June 2023

On Image Response Regression With High-Dimensional Data

Noah Fuerth
University of Windsor, fuert11e@uwindsor.ca

On Image Response Regression With High-Dimensional Data

By

Noah Fuerth

A Major Research Paper
Submitted to the Faculty of Graduate Studies
through the Department of Mathematics and Statistics
in Partial Fulfillment of the Requirements for
the Degree of Master of Science
at the University of Windsor

Windsor, Ontario, Canada

2023

On Image Response Regression With High-Dimensional Data

by

Noah Fuerth

APPROVED BY:

M. Belalia

Department of Mathematics and Statistics

S. Nkurunziza, Advisor

Department of Mathematics and Statistics

May 10, 2023

AUTHOR'S DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this major paper and that no part of this major paper has been published or submitted for publication.

I certify that, to the best of my knowledge, my major paper does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my major paper, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my major paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my major paper, including any final revisions, as approved by my major paper committee and the Graduate Studies office, and that this major paper has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

A recent issue in statistical analysis is modelling data when the effect variable changes at different locations. This can be difficult to accomplish when the dimensions of the covariates are very high, and when the domain of the varying coefficient functions of predictors are not necessarily regular. This research paper will investigate a method to overcome these challenges by approximating the varying coefficient functions using bivariate splines. We do this by splitting the domain of the varying coefficient functions into a number of triangles, and build the bivariate spline functions based on this triangulation. This major paper will outline detailed theoretical results of this method, and provide simulation studies to demonstrate the efficiency of this approach. Finally, to illustrate the application of this method, we analyze heart disease dataset where the given covariates are in spatially varying form.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Dr. Sévérien Nkurunziza, for his continued guidance and support throughout my Master's program. I would also like to thank the University of Windsor committee members for allowing me the opportunity to showcase my research.

I would also like to thank my parents for their constant support and inspiration to pursue my goals.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

In this major paper, we consider addressing the problem of modelling spatially varying data. Spatially varying data are data that changes while moving from one location to another, within a known space. In certain fields of study, including medicine, it can be of great interest to analyze these types of data. With the many advancements of modern technology, there has been increased focus on imaging data, which come from magnetic resonance imaging (MRI) scans, positron emission tomography (PET) scans, etc. When the covariates are images and the observed response is a scalar, the relationship is generally modelled using image-on-scalar regression. There can be scenarios where the functional data are located in a complex domain, in which it can be computationally cumbersome to model the data. This can be an issue because it may be difficult to smooth the functional data if its domain is too complex. Another concern with this idea is the accuracy of the fitted model when the dimension of the predictors is greater than the sample size. The main objective of this research paper is to investigate an appropriate method that is able to regulate these issues.

To analyze spatially varying data, a regression model with functional coefficients and scalar coefficients should be implemented. This would allow one to differentiate between predictor variables which have constant effects and predictor variables which have varying effects. As presented in Li et al. (2021), an efficient method that can be used to work with these functional data is the Partially Linear Spatially Varying Coefficient Model (PLSVCM). The PLSVCM models spatially varying predicted values over a two-dimensional, complex domain. To apply the PLSVCM, we consider a mixture of spatially varying coefficient functions to deal with predictors which have varying effects and constant coefficients to deal with predictors which have constant effects (Li et al. (2021)). To estimate these coefficient functions, approximation of the coefficient functions using bivariate splines should be considered. To work with complex domains, the domain can be split up into a number of triangles. The coefficient functions can then be approximated on this triangulation. To account for the case where the dimension of the predictors exceeds the sample size, adaptive LASSO penalty functions can be included in the model before estimating the scalar coefficients and varying coefficient functions.

The remaining material in this major paper will be organized as follows. In Chapter 2, we give a method to estimate the scalar coefficients and varying coefficient functions of the PLSVCM. In Section 2.1, we provide a detailed method to triangulate the domain of the varying coefficient functions, and approximate them using bivariate splines. In Section 2.2, we give a description of the PLSVCM and state some necessary assumptions given in Li et al. (2021). In Section 2.3, we show how to obtain the estimators from the PLSVCM by minimizing the likelihood function corresponding to the PLSVCM. In Section 2.4, we determine some asymptotic properties of the estimators. In particular, we give the consistency of the estimators of the parameters in the PLSVCM and determine the asymptotic variance-covariance matrix of the estimator of the constant coefficients. In Chapter 3, we consider the case where the number of covariates exceeds the sample size. In this scenario, we model the data using a penalized LASSO regression approach. In Section 3.1, we provide a detailed description of the proposed LASSO model and state some necessary assumptions given in Li et al. (2021). In Section 3.2, we provide some asymptotic properties of the estimators obtained from the LASSO model. In Chapter 4, we look at two applications of the proposed model in this major paper by simulation and real data analysis. In Section 4.1, we run a simulation study to show the efficiency of the estimators obtained from the PLSVCM. In Section 4.2, we apply the PLSVCM to analyze a real dataset, which aims to determine influential predictors that increase the likelihood of having heart disease. In Chapter 5, we conclude by giving a brief summary of the major paper, along with some possible ideas for further research related to this topic. Finally, in the Appendix, we provide detailed proofs of the results in Sections 2.4 and 3.2. We also state and prove some key results used in the proofs of the results in Sections 2.4 and 3.2.

Chapter 2

ESTIMATION OF THE PLSVCM

In this Chapter, we define the PLSVCM and show how to approximate the varying coefficient functions using bivariate splines over triangulations. We also establish some asymptotic results of the estimators of the constant and spatially varying coefficients. This is important because it addresses the potential problem in which the domain of the functional data is not necessarily regular.

2.1 Approximation of the Varying Coefficient Functions

One method to overcome the problem of smoothing over complex domains is to approximate the varying coefficients by bivariate spline basis functions over a triangulated domain, which was proposed in Lai and Schumaker (2007).

To set up some notation, let τ be a triangle such that its three points do not lie along one straight line. Denote $\Delta = \{\tau_1, \tau_2, ..., \tau_N\}$ as a triangulation of an arbitrary domain, $\Omega = \cup_{i=1}^{N} \tau_i$, under the condition that if τ_i and τ_j $(i \neq j)$ intersect, they much share a common vertex or share a common edge. For any triangle, τ, in a triangulation, Δ, define $|\tau|$ as the length of the longest edge of τ. Let R_τ be the radius of the largest circle that can be wholly contained inside of τ, and let $S_\tau = \frac{|\tau|}{R_\tau}$

be the shape parameter of τ. Define the length of the longest edge in the triangulation Δ as $|\Delta| = \max\{|\tau|, \tau \in \Delta\}$.

The Bernstein basis polynomial can be used to create the bivariate spline functions, which can approximate the varying coefficient functions (Li et al. (2021)). Given a triangle, $\tau \in \Delta$ with non-zero area, define its vertices as $< \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 >$ in a counter-clockwise way (Lai and Schumaker (2007)). Then, any point $\boldsymbol{v}$ inside τ can be expressed as $\boldsymbol{v} = b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + b_3\boldsymbol{v}_3$, where $b_1 + b_2 + b_3 = 1$. The scalars (b_1, b_2, b_3) are called the barycentric coordinates of $\boldsymbol{v}$ relative to the triangle τ, and are used to form the Bernstein basis polynomials. To introduce the definition of the Bernstein basis polynomials, let $\boldsymbol{v}$ be a point inside a triangle τ, whose area is non-zero. Let (b_1, b_2, b_3) be the barycentric coordinates of $\boldsymbol{v}$ relative to the triangle τ. For nonnegative integers, i, j, k, the Bernstein basis polynomial of degree d relative to τ is defined as

$$B_{ijk}^{\tau,d}(\boldsymbol{v}) := \frac{d!}{i!j!k!} b_1^i b_2^j b_3^k, \text{ with } \ i + j + k = d$$

Let $\mathcal{P}_d(\tau)$ denote the space of all polynomials with degree less than or equal to d, defined on τ. Then, the set $\{B_{ijk}^{\tau,d}(\boldsymbol{v}) : i, j, k \geq 0, i + j + k = d\}$ forms a basis for $\mathcal{P}_d(\tau)$. This means that any polynomial $p(\boldsymbol{v}) \in \mathcal{P}_d(\tau)$ can be written as $p(\boldsymbol{v})|_\tau = \sum_{i+j+k=d} \gamma_{ijk}^\tau B_{ijk}^{\tau,d}(\boldsymbol{v})$. The coefficients $\{\gamma_{ijk}^\tau\}_{i+j+k=d}$ are called the "B-coefficients of p."

2.2 Description of the Model

To define the data, let Ω be a two-dimensional domain with a complex structure. Let $\mathbf{s}_j = (s_{1j}, s_{2j})^T, j = 1, 2, ..., N_s$, be a vector in Ω, where N_s is the number of elements in Ω. Let n be the sample size, and define $Y_i(\mathbf{s}_j), i = 1, 2, ..., n$, as the actual observed values at the point $\mathbf{s}_j$. Define $\mathbf{X}_{(i)} = (X_{i1}, X_{i2}, ..., X_{ip})^T, i = 1, 2, .., n$, as the vector of covariates for the i^{th} sample, where p is the number of covariates. Let A_c be the index set for constant coefficients and A_v be the index set for varying coefficients.

For all $k \in A_c$, denote the constant coefficient parameter as α_{0k}. For all $k \in A_v$, denote the actual varying coefficient function as β_{0k}. Then, for $i = 1, 2, ..., n$ and $j = 1, 2, ..., N_s$, the PLSVCM is

$$Y_i(\mathbf{s}_j) = \sum_{k \in A_c} X_{ik}\alpha_{0k} + \sum_{k \in A_v} X_{ik}\beta_{0k}(\mathbf{s}_j) + \eta_i(\mathbf{s}_j) + \epsilon_i(\mathbf{s}_j).$$

The within image dependences, $\eta_i, i = 1, 2, ..., n$, are assumed to be independent and identical copies of a stochastic process with mean zero and covariance function $G_\eta(\mathbf{s}, \mathbf{s}')$. The measurement errors, $\epsilon_i, i = 1, 2, ..., n$, are assumed to be independent and identical copies of a random process with mean zero, and covariance $\text{Cov}(\epsilon_i(\mathbf{s}), \epsilon_i(\mathbf{s}'))$. It can be assumed that for all $i = 1, 2, ..., n$ and for all $k = 1, 2, ..., p$, ϵ_i and η_i are independent, ϵ_i and X_{ik} are independent, and η_i and X_{ik} are independent.

Before performing the estimation method for the parameter values, some assumptions must be stated (Li et al. (2021)). For any function f over the closure of the domain Ω, let $||f||_{\infty,\Omega} = \sup_{\mathbf{s}\in\Omega}|f(\mathbf{s})|$. Let $D^k_{s_j} f(\mathbf{s})$ be the k^{th} derivative of f at $\mathbf{s}$ in the direction of s_j, where $j = 1, 2$. Denote $|f|_{q,\infty,\Omega} = \max_{i+j=q} ||D^i_{s_1} D^j_{s_2} f(\mathbf{s})||_{\infty,\Omega}$.

Assumption 1. *For all $k \in A_v$, $\beta_{0k} \in W^{d+1,\infty}(\Omega) = \{f : |f|_{q,\infty,\Omega} < \infty, 0 \le q \le d + 1\}$, where d is a nonnegative integer. Further, $\int_\Omega \beta_{0k}(\mathbf{s})d\mathbf{s} = 0$, for all $k \in A_v$.*

Assumption 2. *For all $k = 1, 2, ...p$, there exists a positive real number C_X, such that $E[|X_k|^6] \le C_X$.*

Assumption 3. *For all $i = 1, 2, ..., n$ and for all $j = 1, 2, ..., N_s$, the errors ϵ_{ij} are independent with mean 0 and variance σ^2_ϵ. For all $i = 1, 2, ..., n$, $j = 1, 2, ..., N_s$, and for all $\mathbf{s} \in \Omega$, $0 < c_G \le G_\eta(\mathbf{s}, \mathbf{s}) \le C_G < \infty$, with $c_G, C_G \in \mathbb{R}$.*

Assumption 4. *The triangulation Δ is π-quasi uniform. That is, for all $\tau \in \Delta$, there exists a positive real number π, such that $S_\tau \le \pi$.*

Assumption 5. *As $N_s \to \infty, n \to \infty$, $N_s|\Delta|^2 \to \infty$ and $n|\Delta|^{2(d+1)} \to 0$.*

2.3 Estimation Method

To perform the estimation, triangulate the domain and let $\mathcal{S}_d^r(\Delta) = \{s \in \mathbb{C}^r(\Omega) : s|_\tau \in \mathcal{P}_d(\tau), \tau \in \Delta\}$ be a spline space with degree d and smoothness parameter r, over a triangulation Δ. $\mathbb{C}^r(\Omega)$ is the the set of all r^{th} continuously differentiable functions over Ω, $s|_\tau$ is the polynomial part of the spline s restricted on $\tau \in \Delta$, and $\mathcal{P}_d$ is defined as the space of all polynomials with degree less than or equal to d. Define I_k as the index set for the k^{th} spline basis function. Then for all $\mathbf{s} \in \Omega$, and for all $k = 1, 2, ..., p$, approximate $\beta_k(\mathbf{s})$ by $\sum_{\ell \in I_k} B_{k\ell}(\mathbf{s}) c_{k\ell}$, where $\mathbf{c}_k = (c_{k\ell}, \ell \in I_k)^T$ is the vector of spline coefficients and for all $\mathbf{s} \in \Omega$, $\mathbf{B}_k(\mathbf{s}) = (B_{k\ell}(\mathbf{s}), \ell \in I_k)^T$ is the vector of bivariate basis functions. Applying the method in Yu et al. (2019) and Li et al. (2021), let $\mathbf{H}_k$ be the constraint matrix on the vectors $\mathbf{c}_k$, such that $\mathbf{H}_k \mathbf{c}_k = \mathbf{0}$. For all $\mathbf{s} \in \Omega$, assume that $\mathbf{B}_1 = \mathbf{B}_2 = ... = \mathbf{B}_p$ and define this to be $\mathbf{B}(\mathbf{s}) = (B_\ell(\mathbf{s}), \ell \in I)^T$. Also, assume that $\mathbf{H}_1 = \mathbf{H}_2 = ... = \mathbf{H}_p$ and define this as $\mathbf{H}$. Let $Y_{ij} = Y_i(\mathbf{s}_j)$, then to obtain estimators $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1 \; \hat{\alpha}_2 \; \ldots \; \hat{\alpha}_{|A_c|})^T$ and $\hat{\mathbf{c}} = (\hat{\mathbf{c}}_1^T \; \hat{\mathbf{c}}_2^T \ldots \; \hat{\mathbf{c}}_{|A_v|}^T)^T$ for $\boldsymbol{\alpha} = (\alpha_1 \; \alpha_2 \; \ldots \; \alpha_{|A_c|})^T$ and $\mathbf{c} = (\mathbf{c}_1^T \; \mathbf{c}_2^T \ldots \; \mathbf{c}_{|A_v|}^T)^T$, respectively, the following likelihood function must be minimized:

$$L_n(\boldsymbol{\alpha}, \mathbf{c}) = \sum_{i=1}^{n} \sum_{j=1}^{N_s} \left[Y_{ij} - \sum_{k \in A_c} X_{ik} \alpha_k - \sum_{k \in A_v} X_{ik} \mathbf{B}^T(\mathbf{s}_j) \mathbf{c}_k \right]^2,$$

under the constraint $\mathbf{H}_k \mathbf{c}_k = \mathbf{0}$. By the QR-decomposition, write $\mathbf{H}^T = \mathbf{QR}$, where $\mathbf{Q} = (\mathbf{Q}_1 \; \mathbf{Q}_2)$ is an orthogonal matrix and $\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}$ is an upper-triangular matrix. Further, $\mathbf{Q}_1$ is a matrix containing the first r columns of $\mathbf{Q}$, where r is the rank of $\mathbf{H}$, and $\mathbf{R}_2$ is a zero matrix. Then under no constraints, the likelihood function above becomes

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \sum_{j=1}^{N_s} \left[Y_{ij} - \sum_{k \in A_c} X_{ik} \alpha_k - \sum_{k \in A_v} X_{ik} (\mathbf{B}^*(\mathbf{s}_j))^T \boldsymbol{\gamma}_k \right]^2,$$

where $\mathbf{B}^*(\mathbf{s}) = \mathbf{Q}_2^T\mathbf{B}(\mathbf{s})$, for all $\mathbf{s} \in \Omega$. Assume that $\mathbf{B}^*(\mathbf{s})$ is the collection of all of the normalized basis functions, and to simplify the notation, denote it as $\mathbf{B}(\mathbf{s})$. Now, to obtain the estimators $\hat{\alpha}_k^0$, $\forall k \in A_c$ and $\hat{\boldsymbol{\gamma}}_k^0$, $\forall k \in A_v$ for the true parameter values α_{0k}, $\forall k \in A_c$ and $\boldsymbol{\gamma}_{0k}$, $\forall k \in A_v$, respectively, the following likelihood function must be minimized:

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{N_s}\left[Y_{ij} - \sum_{k\in A_c} X_{ik}\alpha_k - \sum_{k\in A_v} X_{ik}\mathbf{B}^T(\mathbf{s}_j)\boldsymbol{\gamma}_k\right]^2.$$

As a result, for all $\mathbf{s} \in \Omega$, and for all $k \in A_v$, the estimator of the true parameter $\beta_{0k}(\mathbf{s})$ is $\hat{\beta}_k^0(\mathbf{s}) = \mathbf{B}^T(\mathbf{s})\hat{\boldsymbol{\gamma}}_k^0$.

