N}^d)^\top\}^\top$.

For noise-free data, the approximation order of trivariate spline spaces was studied in Lai and Schumaker (2007) when $d \geq 8r + 1$ and Lai (1989) when $d \geq 6r + 3$. Particularly, Lai (1989) proved that when $d \geq 6r + 3$, the space $\mathcal{S}_d^r(\triangle)$ can attain the optimal converge rate; see Lemma 1 below. For any index $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ of order $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + \alpha_3$, we denote the derivatives $D^{\boldsymbol{\alpha}}s = \partial^{|\boldsymbol{\alpha}|}s/\partial_x^{\alpha_1}\partial_y^{\alpha_2}\partial_z^{\alpha_3}$. For any function f over the closure of domain Ω, denote the $L^q(\Omega)$ norm ($1 \leq q < \infty$) and supremum norm as $\|f\|_{L^q(\Omega)} = \{\int_\Omega |f(\boldsymbol{v})|^q d\boldsymbol{v}\}^{1/q}$ and $\|f\|_{\infty,\Omega} = \sup_{\boldsymbol{v} \in \Omega} |f(\boldsymbol{v})|$, respectively. Next, for any $k \geq 0$, denote $|f|_{k,\infty,\Omega} = \max_{|\boldsymbol{\alpha}|=k} \|D^{\boldsymbol{\alpha}}f\|_{\infty,\Omega}$, and $|f|_{k,q,\Omega} = \sum_{|\boldsymbol{\alpha}|\leq k} \|D^{\boldsymbol{\alpha}}f\|_{L^q(\Omega)}$. Further, we define the Sobolev space $\mathcal{W}^{\ell,q}(\Omega) = \{f : |f|_{k,q,\Omega} < \infty, 0 \leq k \leq \ell\}$ for $1 \leq q \leq \infty$, $\ell \geq 1$.

Lemma 1 *(Theorem 3.5.2 in Lai, 1989) For all $f \in \mathcal{W}^{d+1,q}(\Omega)$ with $1 \leq q \leq \infty$, $r \geq 0$, and $d \geq 6r + 3$, there exists a spline $s_f \in \mathcal{S}_d^r(\triangle)$ such that*

$$\|D^{\boldsymbol{\alpha}}(f - s_f)\|_{L^q(\Omega)} \leq K |\triangle|^{d+1-|\boldsymbol{\alpha}|} |f|_{d+1,q,\Omega},$$

for all $0 \leq |\boldsymbol{\alpha}| \leq m$, where $K > 0$ is a constant independent of f and $|\triangle|$ but is dependent on the geometry of $\triangle$.

Remark 2 *Both d and r are parameters for the spline space $\mathcal{S}_d^r(\triangle)$, which determines the smoothness of the TPST estimator and are usually predetermined by the user. Practically, as long as $d \geq r$, we can construct the spline bases. If $d \geq 6r + 3$, we can achieve the full approximation (approximation with optimal convergence rate) order, as shown in Lemma 1. A higher value of d indicates a higher degree polynomial, which can result in a higher computation burden; see our analysis in Section 6. In practice, the choice of d and r is closely tied to the intended interpretation of the estimated function. If the goal is to enhance the signal-to-noise ratio for visualization or to suggest a simple parametric model, then a slightly oversmoothed function with a subjectively chosen parameter may be appropriate. However, if the focus is on accurately estimating the regression function and preserving local structures, then a slightly undersmoothed function may be more suitable.*

3 Penalized Spline Estimators

With all the preparations in the previous introduction, we apply the trivariate spline over the triangulation to recognize the underlying signal from 3D point clouds in this section.

In the following, for any $i = 1, \ldots, n$, let point $\boldsymbol{p}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ be the location or design point of the ith observation in a point cloud of sample size n. Let W_i be the response variable observed on the ith location $\boldsymbol{p}_i$. Then, we regard any point cloud as a set of n observations $\{(\boldsymbol{p}_i, W_i)\}_{i=1}^n$ in general. To extract the underlying signal, we consider the following nonparametric regression model

$$W_i = m(\boldsymbol{p}_i) + \sigma(\boldsymbol{p}_i)\varepsilon_i, \tag{5}$$

where $m(\cdot)$ is some smooth but unknown 3D function, $\sigma(\cdot)$ is the unknown conditional standard deviation function, and ε_i is the random error term with mean zero and variance one. Assume (W_i, ε_i) are general iid copies of (W, ε).

In nonparametric smoothing, the roughness penalty approach is widely used when smoothing noisy data (Green and Silverman, 1994; Wood, 2003; Lai and Wang, 2013). Including a roughness penalty and choosing a proper tuning parameter can avoid overfitting problems and balance the bias and variance of the estimator of the function. To estimate the underlying function m in Equation 5, we formulate the roughness penalty approach as the following penalized least squares problem:

$$\min_{s \in \mathcal{S}_d^r(\triangle)} \sum_{i=1}^n \{W_i - s(\boldsymbol{p}_i)\}^2 + \rho_n \mathcal{E}(s), \tag{6}$$

where the roughness penalty

$$\mathcal{E}(s) = \sum_{|\boldsymbol{\alpha}|=2} \binom{2}{\alpha_1} \binom{2-\alpha_1}{\alpha_2} \int_\Omega \{D^{\boldsymbol{\alpha}} s(\boldsymbol{p})\}^2 \mathrm{d}\boldsymbol{p}, \tag{7}$$

and ρ_n is the roughness penalty parameter. We aim to find the minimizer of Equation 6 in $\mathcal{S}_d^r(\triangle)$, denoted as $\widehat{m}_{\rho_n}$, which is the *Trivariate Penalized Spline over triangulation* (TPST) estimator of m. The tuning parameter ρ_n controls the smoothness of the fitted spline function. A larger ρ_n leads to a less fluctuating function. If ρ_n goes to infinity, our estimator shrinks to linear functions where the roughness penalty $\mathcal{E}(s) = 0$. On the other hand, when $\rho_n = 0$, the estimator becomes the standard unpenalized least squares spline estimator. A proper penalty parameter ρ_n balances the goodness of fit for the data and the volatility of estimated functions.

The penalty $\mathcal{E}(s)$ in Equation 7 is a commonly used penalty; see Green and Silverman (1994) for the 2D case. For a spline function $s \in \mathcal{S}_d^r(\triangle)$, combining Equation 4, the roughness penalty in Equation 7 can be written as follows:

$$\mathcal{E}(s) = \sum_{T \in \triangle} \mathcal{E}(s|_T) = \sum_{T \in \triangle} \sum_{|\boldsymbol{\alpha}|=2} \binom{2}{\alpha_1} \binom{2-\alpha_1}{\alpha_2} \int_T \{D^{\boldsymbol{\alpha}} s|_T(\boldsymbol{p})\}^2 \mathrm{d}\boldsymbol{p} = \sum_{T \in \triangle} \boldsymbol{\gamma}_T^\top \mathbf{P}_T \boldsymbol{\gamma}_T, \quad (8)$$

where $\mathbf{P}_T$ is the corresponding penalty matrix. It is easy to show that $\mathcal{E}(\mathbf{B}_d^\top \boldsymbol{\gamma}) = \boldsymbol{\gamma}^\top \mathbf{P} \boldsymbol{\gamma}$, where $\mathbf{P}$ is the block diagonal penalty matrix. See Section A.6 for more details in derivations and calculations of penalty matrices.

Since $s \in \mathcal{C}^r(\triangle)$, the coefficients of s satisfy some smoothness conditions across each interior faces of $\triangle$. One can obtain a smoothness constraint matrix $\mathbf{H}$ such that $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$ by repeatedly applying Equation A.7 in the Appendix A over all shared triangular faces, and more details are available in Section A.2 in the supplementary material. A MATLAB implementation is discussed in Awanou et al. (2006). Thus, the objective function can be written as

$$\min_{\boldsymbol{\gamma}} \sum_{i=1}^n \left\{ W_i - \mathbf{B}_d(\boldsymbol{p}_i)^\top \boldsymbol{\gamma} \right\}^2 + \rho_n \boldsymbol{\gamma}^\top \mathbf{P} \boldsymbol{\gamma}, \text{ subject to } \mathbf{H}\boldsymbol{\gamma} = \mathbf{0}, \quad (9)$$

where $\mathbf{H}$ is the matrix for all smoothness conditions across shared edges or faces of tetrahedra, which depends on the smoothness parameter r and the structure of the triangulation. One can use QR decomposition to get rid of the constraint in Equation 9. Specifically, $\mathbf{H}^\top = \mathbf{QR} = (\mathbf{Q}_1 \ \mathbf{Q}_2)\binom{\mathbf{R}_1}{\mathbf{R}_2}$, where $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{R}$ is an upper triangular matrix, $\mathbf{R}_1$ is a full rank matrix with the same rank as $\mathbf{H}$, and $\mathbf{R}_2$ is a matrix of zeros. Note that for any vector $\boldsymbol{\gamma}$ satisfying $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$, there exists some $\boldsymbol{\theta}$ such that $\boldsymbol{\gamma} = \mathbf{Q}_2\boldsymbol{\theta}$. Also, for any $\boldsymbol{\theta}$, $\mathbf{H}(\mathbf{Q}_2\boldsymbol{\theta}) = \mathbf{0}$ holds. Thus, Equation 9 is equivalent to a penalized regression problem without constraint:

$$\sum_{i=1}^n \left\{ W_i - \mathbf{B}_d(\boldsymbol{p}_i)^\top \mathbf{Q}_2 \boldsymbol{\theta} \right\}^2 + \rho_n \boldsymbol{\theta}^\top \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\theta}, \quad (10)$$

which leads to a closed form of the solution. To be specific, let $\mathbf{W} = (W_1, \ldots, W_n)^\top$, then the minimizer of Equation 10 is given by $\widehat{\boldsymbol{\theta}}_{\rho_n} = (\mathbf{Q}_2^\top \mathbf{B}_d^\top \mathbf{B}_d \mathbf{Q}_2 + \rho_n \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2)^{-1} \mathbf{Q}_2^\top \mathbf{B}_d^\top \mathbf{W}$. Consequently, the spline coefficient in Equation 9 can be estimated by $\widehat{\boldsymbol{\gamma}}_{\rho_n} = \mathbf{Q}_2 \widehat{\boldsymbol{\theta}}_{\rho_n}$, which yields the TPST estimator $\widehat{m}_{\rho_n}(\boldsymbol{p}) = \mathbf{B}_d(\boldsymbol{p})^\top \widehat{\boldsymbol{\gamma}}_{\rho_n}$. Several methods can be used for choosing the penalty parameter ρ_n, such as the block cross-validation (block CV) and generalized cross-validation (GCV). Detailed discussion is given in Section 5.2.

4 Theoretical Results

In the previous section, we discuss how to construct TPST and capture the underlying signal using TPST from the point clouds. In the following, we investigate the theoretical support of TPST.

For random variables X_n, $X_n = O_P(b_n)$ if $\lim_{c\to\infty} \limsup_n P(|X_n| \geq cb_n) = 0$, $n \geq 1$. Similarly, for any constant $c > 0$, $X_n = o_P(b_n)$ if $\lim_{n\to\infty} P(|X_n| \geq cb_n) = 0$. And $a_n \asymp b_n$ if there exist two positive constants c_1, c_2 such that $c_1|a_n| \leq |b_n| \leq c_2|a_n|$, for all $n \geq 1$.

4.1 Convergence Rate

Before we state our main results, we make the following assumptions, which are standard in nonparametric literature (Huang, 2003; Lai and Wang, 2013; Yu et al., 2020). Let $\epsilon_i = \sigma(\boldsymbol{p}_i)\varepsilon_i$, and denote the generic variable of ϵ_i as ϵ.

(A1) The trivariate function $m \in \mathcal{W}^{\ell+1,\infty}(\Omega)$ for an integer $\ell \geq 1$.

(A2) The noise ϵ satisfies that $\lim_{\eta\to\infty} \mathrm{E}[\epsilon^2 I(|\epsilon| > \eta)] = 0$ and $\mathrm{E}|\epsilon^{2+\eta}| \leq \upsilon_\eta$ for some $\eta > 0$. The standard deviation function $\sigma(\boldsymbol{p})$ is continuous on Ω and $0 < c_\sigma \leq \inf_{\boldsymbol{p}\in\Omega} \sigma(\boldsymbol{p}) \leq \sup_{\boldsymbol{p}\in\Omega} \sigma(\boldsymbol{p}) \leq C_\sigma < \infty$.

(A3) The density function of the observations is bounded below and above.

(A4) The number of the tetrahedra N and the sample size n satisfy that $N = Cn^\gamma$ for some constant $C > 0$ and $\gamma < \eta/(2 + \eta)$, where η is given in the Assumption (A2).

To obtain the asymptotic analysis of spline estimators, we rely on an important property that the data-driven norm uniformly approaches its expectation uniformly over the entire spline space. To see this, for any function f over the closure of domain Ω, let $\mathrm{E}_n(f) = n^{-1}\sum_{i=1}^n f(\boldsymbol{p}_i)$ and $\mathrm{E}(f) = \mathrm{E}[f(\boldsymbol{p})]$. Define the empirical inner product and norm as $\langle f_1, f_2 \rangle_{n,\Omega} = \mathrm{E}_n(f_1 f_2)$ and $\|f_1\|_{n,\Omega}^2 = \langle f_1, f_1 \rangle_{n,\Omega}$, respectively, for measurable functions f_1 and f_2 on Ω. The theoretical L^2 inner product and the induced norm are given by $\langle f_1, f_2 \rangle_{L^2(\Omega)} = \mathrm{E}(f_1 f_2)$ and $\|f_1\|_{L^2(\Omega)}^2 = \langle f_1, f_1 \rangle_{L^2(\Omega)}$. We illustrate the uniform convergence rate of the empirical inner product to the theoretical one in the following Lemma 3.

Lemma 3 *Denote the basis for $\mathcal{S}_d^r(\triangle)$ constructed in Lai and Schumaker (2007) by $\{B_\xi\}_{\xi\in\mathcal{M}}$, where $\mathcal{M}$ stands for the index set of spline bases. Let $g_1 = \sum_{\xi\in\mathcal{M}} c_\xi B_\xi$, $g_2 = \sum_{\zeta\in\mathcal{M}} \tilde{c}_\zeta B_\zeta$ be any spline functions in $\mathcal{S}_d^r(\triangle)$. Under Assumptions (A3) and (A4), we have*

$$R_n = \sup_{g_1,g_2\in\mathcal{S}_d^r(\triangle)} \left| \frac{\langle g_1, g_2 \rangle_{n,\Omega} - \langle g_1, g_2 \rangle_\Omega}{\|g_1\|_\Omega \|g_2\|_\Omega} \right| = O_P\left\{ (N\log n)^{1/2} n^{-1/2} \right\}.$$

For the purpose of illustrating theoretical development, we rewrite the penalty in Equation 7 in terms of linear operation. Let $\mathcal{B} := \mathcal{B}(\Omega)$ be the space of all bounded real-valued functions over $\Omega = \cup_{T\in\triangle} T$ equipped with the inner product $n\langle f_1, f_2 \rangle_{n,\Omega} + \rho_n\langle f_1, f_2 \rangle_\mathcal{E}$, where

$$\langle f_1, f_2 \rangle_\mathcal{E} = \sum_{|\boldsymbol{\alpha}|=2} \binom{2}{\alpha_1} \binom{2-\alpha_1}{\alpha_2} \sum_{T\in\triangle} \int_T \{D^{\boldsymbol{\alpha}} f_1(\boldsymbol{p})\}\{D^{\boldsymbol{\alpha}} f_2(\boldsymbol{p})\}\mathrm{d}\boldsymbol{p}$$

for $f_1, f_2 \in \mathcal{B}$.

Next, we introduce a measure of the complexity of the spline space $\mathcal{S}_d^r(\triangle)$, and another measure which bounds the size of the derivatives:

$$V_n = \sup_{g \in \mathcal{S}_d^r(\triangle)} \left\{ \frac{\|g\|_{\infty,\Omega}}{\|g\|_{n,\Omega}}, \|g\|_{n,\Omega} \neq 0 \right\}, \quad \overline{V}_n = \sup_{g \in \mathcal{S}_d^r(\triangle)} \left\{ \frac{\|g\|_{\mathcal{E}}}{\|g\|_{n,\Omega}}, \|g\|_{n,\Omega} \neq 0 \right\}. \tag{11}$$

These two measures will play an important role in developing the asymptotic results. We use the following Lemma 4 to demonstrate the upper bounds of V_n and $\overline{V}_n$.

Lemma 4 *Under Assumptions (A3) and (A4), we have $V_n = O_P(|\triangle|^{-3/2}), \overline{V}_n = O_P(|\triangle|^{-2})$.*

We then define a linear operator $P_{\rho_n} : \mathcal{B} \mapsto \mathcal{S}_d^r(\triangle)$ such that $P_{\rho_n} W = \widehat{m}_{\rho_n}$. Note that in general, P_{ρ_n} is not a linear projection. Thus, we have $P_{\rho_n} W = P_{\rho_n} m + P_{\rho_n} \epsilon$, where $P_{\rho_n} m$ and $P_{\rho_n} \epsilon$ are the penalized spline estimators based on $\{m(\boldsymbol{p}_i)\}_{i=1}^n$ and $\{\epsilon_i\}_{i=1}^n$, respectively. Under some conditions (Huang, 2003), P_0 is a bounded operator on $\mathcal{S}_d^r(\triangle)$ in the maximum norm. Denote $s_{\rho_n, m} = P_{\rho_n} m$ and $s_{\rho_n, \epsilon} = P_{\rho_n} \epsilon$. Consequently, for the penalized spline estimator $\widehat{m}_{\rho_n}$ in Equation 6,

$$\widehat{m}_{\rho_n}(\boldsymbol{p}) - m(\boldsymbol{v}) = \{s_{\rho_n, m}(\boldsymbol{p}) - m(\boldsymbol{p})\} + s_{\rho_n, \epsilon}(\boldsymbol{p}), \tag{12}$$

where $s_{\rho_n, m}(\boldsymbol{p}) - m(\boldsymbol{p})$ and $s_{\rho_n, \epsilon}(\boldsymbol{p})$ are referred to as the bias and noise terms, respectively.

According to the error decomposition in Equation 12, to derive the convergence rate of $\widehat{m}_{\rho_n}$ to m, it is sufficient to evaluate the size of the bias and noise terms. The following propositions provide the upper bound of the bias size and noise size.

Proposition 5 *Under Assumptions (A1), (A3) and (A4), if $d \geq 6r + 3$ and $\triangle$ is a β-quasi-uniform triangulation, then we have*

$$\|s_{\rho_n, m} - m\|_{\infty, \Omega} = O_P \left\{ \frac{\rho_n}{n |\triangle|^{7/2}} |m|_{2,\infty,\Omega} + \left(1 + \frac{\rho_n}{n |\triangle|^{11/2}} \right) |\triangle|^{\ell+1} |m|_{\ell+1,\infty,\Omega} \right\}.$$

Proposition 6 *Under Assumptions (A2) and (A4), $\|s_{\rho_n, \epsilon}\|_{L^2(\Omega)} = O_P \left(\frac{1}{\sqrt{n}|\triangle|^{3/2}} \right).$*

Proposition 7 *Under Assumptions (A2) and (A4), $\|s_{\rho_n, \epsilon}\|_{\infty, \Omega} = O_P \left\{ \frac{(\log n)^{1/2}}{\sqrt{n}|\triangle|^{3/2}} + \frac{\rho_n}{n^{3/2}|\triangle|^7} \right\}.$*

Based on Propositions 5—7, we illustrate the convergence rates of the TPST estimator in the following Theorem 8, in terms of both the $L^2(\Omega)$ and supremum norms.

Theorem 8 *Under Assumptions (A1)—(A4), if $d \geq 6r + 3$ and $\triangle$ is a β-quasi-uniform triangulation, we have*

$$\|\widehat{m}_{\rho_n} - m\|_{L^2(\Omega)} = O_P \left\{ \frac{\rho_n}{n|\triangle|^{7/2}} |m|_{2,\infty,\Omega} + \left(1 + \frac{\rho_n}{n|\triangle|^{11/2}} \right) |\triangle|^{\ell+1} |m|_{\ell+1,\infty,\Omega} + \frac{1}{\sqrt{n}|\triangle|^{3/2}} \right\},$$

$$\|\widehat{m}_{\rho_n} - m\|_{\infty, \Omega} = O_P \left\{ \frac{\rho_n}{n|\triangle|^{7/2}} |m|_{2,\infty,\Omega} + \left(1 + \frac{\rho_n}{n|\triangle|^{11/2}} \right) |\triangle|^{\ell+1} |m|_{\ell+1,\infty,\Omega} \right.$$
$$\left. + \frac{(\log n)^{1/2}}{\sqrt{n}|\triangle|^{3/2}} + \frac{\rho_n}{n^{3/2}|\triangle|^7} \right\}.$$

Remark 9 *For the unpenalized spline estimator, that is, $\rho_n = 0$, if one takes $N \asymp n^{3/(2\ell+5)}$, then $\|\widehat{m}_0 - m\|_{L^2(\Omega)} = O_P(n^{-(\ell+1)/(2\ell+5)})$, which achieves the optimal convergence rate as shown in Stone (1982). Similarly, for the supremum norm, one can obtain that $\|\widehat{m}_0 - m\|_{\infty,\Omega} = O_P\{(n^{-1}\log n)^{(\ell+1)/(2\ell+5)}\}$ when $N \asymp (n/\log n)^{3/(2\ell+5)}$, which is also the optimal rate of convergence. When $\rho_n > 0$, one can obtain optimal convergence rates for both L^2 and supremum norms with $\rho_n = o(n^{(\ell+1/2)/(2\ell+5)})$ and same orders of N.*

Remark 10 *Assumption (A3) is standard in nonparametric literature. Even though this assumption may not be satisfied with missing data, we usually can find a proper choice of triangulation by adjusting the size of some of the tetrahedra to obtain a decent penalized spline fitting.*

4.2 Asymptotic Normality

To derive the asymptotic normality of the TPST estimator, we further assume the following conditions.

(A4') The number of the tetrahedra N and the sample size n satisfy that $N = Cn^\gamma$ for some constant $C > 0$ and $1/(\ell+2) < \gamma < \eta/(2+\eta)$, where ℓ and η are given in the Assumptions (A1) and (A2), respectively.

(A5) The penalized parameter ρ_n satisfies $\rho_n = o(n^{1/2}N^{-2/3} \wedge nN^{-4/3})$.

Remark 11 *A sufficient condition for a negligible bias term is provided by Assumptions (A4') and (A5). Compared with Assumption (A4) in Section 4.1, (A4') further assumes that the number of tetrahedra needs to be greater than a lower bound which depends on the degree of the function. A similar assumption for the univariate case has been discussed in Li and Ruppert (2008). Meanwhile, assumption (A5) requires smaller ρ_n, which reduces the bias through under smoothing.*

Theorem 12 *Under Assumptions (A1)—(A3), (A4') and (A5), as $n \to \infty$, for each $\boldsymbol{p} \in \Omega$,*

$$\frac{\widehat{m}_{\rho_n}(\boldsymbol{p}) - m(\boldsymbol{p})}{\sqrt{\mathrm{Var}\{\widehat{m}_{\rho_n}(\boldsymbol{p})|\mathbb{P}\}}} \xrightarrow{D} N(0,1),$$

where $\mathbb{P}$ is the collection of all $\boldsymbol{p}_i$, $i = 1, \ldots, n$.

Remark 13 *The above asymptotic distribution result can be used to construct asymptotic confidence intervals in theory. For example, if we estimate $m(\boldsymbol{p})$ using piecewise constant splines, Lemma B.6 in the Appendix B gives the size of the pointwise variance $\mathrm{Var}\{\widehat{m}_{\rho_n}(\boldsymbol{p})\} = \sigma^2(\boldsymbol{p})\{nf(\boldsymbol{p})V_T\}^{-1}\{1+o(1)\}$, $\boldsymbol{p} \in \Omega$, where V_T is the volume of tetrahedron T. Therefore, an asymptotic $100(1-\alpha)\%$ pointwise confidence envelop for $m(\boldsymbol{p})$ over Ω is $\widehat{m}_{\rho_n}(\boldsymbol{p}) \pm z_{\alpha/2}\sigma(\boldsymbol{p})\{nf(\boldsymbol{p})V_T\}^{-1/2}$, where $f(\cdot)$ stands for the density function of $\boldsymbol{p}_i$. However, it is very difficult to obtain the exact form of the standard error for general TPST estimators due to the characteristic of the trivariate spline basis functions. To overcome this, Section 5.4 proposes using a wild bootstrap method to estimate the standard errors and quantify the uncertainty of the estimators.*

5 Implementation Details

This section provides some implementation details on how to construct the penalty matrix $\mathbf{P}$ and select the penalty parameter ρ_n in Equation 9. To facilitate discussion, we first introduce the directional derivatives for basis functions, followed by the construction details of the penalty matrix, the selection criteria of the penalty parameter, and triangulation selection.

5.1 Construction of Penalty Matrix

We list here the key steps in the construction of the penalty matrix $\mathbf{P}$ in Equation 8. One can refer to Section A.6 in Appendix A for more details.

For a general multivariate smooth function ϕ, the directional derivative at point $\boldsymbol{p}$ with respect to direction $\boldsymbol{u}$ is defined as

$$D_{\boldsymbol{u}}\phi(\boldsymbol{p}) := \frac{\partial}{\partial t}\phi(\boldsymbol{p}+t\boldsymbol{u})\Big|_{t=0} = \lim_{t\to 0}\frac{\phi(\boldsymbol{p}+t\boldsymbol{u}) - \phi(\boldsymbol{p})}{t}.$$

Accordingly, for vector $\boldsymbol{u} := (u_x, u_y, u_z) \in \mathbb{R}^3$ and trivariate function ϕ, the directional derivative at $\boldsymbol{p} = (x, y, z)$ is

$$D_{\boldsymbol{u}}\phi(x, y, z) := \frac{\partial}{\partial t}\phi(x+tu_x, y+tu_y, z+tu_z)\Big|_{t=0}.$$

Then for the introduced Bernstein basis function with degree d, based on Lemma A.3 in Appendix A, we have

$$D_{\boldsymbol{u}}B_{ijkl}^d(\boldsymbol{p}) = d\left\{a_1 B_{i-1,j,k,l}^{d-1}(\boldsymbol{p}) + a_2 B_{i,j-1,k,l}^{d-1}(\boldsymbol{p}) + a_3 B_{i,j,k-1,l}^{d-1}(\boldsymbol{p}) + a_4 B_{i,j,k,l-1}^{d-1}(\boldsymbol{p})\right\},$$

where (a_1, a_2, a_3, a_4) is the barycentric coordinate of direction $\boldsymbol{u}$. Based on this conclusion, for a tetrahedron s_T, the corresponding penalty term can be written as

$$\begin{aligned}
\mathcal{E}(s_T) &= \sum_{|\boldsymbol{\alpha}|=2}\binom{2}{\alpha_1}\binom{2-\alpha_1}{\alpha_2}\int_T\left\{\sum_{i+j+k+l=d}\gamma_{T;ijkl}D^{\boldsymbol{\alpha}}B_{ijkl}^{d,T}(\boldsymbol{p})\right\}^2 \mathrm{d}\boldsymbol{p} \\
&= \sum_{|\boldsymbol{\alpha}|=2}\boldsymbol{\gamma}_T^\top \mathbf{P}_T^{\boldsymbol{\alpha}}\boldsymbol{\gamma}_T = \boldsymbol{\gamma}_T^\top \mathbf{P}_T\boldsymbol{\gamma}_T,
\end{aligned}$$

where each $\mathbf{P}_T^{\boldsymbol{\alpha}}$ is a $\binom{d+3}{3} \times \binom{d+3}{3}$ matrix with entries $\int_T\{D^{\boldsymbol{\alpha}}B_{ijkl}^{d,T}(\boldsymbol{p})\}\{D^{\boldsymbol{\alpha}}B_{i'j'k'l'}^{d,T}(\boldsymbol{p})\}\mathrm{d}\boldsymbol{p}$ for $\boldsymbol{\alpha}$ satisfying $|\boldsymbol{\alpha}| = 2$. As for the penalty term defined in the whole domain, recall that $\mathcal{E}(s) = \sum_{T\in\triangle}\mathcal{E}(s_T)$ and $\mathcal{E}(s_T) = \boldsymbol{\gamma}_T^\top \mathbf{P}_T\boldsymbol{\gamma}_T$. Therefore, $\mathcal{E}(s) = \boldsymbol{\gamma}^\top \mathbf{P}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_N^\top)^\top$, and $\mathbf{P} = \mathrm{diag}(\mathbf{P}_T, T \in \triangle)$ is a block diagonal matrix.

5.2 Penalty Parameter Selection

To balance the bias and variance of the proposed estimator and achieve a good estimation and prediction performance, it is crucial to choose a suitable value of the penalty parameter ρ_n. Since the in-sample fitting errors can not gauge the prediction accuracy of the fitted function, we select a criterion function that attempts to measure the out-of-sample performance of the fitted model. Minimizing the generalized cross-validation (GCV) criterion
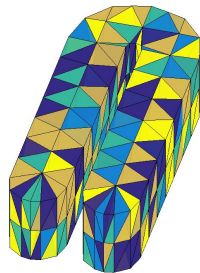
Figure 7: An illustration of block CV using a triangulation. Tetrahedra with the same color belong to the same fold.

is one computationally efficient approach to selecting smoothing parameters that also has good theoretical properties.

Note that 3D object data are often generated with spatial dependence. When performing cross-validation (CV), these dependence structures are usually ignored, leading to underestimating the predictive error (Roberts et al., 2017). To tackle this problem, one can adopt the block CV strategy (Roberts et al., 2017; Valavi et al., 2019). To be specific, all the sample points are first divided into 3D blocks with similar volumes. Then, these blocks are randomly allocated to the CV folds. In this paper, we adopt the triangulation to divide the 3D domain into small 3D blocks. Each tetrahedron is considered as one single 3D block. Figure 7 shows an example of block CV using a triangulation. In Figure 7, we divide the domain into 504 tetrahedra and randomly assign these tetrahedra into five folds with colors indicating different folds.

5.3 Triangulation Selection

As illustrated in Figure 6, to form a proper triangulation, we start with surface reconstruction. A wide range of techniques has been developed to reconstruct the surface from point clouds. Our paper focuses on the methods that generalize a well-sampled point cloud to arbitrary shapes and produce a watertight surface mesh, such as a triangular mesh. Since we are interested in recovering the actual signal within the 3D point clouds, we skip the details of the surface reconstruction here. The triangulation surface representations can be coarse or fine and affect the construction of the triangulation. A very coarse triangulation could give a poor approximation to the object, while a very fine triangulation could introduce a more serious computation burden. Note that Assumption (A4') requires that the number of the tetrahedra, N, is larger than some minimum depending upon the degree of the spline, so theoretically, we can determine the fineness of a triangulation by setting $N \approx \lfloor c_1 n^\gamma \log(n) \rfloor + c_2$, in which $\lfloor \cdot \rfloor$ denotes the integer part, c_1, c_2 are tuning constants, and $1/(\ell + 2) < \gamma < 1$. In practice, we can implement the proposed method in a coarse-to-fine resolution manner with similar criteria discussed in Section 5.2. Our extensive Monte Carlo simulation studies suggest that once a triangulation is fine enough, further refinement usually has little effect on the fitting accuracy. To avoid model over-fitting, we can stop the partition refinement when the model performance on the test set sits flat or even worsens.

5.4 Uncertainty Studies

In this section, we adopt a bootstrap method to quantitatively estimate the uncertainty of the TPST estimator. A great advantage of the bootstrap method is its simplicity, which can be a straightforward way to derive estimates of standard deviations and confidence intervals for estimators of functions over complex domains. As shown in the model in Equation 5, the assumption of homoscedasticity is often invalid when a regression model is used to estimate the point clouds. The wild bootstrap (Mammen, 1993; Hall and Horowitz, 2013) is specifically designed to work when the model is heteroscedastic. We conduct the following wild bootstrap procedure for estimating the standard errors.

Step 1. Based on the data $\{(\boldsymbol{p}_i, W_i)\}_{i=1}^n$, obtain the the TPST estimator $\widehat{m}(\boldsymbol{p}_i)$ described in Section 3 and the following residuals $\widehat{\epsilon}_i = W_i - \widehat{m}(\boldsymbol{p}_i)$, $i = 1, \ldots, n$.

Step 2. Generate the bootstrap residuals $\{\epsilon_i^*\}_{i=1}^n$ by $\epsilon_i^* = \delta_i \widehat{\epsilon}_i$, where $\delta_i = \frac{1 \pm \sqrt{5}}{2}$ with probability $\frac{5 \pm \sqrt{5}}{10}$, respectively. Define $W_i^* = \widehat{W}_i + \epsilon_i^*$.

Step 3. Apply the TPST estimator to the sample $\{(\boldsymbol{p}_i, W_i^*)\}_{i=1}^n$, and obtain the estimated function $\widehat{m}^*(\cdot)$ over the entire domain.

Step 4. Repeat Steps 2 and 3 B times and obtain a bootstrap sample of the TPST estimator as $\{\widehat{m}_b^*(\cdot)\}_{b=1}^B$. Then the standard deviation of $\widehat{m}(\boldsymbol{p})$, $\boldsymbol{p} \in \Omega$, is estimated by

$$\left[\frac{1}{B} \sum_{b=1}^B \left\{ \widehat{m}_b^*(\boldsymbol{p}) - \overline{\widehat{m}}^*(\boldsymbol{p}) \right\}^2 \right]^{1/2},$$

where $\overline{\widehat{m}}^*(\cdot) = B^{-1} \sum_{b=1}^B \widehat{m}_b^*(\cdot)$.

6 Numerical Studies

In this section, we conduct various simulation studies to assess the performance of the proposed TPST method. We use studies in Sections 6.1 and 6.2 to illustrate the capabilities of TPST in handling a variety of complex data structures, Section 6.3 to evaluate the uncertainties of the TPST estimator, and Section 6.4 to show the superiority of TPST in data reduction as well as signal enhancement, compared to other smoothing methods. More specifically, depending on the nature of the point clouds, we conduct experiments with unstructured (random design) and structured (fixed design) point clouds. The case of random design in Section 6.1 mimics the scenario that the point clouds are collections of 3D points distributed randomly in space. In contrast, the case of structured design is studied in Section 6.2. In this structured design setting, the point cloud is reconstructed from grid data; in other words, the locations of the points are deterministic. The evaluation of standard error for the proposed TPST estimator is illustrated in Section 6.3. In Section 6.4, we explore the practical performance of TPST on a structured point cloud by using PET scan data as a reference.

6.1 Unstructured Complete Point Clouds

In this example, we consider an unstructured (random design) point cloud in which the observations are randomly generated over the entire domain. We set the number of observations in each point cloud as $n = 20,000$ and $50,000$. To mimic some complicated scenarios in practice, we generate the point clouds from the following two domains: (i) a cuboid with a hole inside (Ω_1); (ii) a 3D horseshoe (Ω_2). These domains are illustrated in Figure 8. To extract the underlying signal from point clouds, we consider the model in Equation 5, where the random noises, ϵ_i's, are assumed to be independent and identically distributed and follow a normal distribution $N(0, \sigma^2)$. For each domain, we consider two types of underlying functions with different degrees of variation: m_1 and m_2 for Ω_1 (Figure 8 (c) and (e)), and m_3 and m_4 for Ω_2 (Figure 8 (i) and (k)). The noises level, σ, is chosen according to the peak signal-to-noise ratios (PSNR) defined as

$$\text{PSNR} = 20 \log_{10} \left\{ \max_i m(\boldsymbol{p}_i)/\sigma \right\}.$$

In this study, we set PSNR $= 5$ and 10, representing scenarios of high and moderate noise levels, respectively. We use this study to investigate the effect of the size of the point cloud, degree of spline polynomial, and triangulation. We also compare the proposed
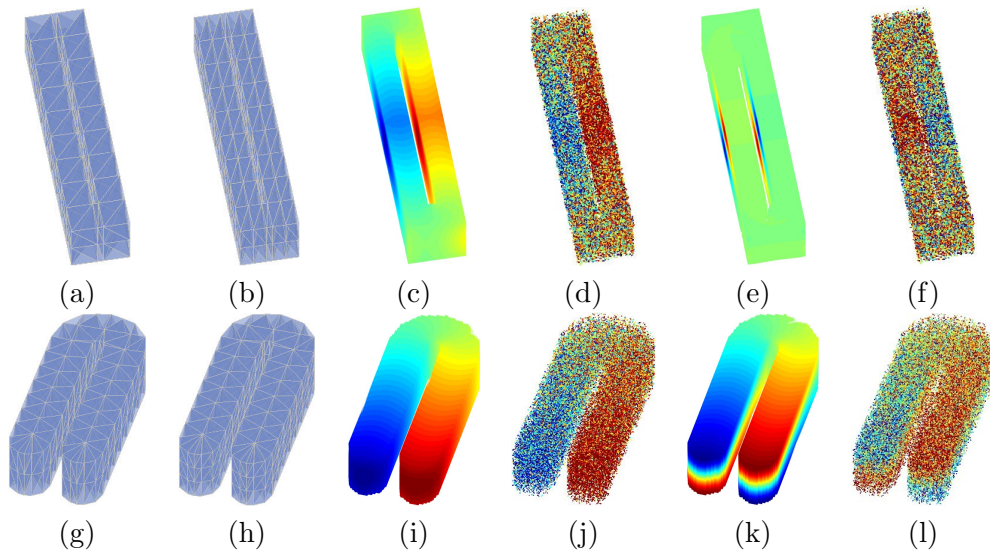

Figure 8: Triangulations, true functions and observed point clouds. Plots (a)—(f), are based on domain Ω_1; Plots (g)—(l), are based on domain Ω_2: (a) $\triangle_{11}$ for Ω_1; (b) $\triangle_{12}$ for Ω_1; (c) true function m_1; (d) observed point clouds with underlying function m_1; (e) true function m_2; (f) observed point clouds with underlying function m_2; (g) $\triangle_{21}$ for Ω_2; (h) $\triangle_{22}$ for Ω_2; (i) true function m_3; (j) observed point clouds with underlying function m_3; (k) true function m_4; (l) observed point clouds with underlying function m_4.
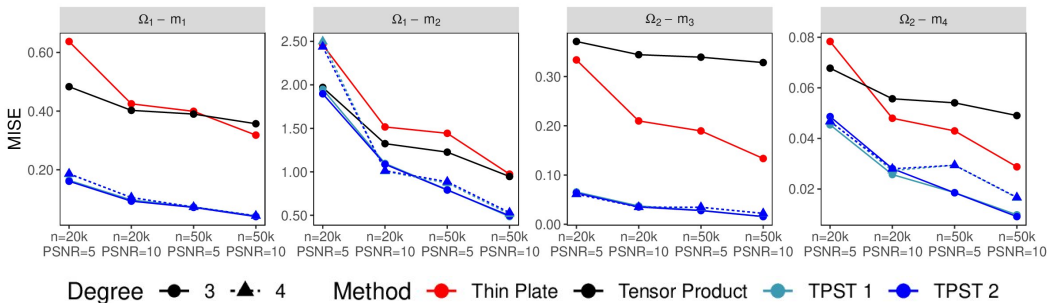
Figure 9: Average MISE plots for different methods on unstructured point clouds.

method with the traditional tensor product spline method (Stone, 1994) and thin plate spline smoothing (Wood, 2003). We implement the tensor product spline and the thin plate spline smoothing using the function "gam" in R package "mgcv". For a fair comparison, we set the dimension of the basis used to represent the underlying function for the thin plate spline and tensor product spline to be very similar to the dimension we used for the TPST method. As a result, the dimension after the numerosity reduction is comparable among different methods.

To evaluate the estimation and prediction performance of each method, we calculate the out-of-sample mean integrated squared error (MISE). Figure 9 presents the average of the MISEs over 200 replications for all different scenarios. Based on Figure 9, one can observe that as the size of the point cloud or the PSNR increases, the estimation and prediction accuracy improves for all the methods. Regardless of the scenarios, the proposed TPST method outperforms the other two traditional techniques with a similar dimension of numerosity reduction. These results indicate that the proposed TPST can better handle the "leakage" problem over the irregular domain than conventional smoothing methods.

To evaluate the effect of triangulations on the TPST, we consider a relatively coarse mesh $\triangle_{11}$ for Ω_1 and $\triangle_{21}$ for Ω_2, and a fine mesh $\triangle_{12}$ for Ω_1 and $\triangle_{22}$ for Ω_2. An illustration of these partitions is given in the Figure 8 (a), (b), (g), and (h), and the number of vertices and the number of tetrahedra are summarized in Table 1. From Figure 9, we can see that for each domain, the TPST estimator has a similar performance based on the two partitions. As discussed in Section 5.3, for TPST, when the number of the tetrahedra is sufficiently large to capture the pattern and features of the underlying function, more delicate triangulations will not benefit the estimation or the data reduction. For example, for domain Ω_2, when we keep $d = 3$, the remaining dimensions after the numerosity reduction are 323 and 340 for the two different triangulations, respectively.

Furthermore, based on Figure 9, one can see that the differences between estimates with $d = 3$ and $d = 4$ are relatively small. When we use a larger d, the estimators are

$\Omega_1 : \triangle_{11}$	$\Omega_1 : \triangle_{12}$	$\Omega_2 : \triangle_{21}$	$\Omega_2 : \triangle_{22}$
240 (120)	456 (180)	504 (207)	756 (276)

Table 1: The number of tetrahedra (vertices) in each triangulation.

usually less biased but with larger variance and more computationally intensive. In this example, $d = 3$ is preferred as they are more efficient in data reduction. In general, the choice of d depends on the smoothness of the underlying function, the strength of signals, and computing resources.

As discussed previously, the spline methods are computationally efficient since they provide a global estimator. In this simulation example, the TPST method takes less than twenty seconds to fit the model for most of the simulation samples on a single Intel E5-2640 v3 core. The speed is comparable to tensor product and thin plate spline estimators.

6.2 Structured Point Clouds with Missing Data

One of the critical aspects of analyzing point cloud data is handling the unevenly distributed point clouds and/or missing values. We use this example to investigate further the proposed method on a structured point cloud at different resolutions with different missing schemes. To be more specific, we mimic three types of missing schemes, including (i) complete data with no missing, (ii) missing at random, and (iii) missing in a contiguous block as well as at random. Figures in Figure 10 illustrates different types of missing data we deal with in this example. Furthermore, we explore various missing rates under different missing mechanisms. For missing at random, we consider the missing rates ranging uniformly from 0 to 0.5, where 0 represents no missing voxels, and 0.5 means half of the voxels are missing. The contiguous block shown in Figure 10 contains 12% of the data.

Based on the first simulation example in Section 6.1, we can see that the performance for various methods is relatively consistent. Thus, we only consider the first domain, Ω_1, and try the same triangulations in this example. We consider two fixed resolutions here: the lower resolution/scale is $60 \times 20 \times 20$ with 22,160 voxels falling within the domain, and the higher resolution/scale is $75 \times 25 \times 25$ with 42,600 voxels inside.

Similar to Section 6.1, we calculate the average MISEs over 200 replications with different missing types and missing rates and illustrated in Figure 11. Based on this Figure, one can see that the prediction accuracy improves for all of the methods as the missing rate decreases. The proposed methods outperform the two traditional methods regardless of the type of missing scheme and the missing rate. Furthermore, the missing type does not affect the
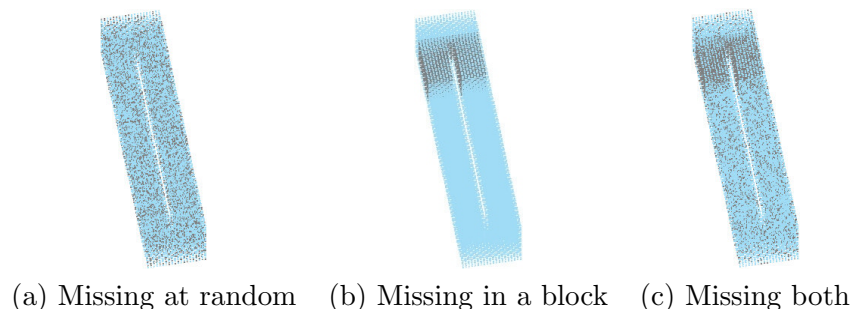

(a) Missing at random   (b) Missing in a block   (c) Missing both

Figure 10: An illustration of three different missing patterns in point clouds. In (a) and (c), missing rate = 30% and in (b), missing rate = 12%.

Random Missing

Tube+Random Missing
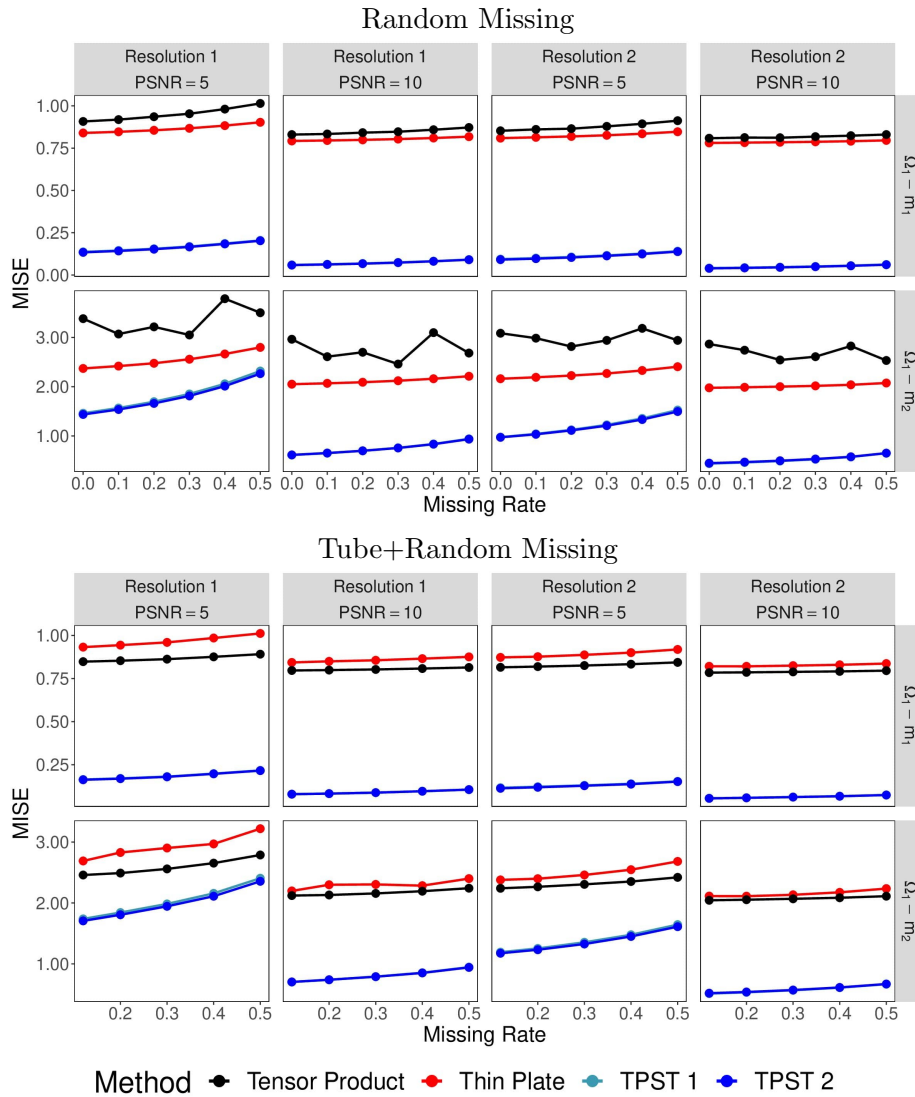
Figure 11: Average MISEs plots for different methods on structured point clouds with/without missing.

proposed method very much. In contrast, the thin plate spline smoothing usually performs better for missing at random, while the tensor product spline is better when missing a contiguous block and at random.

6.3 Standard Error Evaluation

To evaluate the standard error (SE) for the proposed TPST estimator, we conduct another computational experiment using structured points clouds generated from the 3D horseshoe domain (Ω_2). We consider the model in Equation 5, where the underlying signal is illustrated

(a) SE_{mean} (b) SE_{median} (c) SE_{mc}

Figure 12: Standard error plot of the proposed TPST estimators: (a) and (b) are the mean (SE_{mean}) and median (SE_{median}) of these bootstrap SEs across 200 rep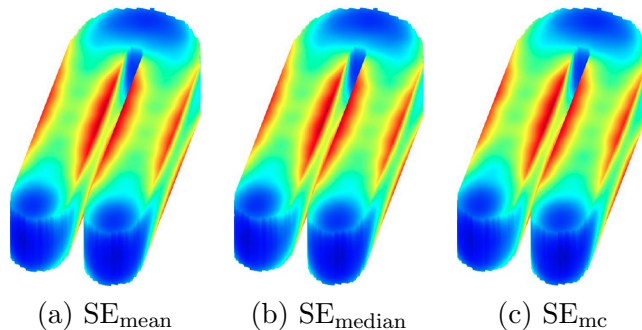lications, respectively; and (c) shows the standard deviation of the TPST estimator based on the 200 Monte Carlo samples (SE_{mc}).

in Figure 8 (i), and the random noises, $\epsilon_i = \sigma(\boldsymbol{p}_i)\varepsilon_i$'s, are generated with $\sigma(\boldsymbol{p}_i) = 6 - \{(x_i - 1.25)^2 + y_i^2 + z_i^2\}$ and $\varepsilon_i \sim N(0,1)$ iid. We consider a fixed solution/scale of $101 \times 65 \times 17$ with $93,449$ voxels falling within the domain.