2.4 Asymptotic Properties of the Estimators

It is of interest to determine the consistency of the estimators $\hat{\alpha}_k^0$ and $\hat{\boldsymbol{\gamma}}_k^0$, under mandatory conditions. Define $\mathbf{X}_{A_c} = (\mathbf{X}_k, k \in A_c)$, where $\mathbf{X}_k$ is the k^{th} column vector of the matrix of predictors, $\mathbf{X}$, and let $\mathbf{Z}_{1,A_c} = \mathbf{X}_{A_c} \otimes \mathbf{1}_{N_s}$, where $\mathbf{1}_{N_s}$ is a column vector of ones with dimension N_s. Similarly, define $\mathbf{X}_{A_v} = (\mathbf{X}_k, k \in A_v)$, and let $\mathbf{Z}_{2,A_v} = \mathbf{X}_{A_v} \otimes \mathbf{B}$. Define $\mathbf{Z}_A = (\mathbf{Z}_m, m \in A_c \cup A_v)$, where $\mathbf{Z}_m = \mathbf{Z}_{1,A_c}$ if $m \in A_c$ and $\mathbf{Z}_m = \mathbf{Z}_{2,A_v}$ if $m \in A_v$. Define the vector of the actual parameter values as $\boldsymbol{\theta}_A = \boldsymbol{\theta}_{0,A} = (\boldsymbol{\alpha}_{0,A}^T\ \boldsymbol{\gamma}_{0,A}^T)^T$. Then, the minimizer for the likelihood function, $L_n(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ defined above, is the ordinary least-squares estimator of $\boldsymbol{\theta}_{0,A}$. This estimator is $\hat{\boldsymbol{\theta}}^0 = (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T\mathbf{Y}$, where $\hat{\boldsymbol{\theta}}^0 = ((\hat{\boldsymbol{\alpha}}^0)^T, (\hat{\boldsymbol{\gamma}}^0)^T)^T$. Further, define $||\cdot||$ as the Euclidean Norm.

The following result from Li et al. (2021) describes the consistency of the estimators $\hat{\boldsymbol{\alpha}}^0$ and $\hat{\boldsymbol{\beta}}^0(\cdot)$.

Theorem 2.1. *Under the assumptions in Section 2.2 and the assumption that* $||\mathbf{C}_A^{-1}||$

is bounded by a positive constant, π_1^{-1}, *where* $\mathbf{C}_A = \frac{1}{nN_s}\mathbf{Z}_A^T\mathbf{Z}_A$, *the following hold:*

$$\text{(a)} \sum_{k\in A_c}(\hat{\alpha}_k^0 - \alpha_{0k})^2 = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)}\right)$$
$$\text{(b)} \sum_{k\in A_v}||\hat{\beta}_k^0 - \beta_{0k}||^2_{L_2(\Omega)} = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)}\right),$$

where $||f||^2_{L_2(\Omega)} = \int_{s\in\Omega} f^2(\mathbf{s})d\mathbf{s}$ *is the* L_2 *norm for a function* f, *over the domain* Ω.

A detailed proof of Theorem 2.1 is given in the Appendix.

Before giving the explicit form of the sample variance-covariance matrix of $\hat{\boldsymbol{\alpha}}^0$, some definitions must be introduced. Let $\mathbf{P}_{\mathbf{Z}_{1,A_c}} = \mathbf{Z}_{1,A_c}(\mathbf{Z}_{1,A_c}^T\mathbf{Z}_{1,A_c})^{-1}\mathbf{Z}_{1,A_c}^T$ be the projection matrix on $\mathbf{Z}_{1,A_c}$, and let $\mathbf{P}_{\mathbf{Z}_{2,A_v}} = \mathbf{Z}_{2,A_v}(\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T$ be the projection matrix on $\mathbf{Z}_{2,A_v}$. Let $\mathbf{D}_c = (\mathbf{I}_{nN_s} - \mathbf{P}_{\mathbf{Z}_{2,A_v}})\mathbf{Z}_{1,A_c}$, where $\mathbf{I}_{nN_s}$ is the identity matrix with dimension $nN_s \times nN_s$. Next, define

$$\Sigma_{i,e} = \{G_\eta(\mathbf{s}_j, \mathbf{s}_{j'})\}_{j,j'=1}^{N_s} + \text{diag}\{\sigma^2(\mathbf{s}_j)\}_{j=1}^{N_s},$$
$$\Sigma_{c,e} = \frac{1}{n^2N_s^2}\mathbf{D}_c^T\text{diag}\{\Sigma_{i,e}\}_{i=1}^n\mathbf{D}_c,$$
$$\Sigma_c = \frac{1}{nN_s}\mathbf{D}_c^T\mathbf{D}_c.$$

The following theorem from Li et al. (2021) gives the sample variance-covariance matrix of $\hat{\boldsymbol{\alpha}}^0$.

Theorem 2.2. *Suppose the assumptions in Section 2.2 hold. Let* $\mathbf{V}_c = \Sigma_c^{-1}\Sigma_{c,e}\Sigma_c^{-1}$. *Then,*

$$\mathbf{V}_c^{-1/2}(\hat{\boldsymbol{\alpha}}^0 - \boldsymbol{\alpha}_{0,A_c}) \xrightarrow[n,N_s\to\infty]{D} N(\mathbf{0}, \mathbf{I}_{|A_c|}),$$

where $\mathbf{I}_{|A_c|}$ *is the identity matrix with dimension* $|A_c|$.

A detailed proof of Theorem 2.2 is given in the Appendix.

Chapter 3

MODELLING HIGH-DIMENSIONAL DATA

In this Chapter, we consider a modified PLSVCM to deal with the case where the data is high-dimensional. To do this, we implement a penalized LASSO regression model to accurately determine which covariates have nonzero constant and varying effects. In Section 3.1, we give a detailed description of the penalized LASSO regression model, along with some assumptions about the proposed estimators. In Section 3.2, we state some asymptotic properties of the estimators obtained from the penalized LASSO regression model.

First, define three index sets:

$$A_c = \{k = 1, 2, ..., p : \alpha_k \neq 0, \beta_k(\cdot) \equiv 0\}$$
$$A_v = \{k = 1, 2, ...p : \beta_k(\cdot) \neq 0\}$$
$$\mathcal{N} = \{k = 1, 2, ..., p : \alpha_k \equiv 0, \beta_k(\cdot) \equiv 0\}$$

Thus, the active index set for $\mathbf{X}$ is $A = A_c \cup A_v$. The main objective is to obtain estimators for the active constant set and active varying set and consequently, the active index set. Recall that for all $k \in A_v$, and for all $\mathbf{s} \in \Omega$, $\hat{\beta}_k(\mathbf{s}) = \mathbf{B}^T(\mathbf{s})\hat{\boldsymbol{\gamma}}_k$. Then, define the estimators for the three index sets above as

$$\hat{A}_c = \{k : |\hat{\alpha}_k| \neq 0, ||\hat{\boldsymbol{\gamma}}_k|| = 0, 1 \leq k \leq p\}$$
$$\hat{A}_v = \{k : ||\hat{\boldsymbol{\gamma}}_k|| \neq 0, 1 \leq k \leq p\}$$
$$\hat{\mathcal{N}} = \{k : |\hat{\alpha}_k| = 0, ||\hat{\boldsymbol{\gamma}}_k|| = 0, 1 \leq k \leq p\}$$

3.1 Description of the Model

For all $k = 1, 2, ..., p$, let $\tilde{\alpha}_k$ and $\tilde{\boldsymbol{\gamma}}_k$ be consistent initial estimators for α_k and $\boldsymbol{\gamma}_k$, respectively. Let $w^c_{n,k} = |\tilde{\alpha}_k|^{-1}$ and $w^v_{n,k} = ||\tilde{\boldsymbol{\gamma}}_k||^{-1}$. Let ρ_{n1} and ρ_{n2} be regularization parameters, with the assumption that $\rho_{n1} \longrightarrow \infty$ and $\rho_{n2} \longrightarrow \infty$, as $n \longrightarrow \infty$ and $N_s \longrightarrow \infty$. Then, for the LASSO regression model, define penalty functions $p_{\rho_{n1}}(|\alpha_k|) = \rho_{n1} w^c_{n,k}|\alpha_k| = \rho_{n1}\frac{|\alpha_k|}{|\tilde{\alpha}_k|}$ and $p_{\rho_{n2}}(||\boldsymbol{\gamma}_k||) = \rho_{n2} w^v_{n,k}||\boldsymbol{\gamma}_k|| = \rho_{n2}\frac{||\boldsymbol{\gamma}_k||}{||\tilde{\boldsymbol{\gamma}}_k||}$.

Under the assumption that $\int_\Omega \beta_k(\mathbf{s})d\mathbf{s} = 0$, then for all $i = 1, 2, ..., n$ and for all $\mathbf{s} \in \Omega$, the Spatially Varying Coefficient Model (SVCM) from Li et al. (2021) is defined as

$$Y_i(\mathbf{s}) = \sum_{k=1}^{p} X_{ik}\alpha_k + \sum_{k=1}^{p} X_{ik}\beta_k(\mathbf{s}) + \eta_i(\mathbf{s}) + \epsilon_i(\mathbf{s}).$$

To accurately perform the model selection for the SVCM above and correctly identify the index sets, the penalized score function given in Li et al. (2021)

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \rho_{n1}, \rho_{n2}) = \sum_{i=1}^{n}\sum_{j=1}^{N_s}\left[Y_i(\mathbf{s}_j) - \sum_{k=1}^{p} X_{ik}\alpha_k - \sum_{k=1}^{p} X_{ik}\mathbf{B}^T(\mathbf{s}_j)\boldsymbol{\gamma}_k\right]^2 + \sum_{k=1}^{p} p_{\rho_{n1}}(|\alpha_k|) + \sum_{k=1}^{p} p_{\rho_{n2}}(||\boldsymbol{\gamma}_k||),$$

must be minimized.

Some further assumptions taken from Li et al. (2021) must be provided before stating certain theoretical results which give some asymptotic properties of the estimators from this model.

Assumption 6. *The cardinalities* $|A_c|$ *and* $|A_v|$ *are fixed. Also, there exists positive real numbers,* c_α, c_β*, such that* $\min_{k\in A_c} |\alpha_{0k}| \geq c_\alpha$ *and* $\min_{k\in A_v} ||\beta_{0k}||_{L_2(\Omega)} \geq c_\beta$*.*

Assumption 7. *For all* $k = 1, 2, ..., p$*, there exists a positive real number* C_X*, such that* $|X_k| < C_X$*, with probability one.*

Assumption 8. *Let* $r_{n\alpha}, r_{n\gamma}$ *be real numbers, such that* $r_{n\alpha}, r_{n\gamma} \longrightarrow \infty$*, as* $n \longrightarrow \infty$*. Then as* $r_{n\alpha}, r_{n\gamma} \longrightarrow \infty$*,*

$$r_{n\alpha} \max_{k\notin A_c} |\tilde{\alpha}_k| = O_p(1)$$
$$r_{n\gamma} \max_{k\notin A_v} ||\tilde{\boldsymbol{\gamma}}_k|| = O_p(1)$$

For real numbers c_α *and* c_β *defined in Assumption 6, there exists positive real numbers* b_α *and* b_γ*, such that*

$$\mathrm{P}\left(\min_{k\in A_c} |\tilde{\alpha}_k| \geq c_\alpha b_\alpha\right) \longrightarrow 1$$
$$\mathrm{P}\left(\min_{k\in A_v} ||\tilde{\boldsymbol{\gamma}}_k|| \geq c_\gamma b_\gamma\right) \longrightarrow 1,$$

as $n, N_s \longrightarrow \infty$*.*

Assumption 9. *Assume that*

$$\frac{\sqrt{nN_s^2\log(\mathrm{p})}}{\rho_{n1}r_{n\alpha}} + \frac{\sqrt{nN_s^2\log(pJ_n)}}{\rho_{n2}r_{n\gamma}} + \frac{nN_s|\Delta|^{d+1}}{\rho_{n1}r_{n\alpha}} = o(1),$$

$$\frac{\sqrt{nN_s^2\log(\mathrm{p})}}{\rho_{n1}r_{n\alpha}} + \frac{\sqrt{nN_s^2\log(pJ_n)}}{\rho_{n2}r_{n\gamma}} + \frac{nN_s|\Delta|^{d+1}}{\rho_{n2}r_{n\gamma}} = o(1),$$

$$\frac{\rho_{n1}^2+\rho_{n2}^2}{nN_s^2} = o(1), \qquad \frac{n}{J_n^{(d+1)}\log(pJ_n)} = o(1).$$

3.2 Asymptotic Results

Theorem 3.1 from Li et al. (2021) gives the asymptotic properties of the constant and varying index sets. Theorem 3.2 from Li et al. (2021) provides the convergence rates of the estimators that are obtained by minimizing the likelihood function L_n above.

Theorem 3.1. *Suppose that the assumptions in Sections 2.2 and 3.1 hold. Then, as* $n \to \infty$ *and* $N_s \to \infty$, $\mathrm{P}(\hat{A}_c = A_c) \longrightarrow 1$ *and* $\mathrm{P}(\hat{A}_v = A_v) \longrightarrow 1$.

Theorem 3.2. *Suppose that the assumptions in Sections 2.2 and 3.1 hold. Let* $\hat{\boldsymbol{\alpha}}$ *and* $\hat{\boldsymbol{\beta}}(\cdot)$ *be estimators that are obtained by minimizing the likelihood function* L_n *above. Then*

$$(a)\ \sum_{k\in A_c}(\hat{\alpha}_k - \alpha_{0k})^2 = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)} + \frac{\rho_{n1}^2+\rho_{n2}^2}{n^2N_s^2}\right)$$

$$(b)\ \sum_{k\in A_v}||\hat{\beta}_k - \beta_{0k}||_{L_2(\Omega)}^2 = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)} + \frac{\rho_{n1}^2+\rho_{n2}^2}{n^2N_s^2}\right).$$

The above results are critical because they provide consistent estimators to accurately predict the image response in a high-dimensional setting. The proofs of these results are outlined in Li et al. (2021). For the convenience of the reader, we also provide a proof with more details in the Appendix.

Chapter 4

SIMULATION RESULTS AND DATA ANALYSIS

4.1 Simulation Study

To demonstrate how the estimation method works, a simulation similar to the study in Section 4.2 in Li et al. (2021) will be conducted. Actual parameter values from the model given in Chapter 2 will be generated first. Two constant coefficients will be estimated (α_{01} and α_{02}) and two varying coefficient functions will be estimated ($\beta_{03}(\cdot)$ and $\beta_{04}(\cdot)$). To differentiate between the SVCM and the PLSVCM, consider the case where α_{01} and α_{02} are both zero and the case where α_{01} and α_{02} are both nonzero. In both of these cases, values from a square domain will be generated. The domain will be triangulated and the mean squared errors of the estimators of the parameters will be computed for different refinements of the triangulation.

First, let $A_c = \{1, 2\}$ and $A_v = \{3, 4\}$. For all $j = 1, 2, ..., N_s$, let $\mathbf{s}_j = (s_{1j}, s_{2j})^T$ be in the domain Ω. Generate s_{1j} and s_{2j} independently from a Uniform(0,1). For all $i = 1, 2, ..., n$, generate $X_{i1}, X_{i2}, X_{i3}, X_{i4}$ independently from a Uniform(-1,1). To simulate the actual varying coefficient functions, let

$$\beta_3(\mathbf{s}_j) = 20\left[(s_{1j} - 0.5)^2 + (s_{j2} - 0.5)^2\right]$$
$$\beta_4(\mathbf{s}_j) = \exp\{-15\left[(s_{1j} - 0.5)^2 + (s_{j2} - 0.5)^2\right]\},$$

for $\mathbf{s}_j \in \Omega, j = 1, 2, ..., N_s$. To simulate the within-image dependence, for all $i = 1, 2, ..., n$ and for all $j = 1, 2, ..., N_s$, let

$$\eta_i(\mathbf{s}_j) = (0.3)Z_{i1}(1.588\sin(\pi s_{1j})) + (0.075)Z_{i2}(2.157\cos(\pi s_{2j}) - 0.039),$$

where Z_{i1} and Z_{i2} are generated independently from a N(0,1). To simulate the errors, generate $\{\epsilon(\mathbf{s}_j) : \mathbf{s}_j \in \Omega\}$ from a Gaussian distribution with mean zero and variance σ_ϵ^2. The values of σ_ϵ^2 are selected in such a way that the signal-noise-ratio, defined as

$$\text{SNR} = \frac{N_s^{-1}\sum_{j=1}^{N_s} \text{ Var}\left[\sum_{k\in A_c} X_{ik}\alpha_{0k} + \sum_{k\in A_v} X_{ik}\beta_{0k}(\mathbf{s}_j)\right]}{N_s^{-1}\sum_{j=1}^{N_s} \text{ Var}\left[\eta_i(\mathbf{s}_j) + \epsilon_i(\mathbf{s}_j)\right]},$$

is approximately equal to either 3 or 5.

The square domain, Ω can be partitioned into many triangles. This simulation will focus on triangulating the domain into 8 triangles with 9 vertices, and then again with 18 triangles with 16 vertices. Both images are depicted below.

(a) 9 vertices and 8 triangles. (b) 16 vertices and 18 triangles.

Figure 4.1: Triangulating a square domain.

In each case, the estimators $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\beta}_3(\cdot)$, and $\hat{\beta}_4(\cdot)$ will be evaluated on both of the domains in Figure 1.

Case 1 ($\alpha_{01} = \alpha_{02} = 0$):

Let Partition 1 represent the triangulated domain split into 8 triangles with 9 vertices and let Partition 2 represent the triangulated domain split into 18 triangles with 16 vertices. The tables below show the mean squared errors of the estimators $\hat{\beta}_3(\cdot)$, and $\hat{\beta}_4(\cdot)$, over 50 simulations on both partitions for different sample sizes ($n = 50, 100$), and different number of points in the domain ($N_s = 1600, 2500$).

Table 4.1: MSE of $\hat{\beta}_3(\cdot)$ (Case 1)

MSE of $\hat{\beta}_3(\cdot)$				
N_s	n	SNR	Partition 1	Partition 2
1600	50	3	24005.60	24308.48
		5	23986.03	24301.29
	100	3	24673.46	24507.35
		5	24671.60	24519.88
2500	50	3	37179.45	38060.01
		5	37139.00	38042.08
	100	3	38498.97	38502.32
		5	38491.34	38511.28

Table 4.2: MSE of $\hat{\beta}_4(\cdot)$ (Case 1)

MSE of $\hat{\beta}_4(\cdot)$				
N_s	n	SNR	Partition 1	Partition 2
1600	50	3	266.89	237.84
		5	258.42	228.48
	100	3	211.51	201.83
		5	209.15	200.04
2500	50	3	333.08	336.11
		5	328.38	327.98
	100	3	291.37	302.79
		5	287.38	299.93

Case 2 ($\alpha_{01} = 1$ and $\alpha_{02} = -1$):

Now, the case where the parameters α_{01} and α_{02} are both nonzero will be considered. Similar to case one, the mean squared errors of the estimators will be given below for different sample sizes, and different number of points in the domain. Note that since α_{01} and α_{02} are constant coefficients, the difference in mean squared errors between the estimators from Partitions 1 and 2 should not differ drastically.

Tables 4.3 and 4.4 below give the point estimates and mean squared errors of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ for 50 simulations from both partitions.

Table 4.3: MSE of $\hat{\alpha}_1$ (Case 2)

			Partition 1		Partition 2	
N_s	n	SNR	$\hat{\alpha}_1$	MSE	$\hat{\alpha}_1$	MSE
1600	50	3	0.0064	0.9919	0.0139	0.9756
		5	0.0075	0.9898	0.0130	0.9774
	100	3	0.0155	0.9708	-0.0006	1.0019
		5	0.0172	0.9675	0.0006	0.9995
2500	50	3	-0.0002	1.0036	0.0159	0.9745
		5	0.0011	1.0008	0.0167	0.9728
	100	3	0.0011	0.9996	0.0054	0.9904
		5	0.0007	1.000	0.0052	0.9908

Table 4.4: MSE of $\hat{\alpha}_2$ (Case 2)

			Partition 1		Partition 2	
N_s	n	SNR	$\hat{\alpha}_2$	MSE	$\hat{\alpha}_2$	MSE
1600	50	3	-0.0206	0.9643	-0.0227	0.9598
		5	-0.02178	0.9621	-0.0242	0.9570
	100	3	-0.0089	0.9839	-0.0118	0.9785
		5	-0.0095	0.9826	-0.0101	0.9817
2500	50	3	-0.0168	0.9699	-0.0434	0.9225
		5	-0.0167	0.9702	-0.0411	0.9267
	100	3	-0.0062	0.9888	-0.0046	0.9917
		5	-0.0065	0.9883	-0.0049	0.9910

Tables 4.5 and 4.6 below give the mean squared errors of $\hat{\beta}_3(\cdot)$ and $\hat{\beta}_4(\cdot)$ for 50 simulations from both partitions.

Table 4.5: MSE of $\hat{\beta}_3(\cdot)$ (Case 2)

MSE of $\hat{\beta}_3(\cdot)$				
N_s	n	SNR	Partition 1	Partition 2
1600	50	3	24621.58	24145.77
		5	24604.53	24148.40
	100	3	24610.97	24389.44
		5	24620.17	24396.94
2500	50	3	38267.63	38190.77
		5	38272.71	38212.02
	100	3	38091.68	38342.35
		5	38093.11	38373.96

Table 4.6: MSE of $\hat{\beta}_4(\cdot)$ (Case 2)

MSE of $\hat{\beta}_4(\cdot)$				
N_s	n	SNR	Partition 1	Partition 2
1600	50	3	246.28	278.42
		5	238.19	267.35
	100	3	211.25	225.95
		5	207.15	221.71
2500	50	3	355.61	344.46
		5	347.85	333.89
	100	3	303.18	320.41
		5	296.75	316.51

4.2 Real Data Analysis

Heart disease is one of the leading causes of mortality in the United States, and it is estimated that the prevalence of cardiovascular disease will continue to increase in the future Madjid and Fatemi (2013). There are many risk factors that are related to heart disease, including age, blood pressure, cigarette smoking, serum cholesterol levels, etc. Kannel et al. (1964). We analyze which variables are influential in increasing the risk of heart disease by applying the spatially varying model discussed in the paper.

We consider a dataset consisting of 303 subjects, 138 of those who have an increased risk of heart disease, and 165 of those who do not have an increased risk of heart disease. We consider the following 11 predictors: age, gender, chest pain type, resting blood pressure, serum cholesterol level, fasting blood sugar (FBS), resting electrocardiographic results, maximum heart rate, exercise-induced angina, electrocardiographic peak and the number of damaged major vessels. We consider the following variables in spatially varying form: age, resting blood pressure, cholesterol level, maximum heart rate, and electrocardiographic peak. The other 6 variables are categorical.

For gender, there are 96 males and 207 females. For chest pain type, we consider four levels: 1 if the subject had typical angina, 2 if the subject had atypical angina, 3 if the subject had nonanginal chest pain, and 4 if the subject was asymptomatic. For FBS, we consider two levels: 1 if the subject's FBS was greater than 120 mg/dl, and 0 if the subject's FBS was less than or equal to 120 mg/dl. For resting electrocardiographic results, we consider three levels: 0 if the subject's resting electrocardiographic results were normal, 1 if the subject had ST-T wave abnormality, and 2 if the subject showed probable of definite left ventricular hypertrophy. For exercise-induced angina, we consider two levels: 1 if the subject had exercise-induced angina, and 0 if the subject did not have exercise-induced angina. For the number of damaged major vessels, we consider four levels: 0 if no major vessel is damaged, 1 if one major vessel is damaged, 2 if two major vessels are damaged, and 3 if three major vessels are damaged.

To set up the model, we generate the Bernstein basis polynomials over a triangulation with 16 vertices and 18 triangles. After we notice that following variables are significant in predicting the likelihood of heart disease: gender, chest pain type, maximum heart rate, exercise-induced angina, electrocardiographic peak and the number of damaged major vessels. When considering level 1 from gender, we obtain a p-value of approximately 4.46×10^{-6} and an estimate of approximately -0.2180. This means that females have about 21.8% less of a chance of suffering from heart disease than males, holding other variables constant. The p-values for chest pain type are approximately 0.0011, 3.97×10^{-6}, and 0.0006 for levels 1,2 and 3, respectively. The estimates for levels 1,2 and 3 are approximately 0.2179, 0.2610 and 0.2939, respectively. Thus, compared to those with no chest pain, those with typical angina have about 21.8% more of a chance of suffering from heart disease, those with atypical angina have about 26.1% more of a chance of suffering from heart disease and those with nonanginal pain have about 29.4% more of a chance of suffering from heart disease, holding other variables constant. For maximum heart rate, we obtain a p-value of approximately 0.0068 and an estimate of approximately 0.0030. Thus, for every one unit increase in maximum heart rate, the likelihood of having heart disease increases by about 0.3%, holding other variables constant. The estimate for exercise-induced angina was approximately -0.1283 for level 1, with a p-value of approximately 0.0138. Thus, holding other variables constant, those who had exercise-induced angina have about 12.8% less of a chance of suffering from heart disease than those who did not have exercise-induced angina. The estimate of electrocardiographic peak was about -0.0732, with a p-value of approximately 0.0004. This means that for every one unit increase in electrocardiographic peak, the chance of having heart disease decreases by about 7.3%, holding other variables constant. Finally, for levels 1,2 and 3 of the number of damaged major vessels, the estimates were approximately -0.2837 with p-value 4.24×10^{-7}, -0.3455 with p-value 1.30×10^{-6} and -0.3124 with p-value 0.0006, respectively. Thus, compared to those with three damaged major vessels, those with no damaged major vessels have 28.4% less of a chance of having heart disease, those with one damaged major vessel have 34.6% less of a chance of having heart disease, and those with two damaged major vessels have 31.2% less of a chance of having heart disease, holding other variables constant.

Chapter 5

CONCLUSION

I this major paper we discussed the difficulties of modelling spatially varying data over complex domains. In this major paper, the method of approximating bivariate varying coefficient functions over a triangulated domain was investigated in depth. Theoretically, we demonstrated how consistent estimators of the constant and varying coefficients are obtained from the PLSVCM when the active constant and active varying index sets are known. Through simulation, numerical values for the estimators from the PLSVCM were calculated. Simulation was used to determine if the mean squared errors of the estimators changed when the refinement of the triangulations varied. In the simulation study, the given domain was partitioned into two different triangulations and the mean squared error of the estimators of the varying coefficient functions were compared.

When the dimension of the covariates are very large, the active index sets need to be estimated. In Chapter 3, the dimension of the covariates are greater than the sample size. Thus, a penalized regression approach was considered to mitigate error . Based on the Karush-Kuhn-Tucker conditions given in Boyd et al. (2004), new estimators for the constant coefficients and varying coefficient functions were obtained and detailed theoretical results related to these estimators were shown.

With today's technological advancements, there has been emphasis on three-dimensional imaging scans. To treat these imaging scans as covariates in a regression model, the domain of the varying coefficient functions must be increased from a two-dimensional domain to a three-dimensional domain. This could beneficial to future research, as it would allow one to consider the whole three-dimensional image as a covariate.

REFERENCES

Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Kannel, W. B., DAWBER, T. R., FRIEDMAN, G. D., GLENNON, W. E., and MCNAMARA, P. M. (1964). Risk factors in coronary heart disease: the framingham study. *Annals of internal medicine*, 61(5_Part_1):888–899.

Lai, M.-J. and Schumaker, L. L. (2007). *Spline functions on triangulations*. Cambridge University Press.

Lai, M.-J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica*, pages 1399–1417.

Li, X., Wang, L., Wang, H. J., and Initiative, A. D. N. (2021). Sparse learning and structure identification for ultrahigh-dimensional image-on-scalar regression. *Journal of the American Statistical Association*, 116(536):1994–2008.

Madjid, M. and Fatemi, O. (2013). Components of the complete blood count as risk predictors for coronary heart disease: in-depth review and update. *Texas Heart Institute Journal*, 40(1):17.

Wei, F. and Huang, J. (2010). Consistent group selection in high-dimensional linear regression. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability*, 16(4):1369.

Yu, S., Wang, G., Wang, L., Liu, C., and Yang, L. (2019). Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association*.

APPENDIX

The purpose of the Appendix is to state and prove some of the lemmas used in Li et al. Li et al. (2021), as well as provide detailed proofs of the Theorems in this paper.

Denote $\mathcal{H}^1 = \{f : \int_\Omega f(\mathbf{s})dQ(\mathbf{s}) = 0, \int_\Omega f^2(\mathbf{s})dQ(\mathbf{s}) < \infty\}$ as the space for centred functions, where $Q(\mathbf{s})$ is a distribution with positive continuous density. Denote $\mathcal{H}^2 = \{f : \int_\Omega f(\mathbf{s})dQ(\mathbf{s}) = 0, \int_\Omega f^2(\mathbf{s})dQ(\mathbf{s}) = 1\}$ as the space for normalized functions.

Lemma A1. *Assume that $d \geq 3r + 2$ and for all $k \in A_v$, $\beta_{0k} \in W^{d+1,\infty}(\Omega) \cap \mathcal{H}^1$. Then for $k = 1, 2, ..., p$, there exists a vector $\boldsymbol{\gamma}_{0k}$, where $||\boldsymbol{\gamma}_{0k}|| \neq 0$ if $k \in A_v$, and $||\boldsymbol{\gamma}_{0k}|| = 0$ if $k \notin A_v$.*

Further, there exists a positive constant C depending on d and π, such that for all $k \in A_v$, and for all normalized Bernstein basis polynomials $\mathbf{B}_k = (B_{k\ell}, \ell \in I_k)^T$ of degree $d \geq 0$,

$$sup_{s\in\Omega}|\beta_{0k}(\mathbf{s}) - \mathbf{B}_k^T(\mathbf{s})\boldsymbol{\gamma}_{0k}| \leq C|\Delta|^{d+1}|\beta_{0k}|_{d+1,\infty}.$$

Proof. Note that from Lai et al. Lai and Wang (2013), there exists a vector $\boldsymbol{\gamma}_{0k}^*$ and a positive constant C, such that for Bernstein basis polynomials $\mathbf{B}_k^*$,

$$\sup_{\mathbf{s}\in\Omega}|\beta_{0k}(\mathbf{s}) - (\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}^*| \leq \frac{C}{2}|\Delta|^{d+1}|\beta_{0k}|_{d+1,\infty}$$

Let

$$C' = ||\mathbf{B}_k^* - \int_\Omega \mathbf{B}_k^*||_{L_2(\Omega)}$$

and define $\boldsymbol{\gamma}_{0k} = C'\boldsymbol{\gamma}_{0k}^*$. Let $\mathbf{B}_k(\mathbf{s}) = \frac{1}{C'}(\mathbf{B}_k^*(\mathbf{s}) - \int_\Omega \mathbf{B}_k^*(\mathbf{s})dQ(\mathbf{s}))$. Then

$$\left|\beta_{0k}(\mathbf{s}) - \mathbf{B}_k^T(\mathbf{s})\boldsymbol{\gamma}_{0k}\right| \leq \left|\beta_{0k}(\mathbf{s}) - (\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}\right| + \left|\int_\Omega (\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}^* dQ(\mathbf{s})\right|$$

Adding and subtracting $\beta_{0k}(\mathbf{s})$ and taking the supremum over Ω gives

$$\begin{aligned}\sup_{\mathbf{s}\in\Omega}\left|\beta_{0k}(\mathbf{s}) - \mathbf{B}_k^T(\mathbf{s})\boldsymbol{\gamma}_{0k}\right| \leq\ & \sup_{\mathbf{s}\in\Omega}\left|\beta_{0k}(\mathbf{s}) - (\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}^*\right| \\ & + \sup_{\mathbf{s}\in\Omega}\left|\int_\Omega \left[(\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}^* + \beta_{0k}(\mathbf{s}) - \beta_{0k}(\mathbf{s})\right] dQ(\mathbf{s})\right|.\end{aligned}$$

Then, by the Triangle Inequality,

$$\begin{aligned}\sup_{\mathbf{s}\in\Omega}\left|\beta_{0k}(\mathbf{s}) - \mathbf{B}_k^T(\mathbf{s})\boldsymbol{\gamma}_{0k}\right| \leq\ & \sup_{\mathbf{s}\in\Omega}\left|\beta_{0k}(\mathbf{s}) - (\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}^*\right| \\ & + \sup_{\mathbf{s}\in\Omega}\left|\int_\Omega \left[(\mathbf{B}_k^*(\mathbf{s}))^T\boldsymbol{\gamma}_{0k}^* - \beta_{0k}(\mathbf{s})\right] dQ(\mathbf{s})\right| \\ & + \sup_{\mathbf{s}\in\Omega}\left|\int_\Omega \beta_{0k}(\mathbf{s})dQ(\mathbf{s})\right|.\end{aligned}$$

Note that $\sup_{\mathbf{s}\in\Omega}\left|\int_\Omega \beta_{0k}(\mathbf{s})dQ(\mathbf{s})\right| = 0$, since $\beta_{0k} \in W^{d+1,\infty}(\Omega)\cap\mathcal{H}^1$, for all $k \in A_v$. Therefore,

$$\begin{aligned}\sup_{\mathbf{s}\in\Omega}\left|\beta_{0k}(\mathbf{s}) - \mathbf{B}_k^T(\mathbf{s})\boldsymbol{\gamma}_{0k}\right| &\leq \left(\frac{C}{2} + \frac{C}{2}\right)|\Delta|^{d+1}|\beta_{0k}|_{d+1,\infty} \\ &= C|\Delta|^{d+1}|\beta_{0k}|_{d+1,\infty}.\end{aligned}$$

□

The following lemma from Li et al. (2021) is used to state some properties of the normalized Bernstein basis polynomials.

Lemma A2. *For any normalized Bernstein basis polynomials $B_\ell, B_{\ell'} \in \mathcal{H}^2$, with* *degree $d \geq 0$ and $\ell, \ell' \in J_n$, the following hold:*

$$\max_{\ell \in J_n} \left| \int_{\Omega} B_\ell^k(\mathbf{s}) dQ(\mathbf{s}) \right| = O(|\Delta|^{2-k}), \tag{A1}$$

$$\max_{\ell, \ell' \in J_n} \left| \int_{\Omega^2} B_\ell^k(\mathbf{s}) B_{\ell'}^k(\mathbf{s}') dQ(\mathbf{s}) dQ(\mathbf{s}') \right| = O(|\Delta|^{4-2k}), \tag{A2}$$

$$\max_{\ell \in J_n} \left| \frac{1}{N_s} \sum_{j,j'=1}^{N_s} B_\ell^k(\mathbf{s}_j) - \int_{\Omega} B_\ell^k(\mathbf{s}) dQ(\mathbf{s}) \right| = O(N_s^{-1/2} |\Delta|^{-k+1}), \ k \geq 1, \tag{A3}$$

$$\max_{\ell, \ell' \in J_n} \left| \frac{1}{N_s^2} \sum_{j,j'=1}^{N_s} G_\eta(\mathbf{s}_j, \mathbf{s}_{j'}) B_\ell(\mathbf{s}_j) B_{\ell'}(\mathbf{s}_{j'}) - \int_{\Omega^2} G_\eta(\mathbf{s}, \mathbf{s}') B_\ell(\mathbf{s}) B_{\ell'}(\mathbf{s}') dQ(\mathbf{s}) dQ(\mathbf{s}') \right|$$
$$= O(N_s^{-1/2} |\Delta|), \tag{A4}$$

$$\max_{\ell \in J_n} \left| \frac{1}{N_s} \sum_{j=1}^{N_s} B_\ell^2(\mathbf{s}_j) \sigma^2(\mathbf{s}_j) - \int_{\Omega} \sigma^2(\mathbf{s}) B_\ell^2(\mathbf{s}) dQ(\mathbf{s}) \right| = O(N_s^{-1/2} |\Delta|^{-1}). \tag{A5}$$

The proof of the above lemma is outlined in Li et al. (2021).