To quantify the uncertainty of the estimator, we generate 200 replications as in Sections 6.1 and 6.2. For each replication, we calculate the bootstrap SEs over 100 bootstrap samples using the wide bootstrap method introduced in Section 5.4. We then compute the the mean (SE_{mean}) and median (SE_{median}) of these bootstrap SEs across 200 replications. Additionally, we calculate the standard deviation of the TPST estimator based on the Monte Carlo samples (SE_{mc}), which serves as the true value for SE. The results are displayed in Figure 12 (a) and (b) for SE_{mean} and SE_{median}, respectively, and in Figure 12 (c) for SE_{mc}. From these plots, one sees that the SE_{mean} and SE_{median} is very close to SE_{mc}, which verifies the accuracy of the proposed bootstrap SE estimation method.

6.4 Biomedical Imaging Analysis

In our increasingly aging societies, Alzheimer's disease (AD) has become the most frequent cause of dementia. Much progress has been made in assisting the early diagnosis of AD with neuroimaging techniques. One widely used neuroimaging technique is PET imaging, which can also be considered as an example of structured point clouds. However, the traditional PET scanning technique limits the overall resolution of the brain image, and there is a lack of effective and efficient image reconstruction methods. In this example, we apply the proposed TPST method to denoise and enhance the resolution of an actual PET image while improving the storage efficiency. A visual representation of transverse, coronal, and sagittal planes is shown in Figure 13 (a). Based on this figure, we can observe that brain images are collected on a rectangular parallelepiped grid with a dimension of $79 \times 95 \times 66$. However, accurate signals are only present within the voxels of the human brain. As traditional methods, such as tensor product spline and thin plate spline, cannot handle complex domains, the model is trained using all the voxels within the entire image. In

contrast, our proposed TPST method can manage complex domains efficiently. Therefore, we use a brain mask to determine the domain of the human brain, resulting in 280,000 voxels within the human brain domain. These masks can be obtained by many methods, such as building a PET template by averaging normalized PET images or testing the quantile value of each voxel to exceed a given threshold.

The results from Section 6.1 suggest that the choice of triangulation has a minimal effect on the performance of the TPST estimator as long as the number of tetrahedra is sufficient to capture the underlying pattern. In our implementation of the TPST method for brain imaging analysis, we use the triangulation depicted in Figure 13 (b) that consists of 317 tetrahedra and 117 vertices. To effectively capture high-frequency oscillations present in human brain scans, we set the polynomial degree d to at least 4. Additionally, we investigate the impact of different levels of global smoothness by considering $r = 0$ and $r = 1$. Figure 13 (c) shows an estimated PET image based on the settings $d = 5$ and $r = 1$, demonstrating the ability of the proposed TPST estimator to denoise the image while preserving the overall brain structure. To quantify the uncertainty of our estimations, we compute the bootstrap standard error as detailed in Section 5.4, and the logarithmic (base 10) transformation of the standard error map is presented in Figure 13 (d).

To compare the estimation accuracy, we consider four different measures: (i) root mean squared error, RMSE $= n^{-1/2}\{\sum_{i=1}^{n}(\widehat{W}_i - W_i)^2\}^{1/2}$, (ii) mean absolute error, MAE $= n^{-1}\sum_{i=1}^{n}|\widehat{W}_i - W_i|$, (iii) peak signal-to-noise ratio, PSNR $= 20\log_{10}\{\max(W_i)/\text{RMSE}\}$, and (iv) mean relative absolute error, MRAE $= n^{-1}\sum_{i=1}^{n}|\widehat{W}_i - W_i|/W_i$. The results of the TPST method are compared to those of other smoothing techniques, such as thin plate smoothing and tensor product spline. Similar to simulation studies in previous sections, we implement the tensor product spline and the thin plate spline smoothing using the function "gam" in R package "mgcv". Since the thin plate and tensor product methods could be affected by the noises outside the actual brain, we calculate all the measurements based on the voxels within the brain domain to ensure a fair comparison. The estimation and numerosity reduction results of different methods are shown in Table 2.

If a lower target for data reduction is acceptable, smoothing methods can achieve higher levels of accuracy. However, this comes at the cost of a larger basis expansion size. Conventional thin plate and tensor product splines, due to their specific functional form, may result in a high-dimensional representation of the data, particularly in the presence of significant noise or random fluctuations. This can increase the risk of overfitting, as the model may become overly complex and struggle to accurately capture the underlying patterns in the data. Additionally, the high dimensionality can result in memory constraints and make it difficult to process larger data sets. The TPST approach, however, addresses these limitations by providing better control over local variations in the data, resulting in a more compact representation compared to thin plate and tensor product splines. In this example, we cannot implement the thin plate and tensor product splines methods when the dimension of the spline basis exceeds 2500 on a regular PC. As demonstrated in Table 2, the TPST method, in contrast, is better equipped to deal with data with complex shapes or patterns due to its unique combination of sparsity, local control, adaptivity, and computational efficiency.

(a) Observed PET image

(b) Triangulation

(c) Estimated PET image

(d) Bootstrap standard error map
(logarithm with base 10)
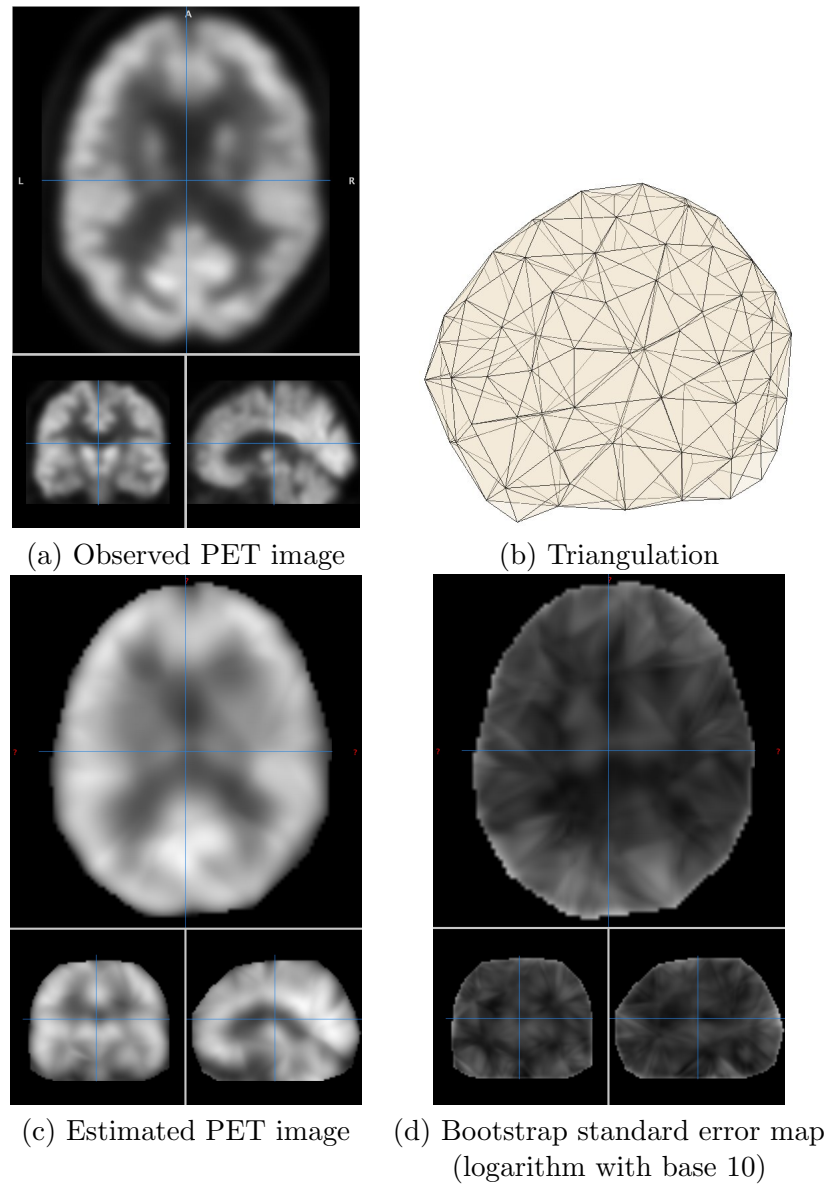
Figure 13: An illustration of an observed PET image, triangulations of the brain domain, estimated PET image, and corresponding bootstrap standard error map.

7 Conclusions and Discussion

Challenges in handling and analyzing the irregularly-shaped 3D point cloud data motivated the advanced statistical methods to uncover the underlying trajectory over the point clouds. Unlike conventional smoothing methods, the proposed TPST methods are proven to perform well for complex objects. Moreover, the proposed approaches are able to handle the irregularly and regularly missing data problem, effectively denoise or deblur the data while

Method	RMSE	MAE	PSNR	MRAE	Dimension
Thin Plate	0.0906	0.0706	26.2514	0.0917	2000
Tensor Product	0.0945	0.0740	25.8799	0.0957	2197
TPST $(d = 5, r = 1)$	0.0860	0.0656	26.7026	0.0877	2212
TPST $(d = 4, r = 0)$	0.0647	0.0490	29.1734	0.0652	4247
TPST $(d = 6, r = 1)$	0.0539	0.0405	30.7608	0.0522	4859
TPST $(d = 5, r = 0)$	0.0461	0.0339	32.1176	0.0444	7941

Table 2: Estimation and numerosity reduction results.

preserving inherent geometric features or spatial structures, and providing multi-resolution reconstruction. The experimental results demonstrate the effectiveness of the proposed approaches compared to existing smoothing techniques.

Modern geoinformation technologies have provided a variety of feasible means for generating point clouds with billions of points. The enormous size of point cloud data always leads to serious time and memory consumption problems in data processing, storage, visualization, and transmission. Two approaches are currently considered the most promising ones: parallel computing and data reduction. As a numerosity reduction method, the proposed method is an efficient data reduction technique that replaces the original big cloud point with a much smaller form of data representation. On the other hand, parallel computing is very appealing to address such "big" data issues. Unlike tensor product spline or thin plate spline, one unique feature of the proposed TPST is its great scalability in computing. Specifically, the spline basis function is generated restricted to each tetrahedron without any overlap, and the smoothness is achieved only by the constraints on the spline coefficients. Based on this feature, we plan to develop a parallel computing algorithm based on domain decomposition in future work. We will extend the ideas in Lai and Schumaker (2009) for the 2D setting to the 3D setting.

Acknowledgments and Disclosure of Funding

Appendix A. Detailed Introduction of Trivariate Splines on Triangulation

In this section, we give a detailed discussion of the trivariate splines on tetrahedra, along with an introduction to some theoretical properties. In Section A.1, we first introduce the barycentric coordinates associated with a tetrahedron and show that a trivariate polynomial can be written in a convenient form using the barycentric coordinates. In Section A.2, we describe the directional derivatives of a polynomial, and smoothness conditions for polynomials on adjoining triangular faces. The introduction of trivariate splines on a triangulation is given in Section 2.3 in the main part. Section A.3 illustrates an example of the constraint matrix. The technical details of the conclusions in Section A.2 were given in Section A.4. Sections A.5 and A.6 give a detailed introduction and induction for the directional derivatives for basis functions and penalty matrix, respectively.

A.1 Barycentric Coordinates and Bernstein Basis Polynomials

Given a tetrahedron $T = \langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$, any fixed point $\boldsymbol{p} := (x, y, z) \in \mathbb{R}^3$ has a unique representation in terms of $\langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$,

$$\boldsymbol{p} = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + b_3\boldsymbol{v}_3 + b_4\boldsymbol{v}_4, \quad \text{with } b_1 + b_2 + b_3 + b_4 = 1,$$

where (b_1, b_2, b_3, b_4) are called the *barycentric coordinates* of $\boldsymbol{p}$ *relative to the tetrahedron* T. When the point $\boldsymbol{p}$ is inside or on the faces of T, all b_1, b_2, b_3 and b_4 are nonnegative. By Cramer's rule, the barycentric coordinate corresponding to vertex $\boldsymbol{v}_i$, satisfies $b_i = \det(\mathbf{M}_i)/\det(\mathbf{M})$, $i = 1, \ldots, 4$, where

$$\mathbf{M} := \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{pmatrix}$$

and $\mathbf{M}_i$ replaces $\mathbf{M}$'s i-th column with $(1 \ x \ y \ z)^\top$.

The barycentric coordinates (b_1, b_2, b_3, b_4) also have an interesting geometric interpretation. As shown in Figure A.1, for any $\boldsymbol{p} \in T$, it divides the tetrahedron T to four sub-tetrahedra, T_1, T_2, T_3 and T_4. Notice that $\det(\mathbf{M}) = 6V_T$, then $b_i = V_{T_i}/V_T$, where T_i replace the vertex $\boldsymbol{v}_i$ in V_T with $\boldsymbol{p}$, $i = 1, \ldots, 4$.

For a nondegenerate tetrahedron T and a point $\boldsymbol{p} \in T$ with barycentric coordinates (b_1, b_2, b_3, b_4), for nonnegative integers i, j, k, l with $i + j + k + l = d$, define *trivariate Bernstein basis polynomial of degree d relative to T* as

$$B_{ijkl}^{d,T}(\boldsymbol{p}) := \frac{d!}{i!j!k!l!} b_1^i b_2^j b_3^k b_4^l, \text{ with } i + j + k + l = d.$$

For any positive integer d and tetrahedron T, let $\mathcal{P}_d(T)$ be the space of all trivariate polynomials defined on T with degrees less than or equal to d. Note that the dimension of $\mathcal{P}_d(T)$ is $\binom{d+3}{3}$.

According to Theorem 15.8 in Lai and Schumaker (2007) and Lemma A.6 in Appendix Section A.4.1, the set of Bernstein basis polynomials

$$\boldsymbol{B}_T^d(\boldsymbol{p}) := \{B_{ijkl}^{d,T}(\boldsymbol{p}) : i, j, k, l \geq 0, i + j + k + l = d\}$$
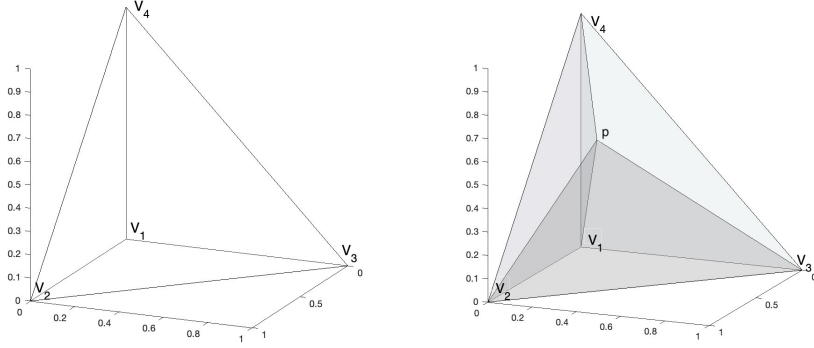
Figure A.1: An illustration of barycentric coordinates of a point $\boldsymbol{p}$ in a tetrahedron T.

forms a basis for the space of polynomials $\mathcal{P}_d(T)$. In addition, Bernstein basis functions $\{B_{ijkl}^{d,T}(\boldsymbol{p})\}_{i+j+k+l=d}$ have the following properties:

1. $\{B_{ijkl}^{d,T}\}$ form a partition of unity, that is, for all $\boldsymbol{p} \in T$, $\sum_{i+j+k+l=d} B_{ijkl}^{d,T}(\boldsymbol{p}) = 1$;

2. for all $\boldsymbol{p} \in T$, $0 \le B_{ijkl}^{d,T}(\boldsymbol{p}) \le 1$;

3. $B_{ijkl}^{d,T}$ has a unique maximum at the point $d^{-1}(i\boldsymbol{v}_1 + j\boldsymbol{v}_2 + k\boldsymbol{v}_3 + l\boldsymbol{v}_4)$.

Remark A.1 *Barycentric coordinates are invariant to linear transformations of Cartesian coordinates, that is, they do not depend on the orientation or location of the tetrahedra. Consequently, the trivariate splines based on Bernstein basis polynomials, which are constructed with barycentric coordinates, are also invariant to linear transformations.*

In order to further illustrate the Bernstein basis functions, we present an example of $\{B_{ijkl}^{d,T}(\boldsymbol{p})\}_{i+j+k+l=d}$ for $d = 4$ in Figure A.2. In this example, there are $\binom{7}{3} = 35$ basis functions in total. Note that all the function values vary between 0 and 1, and the colors scale to the quantiles of function values.

Thus, given Bernstein basis functions $\{B_{ijkl}^{d,T}(\boldsymbol{p})\}_{i+j+k+l=d}$, any polynomial $\phi(\boldsymbol{p}) \in \mathcal{P}_d(T)$ can be written uniquely as *B-form*,

$$\phi(\boldsymbol{p})|_T = \sum_{i+j+k+l=d} \gamma_{T;ijkl} B_{ijkl}^{d,T}(\boldsymbol{p}) = \boldsymbol{B}_T^d(\boldsymbol{p})^\top \boldsymbol{\gamma}_T, \tag{A.1}$$

where the coefficients $\boldsymbol{\gamma}_T = \{\gamma_{T;ijkl}\}_{i+j+k+l=d}$ are called *B-coefficients* of ϕ. For the purpose of computer implementation, in this paper, we employ the lexicographical order for ordering of the coefficients $\boldsymbol{\gamma}_T$. To be specific, $\gamma_{T;ijkl}$ orders ahead of $\gamma_{T;i'j'k'l'}$ either (i) $i > i'$, or (ii) $i = i'$ and $j > j'$, or (iii) $i = i'$, $j = j'$ and $k > k'$, or (iv) $i = i'$, $j = j'$, $k = k'$ and $l > l'$. Consequently, we can express

$$\boldsymbol{B}_T^d(\boldsymbol{p}) = \left(B_{d,0,0,0}^{d,T}(\boldsymbol{p}), B_{d-1,1,0,0}^{d,T}(\boldsymbol{p}), B_{d-1,0,1,0}^{d,T}(\boldsymbol{p}), \dots, B_{0,0,1,d-1}^{d,T}(\boldsymbol{p}), B_{0,0,0,d}^{d,T}(\boldsymbol{p}) \right)^\top$$
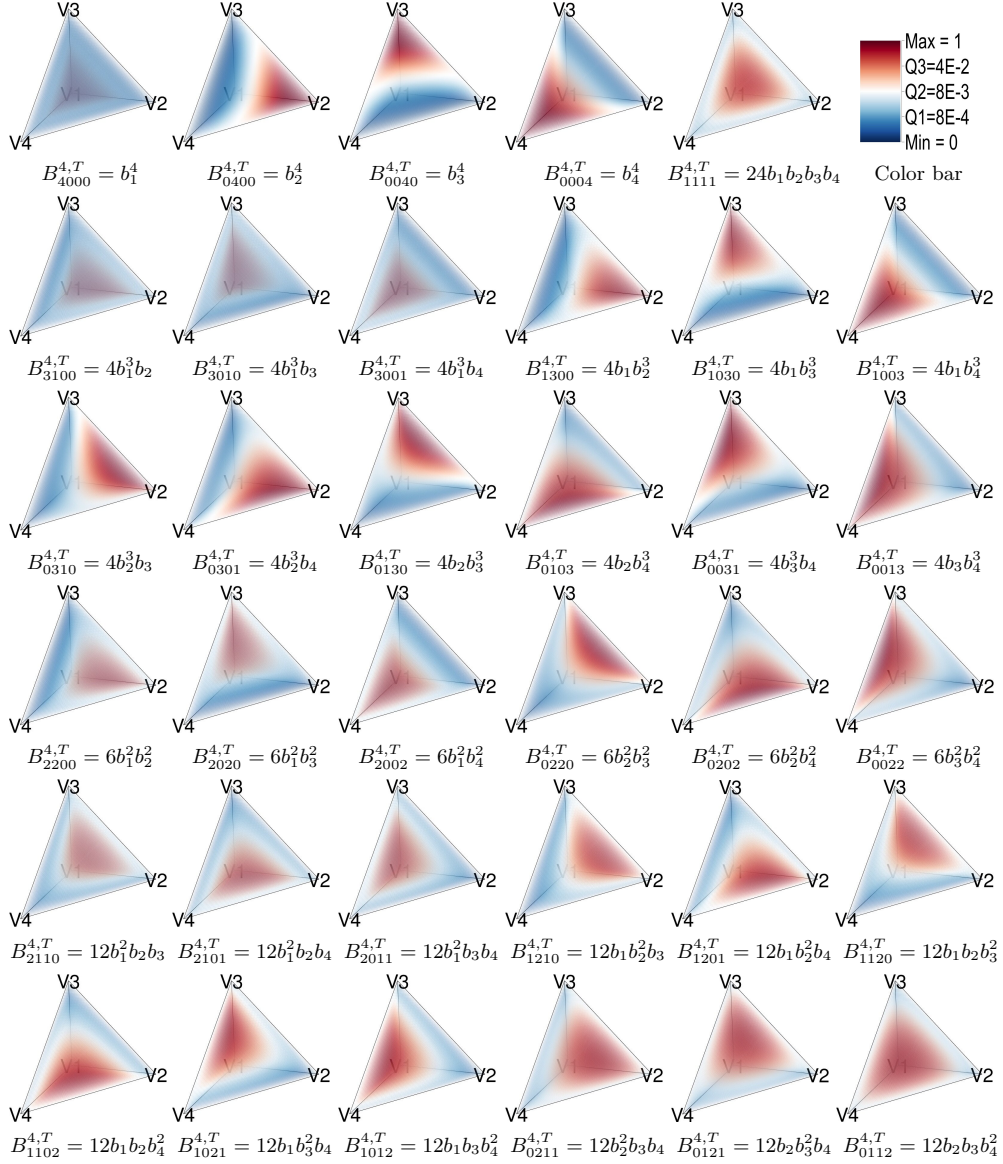
Figure A.2: An illustration of Bernstein basis functions $\{B_{ijkl}^{4,T}(\boldsymbol{p})\}$.

and

$$\boldsymbol{\gamma}_T = \left(\gamma_{T;d,0,0,0}, \gamma_{T;d-1,1,0,0}, \gamma_{T;d-1,0,1,0}, \ldots, \gamma_{T;0,0,1,d-1}, \gamma_{T;0,0,0,d}\right)^\top. \tag{A.2}$$

Accordingly, in Equation A.2, the index of the element $\gamma_{T;ijkl}$ in the vector $\boldsymbol{\gamma}_T$ is:

$$\sum_{m=0}^{d-i} \frac{(m+1)m}{2} + \sum_{n=0}^{d-i-j} (n+1) - k.$$

Note that using a different ordering method will not affect the evaluation results for the trivariate polynomial functions.

It is convenient to derive conditions of continuous connection for polynomials defined on adjacent tetrahedra in using barycentric coordinates and Bernstein basis polynomials. We use the following Example A.1 for illustration.

Example A.1 *Two adjacent tetrahedra $T = \langle \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$ and $\widetilde{T} = \langle \boldsymbol{v}_5, \boldsymbol{v}_1, \boldsymbol{v}_4, \boldsymbol{v}_3 \rangle$ share a common triangular face $F = \langle \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$, as illustrated in Figure A.3, where the Cartesian coordinates of the five vertices are $\boldsymbol{v}_1 = (0,0,0)$, $\boldsymbol{v}_2 = (1,0,0)$, $\boldsymbol{v}_3 = (0,1,0)$, $\boldsymbol{v}_4 = (0,0,1)$, $\boldsymbol{v}_5 = (-1,0,0)$, respectively.*

Denote two sets of Bernstein polynomial basis defined on T and $\widetilde{T}$ using the barycentric coordinates as $\{B_{ijkl}^{d,T}(\boldsymbol{p})\}_{i+j+k+l=d}$ and $\{\widetilde{B}_{ijkl}^{d,\widetilde{T}}(\widetilde{\boldsymbol{p}})\}_{i+j+k+l=d}$, respectively. Consider two degree-d polynomials $\phi(\boldsymbol{p})$ and $\widetilde{\phi}(\widetilde{\boldsymbol{p}})$ defined on T and $\widetilde{T}$, respectively, with B-forms

$$\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{ijkl} B_{ijkl}^{d,T}(\boldsymbol{p}), \quad \widetilde{\phi}(\widetilde{\boldsymbol{p}}) = \sum_{i+j+k+l=d} \widetilde{\gamma}_{ijkl} \widetilde{B}_{ijkl}^{d,\widetilde{T}}(\widetilde{\boldsymbol{p}}).$$

For point $\boldsymbol{p}_F \in F$, the barycentric coordinates with respect to T and $\widetilde{T}$ are $(0, b_1, b_3, 1 - b_1 - b_3)$ and $(0, b_1, 1 - b_1 - b_3, b_3)$, respectively. Accordingly, we have

$$\phi(\boldsymbol{p}_F) = \sum_{j+k+l=d} \gamma_{0jkl} \frac{d!}{j!k!l!} b_1^j b_3^k (1 - b_1 - b_3)^l,$$

$$\widetilde{\phi}(\boldsymbol{p}_F) = \sum_{j+k+l=d} \widetilde{\gamma}_{0jkl} \frac{d!}{j!k!l!} b_1^j (1 - b_1 - b_3)^k b_3^l. \tag{A.3}$$

Therefore, ϕ and $\widetilde{\phi}$ are continuous on F if and only if

$$\gamma_{0jkl} = \widetilde{\gamma}_{0jlk} \tag{A.4}$$

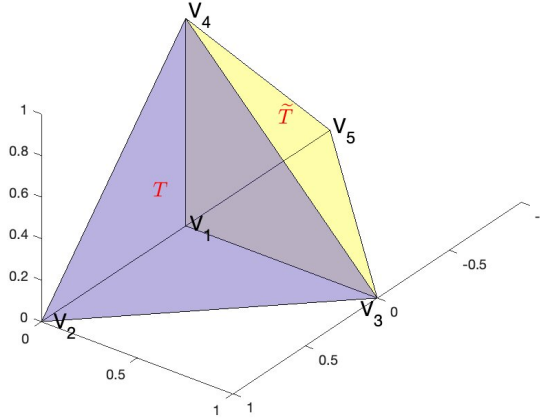for $j, k, l \geq 0$ and $j + k + l = d$.

Figure A.3: An example of two tetrahedra sharing a common face.

A.2 Directional Derivatives and Smoothness

To generalize the smoothness restriction over the joint triangular face for two adjacent tetrahedra, we need to introduce the definitions of directional derivative first. Recall that for a general multivariate smooth function ϕ, the directional derivative at point $\boldsymbol{p}$ with respect to direction $\boldsymbol{u}$ is defined as

$$D_{\boldsymbol{u}}\phi(\boldsymbol{p}) := \frac{\partial}{\partial t}\phi(\boldsymbol{p}+t\boldsymbol{u})\Big|_{t=0} = \lim_{t\to 0}\frac{\phi(\boldsymbol{p}+t\boldsymbol{u})-\phi(\boldsymbol{p})}{t}.$$

Accordingly, for vector $\boldsymbol{u} := (u_x, u_y, u_z) \in \mathbb{R}^3$ and trivariate function ϕ, the directional derivative at $\boldsymbol{p} = (x, y, z)$ is

$$D_{\boldsymbol{u}}\phi(x, y, z) := \frac{\partial}{\partial t}\phi(x+tu_x, y+tu_y, z+tu_z)\Big|_{t=0}.$$

Remark A.2 *If ϕ is a polynomial of degree d, by calculus, $D_{\boldsymbol{u}}\phi(x, y, z) = u_x D_x\phi(x, y, z) + u_y D_y\phi(x, y, z) + u_z D_z\phi(x, y, z)$, so $D_{\boldsymbol{u}}\phi$ is a polynomial of degree $d-1$.*

Consider direction $\boldsymbol{u} = \boldsymbol{p}_1 - \boldsymbol{p}_2$, where for $\jmath = 1, 2$, $\boldsymbol{p}_\jmath \in \mathbb{R}^3$ have barycentric coordinates $(b_{\jmath 1}, b_{\jmath 2}, b_{\jmath 3}, b_{\jmath 4})$ with respect to T. Then $\boldsymbol{u}$ is uniquely described by the *directional coordinates* $\boldsymbol{a} = (a_1, a_2, a_3, a_4) = (b_{11} - b_{21}, b_{12} - b_{22}, b_{13} - b_{23}, b_{14} - b_{24})$. Obviously, $a_1 + a_2 + a_3 + a_4 = 0$. Direct calculation gives the directional derivative of the Bernstein basis polynomial B_{ijkl}^d.

Lemma A.3 *(Lemma 15.12 in Lai and Schumaker, 2007) Consider direction $\boldsymbol{u}$ with directional coordinates $\boldsymbol{a} = (a_1, a_2, a_3, a_4)$. Then*

$$D_{\boldsymbol{u}}B_{ijkl}^d(\boldsymbol{p}) = d\left\{a_1 B_{i-1,j,k,l}^{d-1}(\boldsymbol{p}) + a_2 B_{i,j-1,k,l}^{d-1}(\boldsymbol{p}) + a_3 B_{i,j,k-1,l}^{d-1}(\boldsymbol{p}) + a_4 B_{i,j,k,l-1}^{d-1}(\boldsymbol{p})\right\},$$

for any $\boldsymbol{p} \in T$ and $i + j + k + l = d$.

Consequently, one can obtain the directional derivative for any trivariate polynomial ϕ.

Theorem A.4 *(Theorems 15.13 and 15.14 in Lai and Schumaker, 2007) Consider direction $\boldsymbol{u}$ with directional coordinates $\boldsymbol{a} = (a_1, a_2, a_3, a_4)$. Then for any trivariate polynomial ϕ with B-form in Equation A.1, the directional derivative is*

$$D_{\boldsymbol{u}}\phi(\boldsymbol{p}) = d\sum_{i+j+k+l=d-1}\gamma_{ijkl}^{(1)}(\boldsymbol{a})B_{ijkl}^{d-1}(\boldsymbol{p}), \tag{A.5}$$

where $\gamma_{ijkl}^{(1)}(\boldsymbol{a}) = a_1\gamma_{i+1,j,k,l} + a_2\gamma_{i,j+1,k,l} + a_3\gamma_{i,j,k+1,l} + a_4\gamma_{i,j,k,l+1}$.

In general, given $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ with associated directional coordinates $\boldsymbol{a}^{(\imath)} = (a_1^{(\imath)}, a_2^{(\imath)}, a_3^{(\imath)}, a_4^{(\imath)})$, $\imath = 1, \ldots, m$,

$$D_{\boldsymbol{u}_m}\cdots D_{\boldsymbol{u}_1}\phi(\boldsymbol{p}) = \frac{d!}{(d-m)!}\sum_{i+j+k+l=d-m}\gamma_{ijkl}^{(m)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(m)})B_{ijkl}^{d-m}(\boldsymbol{p}), \tag{A.6}$$

where the coefficients are defined recursively as follows:

$$
\begin{aligned}
\gamma_{ijkl}^{(m)}(\boldsymbol{a}^{(1)},\ldots,\boldsymbol{a}^{(m)}) =& a_1^{(m)}\gamma_{i+1,j,k,l}^{(m-1)}(\boldsymbol{a}^{(1)},\ldots,\boldsymbol{a}^{(m-1)}) + a_2^{(m)}\gamma_{i,j+1,k,l}^{(m-1)}(\boldsymbol{a}^{(1)},\ldots,\boldsymbol{a}^{(m-1)}) \\
&+ a_3^{(m)}\gamma_{i,j,k+1,l}^{(m-1)}(\boldsymbol{a}^{(1)},\ldots,\boldsymbol{a}^{(m-1)}) + a_4^{(m)}\gamma_{i,j,k,l+1}^{(m-1)}(\boldsymbol{a}^{(1)},\ldots,\boldsymbol{a}^{(m-1)}),
\end{aligned}
$$

for $m = 1,\ldots,d$, with $\gamma_{ijkl}^{(0)}(\boldsymbol{a}) = \gamma_{ijkl}$.

In preparation for the discussion of trivariate spline and spline spaces, we need the following conditions for a smooth join between two polynomial on adjoining tetrahedra, like Example A.1 illustrated in Figure A.3.

Theorem A.5 *Suppose $\{\gamma_{ijkl}\}$ and $\{\widetilde{\gamma}_{ijkl}\}$ are B-coefficients of ϕ and $\widetilde{\phi}$ relative to two tetrahedra $T = \langle \boldsymbol{v}_2, \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$ and $\widetilde{T} = \langle \boldsymbol{v}_5, \boldsymbol{v}_1, \boldsymbol{v}_4, \boldsymbol{v}_3 \rangle$, respectively, where T and $\widetilde{T}$ share a common face $F = \langle \boldsymbol{v}_1, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$. Then the following statements are equivalent:*

1. *ϕ and $\widetilde{\phi}$ join together with $\mathcal{C}^r$ continuity across the face F;*

2. *For all $\boldsymbol{p}_F \in F$, $m = 0,\ldots,r$ and for all directions $\boldsymbol{u}$,*

$$
D_{\boldsymbol{u}}^m \phi(\boldsymbol{p}_F) = D_{\boldsymbol{u}}^m \widetilde{\phi}(\boldsymbol{p}_F);
$$

3. *For ϕ and $\widetilde{\phi}$ with B-forms in Equation A.3, for $i + j + k = d - m$, $m = 0,\ldots,r$,*

$$
\widetilde{\gamma}_{mijk} = \sum_{\nu+\mu+\kappa+\delta=m} \gamma_{\nu,i+\mu,k+\kappa,j+\delta} B_{\nu\mu\kappa\delta}^m(\boldsymbol{v}_5). \tag{A.7}
$$

Consider the case of $d = 2$ piecewise polynomial. There are in total 10 Bernstein basis polynomials with coefficients

$$
\{\gamma_{ijkl}\} = (\gamma_{2000}, \gamma_{1100}, \gamma_{1010}, \gamma_{1001}, \gamma_{0200}, \gamma_{0110}, \gamma_{0101}, \gamma_{0020}, \gamma_{0011}, \gamma_{0002})^\top,
$$
$$
\{\widetilde{\gamma}_{ijkl}\} = (\widetilde{\gamma}_{2000}, \widetilde{\gamma}_{1100}, \widetilde{\gamma}_{1010}, \widetilde{\gamma}_{1001}, \widetilde{\gamma}_{0200}, \widetilde{\gamma}_{0110}, \widetilde{\gamma}_{0101}, \widetilde{\gamma}_{0020}, \widetilde{\gamma}_{0011}, \widetilde{\gamma}_{0002})^\top. \tag{A.8}
$$

Note that the barycentric coordinate of $\boldsymbol{v}_5$ with respect to T is $(2, -1, 0, 0)$.

If the trivariate polynomial is continuous over the whole region, then applying Equation A.7 for $m = r = 0$ generates

$$
\widetilde{\gamma}_{0ijk} = \gamma_{0ikj} B_{0000}^0(\boldsymbol{v}_5) = \gamma_{0ikj}, \quad i + j + k = d, \tag{A.9}
$$

which matches the conclusion in Equation A.4.

Based on the de Casteljau Algorithm (Theorem A.7 in Appendix Section A.4.2), we propose the following computationally efficient algorithm in Algorithm A.1 to calculate the derivatives of $\phi(\boldsymbol{p})$.

Algorithm A.1 Algorithm for the derivatives of $\phi(\boldsymbol{p})$

Input: Polynomial with B-form $\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{ijkl} B_{ijkl}^d(\boldsymbol{p})$, directions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ with associated directional coordinates $\boldsymbol{a}^{(\imath)} = (a_1^{(\imath)}, a_2^{(\imath)}, a_3^{(\imath)}, a_4^{(\imath)}), \imath = 1, \ldots, m$.

Initialization: $\imath := 0$, $\gamma_{ijkl}^{(0)} := \gamma_{ijkl}$.

for $\imath = 1, \ldots, m$ **do**

   for $i + j + k + l = d - \imath$ **do**

     Compute

$$
\begin{aligned}
\gamma_{ijkl}^{(\imath)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(m)}) =& a_1^{(\imath)} \gamma_{i+1,j,k,l}^{(\imath-1)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\imath-1)}) + a_2^{(\imath)} \gamma_{i,j+1,k,l}^{(\imath-1)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\imath-1)}) \\
&+ a_3^{(\imath)} \gamma_{i,j,k+1,l}^{(\imath-1)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\imath-1)}) + a_4^{(\imath)} \gamma_{i,j,k,l+1}^{(\imath-1)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(\imath-1)}).
\end{aligned}
$$

   end for

end for

Output: $D_{\boldsymbol{u}_m} \cdots D_{\boldsymbol{u}_1} \phi(\boldsymbol{p}) = \frac{d!}{(d-m)!} \sum_{i+j+k+l=d-m} \gamma_{ijkl}^{(m)}(\boldsymbol{a}^{(1)}, \ldots, \boldsymbol{a}^{(m)}) B_{ijkl}^{d-m}(\boldsymbol{p})$.

A.3 Example of the Constraint Matrix

We illustrate the construction of the constraint matrix $\mathbf{H}$ using the Example A.1 presented with Figure A.3.

Continue our discussions in Section A.2 and consider the case of $d = 2$ piecewise polynomial. Recall that there are in total 10 Bernstein basis polynomials, with coefficients $\{\gamma_{ijkl}\}$ and $\{\widetilde{\gamma}_{ijkl}\}$ in Equation A.8, and the barycentric coordinate of $\boldsymbol{v}_5$ with respect to T is $(2, -1, 0, 0)$.

As shown in Equation A.9, if the trivariate polynomial is continuous over the whole region, then $\widetilde{\gamma}_{0ijk} = \gamma_{0ikj}$ for $i + j + k = d$. In this case, we can write the constraint matrix $\mathbf{H}$ as

$$
\mathbf{H} = \begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0
\end{pmatrix}.
\tag{A.10}
$$

Furthermore, if the trivariate polynomial has continuous first derivatives over the whole region, then Equation A.7 holds for both $m = r = 0$ and $m = 1$. Thus, in addition, we also need for any non-negative integers i, j, k such that $i + j + k = 1$,

$$
\begin{aligned}
\widetilde{\gamma}_{1ijk} =& \gamma_{1ikj} B_{1000}^1(\boldsymbol{v}_5) + \gamma_{0,i+1,j} B_{0100}^1(\boldsymbol{v}_5) + \gamma_{0,i,k+1,j} B_{0010}^1(\boldsymbol{v}_5) + \gamma_{0,i,k,j+1} B_{0001}^1(\boldsymbol{v}_5) \\
=& 2\gamma_{1ikj} - \gamma_{0,i+1,k,j}.
\end{aligned}
\tag{A.11}
$$

Therefore, for $r = 1$ where the trivariate spline has continuous first derivatives over the whole region, we will obtain the following $\mathbf{H}$

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Here, the first six rows match the form in Equation A.10, and the last three rows correspond to Equation A.11.

A.4 Proof of Theorems in Section A.2

In this section, we provide detailed proofs for the theorems in Section A.2.

A.4.1 Theoretical properties of Bernstein bases

Lemma A.6 *For the Bernstein basis functions* $\{B_{ijkl}^{d,T}(\boldsymbol{p})\}_{i+j+k+l=d}$, *each* $B_{ijkl}^{d,T}$ *has a unique maximum at the point* $d^{-1}(i\boldsymbol{v}_1 + j\boldsymbol{v}_2 + k\boldsymbol{v}_3 + l\boldsymbol{v}_4)$.

To show Lemma A.6, we first need to show Lemma A.3.

Proof of Lemma A.3. Suppose $\boldsymbol{b} = (b_1, b_2, b_3, b_4)$ are the barycentric coordinates of $\boldsymbol{p}$. Then the barycentric coordinates of $\boldsymbol{p} + t\boldsymbol{u}$ are $(b_1 + ta_1, b_2 + ta_2, b_3 + ta_3, b_4 + ta_4)$. Thus, for $i + j + k + l = d$,

$$B_{ijkl}^d(\boldsymbol{p} + t\boldsymbol{u}) = \frac{d!}{i!j!k!l!}(b_1 + ta_1)^i(b_2 + ta_2)^j(b_3 + ta_3)^k(b_4 + ta_4)^l.$$

Hence,

$$
\begin{aligned}
D_{\boldsymbol{u}}B_{ijkl}^d(\boldsymbol{p}) &= \frac{\partial}{\partial t}\phi(\boldsymbol{p} + t\boldsymbol{u})\Big|_{t=0} \\
&= \frac{d!}{i!j!k!l!}\left(ia_1 b_1^{i-1}b_2^j b_3^k b_4^l + ja_2 b_1^i b_2^{j-1}b_3^k b_4^l + ka_3 b_1^i b_2^j b_3^{k-1}b_4^l + la_4 b_1^i b_2^j b_3^k b_4^{l-1}\right) \\
&= d\left\{a_1 B_{i-1,j,k,l}^{d-1}(\boldsymbol{p}) + a_2 B_{i,j-1,k,l}^{d-1}(\boldsymbol{p}) + a_3 B_{i,j,k-1,l}^{d-1}(\boldsymbol{p}) + a_4 B_{i,j,k,l-1}^{d-1}(\boldsymbol{p})\right\}.
\end{aligned}
$$

$\blacksquare$

Proof of Lemma A.6 For $T = \langle \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \boldsymbol{v}_4 \rangle$, the barycentric coordinates for $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ and $\boldsymbol{v}_4$ are $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$, respectively. Accordingly, the directional coordinates of $\boldsymbol{u}_1 = \boldsymbol{v}_1 - \boldsymbol{v}_2$, $\boldsymbol{u}_2 = \boldsymbol{v}_1 - \boldsymbol{v}_3$ and $\boldsymbol{u}_3 = \boldsymbol{v}_1 - \boldsymbol{v}_4$ are $(1, -1, 0, 0)$, $(1, 0, -1, 0)$ and $(1, 0, 0, -1)$, respectively. For $\boldsymbol{p} \in T$ with barycentric coordinates $\boldsymbol{b} = (b_1, b_2, b_3, b_4)$, consider derivatives of $B_{ijkl}^d(\boldsymbol{p})$ with respect to directions $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$, then

$$D_{\boldsymbol{u}_1}B_{ijkl}^d(\boldsymbol{p}) = B_{ijkl}^d(\boldsymbol{p})(ib_1^{-1} - jb_2^{-1}), \quad D_{\boldsymbol{u}_2}B_{ijkl}^d(\boldsymbol{p}) = B_{ijkl}^d(\boldsymbol{p})(ib_1^{-1} - kb_3^{-1}),$$
$$D_{\boldsymbol{u}_3}B_{ijkl}^d(\boldsymbol{p}) = B_{ijkl}^d(\boldsymbol{p})(ib_1^{-1} - lb_4^{-1}).$$

Setting these equations to zero and combining with $b_1 + b_2 + b_3 + b_4 = 1$ gives $(b_1, b_2, b_3, b_4) = d^{-1}(i, j, k, l)$. ∎

A.4.2 Proof of Theorems A.4 and A.5

Proof of Theorem A.4 For $\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{ijkl} B_{ijkl}^d(\boldsymbol{p})$, by Lemma A.3,

$$
\begin{aligned}
D_{\boldsymbol{u}}\phi(\boldsymbol{p}) &= \sum_{i+j+k+l=d} \gamma_{ijkl} D_{\boldsymbol{u}} B_{ijkl}^d(\boldsymbol{p}) \\
&= \sum_{i+j+k+l=d} \gamma_{ijkl} d \left\{ a_1 B_{i-1,j,k,l}^{d-1}(\boldsymbol{p}) + a_2 B_{i,j-1,k,l}^{d-1}(\boldsymbol{p}) + a_3 B_{i,j,k-1,l}^{d-1}(\boldsymbol{p}) + a_4 B_{i,j,k,l-1}^{d-1}(\boldsymbol{p}) \right\} \\
&= d \sum_{i+j+k+l=d-1} \left(a_1 \gamma_{i+1,j,k,l} + a_2 \gamma_{i,j+1,k,l} + a_3 \gamma_{i,j,k+1,l} + a_4 \gamma_{i,j,k,l+1} \right) B_{ijkl}^{d-1}(\boldsymbol{p}).
\end{aligned}
$$

Thus, Equation A.5 follows. Consequently, one can obtain Equation A.6 by repeatedly applying Equation A.5 for directions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$. ∎

To show Theorem A.5, we need the following results.

Theorem A.7 *(Theorems 15.10 in Lai and Schumaker, 2007) Suppose $\phi(\boldsymbol{p})$ is a trivariate polynomial with B-form $\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{ijkl} B_{ijkl}^d(\boldsymbol{p})$. Define $\gamma_{ijkl}^{(0)} := \gamma_{ijkl}$, $i+j+k+l = d$. Suppose $\boldsymbol{p}$ has barycentric coordinates $\boldsymbol{b} = (b_1, b_2, b_3, b_4)$. Then*

$$
\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d-m} \gamma_{ijkl}^{(m)} B_{ijkl}^{d-m}(\boldsymbol{p}),
$$

where for $m = 1, \ldots, d$, $\gamma_{ijkl}^{(m)}$ are computed by the recursion

$$
\gamma_{ijkl}^{(m)} = b_1 \gamma_{i+1,j,k,l}^{(m-1)} + b_2 \gamma_{i,j+1,k,l}^{(m-1)} + b_3 \gamma_{i,j,k+1,l}^{(m-1)} + b_4 \gamma_{i,j,k,l+1}^{(m-1)}, \tag{A.12}
$$

for $i + j + k + l = d - m$.

Remark A.8 *The recursive formula in Equation A.12 is also referred as de Casteljau algorithm. See Section 15.6 in Lai and Schumaker (2007) for more details.*

Lemma A.9 *The coefficients in the recursive formula of de Casteljau algorithm in Equation A.12 are given by*

$$
\gamma_{ijkl}^{(m)} = \sum_{i'+j'+k'+l'=m} \gamma_{i+i',j+j',k+k',l+l'} B_{i'j'k'l'}^m(\boldsymbol{p}), \tag{A.13}
$$

where $i + j + k + l = d - m$.