The following lemma is cited from Li et al. (2021).

Lemma A3. *Recall the definition of ζ_{ij} $(i = 1, 2, ..., n, j = 1, 2, ..., N_s)$ from the proof of Theorem 1 and that $\boldsymbol{\zeta}_i = (\zeta_{i1}\ \zeta_{i2} \ldots \zeta_{iN_s})^T$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^T\ \boldsymbol{\zeta}_2^T \ldots \boldsymbol{\zeta}_n^T)^T$. Then, under the assumptions in Sections 2.2 and 3.1 and that $N_s^{1/2}|\Delta| \longrightarrow \infty$, as $N_s \longrightarrow \infty$,*

$$\frac{||\mathbf{Z}_A^T \boldsymbol{\zeta}||^2}{(nN_s)^2} = O_p\left(|\Delta|^{2(d+1)}\right).$$

Proof. By definition,

$$\mathbf{Z}_A^T \boldsymbol{\zeta} = \sum_{i=1}^{n} \sum_{j=1}^{N_s} \left(\mathbf{X}_{(i),A_c}^T, \mathbf{X}_{(i),A_v}^T \otimes \mathbf{B}^T(\mathbf{s}_j)\right)^T \times \sum_{k' \in A_v} X_{ik'} \delta_{jk'}.$$

Taking the Euclidean norm and dividing by $(nN_s)^2$ gives

$$\frac{\|\mathbf{Z}_A^T\boldsymbol{\zeta}\|}{n^2N_s^2} = \frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\left[\sum_{k\in A_c} X_{ik}X_{i'k} + \sum_{k\in A_v} X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_j')\right]$$
$$\times \sum_{k'\in A_v}\sum_{k''\in A_v} X_{ik'}X_{i'k''}\delta_{jk'}\delta_{j'k''}.$$

Recall that for all $j = 1, 2, ..., N_s$ and for all $k \in A_v$, δ_{jk} is nonrandom. Thus for all $k \in A_c$,

$$\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\mathrm{E}\left[X_{ik}X_{i'k}\sum_{k',k''\in A_v} X_{ik'}X_{i'k''}\delta_{jk'}\delta_{j'k''}\right]$$
$$\leq \frac{1}{n^2N_s^2}\max_{j,k'}|\delta_{jk'}|\max_{j',k''}|\delta_{j'k''}|\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\sum_{k'\in A_v}\sum_{k''\in A_v}\mathrm{E}\left[|X_{ik}X_{i'k}X_{ik'}X_{i'k''}|\right].$$

Recall from Lemma A1 that $\sup_{s\in\Omega}|\beta_{0k}(\mathbf{s}) - \mathbf{B}_k^T(\mathbf{s})\boldsymbol{\gamma}_{0k}| \leq C|\Delta|^{d+1}|\beta_{0k}|_{d+1,\infty}$. So,

$$\max_{j,k'}|\delta_{jk'}| = \max_{j,k'}|\beta_{0k'}(\mathbf{s}) - \mathbf{B}_{k'}^T(\mathbf{s}_j)\boldsymbol{\gamma}_{0k'}| = O\left(|\Delta|^{d+1}\right),$$

and then $\max_{j',k''}|\delta_{j'k''}| = O\left(|\Delta|^{d+1}\right)$.

Therefore,

$$\max_{j,k'}|\delta_{jk'}| \times \max_{j',k''}|\delta_{j'k''}| = O\left(|\Delta|^{2(d+1)}\right).$$

Now, a bound for the expected value of the product of the predictor variables must be found. To do this, the Cauchy-Schwarz inequality must be applied twice. So for all $i, i' = 1, 2, ..., n$, for all $k \in A_c$, and for all $k', k'' \in A_v$,

$$\mathrm{E}\left[|X_{ik}X_{i'k}X_{ik'}X_{i'k''}|\right] = \mathrm{E}\left[|X_{ik}X_{i'k}||X_{ik'}X_{i'k''}|\right] \leq \sqrt{\mathrm{E}\left[|X_{ik}X_{ik'}|^2\right]}\sqrt{\mathrm{E}\left[|X_{i'k}X_{i'k''}|^2\right]}.$$

Then,

$$\sqrt{\mathrm{E}\left[|X_{ik}X_{ik'}|^2\right]}\sqrt{\mathrm{E}\left[|X_{i'k}X_{i'k''}|^2\right]} \leq \sqrt{\sqrt{\mathrm{E}\left[|X_{ik}|^4\right]}\sqrt{\mathrm{E}\left[|X_{ik'}|^4\right]}}\sqrt{\sqrt{\mathrm{E}\left[|X_{i'k}|^4\right]}\sqrt{E\left[|X_{i'k''}|^4\right]}}.$$

Raising each term to the power of 3/2 and then to the power of 2/3 gives

$$\left((\mathrm{E}\left[|X_{ik}|^4\right])^{3/2}(\mathrm{E}\left[|X_{ik'}|^4\right])^{3/2}(\mathrm{E}\left[|X_{i'k}|^4\right])^{3/2}(\mathrm{E}\left[|X_{i'k''}|^4\right])^{3/2}\right)^{\frac{1}{4}\times\frac{2}{3}}.$$

Now, let $g(x) = x^{3/2}$, which is convex on the interval $(0, +\infty)$. So for any random variable X,

$$(\mathrm{E}[|X|])^{3/2} \leq \mathrm{E}[|X|^{3/2}],$$

by Jensen's inequality. Applying Jensen's inequality, we get

$$\begin{aligned}
&\left((\mathrm{E}\left[|X_{ik}|^4\right])^{3/2}(\mathrm{E}\left[|X_{ik'}|^4\right])^{3/2}(\mathrm{E}\left[|X_{i'k}|^4\right])^{3/2}(\mathrm{E}\left[|X_{i'k''}|^4\right])^{3/2}\right)^{\frac{1}{6}} \\
&\leq \left(\mathrm{E}[|X_{ik}|^6]\mathrm{E}[|X_{ik'}|^6]\mathrm{E}[|X_{i'k}|^6]\mathrm{E}[|X_{i'k''}|^6]\right)^{\frac{1}{6}} \\
&\leq (C_1C_2C_3C_4)^{\frac{1}{6}},
\end{aligned}$$

where C_1, C_2, C_3, C_4 are all positive real numbers given in Assumption 2.

Define $C = (C_1C_2C_3C_4)^{\frac{1}{6}}$. Then

$$\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\sum_{k'\in A_v}\sum_{k''\in A_v}\mathrm{E}\left[|X_{ik}X_{i'k}X_{ik'}X_{i'k''}|\right] \leq n^2N_s^2|A_v|^2C.$$

Therefore, for all $i, i' = 1, 2, ..., n$ and for all $k \in A_c$,

$$\begin{aligned}
&\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\mathrm{E}\left[X_{ik}X_{i'k}\sum_{k',k''\in A_v}X_{ik'}X_{i'k''}\delta_{jk'}\delta_{j'k''}\right] \\
&\leq \frac{1}{n^2N_s^2}\left|\Delta\right|^{2(d+1)}n^2N_s^2|A_v|^2C.
\end{aligned}$$

Hence,

$$\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\mathrm{E}\left[X_{ik}X_{i'k}\sum_{k',k''\in A_v}X_{ik'}X_{i'k''}\delta_{jk'}\delta_{j'k''}\right] = O\left(\left|\Delta\right|^{2(d+1)}\right).$$

Similarly, for all $i, i' = 1, 2, ..., n$ and for all $k \in A_v$,

$$\frac{1}{n^2 N_s^2} \sum_{i,i'=1}^{n} \sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik} X_{i'k} \mathbf{B}^T(\mathbf{s}_j) \mathbf{B}(\mathbf{s}_{j'}) \sum_{k' \in A_v} \sum_{k'' \in A_v} X_{ik'} X_{i'k''} \delta_{jk'} \delta_{j'k''} \right]$$
$$= \frac{1}{n^2 N_s^2} \sum_{i,i'=1}^{n} \sum_{j,j'=1}^{N_s} \sum_{\ell=1}^{J_n} B_\ell(\mathbf{s}_j) B_\ell(\mathbf{s}_{j'}) \sum_{k' \in A_v} \sum_{k'' \in A_v} \mathrm{E}\left[X_{ik} X_{i'k} X_{ik'} X_{i'k''} \delta_{jk'} \delta_{j'k''} \right]$$
$$\leq \max_{j,k'} |\delta_{jk'}| \max_{j',k''} |\delta_{j'k''}| \sum_{\ell=1}^{J_n} \frac{1}{N_s^2} \sum_{j,j'=1}^{N_s} |B_\ell(\mathbf{s}_j) B_\ell(\mathbf{s}_{j'})|$$
$$\times \frac{1}{n^2} \sum_{i,i'=1} \sum_{k',k'' \in A_v} \mathrm{E}\left[|X_{ik} X_{i'k} X_{ik'} X_{i'k''}| \right].$$

Arguing in the same way as earlier in the proof, for all $i, i' = 1, 2, ..., n$ and for all $k \in A_v$,

$$\mathrm{E}\left[|X_{ik} X_{i'k} X_{ik'} X_{i'k''}| \right]$$
$$\leq \left(\mathrm{E}[|X_{ik}|^6] \mathrm{E}[|X_{ik'}|^6] \mathrm{E}[|X_{i'k}|^6] \mathrm{E}[|X_{i'k''}|^6] \right)^{\frac{1}{6}}$$
$$\leq (C_1' C_2' C_3' C_4')^{\frac{1}{6}},$$

where C_1', C_2', C_3', C_4' are all positive real numbers defined in Assumption 2.

Denote $C' = C_1' C_2' C_3' C_4'$. Then

$$\sum_{i,i'=1} \sum_{k',k'' \in A_v} \mathrm{E}\left[|X_{ik} X_{i'k} X_{ik'} X_{i'k''}| \right] \leq n^2 |A_v|^2 C'.$$

By the definition of Bernstein basis polynomials,

$$\sum_{\ell=1}^{J_n} \frac{1}{N_s^2} \sum_{j,j'=1}^{N_s} |B_\ell(\mathbf{s}_j) B_\ell(\mathbf{s}_{j'})| \leq \sum_{\ell=1}^{J_n} \left(\frac{N_s^2}{N_s^2} C^* \right) = J_n C^*,$$

where C^* is a positive real number. Further, it has already been showed that

$$\max_{j,k'} |\delta_{jk'}| \times \max_{j',k''} |\delta_{j'k''}| = O\left(|\Delta|^{2(d+1)} \right).$$

Putting everything together gives

$$\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\mathrm{E}\left[X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\sum_{k'\in A_v}\sum_{k''\in A_v}X_{ik'}X_{i'k''}\delta_{jk'}\delta_{j'k''}\right]$$
$$\leq \frac{1}{n^2}\left|\Delta\right|^{2(d+1)}J_nC^*n^2|A_v|^2C'.$$

Hence,

$$\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\mathrm{E}\left[X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\sum_{k'\in A_v}\sum_{k''\in A_v}X_{ik'}X_{i'k''}\delta_{jk'}\delta_{j'k''}\right]=O\left(|\Delta|^{2(d+1)}\right).$$

Therefore,

$$\frac{||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2}{(nN_s)^2}=O_p\left(|\Delta|^{2(d+1)}\right),$$

which proves the result.

□

The following lemma is cited from Li et al. (2021).

Lemma A4. *Recall that* $\boldsymbol{\epsilon}_i=(\epsilon_i(\mathbf{s}_1)\ \epsilon_i(\mathbf{s}_2)\dots\epsilon_i(\mathbf{s}_{N_s}))^T$ *and* $\boldsymbol{\epsilon}=(\boldsymbol{\epsilon}_1^T\ \boldsymbol{\epsilon}_2^T\dots\boldsymbol{\epsilon}_n^T)^T$. *Then, under the assumptions in Sections 2.2 and 3.1 and that* $N_s^{1/2}|\Delta|\longrightarrow\infty$, *as* $N_s\longrightarrow\infty$,

$$\frac{||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2}{(nN_s)^2}=O_p\left(\frac{1}{nN_s|\Delta|^2}\right).$$

Proof. By definition of the Euclidean Norm,

$$\frac{||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2}{(nN_s)^2}=\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\left[\sum_{k\in A_c}X_{ik}X_{i'k}+\sum_{k\in A_v}X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\right]\times\epsilon_{ij}\epsilon_{i'j'}$$

Using the condition that every X_{ik} $(i=1,2,...,n,\ k=1,2,...,p)$ and ϵ_{ij} $(i=$

$1,2,...,n,\ j=1,2,...,N_s$) are independent, for all $i=1,2,...,n$ and for all $k \in A_c$,

$$\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik}^2 \epsilon_{ij}\epsilon_{ij'}\right] = \frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik}^2\right] \mathrm{E}\left[\epsilon_{ij}\epsilon_{ij'}\right].$$

Note that for all $i=1,2,...,n$ and for all $j=1,2,...,N_s$, the ϵ_{ij}'s are independent with mean zero and variance $\sigma^2(\mathbf{s}_j)$ by Assumption 3. Thus for $j \neq j'$, $\mathrm{E}\left[\epsilon_{ij}\epsilon_{ij'}\right] = \mathrm{E}\left[\epsilon_{ij}\right]\mathrm{E}\left[\epsilon_{ij'}\right] = 0$. Therefore, write $\mathrm{E}\left[\epsilon_{ij}\epsilon_{ij'}\right] = \sigma^2(\mathbf{s}_j)\mathcal{I}(j=j')$, where $\mathcal{I}(\cdot)$ is the indicator function. Thus, by Cauchy-Schwarz inequality,

$$\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik}^2\right]\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j') \leq \frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} (\mathrm{E}\left[X_{ik}^4\right])^{1/2}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j').$$

By Jensen's inequality and Assumption 2,

$$\begin{aligned}\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} (\mathrm{E}\left[X_{ik}^4\right])^{1/2}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j') &\leq \frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} (\mathrm{E}\left[X_{ik}^6\right])^{1/3}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j') \\ &\leq \frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} c_x^{1/3}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j').\end{aligned}$$

This gives

$$\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} (\mathrm{E}\left[X_{ik}^4\right])^{1/2}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j') \leq \frac{1}{N_s^2}\sum_{j=1}^{N_s} c_x^{1/3}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j'),$$

where $c_x^{1/3}$ is a positive real number. So, for some positive real number c_x',

$$\frac{1}{N_s^2}\sum_{j=1}^{N_s} c_x^{1/3}\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j') \leq \frac{1}{N_s^2}N_s(c_x') = O\left(\frac{1}{N_s}\right),$$

because $\sigma^2(\mathbf{s}_j)$ is bounded, for all $j=1,2,...,N_s$. Thus, for all $i=1,2,...,n$ and for all $k \in A_c$,

$$\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik}^2 \epsilon_{ij}\epsilon_{ij'}\right] = O\left(\frac{1}{N_s}\right).$$

Similar as before, for all $i = 1, 2, ..., n$ and for all $k \in A_v$,

$$\begin{aligned}
&\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik}^2 \mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\epsilon_{ij}\epsilon_{ij'}\right] \\
&= \frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} \mathrm{E}\left[X_{ik}^2\right]\sum_{\ell=1}^{J_n} B_\ell^2(\mathbf{s}_j)\mathrm{E}\left[\epsilon_{ij}\epsilon_{ij'}\right] \\
&\leq \frac{1}{N_s^2}\sum_{j,j'=1}^{N_s} k_x^{1/3}\sum_{\ell=1}^{J_n} B_\ell^2(\mathbf{s}_j)\sigma^2(\mathbf{s}_j)\mathcal{I}(j=j') \\
&= \frac{k_x^{1/3}}{N_s^2}\sum_{j=1}^{N_s}\sum_{\ell=1}^{J_n} B_\ell^2(\mathbf{s}_j)\sigma^2(\mathbf{s}_j),
\end{aligned}$$

where $k_x^{1/3} := k$ is some positive real number.