Proof Define operators $E_1 \gamma_{ijkl} = \gamma_{i+1,j,k,l}$, $E_2 \gamma_{ijkl} = \gamma_{i,j+1,k,l}$, $E_3 \gamma_{ijkl} = \gamma_{i,j,k+1,l}$ and $E_4 \gamma_{ijkl} = \gamma_{i,j,k,l+1}$. Thus, by Equation A.12, for $i + j + k + l = d - m$,

$$
\begin{aligned}
\gamma_{ijkl}^{(m)} &= (b_1 E_1 + b_2 E_2 + b_3 E_3 + b_4 E_4) \gamma_{ijkl}^{(m-1)} = (b_1 E_1 + b_2 E_2 + b_3 E_3 + b_4 E_4)^m \gamma_{ijkl} \\
&= \sum_{i'+j'+k'+l'=m} B_{i'j'k'l'}^m(\boldsymbol{p}) E_1^{i'} E_2^{j'} E_3^{k'} E_4^{l'} \gamma_{ijkl} = \sum_{i'+j'+k'+l'=m} \gamma_{i+i',j+j',k+k',l+l'} B_{i'j'k'l'}^m(\boldsymbol{p}).
\end{aligned}
$$

∎

33

Lemma A.10 *Consdier a trivariate polynomial $\phi(\boldsymbol{p})$ with $\phi(\boldsymbol{p}) = \sum_{i+j+k+l=d} \gamma_{ijkl} B_{ijkl}^d(\boldsymbol{p})$ as its B-form. Then for any $1 \leq n \leq d$, the n-th order directional derivative of ϕ with respect to the direction $\boldsymbol{u} = \boldsymbol{v}_4 - \boldsymbol{v}_2$ is given by*

$$D_{\boldsymbol{u}}^n \phi(\boldsymbol{p}) = \frac{d!}{(d-n)!} \sum_{i+j+k+l=d-n} \gamma_{ijkl}^{(n)} B_{ijkl}^{d-n}(\boldsymbol{p},$$

with

$$\gamma_{ijkl}^{(n)} = \sum_{m=0}^{n} \binom{n}{m}(-1)^m \gamma_{i+m,j,k,l+n-m}, \qquad i+j+k+l = d-n.$$

Proof The directional coordinates of $\boldsymbol{u} = \boldsymbol{v}_4 - \boldsymbol{v}_2$ is $\boldsymbol{a} = (-1,0,0,1)$. By Equation A.13,

$$\gamma_{ijkl}^{(n)} = \sum_{i'+j'+k'+l'=n} \gamma_{i+i',j+j',k+k',l+l'} B_{i'j'k'l'}^n(\boldsymbol{p}) = \sum_{i'+l'=n} \gamma_{i+i',j,k,l+l'} \frac{n!}{i'!l'!}(-1)^{i'}$$

$$= \sum_{m=0}^{n} \binom{n}{m}(-1)^m \gamma_{i+m,j,k,l+n-m}, \qquad i+j+k+l = d-n.$$

∎

Proof of Theorem A.5 The equivalence between Statements 1 and 2 is obvious by definition. Thus, we just show the equivalence between Statements 2 and 3.

i) We start to consider when $r = 0$. It is equivalent to consider directions $\widetilde{\boldsymbol{u}}$ along F, where the directional coordinates are $(0, 1 - \widetilde{b}_3 - \widetilde{b}_4, \widetilde{b}_3, \widetilde{b}_4)$ and $(0, 1 - \widetilde{b}_3 - \widetilde{b}_4, \widetilde{b}_4, \widetilde{b}_3)$ with respect to the tetrahedron T and $\widetilde{T}$, respectively. Thus, ϕ and $\widetilde{\phi}$ join continuously along F if and only if

$$\sum_{j+k+l=d} \gamma_{0jkl} \frac{d!}{j!k!l!} (1 - \widetilde{b}_3 - \widetilde{b}_4)^j (\widetilde{b}_3)^k (\widetilde{b}_4)^l = \sum_{j+k+l=d} \widetilde{\gamma}_{0jkl} \frac{d!}{j!k!l!} (1 - \widetilde{b}_3 - \widetilde{b}_4)^j (\widetilde{b}_4)^k (\widetilde{b}_3)^l.$$

That is, $\gamma_{0jkl} = \widetilde{\gamma}_{0jlk}$, with $j + k + l = d$, which matches our conclusion in Equation A.7.

ii) Then we consider for $r > 0$. First note that

$$D_{\boldsymbol{u}}^n \phi(\boldsymbol{p}_F) = D_{\boldsymbol{u}}^n \widetilde{\phi}(\boldsymbol{p}_F) \tag{A.14}$$

holds for any $\boldsymbol{p}_F \in F$, $n = 0, \ldots, r$, if and only if Equation A.14 holds for the direction $\boldsymbol{u} = \boldsymbol{v}_5 - \boldsymbol{v}_3$. It is a fact because by the argument in i), all derivatives of ϕ and $\widetilde{\phi}$ corresponding to the directions $\boldsymbol{v}_3 - \boldsymbol{v}_1$ and $\boldsymbol{v}_4 - \boldsymbol{v}_1$, agree at every point on F. And derivatives in all other directions can be written as linear combinations of $D_{\boldsymbol{u}}$, $D_{\boldsymbol{v}_3-\boldsymbol{v}_1}$ and $D_{\boldsymbol{v}_4-\boldsymbol{v}_1}$.

Let $\boldsymbol{b} = (b_1, b_2, b_3, b_4)$ be the barycentric coordinates of $\boldsymbol{v}_5$ relative to the tetrahedron T. Correspondingly, the directional coordinates of $\boldsymbol{u}$ are $\boldsymbol{a} = (b_1, b_2, b_3 - 1, b_4)$ and $\widetilde{\boldsymbol{a}} = (1, 0, 0, -1)$ with respect to T and $\widetilde{T}$, respectively.

By Theorem A.4, for each $0 \leq n \leq r$,

$$D_{\boldsymbol{u}}^n \phi(\boldsymbol{p}_F)|_F = \frac{d!}{(d-n)!} \sum_{j+k+l=d-n} \gamma_{0jkl}^{(n)}(\boldsymbol{a}) B_{0jkl}^{d-n}(\boldsymbol{p}_F),$$

$$D_{\boldsymbol{u}}^n \widetilde{\phi}(\boldsymbol{p}_F)|_F = \frac{d!}{(d-n)!} \sum_{j+k+l=d-n} \gamma_{0jkl}^{(n)}(\widetilde{\boldsymbol{a}}) B_{0jkl}^{d-n}(\boldsymbol{p}_F).$$

34

Since for points $\boldsymbol{p}_F \in F$, $\widetilde{B}_{0jkl}^{d-n}(\boldsymbol{p}_F) = B_{0jlk}^{d-n}(\boldsymbol{p}_F)$, it follows that Equation A.14 holds if and only if for $j + k + l = d - n$, $n = 0, \ldots, r$,

$$\widetilde{\gamma}_{0jkl}^{(n)}(\widetilde{\boldsymbol{a}}) = \gamma_{0jlk}^{(n)}(\boldsymbol{a}). \tag{A.15}$$

By Lemma A.10, for $j + k + l = d - n$,

$$\widetilde{\gamma}_{0jkl}^{(n)}(\widetilde{\boldsymbol{a}}) = \sum_{m=0}^{n} (-1)^{n-m} \binom{n}{m} \widetilde{\gamma}_{m,j,k,d-m-j-k}. \tag{A.16}$$

In another direction, following the proof of Lemma A.9, for $j + l + k = d - n$,

$$\begin{aligned}
\gamma_{0jlk}^{(n)}(\boldsymbol{a}) &= \{b_1 E_1 + b_2 E_2 + (b_3 - 1)E_3 + b_4 E_4\}^n \, \gamma_{0jlk} \\
&= (b_1 E_1 + b_2 E_2 + b_3 E_3 + b_4 E_4 - E_3)^n \, \gamma_{0jlk} \\
&= \sum_{m=0}^{n} (-1)^{n-m} \binom{n}{m} (b_1 E_1 + b_2 E_2 + b_3 E_3 + b_4 E_4)^n \, \gamma_{0,j,l+n-m,k} \\
&= \sum_{m=0}^{n} (-1)^{n-m} \binom{n}{m} \gamma_{0,j,d-j-k-m,k}^{(m)}(\boldsymbol{b}). \tag{A.17}
\end{aligned}$$

Combining Equations A.16 and A.17, Equation A.15 holds if and only if for $j+k+l = d-n$ and $n = 0, \ldots, r$, $\widetilde{\gamma}_{njkl} = \gamma_{0jlk}^{(n)}(\boldsymbol{b})$. By Lemma A.9,

$$\gamma_{0jlk}^{(n)}(\boldsymbol{b}) = \sum_{i'+j'+k'+l'=n} \gamma_{i',j+j',k+k',l+l'} B_{i'j'k'l'}^n(\boldsymbol{v}_5).$$

Thus, Equation A.14 holds if and only if

$$\widetilde{\gamma}_{njkl} = \sum_{i'+j'+k'+l'=n} \gamma_{i',j+j',k+k',l+l'} B_{i'j'k'l'}^n(\boldsymbol{v}_5),$$

that is, Equation A.7 follows. ∎

A.5 Directional Derivatives for Basis Functions

Based on the conclusion in Equation A.6 in Appendix A, the mth order directional derivatives for all the Bernstein basis functions with degree d can be written as some linear combination of Bernstein basis functions with degree $d - m$. Specifically, for directions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$, there exists a $\binom{d+3}{3} \times \binom{d+3-m}{3}$ matrix $\mathbf{C}_d^{(m)}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ such that

$$D_{\boldsymbol{u}_m} \cdots D_{\boldsymbol{u}_1} \mathbf{B}_d(\boldsymbol{p}) = \mathbf{C}_d^{(m)}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m) \mathbf{B}_{d-m}(\boldsymbol{p}).$$

Consequently, with $D_{\boldsymbol{u}_{m-1}} \cdots D_{\boldsymbol{u}_1} \mathbf{B}_d(\boldsymbol{p}) = \mathbf{C}_d^{(m-1)}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{m-1}) \mathbf{B}_{d-m+1}(\boldsymbol{p})$, we have

$$\begin{aligned}
D_{\boldsymbol{u}_m} D_{\boldsymbol{u}_{m-1}} \cdots D_{\boldsymbol{u}_1} \mathbf{B}_d(\boldsymbol{p}) &= \mathbf{C}_d^{(m-1)}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{m-1}) D_{\boldsymbol{u}_m} \mathbf{B}_{d-m+1}(\boldsymbol{p}) \\
&= \mathbf{C}_d^{(m-1)}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{m-1}) \mathbf{C}_{d-m+1}^{(1)}(\boldsymbol{u}_m) \mathbf{B}_{d-m}(\boldsymbol{p}),
\end{aligned}$$

which implies $\mathbf{C}_d^{(m)}(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m) = \mathbf{C}_d^{(m-1)}(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{m-1})\mathbf{C}_{d-m+1}^{(1)}(\boldsymbol{u}_m)$. Keep decomposing matrix $\mathbf{C}_d^{(m-1)}(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{m-1})$, we have $\mathbf{C}_d^{(m)}(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m) = \mathbf{C}_d^{(1)}(\boldsymbol{u}_1)\cdots\mathbf{C}_{d-m+1}^{(1)}(\boldsymbol{u}_m)$.

Therefore, we can obtain the explicit form of the mth order directional derivatives for the Bernstein basis functions with degree d, given the explicit form of matrix $\mathbf{C}_d^{(1)}(\boldsymbol{u})$. Based on Lemma A.3,

$$D_{\boldsymbol{u}}B_{ijkl}^d(\boldsymbol{p}) = d\left\{a_1 B_{i-1,j,k,l}^{d-1}(\boldsymbol{p}) + a_2 B_{i,j-1,k,l}^{d-1}(\boldsymbol{p}) + a_3 B_{i,j,k-1,l}^{d-1}(\boldsymbol{p}) + a_4 B_{i,j,k,l-1}^{d-1}(\boldsymbol{p})\right\},$$

where (a_1, a_2, a_3, a_4) is the barycentric coordinate of direction $\boldsymbol{u}$. Then, the $\{\sum_{m=0}^{d-i}(m+1)m/2 + \sum_{n=0}^{d-i-j}(n+1) - k\}$th row of matrix $\mathbf{C}_d^{(1)}(\boldsymbol{u})$ is $d(a_1\boldsymbol{e}_{\mathcal{I}_1} + a_2\boldsymbol{e}_{\mathcal{I}_2} + a_3\boldsymbol{e}_{\mathcal{I}_3} + a_4\boldsymbol{e}_{\mathcal{I}_4})$, where $\boldsymbol{e}_{\mathcal{I}}$ is unit vector with $\mathcal{I}$th element being one, $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$ and $\mathcal{I}_4$ are indexes of basis functions $B_{i-1,j,k,l}^{d-1}(\boldsymbol{p})$, $B_{i,j-1,k,l}^{d-1}(\boldsymbol{p})$, $B_{i,j,k-1,l}^{d-1}(\boldsymbol{p})$, and $B_{i,j,k,l-1}^{d-1}(\boldsymbol{p})$.

In the following, we provide a simple example to illustrate how to calculate the first and second-order derivatives and therefore to construct the penalty matrix $\mathbf{P}_T$ in Equation A.18 in Section A.6. Since the penalty function is calculated through the second-order derivatives in terms of the three-dimensional Cartesian coordinate system $x = (1,0,0)$, $y = (0,1,0)$ and $z = (0,0,1)$, in the following example, we only consider directional derivatives for x, y, z.

Consider the basis functions on the tetrahedron T in the Figure A.1 with $d = 3$. The horizontal axis $x = (1,0,0)$ can be written as $x = \boldsymbol{v}_1 - \boldsymbol{v}_2$, therefore, the barycentric coordinates of x is $(1,-1,0,0)$. Then, the first and second order derivatives of all the Bernstein basis functions $\mathbf{B}_3(\boldsymbol{p})$ are $D^{(1,0,0)}\mathbf{B}_3(\boldsymbol{p}) = \mathbf{C}_3^{(1)}(1,0,0)\mathbf{B}_2(\boldsymbol{p})$ and $D^{(2,0,0)}\mathbf{B}_3(\boldsymbol{p}) = \mathbf{C}_3^{(2)}(x,x)\mathbf{B}_1(\boldsymbol{p}) = \mathbf{C}_3^{(1)}(x)\mathbf{C}_2^{(1)}(x)\mathbf{B}_1(\boldsymbol{p})$, where matrices $\mathbf{C}_2^{(1)}(x)$ and $\mathbf{C}_3^{(1)}(x)$ are

$$\mathbf{C}_2^{(1)}(x) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{C}_3^{(1)}(x) = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

A.6 Details of Constructing Penalty Matrix

Next, we introduce the details of constructing penalty $\mathbf{P}$ matrix. Following the definition of penalty matrix $\mathbf{P}_T$ in the Section 3, we can further implement as

$$
\begin{aligned}
\mathcal{E}(s_T) &= \sum_{|\boldsymbol{\alpha}|=2} \binom{2}{\alpha_1} \binom{2-\alpha_1}{\alpha_2} \int_T \left\{ \sum_{i+j+k+l=d} \gamma_{T;ijkl} D^{\boldsymbol{\alpha}} B_{ijkl}^{d,T}(\boldsymbol{p}) \right\}^2 \mathrm{d}\boldsymbol{p} \\
&= \sum_{|\boldsymbol{\alpha}|=2} \binom{2}{\alpha_1} \binom{2-\alpha_1}{\alpha_2} \sum_{i+j+k+l=d} \sum_{i'+j'+k'+l'=d} \gamma_{T;ijkl}\gamma_{T;i'j'k'l'} \times \\
&\quad \int_T \left\{ D^{\boldsymbol{\alpha}} B_{ijkl}^{d,T}(\boldsymbol{p}) \right\} \left\{ D^{\boldsymbol{\alpha}} B_{i'j'k'l'}^{d,T}(\boldsymbol{p}) \right\} \mathrm{d}\boldsymbol{p} \\
&= \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{\gamma}_T^\top \mathbf{P}_T^{\boldsymbol{\alpha}} \boldsymbol{\gamma}_T = \boldsymbol{\gamma}_T^\top \mathbf{P}_T \boldsymbol{\gamma}_T,
\end{aligned}
\tag{A.18}
$$

where each $\mathbf{P}_T^{\boldsymbol{\alpha}}$ is a $\binom{d+3}{3} \times \binom{d+3}{3}$ matrix with entries $\int_T \{D^{\boldsymbol{\alpha}} B_{ijkl}^{d,T}(\boldsymbol{p})\} \{D^{\boldsymbol{\alpha}} B_{i'j'k'l'}^{d,T}(\boldsymbol{p})\} \mathrm{d}\boldsymbol{p}$ for $\boldsymbol{\alpha}$ satisfying $|\boldsymbol{\alpha}| = 2$. Applying the results of directional derivatives in Section A.5, the second order derivative of $\mathbf{B}_d(\boldsymbol{p})$ at directions $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are

$$
D_{\boldsymbol{u}_2} D_{\boldsymbol{u}_1} \mathbf{B}_d(\boldsymbol{p}) = \mathbf{C}_d^{(2)}(\boldsymbol{u}_1, \boldsymbol{u}_2) \mathbf{B}_{d-2}(\boldsymbol{p}).
$$

For notation simplicity, we denote $\mathbf{C}_d^{\boldsymbol{\alpha}}$ the derivative coefficient matrix for $D^{\boldsymbol{\alpha}} \mathbf{B}_d(\boldsymbol{p})$ with respect to direction(s) of Cartesian coordinate system, that is, $D^{\boldsymbol{\alpha}} \mathbf{B}_d(\boldsymbol{p}) = \mathbf{C}_d^{\boldsymbol{\alpha}} \mathbf{B}_{d-|\boldsymbol{\alpha}|}(\boldsymbol{p})$. Therefore, for $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| = 2$, we have

$$
\int_T \{D^{\boldsymbol{\alpha}} \mathbf{B}_d(\boldsymbol{p})\} \{D^{\boldsymbol{\alpha}} \mathbf{B}_d(\boldsymbol{p})\}^\top \mathrm{d}\boldsymbol{p} = \mathbf{C}_d^{\boldsymbol{\alpha}} \left\{ \int_T \mathbf{B}_{d-2}(\boldsymbol{p}) \mathbf{B}_{d-2}(\boldsymbol{p})^\top \mathrm{d}\boldsymbol{p} \right\} (\mathbf{C}_d^{\boldsymbol{\alpha}})^\top.
$$

By the Lemma 15.29 in Lai and Schumaker (2007), we have

$$
\begin{aligned}
\int_T B_{\nu\mu\kappa\delta}^{d-2,T}(\boldsymbol{p}) B_{\nu'\mu'\kappa'\delta'}^{d-2,T}(\boldsymbol{p}) \mathrm{d}\boldsymbol{p} &= \frac{\binom{\nu+\nu'}{\nu}\binom{\mu+\mu'}{\mu}\binom{\kappa+\kappa'}{\kappa}\binom{\delta+\delta'}{\delta}}{\binom{2d-4}{d-2}} \int_T B_{\nu+\nu',\mu+\mu',\kappa+\kappa',\delta+\delta'}^{2d-4,T}(\boldsymbol{p}) \mathrm{d}\boldsymbol{p} \\
&= \frac{\binom{\nu+\nu'}{\nu}\binom{\mu+\mu'}{\mu}\binom{\kappa+\kappa'}{\kappa}\binom{\delta+\delta'}{\delta}}{\binom{2d-4}{d-2}\binom{2d-1}{3}} V_T,
\end{aligned}
$$

recall that V_T is the volume of tetrahedron T. Let $\mathbf{L}_T^{d-2} = \int_T \mathbf{B}_{d-2}(\boldsymbol{p}) \mathbf{B}_{d-2}(\boldsymbol{p})^\top \mathrm{d}\boldsymbol{p}$ be the $\binom{d+1}{3} \times \binom{d+1}{3}$ matrix with entries $\int_T B_{\nu\mu\kappa\delta}^{d-2,T}(\boldsymbol{p}) B_{\nu'\mu'\kappa'\delta'}^{d-2,T}(\boldsymbol{p}) \mathrm{d}\boldsymbol{p}$. We finally obtain $\mathbf{P}_T^{\boldsymbol{\alpha}} = \mathbf{C}_d^{\boldsymbol{\alpha}} \mathbf{L}_T^{d-2} (\mathbf{C}_d^{\boldsymbol{\alpha}})^\top$ and $\mathbf{P}_T = \sum_{|\boldsymbol{\alpha}|=2} \mathbf{P}_T^{\boldsymbol{\alpha}}$. Consequently, $\mathcal{E}(s) = \boldsymbol{\gamma}^\top \mathbf{P} \boldsymbol{\gamma}$.

Following the same example in Section A.5, similarly, the barycentric coordinates of $y = \boldsymbol{v}_3 - \boldsymbol{v}_2$ and $z = \boldsymbol{v}_4 - \boldsymbol{v}_2$ are $(0, -1, 1, 0)$ and $(0, -1, 0, 1)$, respectively. Following similar inductions, the second directional derivative of the Bernstein basis polynomials are the linear combination of $\{B_{1000}^{1,T}(\boldsymbol{p}), B_{0100}^{1,T}(\boldsymbol{p}), B_{0010}^{1,T}(\boldsymbol{p}), B_{0001}^{1,T}(\boldsymbol{p})\}$. For example, given the formulae of $\mathbf{C}_2^{(1)}$ and $\mathbf{C}_3^{(1)}$ shown above, we have $D^{(1,0,0)} B_{2100}^{3,T}(\boldsymbol{p}) = 3B_{1100}^{2,T}(\boldsymbol{p}) - 3B_{2000}^{2,T}(\boldsymbol{p})$, and $D^{(2,0,0)} B_{2100}^{3,T}(\boldsymbol{p}) = -12B_{1000}^{1,T}(\boldsymbol{p}) + 6B_{0100}^{1,T}(\boldsymbol{p})$. According to the results in Sections A.5

and A.6, the coefficient matrices for second order derivatives are $\mathbf{C}_3(2,0,0)$, $\mathbf{C}_3(0,2,0)$, $\mathbf{C}_3(0,0,2)$, $\mathbf{C}_3(1,1,0)$, $\mathbf{C}_3(1,0,1)$, and $\mathbf{C}_3(0,1,1)$:

$$
\mathbf{C}_3(2,0,0) \quad\quad \mathbf{C}_3(0,2,0) \quad\quad \mathbf{C}_3(0,0,2)
$$

$$
\begin{pmatrix}
6 & 0 & 0 & 0 \\
-12 & 6 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & 0 & 0 & 6 \\
6 & -12 & 0 & 0 \\
0 & 0 & -12 & 0 \\
0 & 0 & 0 & -12 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 6 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
-12 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 6 & 0 & 0 \\
0 & -12 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 6 & -12 & 0 \\
0 & 0 & 0 & -12 \\
0 & 0 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
-12 & 0 & 0 & -12 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
0 & 6 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & -12 & 0 & 6 \\
0 & 0 & 0 & 0 \\
0 & 0 & -12 & 0 \\
0 & 6 & 0 & -12 \\
0 & 0 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 0 & 0 & 0
\end{pmatrix},
$$

$$
\mathbf{C}_3(1,1,0) \quad\quad \mathbf{C}_3(1,0,1) \quad\quad \mathbf{C}_3(0,1,1)
$$

$$
\begin{pmatrix}
0 & 0 & 0 & 0 \\
-6 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & -6 & 0 & 0 \\
-6 & 6 & -6 & 0 \\
0 & 0 & 0 & -6 \\
0 & 0 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 0 & 0 & 0 \\
0 & 6 & 0 & 0 \\
0 & -6 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 0 & -6 & 0 \\
0 & 0 & 0 & -6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 0 & 0 & 0 \\
-6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
6 & -6 & 0 & 0 \\
0 & 0 & -6 & 0 \\
-6 & 6 & 0 & -6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & 0 & 0 & 6 \\
0 & 6 & 0 & 0 \\
0 & 0 & 6 & 0 \\
0 & -6 & 0 & 6 \\
0 & 0 & 0 & 0 \\
0 & 0 & -6 & 0 \\
0 & 0 & 0 & -6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
-6 & 0 & 0 & 0 \\
-6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 6 & 0 & 0 \\
0 & -6 & 6 & 0 \\
0 & -6 & 0 & 6 \\
0 & 0 & -6 & 0 \\
0 & 6 & -6 & -6 \\
0 & 0 & 0 & -6 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
$$

Therefore, we have

$$\mathbf{L}_T^1 = \int_T \mathbf{B}_1(\boldsymbol{p})\mathbf{B}_1(\boldsymbol{p})^\top \mathrm{d}\boldsymbol{p} = \frac{1}{12}\begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \text{ and } \mathbf{P}_T = \sum_{|\boldsymbol{\alpha}|=2} \mathbf{C}_3(\boldsymbol{\alpha})\mathbf{L}_T^1\{\mathbf{C}_3(\boldsymbol{\alpha})\}^\top.$$

Appendix B. Proof of Theoretical Results

B.1 Preliminaries

In this section, we start with the stability condition for trivariate spline bases over triangulations, and establish the uniform rate of the approximation of empirical inner product to the theoretical inner product based on the stability property.

Lemma B.1 *Let* $\{B_\xi\}_{\xi\in\mathcal{M}}$ *be the basis for* $\mathcal{S}_d^r(\triangle)$ *constructed in Lai and Schumaker (2007), where* $\mathcal{M}$ *stands for the index set of spline bases. Then there exist positive constants* C_1, C_2 *depending on degree* d *and partition quasi-uniform parameter* β *such that*

$$C_1|\triangle|^3 \sum_{\xi\in\mathcal{M}} |c_\xi|^2 \leq \left\|\sum_{\xi\in\mathcal{M}} c_\xi B_\xi\right\|_{L^2(\Omega)}^2 \leq C_2|\triangle|^3 \sum_{\xi\in\mathcal{M}} |c_\xi|^2,$$

for all $c_\xi, \xi \in \mathcal{M}$.

Proof The proof of Lemma B.1 follows directly from the Theorem 17.18 in Lai and Schumaker (2007). ∎

We illustrate the connection between the theoretical norm and L^2 norm in Lemma B.2.

Lemma B.2 *Let* $\{B_\xi\}_{\xi\in\mathcal{M}}$ *be the basis for* $\mathcal{S}_d^r(\triangle)$ *as in Lemma* B.1. *Under Assumption (A3), if* $d \geq 6r+3$ *and* $\triangle$ *is a* β-*quasi-uniform triangulation, there exist positive constants* C_1, C_2 *such that*

$$C_1 \left\|\sum_{\xi\in\mathcal{M}} c_\xi B_\xi\right\|_{L^2(\Omega)}^2 \leq \left\|\sum_{\xi\in\mathcal{M}} c_\xi B_\xi\right\|_{\Omega}^2 \leq C_2 \left\|\sum_{\xi\in\mathcal{M}} c_\xi B_\xi\right\|_{L^2(\Omega)}^2,$$

for all $c_\xi, \xi \in \mathcal{M}$.

Proof It is straightforward to obtain upper and lower bounds of $\|\sum_{\xi\in\mathcal{M}} c_\xi B_\xi\|_\Omega^2$ by Assumption (A3). ∎

Proof of Lemma 3. According to the definition of empirical inner product and induced norm of the theoretical L^2 inner product in Section 3,

$$\langle g_1, g_2\rangle_{n,\Omega} = \frac{1}{n}\sum_{i=1}^n \left\{\sum_{\xi\in\mathcal{M}} c_\xi B_\xi(\boldsymbol{p}_i)\right\} \left\{\sum_{\zeta\in\mathcal{M}} \tilde{c}_\zeta B_\zeta(\boldsymbol{p}_i)\right\} = \sum_{\xi,\zeta\in\mathcal{M}} c_\xi \tilde{c}_\zeta \langle B_\xi, B_\zeta\rangle_{n,\Omega},$$

$$\|g_1\|_\Omega^2 = \sum_{\xi,\xi'\in\mathcal{M}} c_\xi c_{\xi'} \left\langle B_\xi, B_{\xi'}\right\rangle_\Omega, \text{ and } \|g_2\|_\Omega^2 = \sum_{\zeta,\zeta'\in\mathcal{M}} \tilde{c}_\zeta \tilde{c}_{\zeta'} \left\langle B_\zeta, B_{\zeta'}\right\rangle_\Omega.$$

By Lemma B.1, we have for g_1 and g_2, $C_1|\triangle|^3 \sum_{\xi\in\mathcal{M}} |c_\xi|^2 \le \|g_1\|_{L^2(\Omega)}^2 \le C_2|\triangle|^3 \sum_{\xi\in\mathcal{M}} |c_\xi|^2$ and $C_1|\triangle|^3 \sum_{\zeta\in\mathcal{M}} |\tilde{c}_\zeta|^2 \le \|g_2\|_{L^2(\Omega)}^2 \le C_2|\triangle|^3 \sum_{\zeta\in\mathcal{M}} |\tilde{c}_\zeta|^2$. Consequently, by Lemma B.2,

$$C_1|\triangle|^3 \left\{\sum_{\xi\in\mathcal{M}} |c_\xi|^2 \sum_{\zeta\in\mathcal{M}} |\tilde{c}_\zeta|^2\right\}^{1/2} \le \|g_1\|_\Omega \|g_2\|_\Omega \le C_2|\triangle|^3 \left\{\sum_{\xi\in\mathcal{M}} |c_\xi|^2 \sum_{\zeta\in\mathcal{M}} |\tilde{c}_\zeta|^2\right\}^{1/2}.$$

Thus,

$$\begin{aligned}
R_n &\le \frac{\sum_{\xi,\zeta\in\mathcal{M}} |c_\xi \tilde{c}_\zeta|}{C_1|\triangle|^3 \left\{\sum_{\xi\in\mathcal{M}} |c_\xi|^2 \sum_{\zeta\in\mathcal{M}} |\tilde{c}_\zeta|^2\right\}^{1/2}} \max_{\xi,\zeta\in\mathcal{M}} \left|\langle B_\xi, B_\zeta\rangle_{n,\Omega} - \langle B_\xi, B_\zeta\rangle_\Omega\right| \\
&\le C_1^{-1} |\triangle|^{-3} \max_{\xi,\zeta\in\mathcal{M}} \left|\langle B_\xi, B_\zeta\rangle_{n,\Omega} - \langle B_\xi, B_\zeta\rangle_\Omega\right|.
\end{aligned} \tag{B.1}$$

To obtain the conclusion, it suffices to show that with probability one,

$$\max_{\xi,\zeta\in\mathcal{M}} \left|\langle B_\xi, B_\zeta\rangle_{n,\Omega} - \langle B_\xi, B_\zeta\rangle_\Omega\right| = O_P\left\{(\log n)^{1/2}(nN)^{-1/2}\right\}. \tag{B.2}$$

Let $R_{\xi,\zeta,i} = B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i) - \mathrm{E}B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i)$. Then we have the second moment $\mathrm{E}R_{\xi,\zeta,i}^2 = \mathrm{E}\{B_\xi^2(\boldsymbol{p}_i) B_\zeta^2(\boldsymbol{p}_i)\} - \{\mathrm{E}B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i)\}^2$, where we have $\mathrm{E}\{B_\xi^2(\boldsymbol{p}_i) B_\zeta^2(\boldsymbol{p}_i)\} \sim |\triangle|^3$ and $\{\mathrm{E}B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i)\}^2 \sim |\triangle|^6$. Hence, $\mathrm{E}R_{\xi,\zeta,i}^2 \sim |\triangle|^3$. Note that the k-th moment is $\mathrm{E}|R_{\xi,\zeta,i}|^k \le 2^{k-1}\{\mathrm{E}|B_\xi(\boldsymbol{p}_i)B_\zeta(\boldsymbol{p}_i)|^k + |\mathrm{E}B_\xi(\boldsymbol{p}_i)B_\zeta(\boldsymbol{p}_i)|^k\}$, where $\mathrm{E}|B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i)|^k \sim |\triangle|^3$, $|\mathrm{E}B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i)|^k \sim |\triangle|^{3k}$. Then there exists a constant $C > 0$ such that $\mathrm{E}|R_{\xi,\zeta,i}|^k \le C2^{k-1}k!\mathrm{E}R_{\xi,\zeta,i}^2$. Thus, $\{R_{\xi,\zeta,i}\}_{i=1}^n$ satisfying Cramer's condition with some constant $C \sim O(1)$. By Bernstein's inequality in Bosq (1998), for $\delta > 0$ large enough,

$$P\left(\frac{1}{n}\left|\sum_{i=1}^n R_{\xi,\zeta,i}\right| \ge \delta\sqrt{\frac{\log n}{nN}}\right) \le 2\exp\left(\frac{-\delta^2 \log n}{4 + 2c\delta\sqrt{N\log n/n}}\right) \le 2n^{-4}. \tag{B.3}$$

It is easy to see that the cardinality of $\mathcal{M}$ is $(d+1)(d+2)(d+3)N/6$ as discussed in Section 2.3. Thus, for the $\delta > 0$ in Equation B.3,

$$\sum_{n=1}^\infty P\left(\max_{\xi,\zeta\in\mathcal{M}} \left|\frac{1}{n}\sum_{i=1}^n R_{\xi,\zeta,i}\right| \ge \delta\sqrt{\frac{\log n}{nN}}\right) \le \frac{1}{6}\sum_{n=1}^\infty \frac{\{(d+1)(d+2)(d+3)N\}^2}{n^4} \le C\sum_{n=1}^\infty n^{-2} < \infty.$$

Borel-Cantelli Lemma entails that $\max_{\xi,\zeta\in\mathcal{M}} |n^{-1}\sum_{l=1}^n R_{\xi,\zeta,i}| = O_P\{(\log n)^{1/2}(nN)^{-1/2}\}$. The desired result follows from Equations B.1 and B.2. $\blacksquare$

As a direct result of Lemma 3, we have

$$\sup_{g\in\mathcal{S}_d^r(\triangle)} \left|\|g\|_{n,\Omega}^2 \Big/ \|g\|_\Omega^2 - 1\right| = O_P\left\{(N\log n)^{1/2}n^{-1/2}\right\}. \tag{B.4}$$

B.2 Size of the Bias and Noise Terms (Propositions 5—7)

In this section, we measure the size of the bias term and show the proofs of Proposition 5—7. We firstly proof Lemma 4.

Proof of Lemma 4. According to Lemma 15.2 of Lai and Schumaker (2007), for any $g \in \mathcal{S}_d^r(\triangle)$,

$$\|g\|_\infty = \|g\|_{\infty,T} \le K_1 V_T^{-1/2} \|g\|_{L^2(T)} \le \frac{K}{\varrho_\triangle^{3/2}} \|g\|_{L^2(T)} \le \frac{K}{\varrho_\triangle^{3/2}} \|g\|_{L^2(\Omega)} \le \frac{K_\beta}{|\triangle|^{3/2}} \|g\|_{L^2(\Omega)},$$

where K_β is a positive constant dependent on β and recall that V_T is the volume of tetrahedron T. Consequently, by the Markov's inequality Theorem 15.28 in Lai and Schumaker (2007), for any $g \in \mathcal{S}_d^r(\triangle)$, $\|g\|_{\infty,\Omega} \le K_\beta |\triangle|^{-3/2} \|g\|_{L^2(\Omega)}$ and $\|g\|_\mathcal{E} \le K_\beta |\triangle|^{-2} \|g\|_{L^2(\Omega)}$. Equation B.4 implies that

$$\sup_{g \in \mathcal{S}_d^r(\triangle)} \left\{ \|g\|_{n,\Omega} \big/ \|g\|_{L^2(\Omega)} \right\} \ge \left[1 - O_P \left\{ (N \log n)^{1/2} n^{-1/2} \right\} \right]^{1/2}.$$

Thus, we have

$$
\begin{aligned}
V_n &\le K_\beta |\triangle|^{-3/2} \left[1 - O_P \left\{ (N \log n)^{1/2} n^{-1/2} \right\} \right]^{-1/2} = O_P \left(|\triangle|^{-3/2} \right), \\
\overline{V}_n &\le K_\beta |\triangle|^{-2} \left[1 - O_P \left\{ (N \log n)^{1/2} n^{-1/2} \right\} \right]^{-1/2} = O_P \left(|\triangle|^{-2} \right).
\end{aligned}
$$

∎

Proof of Proposition 5. By triangle inequality and the definition of V_n,

$$
\begin{aligned}
\|m - s_{\rho_n,m}\|_{\infty,\Omega} &\le \|m - s_{0,m}\|_{\infty,\Omega} + \|s_{0,m} - s_{\rho_n,m}\|_{\infty,\Omega} \\
&\le \|m - s_{0,m}\|_{\infty,\Omega} + V_n \|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega}.
\end{aligned}
\tag{B.5}
$$

By the definition of $s_{\rho_n,m}$, we have $\forall u \in \mathcal{S}_d^r(\triangle)$, $t \in \mathbb{R}_+$,

$$n \|m - s_{\rho_n,m}\|_{n,\Omega}^2 + \rho_n \|s_{\rho_n,m}\|_\mathcal{E}^2 \le n \|m - s_{\rho_n,m} - tu\|_{n,\Omega}^2 + \rho_n \|s_{\rho_n,m} + tu\|_\mathcal{E}^2$$

$$\Rightarrow \quad n \langle m - s_{\rho_n,m}, u \rangle_{n,\Omega} - \rho_n \langle s_{\rho_n,m}, u \rangle_\mathcal{E} \le \frac{t}{2} \left(\|u\|_{n,\Omega}^2 + \rho_n \|u\|_\mathcal{E}^2 \right).$$

Similarly, take $t \in \mathbb{R}_-$ and induct in a similar fashion, we thus obtain

$$\left| n \langle m - s_{\rho_n,m}, u \rangle_{n,\Omega} - \rho_n \langle s_{\rho_n,m}, u \rangle_\mathcal{E} \right| \le \frac{t}{2} \left(\|u\|_{n,\Omega}^2 + \rho_n \|u\|_\mathcal{E}^2 \right).
\tag{B.6}$$

Since Equation B.6 holds for any $t \in \mathbb{R}$, we can conclude that the penalized spline $s_{\rho_n,m}$ of m is characterized by the orthogonality relations

$$n \langle m - s_{\rho_n,m}, u \rangle_{n,\Omega} = \rho_n \langle s_{\rho_n,m}, u \rangle_\mathcal{E}, \quad \text{for all } u \in \mathcal{S}_d^r(\triangle).
\tag{B.7}$$

When $\rho_n = 0$, we have for $s_{0,m}$,

$$\langle m - s_{0,m}, u \rangle_{n,\Omega} = 0, \quad \text{for all } u \in \mathcal{S}_d^r(\triangle).
\tag{B.8}$$

Combining Equations B.7 and B.8, we obtain $n \langle s_{0,m} - s_{\rho_n,m}, u \rangle_{n,\Omega} = \rho_n \langle s_{\rho_n,m}, u \rangle_{\mathcal{E}}$, for all $u \in \mathcal{S}_d^r(\triangle)$. Inserting $u = s_{0,m} - s_{\rho_n,m}$ yields that

$$n \|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega}^2 = \rho_n \langle s_{\rho_n,m}, s_{0,m} - s_{\rho_n,m} \rangle_{\mathcal{E}}. \tag{B.9}$$

Thus, by Cauchy-Schwarz inequality and the definition of $\overline{V}_n$,

$$n \|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega}^2 \leq \rho_n \|s_{\rho_n,m}\|_{\mathcal{E}} \|s_{0,m} - s_{\rho_n,m}\|_{\mathcal{E}} \leq \overline{V}_n \rho_n \|s_{\rho_n,m}\|_{\mathcal{E}} \|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega}.$$

Similarly, using Equation B.9, we have

$$n \|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega}^2 = \rho_n \left\{ \langle s_{\rho_n,m}, s_{0,m} \rangle_{\mathcal{E}} - \langle s_{\rho_n,m}, s_{\rho_n,m} \rangle_{\mathcal{E}} \right\} \geq 0.$$

Thus, by Cauchy-Schwarz inequality, $\|s_{\rho_n,m}\|_{\mathcal{E}}^2 \leq \langle s_{\rho_n,g}, s_{0,m} \rangle_{\mathcal{E}} \leq \|s_{\rho_n,m}\|_{\mathcal{E}} \|s_{0,m}\|_{\mathcal{E}}$, which implies that $\|s_{\rho_n,m}\|_{\mathcal{E}} \leq \|s_{0,m}\|_{\mathcal{E}}$. Therefore

$$\|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega} \leq n^{-1} \overline{V}_n \rho_n \|s_{0,m}\|_{\mathcal{E}}. \tag{B.10}$$

Combining Equations B.5, B.10 and Lemma 1 in Section 2.3 in main part yields that

$$
\begin{aligned}
\|s_{0,m} - s_{\rho_n,m}\|_{\infty,\Omega} &\leq V_n \|s_{0,m} - s_{\rho_n,m}\|_{n,\Omega} \leq n^{-1} V_n \overline{V}_n \rho_n \|s_{0,m}\|_{\mathcal{E}} \\
&\leq n^{-1} V_n \overline{V}_n \rho_n C_1 \left(|m|_{2,\infty,\Omega} + \sum_{|\boldsymbol{\alpha}|=2} \|D^{\boldsymbol{\alpha}} (m - s_{0,m})\|_{\infty,\Omega} \right) \\
&\leq n^{-1} V_n \overline{V}_n \rho_n C_2 \left(|m|_{2,\infty,\Omega} + |\triangle|^{\ell-1} |m|_{\ell+1,\infty,\Omega} \right).
\end{aligned}
$$

Plugging the orders of V_n and $\overline{V}_n$, therefore,

$$\|s_{0,m} - s_{\rho_n,m}\|_{\infty,\Omega} = O_P \left\{ \frac{\rho_n}{n |\triangle|^{7/2}} \left(|m|_{2,\infty,\Omega} + |\triangle|^{\ell-1} |m|_{\ell+1,\infty,\Omega} \right) \right\}.$$

Hence by Equation B.5,

$$\|m - s_{\rho_n,m}\|_{\infty,\Omega} = O_P \left(|\triangle|^{\ell+1} |m|_{\ell+1,\infty,\Omega} \right) + O_P \left\{ \frac{\rho_n}{n |\triangle|^{7/2}} \left(|m|_{2,\infty,\Omega} + |\triangle|^{\ell-1} |m|_{\ell+1,\infty,\Omega} \right) \right\}.$$

Therefore, Proposition 5 is established. $\blacksquare$

Next we give the proof of Proposition 6.

Proof of Proposition 6. It is known from Lai and Schumaker (2007) that there is a locally support basis B_ξ, $\xi \in \mathcal{M}$ for $\mathcal{S}_d^r(\triangle)$. We write $s_{\rho_n,\epsilon}(\boldsymbol{p}) = \sum_{\xi \in \mathcal{M}} c_{\rho_n,\xi} B_\xi(\boldsymbol{p})$ for some coefficients $c_{\rho_n,\xi}$. It is easy to obtain the orthogonal relations for the penalized spline $s_{\rho_n,\epsilon}$ $n \langle s_{\rho_n,\epsilon} - \varepsilon, u \rangle_{n,\Omega} + \rho_n \langle s_{\rho_n,\epsilon}, u \rangle_{\mathcal{E}} = 0$, for all $u \in \mathcal{S}_d^r(\triangle)$. Consequently

$$\sum_{i=1}^n s_{\rho_n,\epsilon}(\boldsymbol{p}_i) B_\xi(\boldsymbol{p}_i) + \rho_n \langle s_{\rho_n,\epsilon}, B_\xi \rangle_{\mathcal{E}} = \sum_{i=1}^n B_\xi(\boldsymbol{p}_i) \sigma(\boldsymbol{p}_i) \varepsilon_i$$

for all $\xi \in \mathcal{M}$. Multiply $c_{\rho_n,\xi}$ to both sides and take the summation of $\xi \in \mathcal{M}$ and use Cauchy-Schwarz's inequality,

$$
\begin{aligned}
\|s_{\rho_n,\epsilon}\|^2_{n,\Omega} &\leq \|s_{\rho_n,\epsilon}\|^2_{n,\Omega} + \frac{\rho_n}{n}\mathcal{E}(s_{\rho_n,\epsilon}) = \frac{1}{n}\sum_{\xi\in\mathcal{M}} c_{\rho_n,\xi}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i \\
&\leq \left(\sum_{\xi\in\mathcal{M}}|c_{\rho_n,\xi}|^2\right)^{1/2}\left[\sum_{\xi\in\mathcal{M}}\left\{\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right\}^2\right]^{1/2} \\
&\leq \frac{1}{K_1|\triangle|^{3/2}}\|s_{\rho_n,\epsilon}\|_{L^2(\Omega)}\left[\sum_{\xi\in\mathcal{M}}\left\{\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right\}^2\right]^{1/2}.
\end{aligned}
$$

It follows from Equation B.4 that

$$
\|s_{\rho_n,\epsilon}\|^2_{L^2(\Omega)} \leq \|s_{\rho_n,\epsilon}\|^2_{n,\Omega} + O_P\left\{n^{-1/2}(\log n)^{1/2}N^{1/2}\|s_{\rho_n,\epsilon}\|^2_{L^2(\Omega)}\right\}
$$

$$
\leq \frac{\|s_{\rho_n,\epsilon}\|_{L^2(\Omega)}}{K_1|\triangle|^{3/2}}\left[\sum_{\xi\in\mathcal{M}}\left\{\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right\}^2\right]^{1/2} + O_P\left\{\frac{(N\log n)^{1/2}}{n^{1/2}}\|s_{\rho_n,\epsilon}\|^2_{L^2(\Omega)}\right\}.
$$

Next we note that $V_\Omega \leq C_1|\triangle|^3 N$. That is, $|\triangle| \geq (C_1 N)^{-1/3}V_\Omega^{1/3}$. We thus have

$$
\begin{aligned}
\|s_{\rho_n,\epsilon}\|_{L^2(\Omega)} &\leq \frac{(C_1 N)^{1/2}}{V_\Omega^{1/2}K_1}\left[\sum_{\xi\in\mathcal{M}}\left\{\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right\}^2\right]^{1/2} \\
&\quad + O_P\left\{n^{-1/2}(N\log n)^{1/2}\|s_{\rho_n,\epsilon}\|_{L^2(\Omega)}\right\}
\end{aligned}
$$

Thus, we have

$$
\left[1 - O_P\left\{\frac{(N\log n)^{1/2}}{n^{1/2}}\right\}\right]\|s_{\rho_n,\epsilon}\|_{L^2(\Omega)}
$$

$$
\leq \frac{CN^{1/2}}{V_\Omega^{1/2}}\left[\sum_{\xi\in\mathcal{M}}\left\{\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right\}^2\right]^{1/2} \tag{B.11}
$$

for a constant $C > 0$. Observing that the two random variables ε_i and $\boldsymbol{p}_i$ are independent, we have $\mathrm{E}\{n^{-1}\sum_{i=1}^n \varepsilon_i B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\}^2 = n^{-1}\mathrm{E}[B_\xi^2(\boldsymbol{p}_i)\sigma^2(\boldsymbol{p}_i)]$, where according to Lemma 15.2 in Lai and Schumaker (2007), the expectation of $B_\xi^2(\boldsymbol{p}_i)\sigma^2(\boldsymbol{p}_i)$ can be estimated as follows:

$$
\begin{aligned}
\mathrm{E}\left[B_\xi^2(\boldsymbol{p}_i)\sigma^2(\boldsymbol{p}_i)\right] &\leq C_\sigma^2\|B_\xi\|^2_{L^2(\Omega)} \\
&\leq c(\beta,\sigma)|\triangle|^3\|B_\xi\|^2_{\infty,\Omega} \leq C(\beta,\sigma)N^{-1}V_\Omega \tag{B.12}
\end{aligned}
$$

for positive constants $c(\beta,\sigma)$ and $C(\beta,\sigma)$ which depend only on β and σ. Thus, Assumption (A2) and Equation B.12 imply that $n^{-1}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i = O_P(n^{-1/2}N^{-1/2})$. Therefore,

we have $\sum_{\xi \in \mathcal{M}} \{n^{-1} \sum_{i=1}^{n} B_{\xi}(\boldsymbol{p}_i) \sigma(\boldsymbol{p}_i) \varepsilon_i\}^2 = O_P(n^{-1})$. Combining Equation B.11, we obtain that $\|s_{\rho_n, \epsilon}\|_{L^2(\Omega)} = O_P\left(N^{1/2} n^{-1/2}\right)$. So Proposition 6 is established. ∎

Let $\boldsymbol{\Gamma}_{\rho_n}$ be the symmetric positive definite matrix

$$\left[\frac{1}{n} \sum_{i=1}^{n} B_{\xi}(\boldsymbol{p}_i) B_{\zeta}(\boldsymbol{p}_i) + \frac{\rho_n}{n} \langle B_{\xi}, B_{\zeta} \rangle_{\mathcal{E}}\right]_{\xi, \zeta \in \mathcal{M}}. \tag{B.13}$$

We firstly show the bounds of $\boldsymbol{\Gamma}_{\rho_n}$.

Lemma B.3 *Suppose Assumption (A3) holds, $d \geq 6r + 3$, $\triangle$ is a β-quasi-uniform triangulation and $n^{-1} N \log n \to 0$ as $n \to \infty$. For $\boldsymbol{\Gamma}_{\rho_n}$ defined in Equation B.13, we have the following asymptotic properties:*

(i) As $n \to \infty$, for some constants $0 < c_\rho < C_\rho < \infty$, with probability approaching one,

$$c_\rho |\triangle|^3 \leq \rho_{\min}(\boldsymbol{\Gamma}_{\rho_n}) \leq \rho_{\max}(\boldsymbol{\Gamma}_{\rho_n}) \leq C_\rho \left(|\triangle|^3 + \frac{\rho_n}{n|\triangle|}\right).$$

(ii) There exists a constant $M_d > 0$ such that $\|\boldsymbol{\Gamma}_{\rho_n}^{-1}\|_\infty \leq M_d |\triangle|^{-3}$.

(iii) For every vector $\boldsymbol{a} = (a_1, \ldots, a_n)^\top$, there exists a constant $C_d > 0$ such that

$$\left\|\mathbf{B}^\top(\boldsymbol{p}) \, \boldsymbol{\Gamma}_{\rho_n}^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{B}(\boldsymbol{p}_i) a_i \right\|_\infty \leq C_d \|\boldsymbol{a}\|_\infty.$$

Proof (i) Let's randomly pick $\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{M}|}$, then for $g(\boldsymbol{p}) = \mathbf{B}^\top(\boldsymbol{p})\boldsymbol{\theta}$, we have $g \in \mathcal{S}_d^r(\triangle)$. Then

$$\boldsymbol{\theta}^\top \boldsymbol{\Gamma}_{\rho_n} \boldsymbol{\theta} = \boldsymbol{\theta}^\top \frac{1}{n} \sum_{i=1}^{n} \mathbf{B}(\boldsymbol{p}_i) \mathbf{B}^\top(\boldsymbol{p}_i) \boldsymbol{\theta} + \boldsymbol{\theta}^\top \frac{\rho_n}{n} [\langle B_m, B_{m'} \rangle_{\mathcal{E}}]_{m,m' \in \mathcal{M}} \boldsymbol{\theta} = \|g\|_{n,\Omega}^2 + \frac{\rho_n}{n} \|g\|_{\mathcal{E}}^2.$$

By Equation B.4, $\left|\|g\|_{n,\Omega}^2 / \|g\|_{L^2(\Omega)}^2 - 1\right| \leq R_n$, and combined with Lemma B.1, we have

$$c(1 - R_n)|\triangle|^3 \|\boldsymbol{\theta}\|^2 \leq (1 - R_n)\|g\|_{L^2(\Omega)}^2 \leq \|g\|_{n,\Omega}^2 \leq (1 + R_n)\|g\|_{L^2(\Omega)}^2 \leq C(1 + R_n)|\triangle|^3 \|\boldsymbol{\theta}\|^2.$$

Thus, $\rho_{\min}(\boldsymbol{\Gamma}_{\rho_n}) \geq c_\rho |\triangle|^3$ for some positive constant c_ρ.