Using Equation (A5), we get

$$\begin{aligned}
\frac{k}{N_s^2}\sum_{j=1}^{N_s}\sum_{\ell=1}^{J_n} B_\ell^2(\mathbf{s}_j)\sigma^2(\mathbf{s}_j) &= \frac{k}{N_s}\sum_{\ell=1}^{J_n}\left[\frac{1}{N_s}\sum_{j=1}^{N_s} B_\ell^2(\mathbf{s}_j)\sigma^2(\mathbf{s}_j)\right] \\
&= \frac{k}{N_s}\sum_{\ell=1}^{J_n}\int_\Omega B_\ell^2(\mathbf{s})\sigma^2(\mathbf{s})dQ(\mathbf{s}) \\
&\quad + \frac{k}{N_s}\sum_{\ell=1}^{J_n}\int_\Omega B_\ell^2(\mathbf{s})\sigma^2(\mathbf{s})dQ(\mathbf{s}) \cdot O\left(N_s^{-1/2}|\Delta|^{-1}\right).
\end{aligned}$$

The assumption that $N_s^{1/2}|\Delta| \longrightarrow \infty$, as $N_s \to \infty$ implies that $N_s^{-1/2}|\Delta|^{-1} \longrightarrow 0$, as $N_s \to \infty$. Therefore,

$$\begin{aligned}
&\frac{k}{N_s}\sum_{\ell=1}^{J_n}\int_\Omega B_\ell^2(\mathbf{s})\sigma^2(\mathbf{s})dQ(\mathbf{s}) + \frac{k}{N_s}\sum_{\ell=1}^{J_n}\int_\Omega B_\ell^2(\mathbf{s})\sigma^2(\mathbf{s})dQ(\mathbf{s}) \cdot O\left(N_s^{-1/2}|\Delta|^{-1}\right) \\
&= \frac{k}{N_s}\sum_{\ell=1}^{J_n}\left[\int_\Omega B_\ell^2(\mathbf{s})\sigma^2(\mathbf{s})dQ(\mathbf{s})\right] = \frac{k}{N_s}J_n|\Delta|^{-2} = O\left(\frac{1}{N_s|\Delta|^2}\right).
\end{aligned}$$

For all $i \neq i'$, $\mathrm{E}\left[X_{ik}X_{i'k}\epsilon_{ij}\epsilon_{i'j'}\right] = 0$ and $\mathrm{E}\left[X_{ik}X_{i'k}\right] = 0$, so

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i,i'=1}^{n}\mathrm{E}\left[\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s}\left(\sum_{k\in A_c}X_{ik}X_{i'k}+\sum_{k\in A_v}X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\right)\epsilon_{ij}\epsilon_{ij'}\right]\\
&=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{E}\left[\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s}\left(\sum_{k\in A_c}X_{ik}X_{i'k}+\sum_{k\in A_v}X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\right)\epsilon_{ij}\epsilon_{ij'}\right]\\
&=\frac{1}{n^2}\times n\times O\left(\frac{1}{N_s|\Delta|^2}\right).
\end{aligned}$$

This gives

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i,i'=1}^{n}\mathrm{E}\left[\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s}\left(\sum_{k\in A_c}X_{ik}X_{i'k}+\sum_{k\in A_v}X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\right)\epsilon_{ij}\epsilon_{ij'}\right]\\
&=O\left(\frac{1}{nN_s|\Delta|^2}\right).
\end{aligned}$$

Therefore,

$$\frac{||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2}{(nN_s)^2} = O_p\left(\frac{1}{nN_s|\Delta|^2}\right),$$

which proves the result. □

The following lemma is cited from Li et al. (2021).

Lemma A5. *Recall from definition that* $\boldsymbol{\eta}_i = (\eta_i(\mathbf{s}_1)\ \eta_i(\mathbf{s}_2)\dots\eta_i(\mathbf{s}_{N_s}))^T$ *and* $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T\ \boldsymbol{\eta}_2^T\dots\boldsymbol{\eta}_n^T)^T$*. Then, under the assumptions in Sections 2.2 and 3.1 and that* $N_s^{1/2}|\Delta| \longrightarrow \infty$*, as* $N_s \longrightarrow \infty$*,*

$$\frac{||\mathbf{Z}_A^T\boldsymbol{\eta}||^2}{(nN_s)^2} = O_p\left(\frac{1}{n}\right).$$

Proof. By defiintion of the Euclidean norm,

$$\frac{||\mathbf{Z}_A^T\boldsymbol{\eta}||^2}{(nN_s)^2} = \frac{1}{(nN_s)^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\left[\sum_{k\in A_c} X_{ik}X_{i'k} + \sum_{k\in A_v} X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\right]\times \eta_i(\mathbf{s}_j)\eta_{i'}(\mathbf{s}_{j'})$$

Note that for all $i \neq i'$, and $k \in A_c$, $\mathrm{E}\left[X_{ik}X_{i'k}\right] = 0$. Thus, by the condition that every X_{ik} $(i = 1, 2, ..., n,\ k = 1, 2, ..., p)$ and $\eta_i(\mathbf{s}_j)$ $(i = 1, 2, ..., n,\ j = 1, 2, ..., N_s)$ are independent,

$$\begin{aligned}\frac{1}{n^2N_s^2}\sum_{j,j'=1}^{N_s}\sum_{i,i'=1}^{n}\mathrm{E}\left[X_{ik}X_{i'k}\eta_i(\mathbf{s}_j)\eta_{i'}(\mathbf{s}_{j'})\right] &= \frac{1}{n^2N_s^2}\sum_{j,j'=1}^{N_s}\sum_{i=1}^{n}\mathrm{E}\left[X_{ik}^2\eta_i(\mathbf{s}_j)\eta_{i'}(\mathbf{s}_{j'})\right]\\ &= \frac{1}{n^2N_s^2}\sum_{i=1}^{n}\mathrm{E}\left[X_{ik}^2\right]\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})\end{aligned}$$

By the Cauchy-Schwarz inequality, for all $k \in A_c$,

$$\begin{aligned}\frac{1}{n^2N_s^2}\sum_{i=1}^{n}\mathrm{E}\left[X_{ik}^2\right]\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'}) &\leq \frac{1}{n^2N_s^2}\sum_{i=1}^{n}(\mathrm{E}\left[X_{ik}^4\right])^{1/2}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})\\ &\leq \frac{1}{n^2N_s^2}\sum_{i=1}^{n}(\mathrm{E}\left[X_{ik}^6\right])^{1/3}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'}).\end{aligned}$$

By Jensen's Inequality, for all $k \in A_c$,

$$\frac{1}{n^2N_s^2}\sum_{i=1}^{n}(\mathrm{E}\left[X_{ik}^4\right])^{1/2}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'}) \leq \frac{1}{n^2N_s^2}\sum_{i=1}^{n}(\mathrm{E}\left[X_{ik}^6\right])^{1/3}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'}).$$

Recall from Assumption 2 that there exists a positive real number C_X, such that $\mathrm{E}[|X_k|^6] \leq C_X$. Recall from Assumption 3 that there exists a positive real number C_G, such that $G_\eta(\mathbf{s},\mathbf{s}) \leq C_G$, for all $\mathbf{s} \in \Omega$. So for all $k \in A_c$,

$$\begin{aligned}\frac{1}{n^2N_s^2}\sum_{i=1}^{n}(\mathrm{E}\left[X_{ik}^6\right])^{1/3}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'}) &\leq \frac{1}{n^2N_s^2}\sum_{i=1}^{n}(C_X)^{1/3}\sum_{j,j'=1}^{N_s}C_G\\ &= \frac{1}{n^2N_s^2}(nC_X)(N_s^2C_G).\end{aligned}$$

Thus,

$$\frac{1}{n^2N_s^2}\sum_{i=1}^{n}(\mathrm{E}\left[X_{ik}^6\right])^{1/3}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})=O\left(\frac{1}{n}\right).$$

Again since $\mathrm{E}\left[X_{ik}X_{i'k}\right]=0$, for all $i\neq i'$, and $k\in A_v$,

$$\begin{aligned}
&\frac{1}{n^2N_s^2}\sum_{i,i'=1}^{n}\sum_{j,j'=1}^{N_s}\mathrm{E}\left[X_{ik}X_{i'k}\mathbf{B}^T(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_{j'})\eta_i(\mathbf{s}_j)\eta_{i'}(\mathbf{s}_{j'})\right]\\
&=\frac{1}{n^2N_s^2}\sum_{i=1}^{n}\mathrm{E}\left[X_{ik}^2\right]\sum_{\ell=1}^{J_n}\sum_{j,j'=1}^{N_s}B_\ell(\mathbf{s}_j)B_\ell(\mathbf{s}_{j'})G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})\\
&\leq\frac{1}{n^2N_s^2}\sum_{i=1}^{n}C'_X\sum_{\ell=1}^{J_n}\sum_{j,j'=1}^{N_s}B_\ell(\mathbf{s}_j)B_\ell(\mathbf{s}_{j'})G_\eta(\mathbf{s}_j,\mathbf{s}_{j'}),
\end{aligned}$$

where C'_X is a positive real number. For all $1\leq\ell\leq J_n$, $B_\ell^2(\mathbf{s})B_\ell^2(\mathbf{s}')\neq 0$ only if $\mathbf{s}$ and $\mathbf{s}'$ are in the same triangle $\tau_{\lceil\ell/d'\rceil}$, where $d'=\frac{(d+1)(d+2)}{2}$ is the number of Bernstein basis polynomials on each triangle.

Recall from Equation (A4) that

$$\begin{aligned}
&\max_{\ell,\ell'\in J_n}\left|\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s}G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})B_\ell(\mathbf{s}_j)B_{\ell'}(\mathbf{s}_{j'})-\int_{\Omega^2}G_\eta(\mathbf{s},\mathbf{s}')B_\ell(\mathbf{s})B_{\ell'}(\mathbf{s}')dQ(\mathbf{s})dQ(\mathbf{s}')\right|\\
&=O(N_s^{-1/2}|\Delta|).
\end{aligned}$$

This implies that

$$\begin{aligned}
&\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s}B_\ell(\mathbf{s}_j)B_\ell(\mathbf{s}_{j'})G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})\\
&=\left(1+O(N_s^{-1/2}|\Delta|)\right)\int_{\Omega^2}G_\eta(\mathbf{s},\mathbf{s}')B_\ell(\mathbf{s})B_\ell(\mathbf{s}')dQ(\mathbf{s})dQ(\mathbf{s}')\\
&\leq\left(1+O(N_s^{-1/2}|\Delta|)\right)\int_{\tau_{\lceil\ell/d'\rceil}\times\tau_{\lceil\ell/d'\rceil}}G_\eta(\mathbf{s},\mathbf{s}')B_\ell(\mathbf{s})B_\ell(\mathbf{s}')dQ(\mathbf{s})dQ(\mathbf{s}').
\end{aligned}$$

Therefore,

$$\frac{1}{N_s^2}\sum_{j,j'=1}^{N_s}B_\ell(\mathbf{s}_j)B_\ell(\mathbf{s}_{j'})G_\eta(\mathbf{s}_j,\mathbf{s}_{j'})=O\left(|\Delta|^2\right).$$

So, for all $k \in A_v$,

$$= \frac{1}{n^2 N_s^2} \sum_{i=1}^{n} \mathrm{E}\left[X_{ik}^2\right] \sum_{\ell=1}^{J_n} \sum_{j,j'=1}^{N_s} B_\ell(\mathbf{s}_j) B_\ell(\mathbf{s}_{j'}) G_\eta(\mathbf{s}_j, \mathbf{s}_{j'})$$
$$\leq \frac{1}{n^2 N_s^2}(n)(C)(J_n)(N_s^2),$$

where C is a positive real number. Thus,

$$\frac{1}{n^2 N_s^2} \sum_{i=1}^{n} \mathrm{E}\left[X_{ik}^2\right] \sum_{\ell=1}^{J_n} \sum_{j,j'=1}^{N_s} B_\ell(\mathbf{s}_j) B_\ell(\mathbf{s}_{j'}) G_\eta(\mathbf{s}_j, \mathbf{s}_{j'}) = O\left(\frac{1}{n}\right).$$

Therefore,

$$\frac{||\mathbf{Z}_A^T \boldsymbol{\eta}||^2}{(nN_s)^2} = O_p\left(\frac{1}{n}\right),$$

which proves the result. □

Proof of Theorem 2.1. Define $\delta_{jk} = \beta_{0k}(\mathbf{s}_j) - \mathbf{B}^T(\mathbf{s}_j)\boldsymbol{\gamma}_{0k}$ to be the best spline approximation error of β_{0k} at the point $\mathbf{s}_j$. Define $\boldsymbol{\zeta}_i = (\zeta_{i1}\ \zeta_{i2} \ldots \zeta_{iN_s})^T$, where $\zeta_{ij} = \sum_{k \in A_v} X_{ik}\delta_{jk}$, and denote $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1^T\ \boldsymbol{\zeta}_2^T \ldots \boldsymbol{\zeta}_n^T)^T$, which is a vector with length nN_s. Further, define $\boldsymbol{\eta}_i = (\eta_i(\mathbf{s}_1)\ \eta_i(\mathbf{s}_2) \ldots \eta_i(\mathbf{s}_{N_s}))^T$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T\ \boldsymbol{\eta}_2^T \ldots \boldsymbol{\eta}_n^T)^T$ and $\boldsymbol{\epsilon}_i = (\epsilon_i(\mathbf{s}_1)\ \epsilon_i(\mathbf{s}_2) \ldots \epsilon_i(\mathbf{s}_{N_s}))^T$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T\ \boldsymbol{\epsilon}_2^T \ldots \boldsymbol{\epsilon}_n^T)^T$. So, $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ are vectors with length nN_s. Thus, $\mathbf{Y} - \mathbf{Z}_A\boldsymbol{\theta}_{0,A} = \boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}$. Taking the difference between $\hat{\boldsymbol{\theta}}^0$ and $\boldsymbol{\theta}_{0,A}$ gives

$$\begin{aligned}
\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A} &= (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T\mathbf{Y} - \boldsymbol{\theta}_{0,A} \\
&= (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T(\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta} + \mathbf{Z}_A\boldsymbol{\theta}_{0,A}) - \boldsymbol{\theta}_{0,A} \\
&= (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T(\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}) + \mathbf{I}_A\boldsymbol{\theta}_{0,A} - \boldsymbol{\theta}_{0,A} \\
&= (nN_s\mathbf{C}_A)^{-1}\mathbf{Z}_A^T(\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}).
\end{aligned}$$

Then,

$$\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A} = \frac{1}{nN_s}\mathbf{C}_A^{-1}\mathbf{Z}_A^T(\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}),$$

and then, by the Cauchy-Schwarz Inequality,

$$\begin{aligned}||\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A}||^2 &\leq ||\mathbf{C}_A^{-1}||^2 \frac{1}{(nN_s)^2}||\mathbf{Z}_A^T(\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta})||^2 \\ &\leq \pi_1^{-2}\frac{1}{(nN_s)^2}||\mathbf{Z}_A^T(\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta})||^2.\end{aligned}$$

By applying the Triangle Inequality twice,

$$\begin{aligned}||\mathbf{Z}_A^T\boldsymbol{\eta} + \mathbf{Z}_A^T\boldsymbol{\epsilon} + \mathbf{Z}_A^T\boldsymbol{\zeta}||^2 &\leq (||\mathbf{Z}_A^T\boldsymbol{\eta}|| + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}|| + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||)^2 \\ &= ||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2 + 2||\mathbf{Z}_A^T\boldsymbol{\eta}||||\mathbf{Z}_A^T\boldsymbol{\epsilon}|| \\ &\quad + 2||\mathbf{Z}_A^T\boldsymbol{\eta}||||\mathbf{Z}_A^T\boldsymbol{\zeta}|| + 2||\mathbf{Z}_A^T\boldsymbol{\epsilon}||||\mathbf{Z}_A^T\boldsymbol{\zeta}||.\end{aligned}$$

Further,

$$\begin{aligned}&||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2 + 2||\mathbf{Z}_A^T\boldsymbol{\eta}||||\mathbf{Z}_A^T\boldsymbol{\epsilon}|| + 2||\mathbf{Z}_A^T\boldsymbol{\eta}||||\mathbf{Z}_A^T\boldsymbol{\zeta}|| \\ &\quad + 2||\mathbf{Z}_A^T\boldsymbol{\epsilon}||||\mathbf{Z}_A^T\boldsymbol{\zeta}|| \\ &\leq ||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2 + (||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2) \\ &\quad + (||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2) + (||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2) \\ &= 3(||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2 + ||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2).\end{aligned}$$

Thus, for some positive constant c,

$$\begin{aligned}||\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A}||^2 &\leq c\left(\frac{||\mathbf{Z}_A^T\boldsymbol{\eta}||^2}{(nN_s)^2} + \frac{||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2}{(nN_s)^2} + \frac{||\mathbf{Z}_A^T\boldsymbol{\zeta}||^2}{(nN_s)^2}\right) \\ &= O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{nN_s|\Delta|^2}\right) + O_p\left(|\Delta|^{2(d+1)}\right) \\ &= O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)}\right),\end{aligned}$$

by Lemmas A3, A4, and A5.

Define the block matrix $\begin{bmatrix}\mathbf{I}_{|A_c|} & \mathbf{0}_{|A_c|\times|A_v|J_n}\end{bmatrix}$, where $\mathbf{I}_{|A_c|}$ is the $|A_c|$-dimensional identity matrix, and $\mathbf{0}_{|A_c|\times|A_v|J_n}$ is a matrix of zeros with dimension $|A_c| \times |A_v|J_n$. Similarly, define the block matrix $\begin{bmatrix}\mathbf{0}_{|A_v|J_n\times|A_c|} & \mathbf{I}_{|A_v|J_n}\end{bmatrix}$. Write

$$\hat{\boldsymbol{\alpha}}^0 - \boldsymbol{\alpha}_{0,A_c} = \begin{bmatrix}\mathbf{I}_{|A_c|} & \mathbf{0}_{|A_c|\times|A_v|J_n}\end{bmatrix}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A}).$$

This implies that

$$\sum_{k\in A_c}(\hat{\alpha}_k^0 - \alpha_{0k})^2 = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)}\right),$$

which proves (a). Write

$$(\hat{\boldsymbol{\gamma}}^0 - \boldsymbol{\gamma}_{0,A_v}) = \begin{bmatrix}\mathbf{0}_{|A_v|J_n\times|A_c|} & \mathbf{I}_{|A_v|J_n}\end{bmatrix}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A}).$$

Further, for all $k \in A_v$, $\hat{\beta}_k^0(\mathbf{s}) - \beta_{0k}(\mathbf{s}) = \mathbf{B}^T(\mathbf{s})(\hat{\gamma}_k^0 - \gamma_{0,k})$ This implies that

$$\sum_{k\in A_v}||\hat{\beta}_k^0 - \beta_{0k}||^2_{L_2(\Omega)} = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)}\right),$$

which proves (b). □

To consider bounds for the normalized Bernstein basis functions, the following lemma from Lai and Wang (2013) and Li et al. (2021) is cited below.