On the other side, as shown in the proof of Proposition 5 and by Lemma B.1, $\|g\|_{\mathcal{E}}^2 \leq C|\triangle|^{-4} \|g\|_{L^2(\Omega)}^2 \leq C|\triangle|^{-1} \|\boldsymbol{\theta}\|^2$. Thus,

$$\rho_{\max}(\boldsymbol{\Gamma}_{\rho_n}) \leq C \left\{(1 + R_n)|\triangle|^3 + \frac{\rho_n}{n} \frac{1}{|\triangle|}\right\} \leq C_\rho \left(|\triangle|^3 + \frac{\rho_n}{n|\triangle|}\right),$$

for some positive constant C_ρ.

(ii) By (i), $\boldsymbol{\Gamma}_{\rho_n}$ is an invertible symmetric matrix, and its condition number $c_d = \rho_{\max}(\boldsymbol{\Gamma}_{\rho_n})/\rho_{\min}(\boldsymbol{\Gamma}_{\rho_n})$ satisfies $1 < c_d \leq c_\rho^{-1} C_\rho$.

According to the definition of banded matrix in DeVore and Lorentz (1993), a matrix $\mathbf{A} = (a_{ij})$ is said banded with bandwidth b if $a_{ij} = 0, |i - j| \geq b$, and if b is the smallest

integer with this property. Based on the construction of trivariate splines, $\mathbf{\Gamma}_{\rho_n}$ is a banded matrix with bandwidth $b = \binom{d+3}{3}$. By Theorem 13.4.3 in DeVore and Lorentz (1993), $\|\mathbf{\Gamma}_{\rho_n}^{-1}\|_\infty \leq 2\tau^{-2b}\|\mathbf{\Gamma}_{\rho_n}^{-1}\|_2(1-\tau)^{-1}$, where $\tau = (c_d^2 - 1/c_d^2 + 1)^{1/4b} < 1$. Therefore, there exist some positive constant M_d such that $\|\mathbf{\Gamma}_{\rho_n}^{-1}\|_\infty \leq M_d|\triangle|^{-3}$.

(iii) Combining (i) and (ii), it is straightforward to have

$$\left\|\mathbf{B}^\top(\boldsymbol{p})\mathbf{\Gamma}_{\rho_n}^{-1}\frac{1}{n}\sum_{i=1}^n \mathbf{B}(\boldsymbol{p}_j)a_j\right\|_\infty \leq \left\|\mathbf{B}^\top\right\|_\infty \|\mathbf{\Gamma}_{\rho_n}^{-1}\|_\infty \left\|\frac{1}{n}\sum_{i=1}^n \mathbf{B}(\boldsymbol{p}_i)\right\|_\infty \|\boldsymbol{a}\|_\infty \leq C_d\|\boldsymbol{a}\|_\infty.$$

∎

Lemma B.4 *Under Assumptions (A2) and (A4), $\|s_{0,\epsilon}\|_{\infty,\Omega} = O_P\{(\log n)^{1/2}n^{-1/2}|\triangle|^{-3/2}\}$.*

Proof Note that $s_{0,\epsilon}(\boldsymbol{p}) = \sum_{\xi\in\mathcal{M}} \widehat{c}_{0,\xi}B_\xi(\boldsymbol{p})$ for some coefficients $\widehat{c}_{0,\xi}$, so the order of $s_{\rho_n,\epsilon}(\boldsymbol{p})$ is related to that of $\widehat{c}_{0,\xi}$. In fact

$$\|s_{0,\epsilon}\|_{\infty,\Omega} \leq c\|\widehat{\mathbf{c}}_0\|_\infty = \left\|\mathbf{\Gamma}_0^{-1}\left[n^{-1}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right]_{\xi\in\mathcal{M}}\right\|_{\infty,\Omega},$$

where $\widehat{\mathbf{c}}_0 = (\widehat{c}_{0,\xi})_{\xi\in\mathcal{M}}$ and $\mathbf{\Gamma}_0$ is the symmetric positive definite matrix

$$\left[\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)B_\zeta(\boldsymbol{p}_i)\right]_{\xi,\zeta\in\mathcal{M}}$$

defined in Equation B.13. Thus,

$$\|s_{0,\epsilon}\|_{\infty,\Omega} \leq CN\max_{\xi\in\mathcal{M}}\left|n^{-1}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right|, \quad a.s.$$

Next, we show that with probability one,

$$\max_{\xi\in\mathcal{M}}\left|n^{-1}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\varepsilon_i\right| = O\left\{(\log n)^{1/2}(nN)^{-1/2}\right\}. \tag{B.14}$$

To prove Equation B.14, we decompose the noise variable ε_i into a truncated part and a tail part $\varepsilon_i = \varepsilon_{i,1}^{D_n} + \varepsilon_{i,2}^{D_n} + \mu^{D_n}$, where $D_n = n^\alpha$ with $\max\{1/(2+\eta), (1+\gamma)/2(1+\eta)\} < \alpha < (1-\gamma)/2$, and $\varepsilon_{i,1}^{D_n} = \varepsilon_i I\{|\varepsilon_i| > D_n\}$,

$$\varepsilon_{i,2}^{D_n} = \varepsilon_i I\{|\varepsilon_i| \leq D_n\} - \mu^{D_n}, \quad \mu^{D_n} = \mathrm{E}\left[\varepsilon_i I\{|\varepsilon_i| \leq D_n\}\right].$$

It is straightforward to see $\varepsilon_{i,2}^{D_n}$ has mean 0, thus is uniformly bounded by D_n^{-2}. And it is also straightforward to verify that μ^{D_n} is uniformly bounded by $D_n^{-(1+\eta)}$, so the boundedness of trivariate spline basis and of the function σ^2 entail that

$$\left|B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\mu^{D_n}\right| = O\left(D_n^{-(1+\eta)}\right). \tag{B.15}$$

Next we show that tail part vanishes almost surely. Recall that $E|\varepsilon_n^{2+\eta}| \leq v_\eta$ in Assumption (A2), so

$$\sum_{n=1}^{\infty} P\{|\varepsilon_n| > D_n\} \leq \sum_{n=1}^{\infty} \frac{E|\varepsilon_n|^{2+\eta}}{D_n^{2+\eta}} \leq v_\eta \sum_{n=1}^{\infty} D_n^{-(2+\eta)} < \infty.$$

By the Borel-Cantelli Lemma, we have

$$P\{\omega|\exists N(\omega), |\varepsilon_n(\omega)| \leq D_n \text{ for } n > N(\omega)\} = 1.$$

Let $v_\varepsilon = \max\{|\varepsilon_1|, |\varepsilon_2|, ..., |\varepsilon_{N(\omega)}|\}$ and there exists $N_1(\omega) > N(\omega)$, $D_{N_1(\omega)} > v_\varepsilon$. Since $D_n = n^\alpha$ is an increasing function, we have $D_n > D_{N_1(\omega)} > v_\varepsilon$, for $n > N_1(\omega)$. Thus,

$$P\{\omega|\exists N(\omega), |\varepsilon_i(\omega)| \leq D_n, \ 1 \leq i \leq n, \text{ for } n > N(\omega)\} = 1,$$

which implies that

$$P\left\{\omega|\exists N(\omega), \left|\varepsilon_{i,1}^{D_n}\right| = 0, \ 1 \leq i \leq n, \text{ for } n > N(\omega)\right\} = 1.$$

The boundedness of the spline basis implies that

$$\left|\frac{1}{n}\sum_{i=1}^{n} \sigma(\boldsymbol{p}_i) \varepsilon_{i,1}^{D_n} B_\xi(\boldsymbol{p}_i)\right| = O_{a.s.}\left(n^{-k}\right), \text{ for any } k > 0. \tag{B.16}$$

Next let $Z_i = n^{-1}\varepsilon_{i,2}^{D_n}\sigma(\boldsymbol{p}_i) B_\xi(\boldsymbol{p}_i)$. Since $\mathrm{E}(\varepsilon_{i,2}^{D_n}) = 0$, we have

$$\mathrm{Var}\left(\varepsilon_{i,2}^{D_n}\right) = \mathrm{E}\left(\varepsilon_i^2\right) - \mathrm{E}\left[\varepsilon_i^2 I\{|\varepsilon_i| > D_n\}\right] - \left(\mu^{D_n}\right)^2 = 1 + O_P\left\{D_n^{-\eta} + D_n^{-2(1+\eta)}\right\}.$$

According to Equation B.12, $\mathrm{E}[B_\xi^2(\boldsymbol{p}_i)\sigma^2(\boldsymbol{p}_i)] \leq C(\beta, \sigma)N^{-1}V_\Omega$. Note the independence of $\varepsilon_{i,2}^{D_n}, i = 1, \cdots, n$ and the independence between $\varepsilon_{i,2}^{D_n}$ and $B_\xi(\boldsymbol{p}_i)$, we therefore obtain that $V_n^2 = \mathrm{Var}\left(\sum_{i=1}^{n} Z_i\right) = c(nN)^{-1}$ for some $c > 0$. By the fact that $|\varepsilon_{i,2}^{D_n}| < 2D_n$, we have

$$\mathrm{E}\left|\varepsilon_{i,2}^{D_n}\right|^k \leq 2^{k-2}D_n^{k-2}\mathrm{E}\left|\varepsilon_{i,2}^{D_n}\right|^2, \quad k \geq 2.$$

Note that

$$\mathrm{E}|Z_i|^k = n^{-k}\mathrm{E}\left|\varepsilon_{i,2}^{D_n}\right|^k \mathrm{E}|\sigma(\boldsymbol{p}_i) B_\xi(\boldsymbol{p}_i)|^k$$

$$\leq n^{-k}\mathrm{E}\left|\varepsilon_{i,2}^{D_n}\right|^k C_\sigma^{k-2}\|B_\xi\|_{\infty,\Omega}^{k-2}\mathrm{E}|\sigma(\boldsymbol{p}_i) B_\xi(\boldsymbol{p}_i)|^2 \leq (2D_n n^{-1})^{k-2}k!\mathrm{E}(Z_i^2).$$

Thus, $\{Z_i\}_{i=1}^{n}$ satisfies the Cramer condition with constant $c^* = 2n^{-1}D_n$. By the Bernstein inequality, for any $\delta > 0$ large enough,

$$P\left(\left|\sum_{i=1}^{n} Z_i\right| \geq \delta\sqrt{\frac{\log n}{nN}}\right) \leq 2\exp\left(\frac{-\delta^2\frac{\log n}{nN}}{4V_n^2 + 2c^*\delta\sqrt{\frac{\log n}{nN}}}\right) = 2\exp\left(\frac{-\delta^2\log n}{4c + 4D_n\delta\sqrt{\frac{N\log n}{n}}}\right) \leq 2n^{-3}.$$

46

Therefore,

$$\sum_{n=1}^{\infty} P\left(\max_{\xi \in \mathcal{M}} \left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i,2}^{D_n}\sigma\left(\boldsymbol{p}_i\right)B_\xi\left(\boldsymbol{p}_i\right)\right| \geq \delta\sqrt{\frac{\log n}{nN}}\right) \leq \frac{N}{3}(d+1)(d+2)(d+3)\sum_{n=1}^{\infty}n^{-3} < \infty$$

for such $\delta > 0$. Thus, the conclusion follows by combining Borel-Cantelli's lemma with Equations B.15 and B.16. ∎

Proof of Proposition 7. Note that the penalized spline $s_{\rho_n,\epsilon}$ of ε is characterized by the orthogonality relations

$$n\left\langle \varepsilon - s_{\rho_n,\varepsilon}, u\right\rangle_{n,\Omega} = \rho_n\left\langle s_{\rho_n,\varepsilon}, u\right\rangle_{\mathcal{E}}, \quad \text{for all } u \in \mathcal{S}_d^r(\triangle). \tag{B.17}$$

In particular, $s_{0,\varepsilon}$ is characterized by

$$\left\langle \varepsilon - s_{0,\varepsilon}, u\right\rangle_{n,\Omega} = 0, \quad \text{for all } u \in \mathcal{S}_d^r(\triangle). \tag{B.18}$$

Inserting $u = s_{0,\varepsilon} - s_{\rho_n,\epsilon}$ in Equation B.17 and using Equation B.18 with this u yield that

$$\begin{aligned} n\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{n,\Omega}^2 &= \rho_n\left\langle s_{\rho_n,\varepsilon}, s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\rangle_{\mathcal{E}} \\ &= \rho_n\left(\left\langle s_{\rho_n,\epsilon}, s_{0,\epsilon}\right\rangle_{\mathcal{E}} - \left\langle s_{\rho_n,\epsilon}, s_{\rho_n,\epsilon}\right\rangle_{\mathcal{E}}\right). \end{aligned}$$

It follows, by Cauchy-Schwarz inequality, that

$$\left\|s_{\rho_n,\varepsilon}\right\|_{\mathcal{E}}^2 \leq \left\langle s_{\rho_n,\varepsilon}, s_{0,\varepsilon}\right\rangle_{\mathcal{E}} \leq \left\|s_{\rho_n,\varepsilon}\right\|_{\mathcal{E}}\left\|s_{0,\varepsilon}\right\|_{\mathcal{E}},$$

which implies that $\left\|s_{\rho_n,\varepsilon}\right\|_{\mathcal{E}} \leq \left\|s_{0,\varepsilon}\right\|_{\mathcal{E}}$. Thus, by Cauchy-Schwarz inequality and the definition of $\overline{V}_n$.

$$n\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{n,\Omega}^2 \leq \rho_n\left\|s_{\rho_n,\varepsilon}\right\|_{\mathcal{E}}\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{\mathcal{E}} \leq \overline{V}_n\rho_n\left\|s_{\rho_n,\varepsilon}\right\|_{\mathcal{E}}\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{n,\Omega}.$$

Hence, we have

$$\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{n,\Omega} \leq n^{-1}\overline{V}_n\rho_n\left\|s_{0,\varepsilon}\right\|_{\mathcal{E}}. \tag{B.19}$$

Combining Equations 11 and B.19 yields that

$$\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{\infty,\Omega} \leq V_n\left\|s_{0,\varepsilon} - s_{\rho_n,\varepsilon}\right\|_{n,\Omega} \leq n^{-1}V_n\overline{V}_n\rho_n\left\|s_{0,\varepsilon}\right\|_{\mathcal{E}}.$$

Again, applying Markov's inequality Theorem 15.28 in Lai and Schumaker (2007), we get

$$\left\|s_{0,\varepsilon}\right\|_{\mathcal{E}} \leq \frac{C_1}{|\triangle|^2}\left\|s_{0,\varepsilon}\right\|_{L^2(\Omega)}.$$

It therefore follows

$$\left\|s_{\rho_n,\varepsilon}\right\|_{\infty,\Omega} \leq \left\|s_{0,\epsilon}\right\|_{\infty,\Omega} + \left\|s_{0,\varepsilon} - s_{\rho_n,\epsilon}\right\|_{\infty,\Omega} \leq \left\|s_{0,\varepsilon}\right\|_{\infty,\Omega} + \frac{\rho_n}{n}V_n\bar{V}_n\frac{C_1}{|\triangle|^2}\left\|s_{0,\varepsilon}\right\|_{L^2(\Omega)}.$$

According to Lemma B.4, we have $\left\|s_{0,\varepsilon}\right\|_{\infty,\Omega} = O_P\left\{n^{-1/2}|\triangle|^{-3/2}(\log n)^{1/2}\right\}$. The conclusion of Proposition 7 follows from Proposition 6. ∎

B.3 Variance of the Noise Term

In this section, we first derive the size of the asymptotic conditional variance given in Theorem B.5.

Theorem B.5 *Under Assumptions (A1)—(A4), we have with probability approaching one as $n \to \infty$*

$$\frac{C_1 c_\sigma^2}{n(1 + n^{-1}|\triangle|^{-4}\rho_n)^2 |\triangle|^3} \le \mathrm{Var}\left\{s_{\rho_n,\epsilon}(\boldsymbol{p}) | \mathbb{P}\right\} \le \frac{C_2 C_\sigma^2}{n|\triangle|^3}, \ \boldsymbol{p} \in \Omega,$$

for positive constants C_1 and C_2.

Proof Note that

$$s_{\rho_n,\epsilon}(\boldsymbol{p}) = \mathbf{B}(\boldsymbol{p})^\top \widehat{\mathbf{c}}_{\rho_n,\varepsilon}, \tag{B.20}$$

where $\widehat{\mathbf{c}}_{\rho_n,\varepsilon}$ is the coefficient vector for $s_{\rho_n,\epsilon}$ using basis functions B_ξ, $\xi \in \mathcal{M}$ and $\mathbf{B}(\boldsymbol{p}) = [B_\xi(\boldsymbol{p}), \xi \in \mathcal{M}]^\top$ is the vector of basis functions. Note that

$$\widehat{\mathbf{c}}_{\rho_n,\varepsilon} = \left[\sum_{i=1}^n B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i) + \rho_n \langle B_\xi, B_\zeta \rangle_{\mathcal{E}}\right]^{-1}_{\xi,\zeta \in \mathcal{M}} \left[\sum_{i=1}^n B_\xi(\boldsymbol{p}_i) \sigma(\boldsymbol{p}_i) \varepsilon_i\right]_{\xi \in \mathcal{M}}$$

and $\mathrm{Var}\{s_{\rho_n,\epsilon}(\boldsymbol{p}) | \mathbb{P}\} = \mathbf{B}(\boldsymbol{p})^\top \mathrm{E}\left(\widehat{\mathbf{c}}_{\rho_n,\varepsilon} \widehat{\mathbf{c}}_{\rho_n,\varepsilon}^\top | \mathbb{P}\right) \mathbf{B}(\boldsymbol{p})$. Recall the definition of $\boldsymbol{\Gamma}_{\rho_n}$ in Equation B.13 in main part, we have $\mathrm{E}\left(\widehat{\mathbf{c}}_{\rho_n,\varepsilon} \widehat{\mathbf{c}}_{\rho_n,\varepsilon}^\top | \mathbb{P}\right)$ equal to

$$\boldsymbol{\Gamma}_{\rho_n}^{-1} \mathrm{E}\left(\left[\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i) \sigma(\boldsymbol{p}_i) \varepsilon_i\right]_{\xi \in \mathcal{M}} \left[\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i) \sigma(\boldsymbol{p}_i) \varepsilon_i\right]^\top_{\xi \in \mathcal{M}} \middle| \mathbb{P}\right) \boldsymbol{\Gamma}_{\rho_n}^{-1}.$$

The central conditional expectation term in the above line satisfies that

$$\frac{c_\sigma^2}{n}\boldsymbol{\Gamma}_0 \le \frac{1}{n^2}\left[\sum_{i=1}^n B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i) \sigma^2(\boldsymbol{p}_i)\right]_{\xi,\zeta \in \mathcal{M}} \le \frac{C_\sigma^2}{n}\boldsymbol{\Gamma}_0.$$

That is,

$$n^{-1}c_\sigma^2 \mathbf{B}(\boldsymbol{p})^\top \boldsymbol{\Gamma}_{\rho_n}^{-1} \boldsymbol{\Gamma}_0 \boldsymbol{\Gamma}_{\rho_n}^{-1} \mathbf{B}(\boldsymbol{p}) \le \mathrm{Var}\left\{s_{\rho_n,\epsilon}(\boldsymbol{p}) | \mathbb{P}\right\} \le n^{-1}C_\sigma^2 \mathbf{B}(\boldsymbol{p})^\top \boldsymbol{\Gamma}_{\rho_n}^{-1} \boldsymbol{\Gamma}_0 \boldsymbol{\Gamma}_{\rho_n}^{-1} \mathbf{B}(\boldsymbol{p}).$$

Let $\alpha_{\min}(\rho_n)$ and $\alpha_{\max}(\rho_n)$ be the smallest and largest eigenvalues of the positive definite matrix $\boldsymbol{\Gamma}_{\rho_n}$. It follows easily that with probability approaching one,

$$\frac{c_\sigma^2}{n}\alpha_{\max}(\rho_n)^{-2}\alpha_{\min}(0)\|\mathbf{B}(\boldsymbol{p})\|^2 \le \mathrm{Var}\left\{s_{\rho_n,\epsilon}(\boldsymbol{p}) | \mathbb{P}\right\} \le \frac{C_\sigma^2}{n}\alpha_{\min}(\rho_n)^{-2}\alpha_{\max}(0)\|\mathbf{B}(\boldsymbol{p})\|^2.$$

Note that $\|\mathbf{B}(\boldsymbol{p})\|^2 = \sum_{\xi \in \mathcal{M}} B_\xi^2(\boldsymbol{p})$ is bounded above by a constant $C_2 < \infty$ and below by $C_1 > 0$ for any point $\boldsymbol{p} \in \Omega$. Indeed, if $C_1 = 0$ for a point $\boldsymbol{p}$, then $B_\xi(\boldsymbol{p}) = 0$ for all $\xi \in \mathcal{M}$. That is, we have $0 = \sum_{\xi \in \mathcal{M}} B_\xi(\boldsymbol{p})$. It follows that these basis functions are linearly dependent which is a contradiction. On the other hand, for any fixed point $\boldsymbol{p} \in \Omega$,

there is at most C_2 nonzero terms in the above summation, where C_2 is dependent on the smallest angle of the triangulation $\triangle$. Note that B_ξ is uniformly bounded for all $\xi \in \mathcal{M}$, say bounded by 1. Thus, we know $C_2 < \infty$. We summarize the above discussion to get

$$n^{-1}C_1 c_\sigma^2 \alpha_{\max}(\rho_n)^{-2}\alpha_{min}(0) \leq \text{Var}\left\{s_{\rho_n,\epsilon}(\boldsymbol{p})|\mathbb{P}\right\} \leq n^{-1}C_2 C_\sigma^2 \alpha_{\min}(\rho_n)^{-2}\alpha_{\max}(0)$$

with probability approaching one.