Lemma A6. *Recall the spline space $\mathcal{S}_d^r(\Delta)\cap\mathcal{H}^2$ and let $\{B_\ell\}_{\ell\in I}$ be the normalized Bernstein basis polynomials for $\mathcal{S}_d^r(\Delta)\cap\mathcal{H}^2$, where $I \subseteq \{1,2,...,p\}$ is any index set. Then there exists positive real numbers c and C, depending on the smoothness parameter r and the shape parameter π of Δ, such that*

$$c\sum_{\ell\in I}\gamma_\ell^2 \le \|\sum_{\ell\in I}\gamma_\ell B_\ell\|^2_{L_2(\Omega)} \le C\sum_{\ell\in I}\gamma_\ell^2.$$

The proof of this lemma can be found in Lai and Wang (2013).

Now, some new definitions and notations will be cited from Li et al. (2021). For all $k = 1, 2, ..., p$, let $x_k(\mathbf{x}, \mathbf{s}) = x_k$ be a functions which maps $(\mathbf{x}, \mathbf{s})$ to the k^{th} element of $\mathbf{x}$. Define

$$\mathcal{F}_+ = \{F(\mathbf{x}, \mathbf{s}) = \sum_{k \in A_v} x_k g_k(\mathbf{s}) : \int_\Omega g_k(\mathbf{s}) dQ(\mathbf{s}) = 0\}$$

and for all $k \in A_c$, let

$$\begin{aligned} \Gamma_k(\cdot, \cdot) &= \underset{F(\cdot,\cdot) \in \mathcal{F}_+}{\arg\min} \ \mathrm{E}\left[\int_\Omega (X_{ik} - F(\mathbf{X}_{(i)}, \mathbf{s}))^2 dQ(\mathbf{s})\right] \\ &= \underset{F(\cdot,\cdot) \in \mathcal{F}_+}{\arg\min} \ \|x_k - F\|^2 \end{aligned}$$

be the orthogonal projection of x_k onto $\mathcal{F}_+$ relative to the theoretical inner product defined as

$$< g_1, g_2 >= \mathrm{E}\left[\int_\Omega g_1(\mathbf{X}, \mathbf{s}) g_2(\mathbf{X}, \mathbf{s}) dQ(\mathbf{s})\right].$$

Define the corresponding theoretical norm as $|| \cdot ||$. Let $\mathbf{\Gamma}_{A_c}(\mathbf{X}, \mathbf{s}) = \{\Gamma_k(\mathbf{X}, \mathbf{s}), k \in A_c\}^T$ and define

$$\mathcal{F}_{n,+} = \{F(\mathbf{x}, \mathbf{s}) = \sum_{k \in A_v} x_k g_k(\mathbf{s}) : g_k(\mathbf{s}) \in \mathcal{S}_d^r(\Delta) \cap \mathcal{H}^2\}.$$

For all $k \in A_c$, define

$$\Gamma_{n,k}(\cdot, \cdot) = \underset{F(\cdot,\cdot) \in \mathcal{F}_{n,+}}{\arg\min} \ \frac{1}{nN_s} \sum_{i=1}^{n} \sum_{j=1}^{N_s} (X_{ik} - F(\mathbf{X}_{(i)}, \mathbf{s}_j))^2$$

as the orthogonal projection of x_k onto $\mathcal{F}_{n,+}$ relative to the empirical inner product defined as

$$< g_1, g_2 >_{n,N_s} = \frac{1}{nN_s} \sum_{i=1}^{n} \sum_{j=1}^{N_s} g_1(\mathbf{X}_{(i)}, \mathbf{s}_j) \ g_2(\mathbf{X}_{(i)}, \mathbf{s}_j).$$

Define the corresponding empirical norm as $||\cdot||_{n,N_s}$. Next, define two matrices

$$\mathbf{\Xi} = E\int_{\Omega}[\mathbf{X}_{A_c} - \mathbf{\Gamma}_{A_c}(\mathbf{X},\mathbf{s})][\mathbf{X}_{A_c} - \mathbf{\Gamma}_{A_c}(\mathbf{X},\mathbf{s})]^T d\mathbf{s},$$

$$\mathbf{\Xi}_e = E\int_{\Omega^{\otimes 2}}[\mathbf{X}_{A_c} - \mathbf{\Gamma}_{A_c}(\mathbf{X},\mathbf{s})]\Sigma_e(\mathbf{s},\mathbf{s}')[\mathbf{X}_{A_c} - \mathbf{\Gamma}_{A_c}(\mathbf{X},\mathbf{s}')]^T d\mathbf{s}d\mathbf{s}',$$

where $\Sigma_e(\mathbf{s},\mathbf{s}') = G_\eta(\mathbf{s},\mathbf{s}') + \sigma(\mathbf{s})I(\mathbf{s}=\mathbf{s}')$.

The following Theorem from Li et al. (2021) is cited below.

Theorem A1. *Suppose that the assumptions in Sections 2.2 and 3.1 hold. Then for all $k \in A_c$,*

$$||\hat{\Gamma}_{n,k} - \Gamma_k||^2_{n,N_s} = o_p(1).$$

Proof. For any $k \in A_c$, define

$$\tilde{\Gamma}_{n,k} = \underset{F(\cdot,\cdot)\in\mathcal{F}_{n,+}}{\arg\min}\ \mathrm{E}\left[\int_{\Omega}\left(X_{ik} - F(\mathbf{X}_{(i)},\mathbf{s})\right)^2 dQ(\mathbf{s})\right] = \underset{F(\cdot,\cdot)\in\mathcal{F}_{n,+}}{\arg\min}\ ||x_k - F||^2$$

as the orthogonal projection of x_k onto $\mathcal{F}_{n,+}$, relative to the theoretical norm, $||\cdot||$. Then, $\tilde{\Gamma}_{n,k} = \Pi_n x_k$, where Π_n is the projection operator onto $\mathcal{F}_{n,+}$, relative to the theoretical norm. Define $\hat{\Pi}_n$ as the projection operator onto $\mathcal{F}_{n,+}$, relative to the empirical norm. Then, by the triangle inequality,

$$||\hat{\Gamma}_{n,k} - \Gamma_k||_{n,N_s} \leq ||\tilde{\Gamma}_{n,k} - \Gamma_k||_{n,N_s} + ||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||_{n,N_s}.$$

By the definition of Γ_k, there exists $\{g^0_{k,k'} : \int_\Omega g_{k'}(\mathbf{s})dQ(\mathbf{s}) = 0\}_{k'\in A_v}$, such that $\Gamma_k = \sum_{k'\in A_v} x_{k'}g^0_{k,k'}$. Since $\tilde{\Gamma}_{n,k} = \Pi_n x_k$, we have

$$||\tilde{\Gamma}_{n,k} - \Gamma_k||^2 = ||\Pi_n\Gamma_k - \Gamma_k||^2.$$

So,

$$
\begin{aligned}
||\tilde{\Gamma}_{n,k} - \Gamma_k||^2 &= \inf_{F(\cdot,\cdot)\in\mathcal{F}_{n,+}} ||F - \Gamma_k||^2 = \inf_{F(\cdot,\cdot)\in\mathcal{F}_{n,+}} \left\| F - \sum_{k'\in A_v} X_{k'} g^0_{k,k'} \right\|^2 \\
&= \inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} \left\| \sum_{k'\in A_v} X_{k'} g_{k,k'} - \sum_{k'\in A_v} X_{k'} g^0_{k,k'} \right\|^2 \\
&= \inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} \left\| \sum_{k'\in A_v} X_{k'} (g_{k,k'} - g^0_{k,k'}) \right\|^2 .
\end{aligned}
$$

By the Cauchy-Schwarz inequality, $\forall k \in A_c$, we get

$$
\inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} \left\| \sum_{k'\in A_v} X_{k'} (g_{k,k'} - g^0_{k,k'}) \right\|^2 \leq \inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} \sum_{k'\in A_v} ||X_{k'}||^2 \sum_{k'\in A_v} ||g_{k,k'} - g^0_{k,k'}||^2,
$$

and so,

$$
||\tilde{\Gamma}_{n,k} - \Gamma_k||^2 \leq \sum_{k'\in A_v} \mathrm{E}[X^2_{k'}] \sum_{k'\in A_v} \inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} ||g_{k,k'} - g^0_{k,k'}||^2.
$$

By Jensen's inequality, $\forall k' \in A_v$,

$$
\left(\mathrm{E}[X^2_{k'}]\right)^3 \leq \mathrm{E}\left[(X^2_{k'})^3\right],
$$

which implies

$$
\mathrm{E}[X^2_{k'}] \leq (C_X)^{1/3},
$$

where C_X is defined in Assumption 2. So, we have

$$
\begin{aligned}
&\sum_{k'\in A_v} \mathrm{E}[X^2_{k'}] \sum_{k'\in A_v} \inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} ||g_{k,k'} - g^0_{k,k'}||^2 \\
&\leq |A_v|(C_X)^{1/3} \sum_{k'\in A_v} \inf_{g_{k,k'}\in\mathcal{S}^r_d\cap\mathcal{H}^2} ||g_{k,k'} - g^0_{k,k'}||^2_{L_2(\Omega)},
\end{aligned}
$$

and thus, by Lemma A1,

$$
||\tilde{\Gamma}_{n,k} - \Gamma_k||^2 = O\left(|\Delta|^{2(d+1)}\right).
$$

Since $\mathrm{E}\left[||\tilde{\Gamma}_{n,k} - \Gamma_k||_{n,N_s}\right] = ||\tilde{\Gamma}_{n,k} - \Gamma_k||$, we get $||\tilde{\Gamma}_{n,k} - \Gamma_k||_{n,N_s} = O_p\left(|\Delta|^{(d+1)}\right) = o_p(1)$.

Recall that $\hat{\Gamma}_{n,k} = \hat{\Pi}_n x_k$ and $\tilde{\Gamma}_{n,k} = \Pi_n x_k$. Since $(\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k})$ and $(x_k - \tilde{\Gamma}_{n,k})$ are orthogonal in the space $\mathcal{F}_{n,+}$ with respect to the theoretical norm, we have

$$||\tilde{\Gamma}_{n,k} - \Gamma_k||^2 = ||x_k - \hat{\Gamma}_{n,k}||^2 - ||x_k - \tilde{\Gamma}_{n,k}||^2.$$

For the empirical norm, we have

$$||x_k - \hat{\Gamma}_{n,k}||^2_{n,N_s} \leq ||x_k - \tilde{\Gamma}_{n,k}||^2_{n,N_s}. \tag{A6}$$

It is shown in Li et al. (2021) that for any vector of spline functions, $\mathbf{g}(\mathbf{s}) = (g_1(\mathbf{s}), g_2(\mathbf{s}), ..., g_p(\mathbf{s}))^T$ in $\mathcal{S}_d^r \cap \mathcal{H}^2$,

$$\frac{||\mathbf{g}||^2_{n,N_s}}{||\mathbf{g}||^2} - 1 = O_p\left(n^{-1/2}(\log(n))^{-1/2} + N_s^{-1/2}|\Delta|^{-1}\right) = o_p(1),$$

so,

$$\frac{||x_k - \hat{\Gamma}_{n,k}||^2_{n,N_s}}{||x_k - \hat{\Gamma}_{n,k}||^2} = o_p(1) + 1.$$

This gives

$$||x_k - \hat{\Gamma}_{n,k}||^2_{n,N_s} = ||x_k - \hat{\Gamma}_{n,k}||^2\left(o_p(1) + 1\right).$$

Similarly,

$$||x_k - \tilde{\Gamma}_{n,k}||^2_{n,N_s} = ||x_k - \tilde{\Gamma}_{n,k}||^2\left(o_p(1) + 1\right). \tag{A7}$$

From Equations (A6) and (A7), we get

$$\begin{aligned} ||x_k - \hat{\Gamma}_{n,k}||^2_{n,N_s} - ||x_k - \tilde{\Gamma}_{n,k}||^2 &\leq ||x_k - \tilde{\Gamma}_{n,k}||^2_{n,N_s} - ||x_k - \tilde{\Gamma}_{n,k}||^2 \\ &= o_p\left(||x_k - \tilde{\Gamma}_{n,k}||^2\right). \end{aligned}$$

Further,

$$
\begin{aligned}
||x_k - \hat{\Gamma}_{n,k}||^2_{n,N_s} - ||x_k - \tilde{\Gamma}_{n,k}||^2 &= ||x_k - \hat{\Gamma}_{n,k}||^2 \left(o_p(1) + 1\right) - ||x_k - \tilde{\Gamma}_{n,k}||^2 \\
&= ||x_k - \hat{\Gamma}_{n,k}||^2 + o_p\left(||x_k - \tilde{\Gamma}_{n,k}||^2\right) - ||x_k - \tilde{\Gamma}_{n,k}||^2 \\
&= ||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||^2 + o_p\left(||x_k - \tilde{\Gamma}_{n,k}||^2\right),
\end{aligned}
$$

and thus,

$$
||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||^2 = o_p\left(||x_k - \hat{\Gamma}_{n,k}||^2\right) + o_p\left(||x_k - \tilde{\Gamma}_{n,k}||^2\right).
$$

Since $||x_k - \tilde{\Gamma}_{n,k}||^2 = O_p(1)$, we have

$$
||x_k - \hat{\Gamma}_{n,k}|| \leq ||x_k - \tilde{\Gamma}_{n,k}|| + ||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}|| = O_p(1) + ||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||.
$$

Then, we have

$$
||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||^2 = o_p\left(||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||^2\right) + o_p(1).
$$

Hence,

$$
||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||^2 = o_p(1).
$$

Therefore, for all $k \in A_c$,

$$
||\hat{\Gamma}_{n,k} - \tilde{\Gamma}_{n,k}||^2_{n,N_s} = o_p(1),
$$

which proves that

$$
||\hat{\Gamma}_{n,k} - \Gamma_k||^2_{n,N_s} = o_p(1).
$$

□

The following lemma from Li et al. (2021) will be used to prove Theorem 2.2.

Lemma A7. *Suppose that the assumptions in Chapter 2 hold and that for all* $k \in A_c$*, there exists a positive real number* C_X *such that* $|X_{ik}| \leq C_X$*. Then for all* $\mathbf{b} \in \mathbb{R}^{|A_c|}$*, with* $\|\mathbf{b}\|_2 = \max_{\|\mathbf{x}\|=1}\|\mathbf{bx}\| = 1$*,*

$$(\text{Var}(\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e))^{-1/2}(\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e) \xrightarrow[n,N_s\to\infty]{D} N(0,1),$$

where

$$\hat{\boldsymbol{\alpha}}_0^e = \frac{1}{nN_s}\mathbf{U}_{11}\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_{2,A_v}})(\boldsymbol{\eta}+\boldsymbol{\epsilon}),$$

and

$$\mathbf{U}_{11} = (nN_s)\left[\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_{2,A_v}})\mathbf{Z}_{1,A_c}\right]^{-1}.$$

Proof. Write $\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e = (nN_s)^{-1}\mathbf{b}^T\mathbf{U}_{11}\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_{2,A_v}})(\boldsymbol{\eta}+\boldsymbol{\epsilon})$. Then

$$\begin{aligned}(\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e)^T &= \frac{1}{nN_s}\left[(\boldsymbol{\eta}+\boldsymbol{\epsilon})^T(\mathbf{Z}_{1,A_c}^T - \mathbf{Z}_{1,A_c}^T\mathbf{Z}_{2,A_v}(\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T)^T\mathbf{U}_{11}\mathbf{b}\right] \\ &= \frac{1}{nN_s}\left[(\boldsymbol{\eta}+\boldsymbol{\epsilon})^T(\mathbf{Z}_{1,A_c} - \mathbf{Z}_{2,A_v}(\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{1,A_c})\mathbf{U}_{11}\mathbf{b}\right] \\ &= \mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e,\end{aligned}$$

because $\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e$ is a scalar.

For all $i = 1, 2, ..., n$, take the i^{th} row vectors of $\mathbf{Z}_{1,A_c}$ and $\mathbf{Z}_{2,A_v}$, and let

$$a_i = \frac{1}{nN_s}\left[(\boldsymbol{\eta}_i+\boldsymbol{\epsilon}_i)^T(\mathbf{X}_{(i),A_c}^T \otimes \mathbf{1}_{N_s} - \mathbf{X}_{(i),A_v}^T \otimes \mathbf{B}\ (\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{1,A_c})\mathbf{U}_{11}\mathbf{b}\right].$$

Then, write

$$\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e = \sum_{i=1}^{n} a_i W_i,$$

where conditional on $\{\mathbf{X}_i\}_{i=1}^n$, W_i $(i = 1, 2, ..., n)$ are independent with mean zero and variance one.

Note that a_i is a scalar, so write

$$\begin{aligned} a_i^2 &= a_i^T a_i \\ &= \frac{1}{n^2 N_s^2} \mathbf{b}^T \mathbf{U}_{11} \left(\mathbf{X}_{(i),A_c}^T \otimes \mathbf{1}_{N_s} - \mathbf{X}_{(i),A_v}^T \otimes \mathbf{B} \ (\mathbf{Z}_{2,A_v}^T \mathbf{Z}_{2,A_v})^{-1} \mathbf{Z}_{2,A_v}^T \mathbf{Z}_{1,A_c} \right)^T \\ &\quad \times \boldsymbol{\Sigma}_e \left(\mathbf{X}_{(i),A_c}^T \otimes \mathbf{1}_{N_s} - \mathbf{X}_{(i),A_v}^T \otimes \mathbf{B} \ (\mathbf{Z}_{2,A_v}^T \mathbf{Z}_{2,A_v})^{-1} \mathbf{Z}_{2,A_v}^T \mathbf{Z}_{1,A_c} \right) \mathbf{U}_{11} \mathbf{b}, \end{aligned}$$

where $\boldsymbol{\Sigma}_e = \{\Sigma_e(\mathbf{s}_j, \mathbf{s}_{j'})\}_{j,j'=1}^{N_s}$.

Let $\mathbf{X}_{(i),A_c}^{\perp} = \mathbf{X}_{(i),A_c}^T \otimes \mathbf{1}_{N_s} - \mathbf{X}_{(i),A_v}^T \otimes \mathbf{B} \ (\mathbf{Z}_{2,A_v}^T \mathbf{Z}_{2,A_v})^{-1} \mathbf{Z}_{2,A_v}^T \mathbf{Z}_{1,A_c}$. Then for all $i = 1, 2, ..., n$,

$$\begin{aligned} a_i^2 &= \frac{1}{n^2 N_s^2} \mathbf{b}^T \mathbf{U}_{11} (\mathbf{X}_{(i),A_c}^{\perp})^T \boldsymbol{\Sigma}_e (\mathbf{X}_{(i),A_c}^{\perp}) \mathbf{U}_{11} \mathbf{b} \\ &= \frac{1}{n^2 N_s^2} \mathbf{b}^T \mathbf{U}_{11} \left(\sum_{j,j'=1}^{N_s} X_{ijk}^{\perp} \Sigma_e(\mathbf{s}_j, \mathbf{s}_{j'}) X_{ij'k'}^{\perp} \right)_{k,k' \in A_c} \mathbf{U}_{11} \mathbf{b}, \end{aligned}$$

where $X_{ijk}^{\perp}$ is the $(j,k)^{th}$ entry of $\mathbf{X}_{(i),A_c}^{\perp}$.

The eigenvalues of G_η are strictly positive, so by Theorem A1,

$$\frac{1}{n N_s^2} \sum_{i=1}^{n} \left(\sum_{j,j'=1}^{N_s} X_{ijk}^{\perp} \Sigma_e(\mathbf{s}_j, \mathbf{s}_{j'}) X_{ij'k'}^{\perp} \right)_{k,k' \in A_c} \xrightarrow[n, N_s \to \infty]{P} \boldsymbol{\Xi}_e,$$

where $\boldsymbol{\Xi}_e$ is positive definite.

Now,

$$
\begin{aligned}
\mathbf{U}_{11}^{-1} &= \frac{1}{nN_s}\mathbf{Z}_{1,A_c}^T\left(\mathbf{I}_{nN_s} - \mathbf{Z}_{2,A_v}(\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T\right)\mathbf{Z}_{1,A_c} \\
&= \frac{1}{nN_s}\mathbf{Z}_{1,A_c}^T\left(\mathbf{Z}_{1,A_c} - \mathbf{Z}_{2,A_v}(\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{1,A_c}\right) \\
&= \frac{1}{nN_s}(\mathbf{X}_{A_c}\otimes\mathbf{1}_{N_s})^T\left(\mathbf{X}_{A_c}\otimes\mathbf{1}_{N_s} - \mathbf{X}_{A_v}\otimes\mathbf{B}\ (\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{2,A_v})^{-1}\mathbf{Z}_{2,A_v}^T\mathbf{Z}_{1,A_c}\right).
\end{aligned}
$$

So,

$$
\begin{aligned}
\mathbf{U}_{11}^{-1} &= \frac{1}{nN_s}\sum_{i=1}^{n}(\mathbf{X}_{(i),A_c}^{\perp})^T(\mathbf{X}_{(i),A_c}^{\perp}) \\
&= \frac{1}{nN_s}\sum_{i=1}^{n}\left(\sum_{j,j'=1}^{N_s} X_{ijk}^{\perp}X_{ij'k'}^{\perp}\right)_{k,k'\in A_c}.
\end{aligned}
$$

By Theorem 3,

$$
\frac{1}{nN_s}\sum_{i=1}^{n}\left(\sum_{j,j'=1}^{N_s} X_{ijk}^{\perp}X_{ij'k'}^{\perp}\right)_{k,k'\in A_c} \xrightarrow[n,N_s\to\infty]{P} \mathbf{\Xi},
$$

where $\mathbf{\Xi}$ is positive definite.

By Assumption 2,

$$
\begin{aligned}
\sum_{i=1}^{n} a_i^2 &\geq c_1 n^{-1}\mathbf{b}^T\mathbf{U}_{11}\mathbf{b}\left(1 + O(N_s^{-1})\right) \\
&\geq c\ n^{-1}\|\mathbf{b}\|^2,
\end{aligned}
$$

for some positive real numbers c_1 and c. Again by Assumption 2,

$$
\begin{aligned}
\max_{1\leq i\leq n} a_i^2 &\leq (nN_s)^{-2}\ \mathbf{b}^T\mathbf{U}_{11}\left(\sum_{j,j'=1}^{N_s} X_{ijk}^{\perp}\Sigma_e(\mathbf{s}_j,\mathbf{s}_{j'})X_{ij'k'}^{\perp}\right)_{k,k'\in A_c}\mathbf{U}_{11}\mathbf{b} \\
&\leq C\ n^{-2}\|\mathbf{b}\|^2,
\end{aligned}
$$

where C is a positive real number.

Thus,

$$\frac{\max_{1\leq i\leq n} a_i^2}{\sum_{i=1}^n a_i^2} = O_p\left(\frac{1}{n}\right) = o_p(1),$$

so $\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e$ satisfies Lindeberg's condition.

Therefore, by the Lindeberg-Feller Central Limit Theorem,

$$(\mathrm{Var}(\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e))^{-1/2}(\mathbf{b}^T\hat{\boldsymbol{\alpha}}_0^e) \xrightarrow[n,N_s\to\infty]{D} N(0,1).$$

□

Proof of Theorem 2.2. Recall that

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}^0 - \boldsymbol{\alpha}_{0,A_c} &= \begin{bmatrix}\mathbf{I}_{|A_c|} & \mathbf{0}_{|A_c|\times|A_v|J_n}\end{bmatrix}(\hat{\boldsymbol{\theta}}^0 - \boldsymbol{\theta}_{0,A}) \\
&= \begin{bmatrix}\mathbf{I}_{|A_c|} & \mathbf{0}_{|A_c|\times|A_v|J_n}\end{bmatrix}(nN_s)^{-1}\mathbf{C}_A^{-1}\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\zeta}+\boldsymbol{\epsilon}) \\
&= (nN_s)^{-1}\mathbf{U}_{11}\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_2,A_v})((\boldsymbol{\eta}+\boldsymbol{\epsilon})+\boldsymbol{\zeta}).
\end{aligned}$$

Then, let

$$\hat{\boldsymbol{\alpha}}^0 - \boldsymbol{\alpha}_{0,A_c} := \hat{\boldsymbol{\alpha}}_e^0 + \hat{\boldsymbol{\alpha}}_\zeta^0,$$

where

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_e^0 &= (nN_s)^{-1}\mathbf{U}_{11}\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_2,A_v})(\boldsymbol{\eta}+\boldsymbol{\epsilon}), \\
\hat{\boldsymbol{\alpha}}_\zeta^0 &= (nN_s)^{-1}\mathbf{U}_{11}\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_2,A_v})\boldsymbol{\zeta}.
\end{aligned}$$

Recall that

$$\mathbf{V}_c = (nN_s)^{-2}\mathbf{U}_{11}\mathbf{Z}_{1,A_c}^T(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_2,A_v})\mathrm{diag}\{\boldsymbol{\Sigma}_{i,e}\}_{i=1}^n(\mathbf{I}_{nN_s} - \mathbf{P}_{Z_2,A_v})\mathbf{Z}_{1,A_c}\mathbf{U}_{11}$$

is the variance-covariance matrix of $\hat{\boldsymbol{\alpha}}_e^0$, where $\boldsymbol{\Sigma}_{i,e} = \mathrm{Var}(\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i)$.

Denote $\hat{\boldsymbol{\alpha}}_e^0 = (\hat{\alpha}_{e_1}^0 \ \hat{\alpha}_{e_2}^0 \ ... \ \hat{\alpha}_{e_{|A_c|}}^0)^T$ and let $\mathbf{b} = (b_1 \ b_2 \ ... \ b_{|A_c|})^T$ be such that $\|\mathbf{b}\|_2 = 1$. Then by Lemma A7,

$$\left[\operatorname{Var}\left(\sum_{i=1}^{|A_c|} b_i \hat{\alpha}_{e_i}^0\right)\right]^{-1/2} \left(\sum_{i=1}^{|A_c|} b_i \hat{\alpha}_{e_i}^0\right) \xrightarrow[n, N_s \to \infty]{D} \sum_{i=1}^{|A_c|} b_i W_i,$$

where $\sum_{i=1}^{|A_c|} b_i W_i \sim N(0,1)$ for all $i = 1, 2, ..., n$.

By Cramer-Wold device,

$$(\mathbf{V}_c)^{-1/2} \hat{\boldsymbol{\alpha}}_e^0 \xrightarrow[n, N_s \to \infty]{D} N(0, \mathbf{I}_{|A_c|}).$$

Now,

$$\|\hat{\boldsymbol{\alpha}}_\zeta^0\|^2 = \left\| \begin{bmatrix} \mathbf{I}_{|A_c|} & \mathbf{0}_{|A_c| \times |A_v| J_n} \end{bmatrix} (nN_s)^{-1} \mathbf{C}_A^{-1} \mathbf{Z}_A^T \boldsymbol{\zeta} \right\|^2 \leq (nN_s)^{-2} \pi^{-1} \|\mathbf{Z}_A^T \boldsymbol{\zeta}\|^2.$$

Then, by Lemma A3,

$$\|\hat{\boldsymbol{\alpha}}_\zeta^0\|^2 = O_p\left(|\Delta|^{2(d+1)}\right).$$

Thus,

$$\|\hat{\boldsymbol{\alpha}}_\zeta^0\| = O_p\left(|\Delta|^{(d+1)}\right).$$

Then, since $\|\mathbf{V}_c\| \xrightarrow{n, N_s \to \infty} 0$,

$$\mathbf{V}_c^{-1/2} (\hat{\boldsymbol{\alpha}}^0 - \boldsymbol{\alpha}_{0, A_c}) \xrightarrow[n, N_s \to \infty]{D} N(\mathbf{0}, \mathbf{I}_{|A_c|}).$$

□

Some definitions cited from Li et al. (2021) are needed before providing detailed proofs of the results in Chapter 3.

Let

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \tilde{\rho}_{n1}, \tilde{\rho}_{n2}) = \sum_{i=1}^{n}\sum_{j=1}^{N_s}\left(Y_{ij} - \sum_{k=1}^{p} X_{ik}\alpha_k - \sum_{k=1}^{p} X_{ik}\mathbf{B}^T(\mathbf{s}_j)\boldsymbol{\gamma}_k\right)^2 + \tilde{\rho}_{n1}\sum_{k=1}^{p}|\alpha_k| + \tilde{\rho}_{n2}\sum_{k=1}^{p}||\boldsymbol{\gamma}_k||_2,$$

and define the group LASSO estimator $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\alpha}}^T\ \tilde{\boldsymbol{\gamma}}^T)^T$ as the minimizer of L_n.

Define the constant and varying index sets as

$$A_c^* = \{k : \alpha_k \neq 0, 1 \leq k \leq p\},$$
$$\tilde{A}_c^* = \{k : |\tilde{\alpha}_k| \neq 0, 1 \leq k \leq p\},$$
$$\tilde{A}_v = \{k : \|\tilde{\boldsymbol{\gamma}}_k\| \neq 0, 1 \leq k \leq p\},$$
$$\tilde{A} = \tilde{A}_c^* \cup \tilde{A}_v.$$

The following theorem from Li et al. (2021) will be used to determine some properties of the group LASSO estimator $\tilde{\boldsymbol{\theta}}$ defined above.

Theorem A2. *Suppose that the assumptions in Sections 2.2 and 3.1 hold. Then the following statements hold.*

1. *With probability approaching one, $|\tilde{A}_c^*| \leq M|A_c^*|$ and $|\tilde{A}_v| \leq M|A_v|$, for some $1 < M < \infty$.*

2.
$$\sum_{k\in A_c^*}|\tilde{\alpha}_k - \alpha_{0k}|^2 = O_p\left(\frac{\log(pJ_n)}{n} + |\Delta|^{2(d+1)} + \frac{\tilde{\rho}_{n1}^2 + \tilde{\rho}_{n2}^2}{n^2N_s^2}\right),$$
$$\sum_{k\in A_v}\|\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}\|^2 = O_p\left(\frac{\log(pJ_n)}{n} + |\Delta|^{2(d+1)} + \frac{\tilde{\rho}_{n1}^2 + \tilde{\rho}_{n2}^2}{n^2N_s^2}\right).$$

3. *If* $n^{-1}\log(pJ_n) \longrightarrow 0$ *and* $(nN_s)^{-2}(\tilde{\rho}_{n1}^2 + \tilde{\rho}_{n2}^2) \longrightarrow 0$, *as* $n \to \infty$ *and* $N_s \to \infty$, *then with probability converging to one, all of the nonzero parameters* α_{0k}, $k \in A_c^*$ *and* $\beta_{0k}(\cdot)$, $k \in A_v$, *are selected.*

Proof. To set up the proof of Part 1, pick index sets $\mathcal{I}_1 \subseteq \{1, 2, ..., p\}$ and $\mathcal{I}_2 \subseteq \{1, 2, ..., p\}$, such that $|\mathcal{I}_1| = q_1$ and $|\mathcal{I}_2| = q_2$, where q_1 and q_2 are positive real numbers. Define $\mathbf{S}_\mathcal{I} = \left(\tilde{\rho}_{n1} u_{q_1}^T,\ \tilde{\rho}_{n2}\sqrt{J_n}(\mathbf{U}_1^\mathcal{I})^T,\ ...,\ \tilde{\rho}_{n2}\sqrt{J_n}(\mathbf{U}_{q_2}^\mathcal{I})^T\right)^T$, where $u_{q_1} \in \mathbb{R}^{J_n}$ and for all $k = 1, 2, ..., q_2$, $\mathbf{U}_k^\mathcal{I}$ is in a unit ball with dimension J_n. Let $\mathbf{P}_\mathcal{I} = \mathbf{Z}_\mathcal{I}(\mathbf{Z}_\mathcal{I}^T\mathbf{Z}_\mathcal{I}^T)^{-1}\mathbf{Z}_\mathcal{I}^T$ be the projection matrix of $\mathbf{Z}_\mathcal{I}$, and define

$$\mathbf{V}_\mathcal{I} = \mathbf{Z}_\mathcal{I}(\mathbf{Z}_\mathcal{I}^T\mathbf{Z}_\mathcal{I}^T)^{-1}\mathbf{S}_\mathcal{I} - (\mathbf{I} - \mathbf{P}_\mathcal{I})\mathbf{Z}\boldsymbol{\theta}_0.$$

Let $\boldsymbol{\xi} = \boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}$ and define

$$\chi_{q_1,q_2} = \max_{\mathcal{I}=\mathcal{I}_1\cup\mathcal{I}_2} \max_{u_{q_1}\in\{\pm 1\}^{q_1} ||\mathbf{U}_k^\mathcal{I}||_2=1, 1\le k\le q_2} \frac{|\boldsymbol{\xi}^T\mathbf{V}_\mathcal{I}|}{\|\mathbf{V}_\mathcal{I}\|},$$

$$\begin{aligned}\Xi_{|A_c^*|,|A_v|} = \{(\mathbf{Z}, \boldsymbol{\xi}) : \chi_{q_1,q_2} \le (\sqrt{N_s} + \sigma)C_1\sqrt{q_1\log(p) \vee q_2\log(pJ_n)},\\ \forall q_1 \ge |A_c^*|, \forall q_2 \ge |A_v|\},\end{aligned}$$

where C_1 is a sufficiently large enough constant.

As shown in Wei and Huang (2010), if $(\mathbf{Z}, \boldsymbol{\xi}) \in \Xi_{|A_c^*|,|A_v|}$, then $|\tilde{A}_c^*| \le M|A_c^*|$ and $|\tilde{A}_v| \le M|A_v|$, for some $1 < M < \infty$. Write

$$\frac{|\boldsymbol{\xi}^T\mathbf{V}_\mathcal{I}|}{\|\mathbf{V}_\mathcal{I}\|} = \frac{|(\boldsymbol{\eta} + \boldsymbol{\epsilon})^T\mathbf{V}_\mathcal{I} + \boldsymbol{\zeta}^T\mathbf{V}_\mathcal{I}|}{\|\mathbf{V}_\mathcal{I}\|}.$$

Then, by the triangle inequality,

$$\frac{|\boldsymbol{\xi}^T\mathbf{V}_\mathcal{I}|}{\|\mathbf{V}_\mathcal{I}\|} \le \frac{|(\boldsymbol{\eta} + \boldsymbol{\epsilon})^T\mathbf{V}_\mathcal{I}|}{\|\mathbf{V}_\mathcal{I}\|} + \frac{|\boldsymbol{\zeta}^T\mathbf{V}_\mathcal{I}|}{\|\mathbf{V}_\mathcal{I}\|}.$$

By applying the Cauchy-Schwarz inequality on the second term, we get

$$\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})^T\mathbf{V}_{\mathcal{I}}|}{\|\mathbf{V}_{\mathcal{I}}\|}+\frac{|\boldsymbol{\zeta}^T\mathbf{V}_{\mathcal{I}}|}{\|\mathbf{V}_{\mathcal{I}}\|}\leq\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})^T\mathbf{V}_{\mathcal{I}}|}{\|\mathbf{V}_{\mathcal{I}}\|}+\frac{||\boldsymbol{\zeta}^T||\;||\mathbf{V}_{\mathcal{I}}||}{\|\mathbf{V}_{\mathcal{I}}\|},$$

and thus,

$$\frac{|\boldsymbol{\xi}^T\mathbf{V}_{\mathcal{I}}|}{\|\mathbf{V}_{\mathcal{I}}\|}\leq\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})^T\mathbf{V}_{\mathcal{I}}|}{\|\mathbf{V}_{\mathcal{I}}\|}+||\boldsymbol{\zeta}||.$$

Next, define

$$\chi^*_{q_1,q_2}=\max_{\mathcal{I}=\mathcal{I}_1\cup\mathcal{I}_2}\;\max_{u_{q_1}\in\{\pm1\}^{q_1}||\mathbf{U}_k^{\mathcal{I}}||_2=1,1\leq k\leq q_2}\frac{|(\boldsymbol{\eta}+\boldsymbol{\epsilon})^T\mathbf{V}_{\mathcal{I}}|}{\|\mathbf{V}_{\mathcal{I}}\|},$$

$$\begin{aligned}\Xi^*_{|A_c^*|,|A_v|}=\{(\mathbf{Z},\boldsymbol{\eta}+\boldsymbol{\epsilon}):\chi^*_{q_1,q_2}\leq(\sqrt{N_s}+\sigma)C_2\sqrt{q_1\log(p)\vee q_2\log(pJ_n)},\\ \forall q_1\geq|A_c^*|,\forall q_2\geq|A_v|\},\end{aligned}$$

where C_2 is a sufficiently large enough constant.