We now spend some effort to estimate the largest and smallest eigenvalues of $\boldsymbol{\Gamma}_{\rho_n}$. It is easy to see that for any vector $\mathbf{a} = [a_\xi, \xi \in \mathcal{M}]^\top$,

$$\mathbf{a}^\top \boldsymbol{\Gamma}_{\rho_n} \mathbf{a} = \mathbf{a}^\top \left[\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i) B_\zeta(\boldsymbol{p}_i) + \frac{\rho_n}{n}\langle B_\xi, B_\zeta\rangle_\mathcal{E}\right]_{\xi,\zeta \in \mathcal{M}} \mathbf{a}$$

$$= \left[\frac{1}{n}\sum_{i=1}^n \left\{\sum_{\xi \in \mathcal{M}} a_\xi B_\xi(\boldsymbol{p}_i)\right\}^2\right] + \frac{\rho_n}{n}\mathcal{E}\left(\sum_{\xi \in \mathcal{M}} a_\xi B_\xi\right).$$

Let $s_a = \sum_{\xi \in \mathcal{M}} a_\xi B_\xi \in \mathcal{S}_d^r(\triangle)$ be the spline associated with vector $\mathbf{a} = (a_\xi, \xi \in \mathcal{M})^T$. By Equation B.4, we have

$$\|s_a\|_{n,\Omega}^2 = \left\{1 + O_P\left(\sqrt{\frac{\log n}{n/N}}\right)\right\}\|s_a\|_{L^2(\Omega)}^2 \leq K_2\left\{1 + O_P\left(\sqrt{\frac{\log n}{n/N}}\right)\right\}|\triangle|^3\|\mathbf{a}\|^2.$$

Here we have used the stability conditions in Lemma B.1. Furthermore, using Markov's inequality, we have

$$\frac{\rho_n}{n}\mathcal{E}\left(\sum_{\xi \in \mathcal{M}} a_\xi B_\xi\right) \leq \frac{\rho_n}{n}\frac{C}{|\triangle|^4}\|s_a\|_{L^2(\Omega)}^2 \leq \frac{\rho_n}{n}\frac{C}{|\triangle|}K_2\|\mathbf{a}\|^2.$$

Thus, the largest eigenvalue $\alpha_{\max}(\rho_n)$ of the matrix $\boldsymbol{\Gamma}_{\rho_n}$ in Equation B.13 is less than or equal to

$$K_2\left\{1 + O_P\left(\sqrt{\frac{\log n}{n/N}}\right)\right\}|\triangle|^3 + K_2\frac{\rho_n}{n}\frac{C}{|\triangle|}.$$

Thus, we have with probability approaching one

$$\begin{cases} \alpha_{\max}(0) \leq C|\triangle|^3, & \text{if } \rho_n = 0 \\ \alpha_{\max}(\rho_n) \leq C\left(|\triangle|^3 + \frac{\rho_n}{n|\triangle|}\right), & \text{if } \rho_n > 0 \end{cases}$$

for positive constant C. On the other hand, we use Lemma B.1 and Equation B.4 to have

$$\|s_a\|_{n,\Omega}^2 = \left\{1 + O_P\left(\sqrt{\frac{\log n}{n/N}}\right)\right\}\|s_a\|_{L^2(\Omega)}^2 \geq K_1\left\{1 + O_P\left(\sqrt{\frac{\log n}{n/N}}\right)\right\}|\triangle|^3\|\mathbf{a}\|^2.$$

Therefore, the smallest eigenvalue $\alpha_{\min}(\rho_n)$ of the matrix $\boldsymbol{\Gamma}_{\rho_n}$ in Equation B.13 is greater than $K_1\{1 + O_P(\sqrt{Nn^{-1}\log n})\}|\triangle|^3 = C|\triangle|^3$. Summarizing the above discussions to conclude that for $\rho_n = 0$, we have with probability approaching one,

$$c_\beta c_\sigma^2 n^{-1}|\triangle|^{-3} \leq \text{Var}\left\{s_{\rho_n,\epsilon}(\boldsymbol{p})|\mathbb{P}\right\} \leq C_\beta C_\sigma^2 n^{-1}|\triangle|^{-3}.$$

This establishes the result in the case for $\rho_n = 0$. Similar for the case $\rho_n > 0$. We have therefore completed the proof. ∎

The above variance result can be more precise when spline space $\mathcal{S}_0^{-1}(\triangle)$ is considered. The next lemma provides the pointwise variance of $\widehat{s}_{\rho_n,\varepsilon}$ when using spline space $\mathcal{S}_0^{-1}(\triangle)$.

Lemma B.6 *Consider piecewise constant spline space $\mathcal{S}_0^{-1}(\triangle)$ and suppose that the tetrahedra are of equal size. Under Assumptions (A1)—(A4), if the density function $f(\boldsymbol{p})$ of $\boldsymbol{p}$ is continuous and positive on Ω, then*

$$\mathrm{Var}\left\{s_{\rho_n,\epsilon}(\boldsymbol{p})\right\} = \frac{1}{n}\frac{\sigma^2(\boldsymbol{p})}{f(\boldsymbol{p})V_T}\{1 + o(1)\},$$

where V_T is the volume of the tetrahedron T as defined before.

Proof When using spline space $\mathcal{S}_0^{-1}(\triangle)$, the space of piecewise constant functions over $\triangle$, we have $\mathcal{E}(s) = 0$ for all $s \in \mathcal{S}_0^{-1}(\triangle)$. Thus, following Equation B.20, we have $s_{\rho_n,\varepsilon}(\boldsymbol{p}) = \mathbf{B}(\boldsymbol{p})^\top \widehat{\mathbf{c}}_{\rho_n,\varepsilon}$, where

$$\widehat{\mathbf{c}}_{\rho_n,\varepsilon} = \left[\|B_\xi\|_{n,\Omega}^{-2}\frac{1}{n}\sum_{i=1}^n \varepsilon_i B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\right]_{\xi\in\mathcal{M}}.$$

Let $\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p}) = \mathbf{B}(\boldsymbol{p})^\top\tilde{\mathbf{c}}_{\rho_n,\varepsilon}$, where

$$\tilde{\mathbf{c}}_{\rho_n,\varepsilon} = \left[\|B_\xi\|_{L^2(\Omega)}^{-2}\frac{1}{n}\sum_{i=1}^n \varepsilon_i B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\right]_{\xi\in\mathcal{M}}.$$

For any $\boldsymbol{p}\in\Omega$,

$$\left|s_{\rho_n,\epsilon}(\boldsymbol{p}) - \tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})\right| \leq |\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})|\max_{\xi\in\mathcal{M}}\left|\|B_\xi\|_{L^2(\Omega)}^2/\|B_\xi\|_{n,\Omega}^2 - 1\right|.$$

According to Lemma 3, we have

$$\max_{\xi\in\mathcal{M}}\left|\frac{\|B_\xi\|_{L^2(\Omega)}^2}{\|B_\xi\|_{n,\Omega}^2} - 1\right| \leq \frac{R_n}{1 - R_n} = O_P\left\{\sqrt{N\log n/n}\right\}.$$

Thus,

$$\left|s_{\rho_n,\epsilon}(\boldsymbol{p}) - \tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})\right| \leq O_P\left\{(N\log n/n)^{1/2}|\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})|\right\}.$$

Hence, finding the asymptotic variance of $\widehat{s}_{\rho_n,\varepsilon}(\boldsymbol{p})$ is equivalent to finding the asymptotic variance of $\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})$. Next we calculate the pointwise variance of $\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})$. Note that

$$\begin{aligned}\mathrm{Var}\left\{\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})\right\} &= \mathbf{B}(\boldsymbol{p})^\top\mathrm{E}\left(\tilde{\mathbf{c}}_{\rho_n,\varepsilon}\tilde{\mathbf{c}}_{\rho_n,\varepsilon}^\top\right)\mathbf{B}(\boldsymbol{p}) \\ &= \sum_{\xi\in\mathcal{M}}B_\xi^2(\boldsymbol{p})\mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^n\varepsilon_i B_\xi(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\|B_\xi\|_{L^2(\Omega)}^{-2}\right\}^2.\end{aligned}$$

50

For any $\boldsymbol{p} \in \Omega$, let $T_{\xi(\boldsymbol{p})}$ be the tetrahedron that contains $\boldsymbol{p}$. It is easy to see that

$$\|B_\xi\|_{L^2(\Omega)}^2 = \int_\Omega B_\xi^2 (\boldsymbol{p}) f (\boldsymbol{p}) \, \mathrm{d}\boldsymbol{p} = \int_{T_{\xi(\boldsymbol{p})}} f (\boldsymbol{p}) \, \mathrm{d}\boldsymbol{p}.$$

For any continuous function g, let $\omega(g, \varrho) = \sup_{\boldsymbol{p},\boldsymbol{p}'\in\Omega, \|\boldsymbol{p}-\boldsymbol{p}'\|\leq\varrho} |g(\boldsymbol{p}) - g(\boldsymbol{p}')|$ be the moduli of continuity of g on Ω. Then for any $\boldsymbol{p}' \in T_{\xi(\boldsymbol{p})}$, we can write $\{f (\boldsymbol{p}') - \omega(f, d_T)\}V_T \leq \|B_\xi\|_{L^2(\Omega)}^2 \leq \{f (\boldsymbol{p})+\omega(f, d_T)\}V_T$ with d_T being the diameter of the smallest ball containing T. Next, we have

$$
\begin{aligned}
n\mathrm{Var}\{\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})\} &= n \sum_{\xi\in\mathcal{M}} \|B_\xi\|_{L^2(\Omega)}^{-4} B_\xi^2 (\boldsymbol{p}) \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^n \varepsilon_i B_\xi (\boldsymbol{p}_i) \sigma (\boldsymbol{p}_i)\right\}^2 \\
&= \sum_{\xi\in\mathcal{M}} \|B_\xi\|_{L^2(\Omega)}^{-4} B_\xi^2 (\boldsymbol{p}) \mathrm{E}\left\{B_\xi^2 (\boldsymbol{p}) \sigma^2 (\boldsymbol{p})\right\} \\
&= \sum_{\xi\in\mathcal{M}} \|B_\xi\|_{L^2(\Omega)}^{-4} B_\xi^2 (\boldsymbol{p}) \int_\Omega B_\xi^2 (\boldsymbol{p}') \sigma^2 (\boldsymbol{p}') f (\boldsymbol{p}') \, \mathrm{d}\boldsymbol{p}'.
\end{aligned}
$$

By the continuity of functions $\sigma^2(\boldsymbol{p})$ and $f(\boldsymbol{p})$, we have

$$
\begin{aligned}
\mathrm{Var}\{\tilde{s}_{\rho_n,\varepsilon}(\boldsymbol{p})\} &= \frac{1}{n}\sum_{\xi\in\mathcal{M}} \|B_\xi\|_{L^2(\Omega)}^{-4} B_\xi^2 (\boldsymbol{p}) \\
&\quad \times \left[\sigma^2 (\boldsymbol{p}) f (\boldsymbol{p}) V_T + \int_{T_{\xi(\boldsymbol{p})}} \left\{\sigma^2 (\boldsymbol{p}') f (\boldsymbol{p}') - \sigma^2 (\boldsymbol{p}) f (\boldsymbol{p})\right\} \mathrm{d}\boldsymbol{p}'\right] \\
&\leq \frac{1}{n}\frac{\left\{\sigma^2 (\boldsymbol{p}) f (\boldsymbol{p}) + \omega(\sigma^2 f, |T|)\right\} V_T}{[\{f (\boldsymbol{p}) - \omega(f, |T|)\} V_T]^2} = \frac{1}{n}\frac{\sigma^2 (\boldsymbol{p})}{f (\boldsymbol{v}) V_T}\{1 + o(1)\}.
\end{aligned}
$$

The conclusion follows. ∎

B.4 Proof of Theorem 12

Proof of Theorem 12. To prove Theorem 12, we first show that under Assumptions (A1)—(A4),

$$\frac{s_{\rho_n,\epsilon}(\boldsymbol{p})}{\sqrt{\mathrm{Var}\{s_{\rho_n,\epsilon}(\boldsymbol{p})|\mathbb{P}\}}} \xrightarrow{D} N(0,1), \quad n\to\infty. \tag{B.21}$$

Let $\{B_\xi\}_{\xi\in\mathcal{M}}$ be the constructed spline basis functions for $\mathcal{S}$, where $\mathcal{M}$ stands for the index set for spline bases. Note that $s_{\rho_n,\epsilon}(\boldsymbol{p}) = \mathbf{B}(\boldsymbol{p})^\top \hat{\mathbf{c}}_{\rho_n,\varepsilon}$, where $\hat{\mathbf{c}}_{\rho_n,\varepsilon}$ is the coefficient vector for $s_{\rho_n,\epsilon}$ with basis functions B_ξ, $\xi \in \mathcal{M}$ and $\mathbf{B}(\boldsymbol{p}) = [B_\xi(\boldsymbol{p}), \xi \in \mathcal{M}]^\top$ is the vector of basis functions. Then, we have

$$\hat{\mathbf{c}}_{\rho_n,\varepsilon} = \left[\sum_{i=1}^n B_\xi (\boldsymbol{p}_i) B_\zeta (\boldsymbol{p}_i) + \rho_n\langle B_\xi, B_\zeta\rangle_\mathcal{E}\right]_{\xi,\zeta\in\mathcal{M}}^{-1} \left[\sum_{i=1}^n B_\xi (\boldsymbol{p}_i) \sigma (\boldsymbol{p}_i) \varepsilon_i\right]_{\xi\in\mathcal{M}}.$$

Recall the definition of $\boldsymbol{\Gamma}_{\rho_n}$ in Equation B.13, then we have

$$
s_{\rho_n,\epsilon}(\boldsymbol{p}) \;=\; \mathbf{B}(\boldsymbol{p})^\top \boldsymbol{\Gamma}_{\rho_n}^{-1}\left[\frac{1}{n}\sum_{i=1}^n B_\xi(\boldsymbol{p}_i)\,\sigma(\boldsymbol{p}_i)\,\varepsilon_i\right]_{\xi\in\mathcal{M}} \;=\; \frac{1}{n}\sum_{i=1}^n \mathbf{B}(\boldsymbol{p})^\top \boldsymbol{\Gamma}_{\rho_n}^{-1}\mathbf{B}(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)\,\varepsilon_i.
$$

Let $a_i = n^{-1}\mathbf{B}(\boldsymbol{p})^\top\boldsymbol{\Gamma}_{\rho_n}^{-1}\mathbf{B}(\boldsymbol{p}_i)\sigma(\boldsymbol{p}_i)$. Obviously, $s_{\rho_n,\epsilon}(\boldsymbol{p}) = \sum_{i=1}^n a_i\varepsilon_i$, and

$$
a_i^2 = n^{-2}\mathbf{B}(\boldsymbol{p})^\top\boldsymbol{\Gamma}_{\rho_n}^{-1}\mathbf{B}(\boldsymbol{p}_i)\mathbf{B}(\boldsymbol{p}_i)^\top\boldsymbol{\Gamma}_{\rho_n}^{-1}\mathbf{B}(\boldsymbol{p})\sigma^2(\boldsymbol{p}_i).
$$

For any vector $\mathbf{c} = (c_\xi, \xi \in \mathcal{M})^T$, we have

$$
\mathbf{c}^\top\mathbf{B}(\boldsymbol{p}_i)\mathbf{B}(\boldsymbol{p}_i)^\top\mathbf{c} = \left\{\sum_{\xi\in\mathcal{M}} c_\xi B_\xi(\boldsymbol{p}_i)\right\}^2 \leq \|\mathbf{c}\|_2^2\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i),
$$

which means the maximal eigenvalue of the matrix $\mathbf{B}(\boldsymbol{p}_i)\mathbf{B}(\boldsymbol{p}_i)^\top$ is bounded by $\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i)$. Consequently, according to the proof of Theorem B.5 in Section B.3, with probability approaching one,

$$
\begin{aligned}
a_i^2 \;&\leq\; n^{-2}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i)\mathbf{B}(\boldsymbol{p})^\top\boldsymbol{\Gamma}_{\rho_n}^{-1}\boldsymbol{\Gamma}_{\rho_n}^{-1}\mathbf{B}(\boldsymbol{p})\sigma^2(\boldsymbol{p}_i)\\
&\leq\; \frac{C_\sigma^2}{n^2\,|\triangle|^6}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i)\mathbf{B}(\boldsymbol{p})^\top\mathbf{B}(\boldsymbol{p}) = \frac{C_\sigma^2}{n^2\,|\triangle|^6}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p})\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i).
\end{aligned}
$$

On the other hand, $\sum a_i^2 = \mathrm{Var}\{s_{\rho_n,\epsilon}(\boldsymbol{p})|\mathbb{P}\}$. Using similar arguments in the proof of Theorem B.5, for $\rho_n = 0$, we have $\sum a_i^2 \geq c_\sigma^2 n^{-1}|\triangle|^{-3}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p})$ with probability approaching one. Therefore, there exists some positive constant C, such that

$$
\frac{\max\limits_{1\leq i\leq n} a_i^2}{\sum a_i^2} \leq \frac{CC_\sigma^2}{n\,|\triangle|^3\,c_\sigma^2}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i) = O_P\left(Nn^{-1}\right) = o_P(1).
$$

Hence, $\sum a_i\varepsilon_i/\left(\sum a_i^2\right)^{1/2} \to N(0,1)$ by Linderberg-Feller CLT. Similarly for $\rho_n > 0$, we have with probability approaching one,

$$
\sum a_i^2 \geq \frac{c_\sigma^2}{n\left(1+n^{-1}|\triangle|^{-4}\rho_n\right)^2|\triangle|^3}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}),
$$

and

$$
\frac{\max\limits_{1\leq i\leq n} a_i^2}{\sum a_i^2} \leq \frac{CC_\sigma^2\left(1+n^{-1}|\triangle|^{-4}\rho_n\right)^2}{n\,|\triangle|^3\,c_\sigma^2}\sum_{\xi\in\mathcal{M}} B_\xi^2(\boldsymbol{p}_i) = O_P\left(Nn^{-1}\right) = o_P(1),
$$

which gives Equation B.21.

Note that under Assumptions (A3), (A4') and (A5), the bias term in Equation 12 is negligible compared to the order of $[\mathrm{Var}\{s_{\rho_n,\epsilon}(\boldsymbol{p})|\mathbb{P}\}]^{1/2}$ given in Theorem B.5 in the supplementary material (Section B). Thus, Theorem 12 follows directly from Equation B.21. ∎

References

Eleonora Arnone, Alois Kneip, Fabio Nobile, and Laura M. Sangalli. Some first results on the consistency of spatial regression with partial differential equation regularization. *Statistica Sinica*, 32:209–238, 2022.

Gerard Awanou, Ming-Jun Lai, and Paul Wenston. The multivariate spline method for scattered data fitting and numerical solutions of partial differential equations. *Wavelets and splines: Athens 2005*, pages 24–74, 2006.

Matthew Berger, Andrea Tagliasacchi, Lee M. Seversky, Pierre Alliez, Gaël Guennebaud, Joshua A. Levine, Andrei Sharf, and Claudio T. Silva. A survey of surface reconstruction from point clouds. *Computer Graphics Forum*, 36(1):301–329, 2017.

Denis Bosq. *Nonparametric statistics for stochastic processes*. Springer-Verlag, New York, 1998.

Moo K. Chung, Yanli Wang, and Guorong Wu. Discrete heat kernel smoothing in irregular image domains. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5101–5104, 2018.

Ronald A DeVore and George G Lorentz. *Constructive approximation*, volume 303. Berlin: Springer-Verlag, 1993.

Jean Duchon. Fonctions-spline à énergie invariante par rotation. Technical Report 27, Université de Grenoble, 1976.

Paul HC Eilers and Brian D Marx. Flexible smoothing with b-splines and penalties. *Statistical science*, 11(2):89–121, 1996.

Qianqian Fang and David A. Boas. Tetrahedral mesh generation from volumetric binary and grayscale images. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1142–1145, 2009.

Jeff Goldsmith, Lei Huang, and Ciprian M Crainiceanu. Smooth scalar-on-image regression via spatial bayesian variable selection. *Journal of Computational and Graphical Statistics*, 23(1):46–64, 2014.

Peter J Green and Bernard W Silverman. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, London, 1994.

Junyi Guo, Xiaojun Duan, and C K Shum. Non-isotropic gaussian smoothing and leakage reduction for determining mass changes over land and ocean using GRACE data. *Geophysical Journal International*, 181(1):290–302, 2010.

Peter Hall and Joel Horowitz. A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics*, 41(4):1892–1921, 2013.

Jiawei Han, Micheline Kamber, and Jian Pei. Data mining: Concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, 5(4):83–124, 2011.

Jianhua Z Huang. Local asymptotics for polynomial spline regression. *The Annals of Statistics*, 31(5):1600–1635, 2003.

Shuhao Huo, Gui Rong Liu, Junqi Zhang, and Chongmin Song. A smoothed finite element method for octree-based polyhedral meshes with large number of hanging nodes and irregular elements. *Computer Methods in Applied Mechanics and Engineering*, 359:112646, 2020.

Clément Jamin, Pierre Alliez, Mariette Yvinec, and Jean-Daniel Boissonnat. Cgalmesh: A generic framework for delaunay mesh generation. *ACM Transactions on Mathematical Software (TOMS)*, 41(4):1–24, 2015.

Yicheng Kang. Consistent blind image deblurring using jump-preserving extrapolation. *Journal of Computational and Graphical Statistics*, 29(2):372–382, 2020.

Myungjin Kim and Li Wang. Generalized spatially varying coefficient models. *Journal of Computational and Graphical Statistics*, 30(1):1–10, 2021.

Ming-Jun Lai. *On construction of bivariate and trivariate vertex splines on arbitrary mixed grid partitions*. Phd thesis, Texas A&M University, College Station, Texas, 1989.

Ming-Jun Lai and Larry L Schumaker. *Spline functions on triangulations*. Cambridge University Press, 2007.

Ming-Jun Lai and Larry L Schumaker. A domain decomposition method for computing bivariate spline fits of scattered data. *SIAM journal on numerical analysis*, 47(2):911–928, 2009.

Ming-Jun Lai and Li Wang. Bivariate penalized splines for regression. *Statistica Sinica*, 23 (3):1399–1417, 2013.

Honggu Lee, Soohwan Song, and Sungho Jo. 3d reconstruction using a sparse laser scanner and a single camera for outdoor autonomous vehicle. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 629–634, 2016.

Yingxing Li and David Ruppert. On the asymptotics of penalized splines. *Biometrika*, 95 (2):415–436, 2008.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 73(4):423–498, 2011.

Rong Liu and Yichuan Zhao. Empirical likelihood inference for generalized additive partially linear models. *TEST*, 30(3):569–585, 2021.

Shujie Ma. Two-step spline estimating equations for generalized additive partially linear models with large cluster sizes. *Annals of Statistics*, 40:2943–2972, 2012.

Enno Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21:255–285, 1993.

Jeffrey S Morris and Raymond J Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 68(2):179–199, 2006.

Per-Olof Persson and Gilbert Strang. A simple mesh generator in matlab. *SIAM Review*, 46(2):329–345, 2004.

The CGAL Project. *CGAL user and reference manual*. CGAL Editorial Board, 5.1.1 edition, 2020. URL `https://doc.cgal.org/5.1.1/Manual/packages.html`.

Tim Ramsay. Spline smoothing over difficult regions. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 64(2):307–319, 2002.

Philip T Reiss and R Todd Ogden. Functional generalized linear models with images as predictors. *Biometrics*, 66(1):61–69, 2010.

David R Roberts, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.

Szymon Rusinkiewicz and Marc Levoy. Qsplat: A multiresolution point rendering system for large meshes. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 343–352, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

Jinyang Shao. Testing object detection for autonomous driving systems via 3d reconstruction. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*, pages 117–119, 2021.

Jonathan Richard Shewchuk. A condition guaranteeing the existence of higher-dimensional constrained delaunay triangulations. In *Proceedings of the Fourteenth Annual Symposium on Computational Geometry*, SCG '98, pages 76–85, New York, NY, USA, 1998. Association for Computing Machinery.

Hang Si. Tetgen, a delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software (TOMS)*, 41(2):1–36, 2015.

Charles J Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10:1040–1053, 1982.

Charles J Stone. The use of polynomial splines and their tensor products in multivariate function estimation. *The Annals of Statistics*, 22(1):118–171, 1994.

Roozbeh Valavi, Jane Elith, José J Lahoz-Monfort, and Gurutzeta Guillera-Arroita. blockcv: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2):225–232, 2019.

Haonan Wang and M Giovanna Ranalli. Low-rank smoothing splines on complicated domains. *Biometrics*, 63(1):209–217, 2007.

Jing Wang and Lijian Yang. Polynomial spline confidence bands for regression curves. *Statistica Sinica*, 19(1):325–342, 2009.

Li Wang, Guannan Wang, Ming-Jun Lai, and Lei Gao. Efficient estimation of partially linear models for data on complicated domains via bivariate penalized splines over triangulations. *Statistica Sinica*, 30(1):347–369, 2020a.

Lu Wang, Lan Xue, and Lijian Yang. Estimation of additive frontier functions with shape constraints. *Journal of Nonparametric Statistics*, 32(2):262–293, 2020b.

Matthieu Wilhelm, Luca Dedè, Laura M Sangalli, and Pierre Wilhelm. Igs: an isogeometric approach for smoothing on surfaces. *Computer Methods in Applied Mechanics and Engineering*, 302:70–89, 2016.

Raymond KW Wong, Thomas CM Lee, Debashis Paul, and Jie Peng. Fiber direction estimation, smoothing and tracking in diffusion mri. *The Annals of Applied Statistics*, 10 (3):1137–1156, 2016.

Simon N Wood. Thin plate regression splines. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 65(1):95–114, 2003.

Yidong Xu. *Multivariate spline method for scattered data fitting, curve and surface reconstruction, and numerical solution to Poisson equations via domain decomposition method.* PhD thesis, University of Georgia, Athens, Georgia, 2019.

Shan Yu, Guannan Wang, Li Wang, Chenhui Liu, and Lijian Yang. Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association*, 115 (530):761–774, 2020.

Yu Yue and Paul L Speckman. Nonstationary spatial gaussian markov random fields. *Journal of Computational and Graphical Statistics*, 19(1):96–116, 2010.

Hongtu Zhu, Jianqing Fan, and Linglong Kong. Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association*, 109(507):1084–1098, 2014.