As shown in Wei and Huang (2010),

$$\mathrm{P}(\Xi^*_{|A_c^*|,|A_v|})\xrightarrow{n,N_s\to\infty}1.$$

Further, for sufficiently large n,

$$||\boldsymbol{\zeta}||\leq(\sqrt{N_s}+\sigma)C_2\sqrt{|A_v|\log(pJ_n)}.$$

Therefore,

$$\mathrm{P}(\Xi_{|A_c^*|,|A_v|})\xrightarrow{n,N_s\to\infty}1,$$

and hence,

$$(\mathbf{Z},\boldsymbol{\xi})\in\Xi_{|A_c^*|,|A_v|}.$$

Therefore, $|\tilde{A}_c^*|\leq M|A_c^*|$ and $|\tilde{A}_v|\leq M|A_v|$, with probability approaching one, for some $1<M<\infty$. This proves Part 1.

Now, define $\tilde{\boldsymbol{\theta}}^T = (\tilde{\boldsymbol{\theta}}_1^T, \tilde{\boldsymbol{\theta}}_2^T, ..., \tilde{\boldsymbol{\theta}}_{2p}^T) = (\tilde{\alpha}_1, \tilde{\alpha}_2, ..., \tilde{\alpha}_p, \tilde{\boldsymbol{\gamma}}_1, \tilde{\boldsymbol{\gamma}}_2, ..., \tilde{\boldsymbol{\gamma}}_{2p})$, and denote

$$\begin{aligned}A_c' &= A_c^* \cup \tilde{A}_c^*,\\ A_v' &= A_v \cup \tilde{A}_v,\\ A' &= A_c' \cup A_v'.\end{aligned}$$

Let $d' = |A'| = O(|A|)$, and denote $\mathbf{Z}\tilde{\boldsymbol{\theta}} = \mathbf{Z}_{A'}\tilde{\boldsymbol{\theta}}_{A'}$ and $\mathbf{Z}\boldsymbol{\theta}_0 = \mathbf{Z}_{A'}\boldsymbol{\theta}_{0,A'}$. Define

$$\boldsymbol{\nu} = \mathbf{Z}_{A'}(\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}).$$

Then,

$$\begin{aligned}\boldsymbol{\xi} - \boldsymbol{\nu} &= (\mathbf{Y} - \mathbf{Z}\boldsymbol{\theta}_0)\mathbf{Z}_{A'}(\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}) = (\mathbf{Y} - \mathbf{Z}\boldsymbol{\theta}_0) - \mathbf{Z}_{A'}\tilde{\boldsymbol{\theta}}_{A'} + \mathbf{Z}_{A'}\boldsymbol{\theta}_{0,A'})\\ &= \mathbf{Y} - \mathbf{Z}_{A'}\tilde{\boldsymbol{\theta}}_{A'}.\end{aligned}$$

Now,

$$\begin{aligned}&(\boldsymbol{\xi} - \boldsymbol{\nu})^T(\boldsymbol{\xi} - \boldsymbol{\nu}) = \boldsymbol{\xi}^T\boldsymbol{\xi} - \boldsymbol{\xi}^T\boldsymbol{\nu} - \boldsymbol{\nu}^T\boldsymbol{\xi} + \boldsymbol{\nu}^T\boldsymbol{\nu}\\ \iff &\|\mathbf{Y} - \mathbf{Z}_{A'}\tilde{\boldsymbol{\theta}}_{A'}\|^2 = \boldsymbol{\xi}^T\boldsymbol{\xi} - 2\boldsymbol{\xi}^T\boldsymbol{\nu} + \boldsymbol{\nu}^T\boldsymbol{\nu}\\ \iff &\|\mathbf{Y} - \mathbf{Z}_{A'}\tilde{\boldsymbol{\theta}}_{A'}\|^2 - \|\mathbf{Y} - \mathbf{Z}_{A'}\boldsymbol{\theta}_{0,A'}\|^2 = \boldsymbol{\nu}^T\boldsymbol{\nu} - 2\boldsymbol{\xi}^T\boldsymbol{\nu}.\end{aligned}$$

By the definitions of $\tilde{\boldsymbol{\theta}}$, A_c' and A_v',

$$\begin{aligned}&||\mathbf{Y} - \mathbf{Z}_{A'}\tilde{\boldsymbol{\theta}}_{A'}||^2 + \sum_{k\in A_c'} \tilde{\rho}_{n_1}|\tilde{\alpha}_k| + \sum_{k\in A_v'} \tilde{\rho}_{n_2}||\tilde{\boldsymbol{\gamma}}_k||\\ &\leq ||\mathbf{Y} - \mathbf{Z}_{A'}\boldsymbol{\theta}_{0,A'}||^2 + \sum_{k\in A_c'} \tilde{\rho}_{n_1}|\alpha_{0k}| + \sum_{k\in A_v'} \tilde{\rho}_{n_2}||\boldsymbol{\gamma}_{0k}||,\end{aligned}$$

which means

$$||\boldsymbol{\nu}||^2 - 2\boldsymbol{\xi}^T\boldsymbol{\nu} \leq \sum_{k\in A_c'} \tilde{\rho}_{n_1}(|\alpha_{0k}| - |\tilde{\alpha}_k|) + \sum_{k\in A_v'} \tilde{\rho}_{n_2}(||\boldsymbol{\gamma}_{0k}|| - ||\tilde{\boldsymbol{\gamma}}_k||).$$

Note that $|A'_c| \leq d'$ and $|A'_v| \leq d'$, so

$$\sum_{k \in A'_c} \tilde{\rho}_{n_1}^2 = |A'_c| \tilde{\rho}_{n_1}^2 \leq d' \tilde{\rho}_{n_1}^2,$$

$$\sum_{k \in A'_v} \tilde{\rho}_{n_2}^2 = |A_c v| \tilde{\rho}_{n_2}^2 \leq d' \tilde{\rho}_{n_2}^2.$$

By the triangle inequality, we have

$$\begin{aligned}
&\sum_{k \in A'_c} \tilde{\rho}_{n_1}(|\alpha_{0k}| - |\tilde{\alpha}_k|) + \sum_{k \in A'_v} \tilde{\rho}_{n_2}(||\boldsymbol{\gamma}_{0k}|| - ||\tilde{\boldsymbol{\gamma}}_k||) \\
&\leq \sum_{k \in A'_c} \tilde{\rho}_{n_1}|\alpha_{0k} - \tilde{\alpha}_k| + \sum_{k \in A'_v} \tilde{\rho}_{n_2}||\boldsymbol{\gamma}_{0k} - \tilde{\boldsymbol{\gamma}}_k||,
\end{aligned}$$

and then

$$\begin{aligned}
&\sum_{k \in A'_c} \tilde{\rho}_{n_1}|\alpha_{0k} - \tilde{\alpha}_k| + \sum_{k \in A'_v} \tilde{\rho}_{n_2}||\boldsymbol{\gamma}_{0k} - \tilde{\boldsymbol{\gamma}}_k|| \\
&= \sum_{k \in A'_c \cup A'_v} (\tilde{\rho}_{n_1}|\alpha_{0k} - \tilde{\alpha}_k| I(k \in A'_c) + \tilde{\rho}_{n_2}||\boldsymbol{\gamma}_{0k} - \tilde{\boldsymbol{\gamma}}_k|| I(k \in A'_v)) \\
&\leq \sum_{k \in A'} \max(\tilde{\rho}_{n_1}, \tilde{\rho}_{n_2}) \left(|\tilde{\alpha}_k - \alpha_{0k}| I(k \in A'_c) + ||\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}|| I(k \in A'_v)\right) \\
&\leq \sum_{k \in A'} \sqrt{\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n_2}^2} \left(|\tilde{\alpha}_k - \alpha_{0k}| I(k \in A'_c) + ||\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}|| I(k \in A'_v)\right).
\end{aligned}$$

Further, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\sum_{k \in A'} \sqrt{\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n_2}^2} \left(|\tilde{\alpha}_k - \alpha_{0k}| I(k \in A'_c) + ||\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}|| I(k \in A'_v)\right) \\
&\leq \sqrt{\sum_{k \in A'} (\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n_2}^2)} \sqrt{\sum_{k \in A'} \left(|\tilde{\alpha}_k - \alpha_{0k}| I(k \in A'_c) + ||\tilde{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_{0k}|| I(k \in A'_v)\right)^2},
\end{aligned}$$

and then,

$$
\begin{aligned}
&\sqrt{\sum_{k\in A'}(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}\sqrt{\sum_{k\in A'}\left(|\tilde{\alpha}_k-\alpha_{0k}|I(k\in A'_c)+||\tilde{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_{0k}||I(k\in A'_v)\right)^2}\\
&\leq \sqrt{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}\sqrt{\sum_{k\in A'}\left(|\tilde{\alpha}_k-\alpha_{0k}|^2 I(k\in A'_c)+||\tilde{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_{0k}||^2 I(k\in A'_v)\right)}\\
&= \sqrt{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}\sqrt{\sum_{k\in A'_c}|\tilde{\alpha}_k-\alpha_{0k}|^2+\sum_{k\in A'_v}||\tilde{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_{0k}||^2}\\
&= \sqrt{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}||.
\end{aligned}
$$

Denote c as the lower bound of the eigenvalues of $(nN_s)^{-1}\mathbf{Z}_{A'}^T\mathbf{Z}_{A'}$, and write

$$
\sqrt{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}|| = 2\sqrt{\frac{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}{nN_s c}}\frac{1}{2}\sqrt{nN_s c}||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}||,
$$

and thus,

$$
\sqrt{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}|| \leq \frac{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}{nN_s c}+\frac{1}{4}(nN_s c)||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}||.
$$

Therefore, we have

$$
||\boldsymbol{\nu}||^2-2\boldsymbol{\xi}^T\boldsymbol{\nu} \leq \frac{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}{nN_s c}+\frac{1}{4}(nN_s c)||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}||.
$$

Recall that $\boldsymbol{\nu}=\mathbf{Z}_{A'}(\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'})$, so,

$$
||\boldsymbol{\nu}||^2 \geq (nN_s c)||\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}||^2.
$$

Define $\boldsymbol{\xi}^*=\mathbf{Z}_{A'}(\mathbf{Z}_{A'}^T\mathbf{Z}_{A'})^{-1}\mathbf{Z}_{A'}^T\boldsymbol{\xi}$ as the projection of $\boldsymbol{\xi}$ onto the column space of $\mathbf{Z}_{A'}$. Then,

$$
\begin{aligned}
(\boldsymbol{\xi}^*)^T\boldsymbol{\nu} &= \boldsymbol{\xi}^T\mathbf{Z}_{A'}[(\mathbf{Z}_{A'}^T\mathbf{Z}_{A'})^{-1}]\mathbf{Z}_{A'}^T\boldsymbol{\nu} = \boldsymbol{\xi}^T\mathbf{Z}_{A'}(\mathbf{Z}_{A'}^T\mathbf{Z}_{A'})^{-1}\mathbf{Z}_{A'}^T[\mathbf{Z}_{A'}(\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'})]\\
&= \boldsymbol{\xi}^T\mathbf{Z}_{A'}(\tilde{\boldsymbol{\theta}}_{A'}-\boldsymbol{\theta}_{0,A'}) = \boldsymbol{\xi}^T\boldsymbol{\nu}.
\end{aligned}
$$

By the Cauchy-Schwarz inequality,

$$2|\boldsymbol{\xi}^T\boldsymbol{\nu}| = 2|(\boldsymbol{\xi}^*)^T\boldsymbol{\nu}| \leq 2||\boldsymbol{\xi}^*||\ ||\boldsymbol{\nu}|| \leq 2||\boldsymbol{\xi}^*||^2 + \frac{1}{2}||\boldsymbol{\nu}||^2.$$

Then, we have

$$||\boldsymbol{\nu}||^2 - \frac{1}{4}(nN_sc)||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 - \frac{d'(\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n2}^2)}{nN_sc} \leq 2||\boldsymbol{\xi}^*||^2 + \frac{1}{2}||\boldsymbol{\nu}||^2$$

$$\Longrightarrow \frac{1}{2}||\boldsymbol{\nu}||^2 - \frac{1}{4}(nN_sc)||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 \leq 2||\boldsymbol{\xi}^*||^2 + \frac{d'(\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n2}^2)}{nN_sc}.$$

Since

$$\frac{1}{2}(nN_sc)||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 \leq \frac{1}{2}||\boldsymbol{\nu}||^2,$$

we have

$$\left(\frac{1}{2} - \frac{1}{4}\right)(nN_sc)||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 \leq 2||\boldsymbol{\xi}^*||^2 + \frac{d'(\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n2}^2)}{nN_sc},$$

and hence,

$$||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 \leq \frac{8||\boldsymbol{\xi}^*||^2}{nN_sc} + \frac{4d'(\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n2}^2)}{(nN_sc)^2}.$$

Let $\boldsymbol{\eta}^* \equiv \mathbf{P}_{\mathbf{Z}_{A'}}\boldsymbol{\eta}$, $\boldsymbol{\epsilon}^* \equiv \mathbf{P}_{\mathbf{Z}_{A'}}\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}_{A'}^* \equiv \mathbf{P}_{\mathbf{Z}_{A'}}\boldsymbol{\zeta}_{A'}$ be the projections of $\boldsymbol{\eta}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\zeta}_{A'}$ onto the column space of $\mathbf{Z}_{A'}$, respectively. Then. by the triangle inequality,

$$\begin{aligned}||\boldsymbol{\xi}^*||^2 = ||(\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*) + \boldsymbol{\zeta}_{A'}^*||^2 &\leq (||(\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*|| + ||\boldsymbol{\zeta}_{A'}^*||)^2 \\ &= ||(\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 + ||\boldsymbol{\zeta}_{A'}^*||^2 + 2||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||\ ||\boldsymbol{\zeta}_{A'}^*|| \\ &\leq ||(\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 + ||\boldsymbol{\zeta}_{A'}^*||^2 + \left(||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 + ||\boldsymbol{\zeta}_{A'}^*||^2\right),\end{aligned}$$

and hence,

$$||\boldsymbol{\xi}^*||^2 \leq 2\left(||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 + ||\boldsymbol{\zeta}_{A'}^*||^2\right).$$

Note that for some positive real number C,

$$||\boldsymbol{\zeta}_{A'}^*||^2 \leq C(nN_s)|A'|^2|\Delta|^{2(d+1)},$$

so,

$$||\boldsymbol{\xi}^*||^2 \le 2||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 + 2C(nN_s)(d')^2|\Delta|^{2(d+1)}.$$

Further, we have

$$||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 = ||\mathbf{Z}_{A'}(\mathbf{Z}_{A'}^T\mathbf{Z}_{A'})^{-1}\mathbf{Z}_{A'}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 = ||(\mathbf{Z}_{A'}^T\mathbf{Z}_{A'})^{-1/2}\mathbf{Z}_{A'}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2.$$

By the Cauchy-Schwarz inequality,

$$||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 \le ||(\mathbf{Z}_{A'}^T\mathbf{Z}_{A'})^{-1/2}||^2 \; ||\mathbf{Z}_{A'}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 = ||(nN_s)^{-1/2}\mathbf{C}_{A'}^{-1/2}||^2 \; ||\mathbf{Z}_{A'}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2.$$

Hence,

$$||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 \le (nN_sc)^{-1}||\mathbf{Z}_{A'}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2.$$

For any index set $\mathcal{I} \subseteq \{1, 2, ..., p\}$, we have

$$\begin{aligned}\max_{\mathcal{I}:|\mathcal{I}|\le d'} ||\mathbf{Z}_{\mathcal{I}}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 &= \max_{\mathcal{I}:|\mathcal{I}|\le d'} \sum_{m\in\mathcal{I}} \left(||\mathbf{Z}_m^T\boldsymbol{\eta}||^2 + ||\mathbf{Z}_m^T\boldsymbol{\epsilon}||^2\right) \\ &\le (nN_sd')\left((R_1^\eta)^2 + (R_1^\epsilon)^2\right) \vee (nN_sd')\left((R_2^\eta)^2 + (R_2^\epsilon)^2\right),\end{aligned}$$

where

$$R_1^\eta = \max_{1\le k\le p}\left|(nN_s)^{-1/2}\sum_{i=1}^n X_{ik}\sum_{j=1}^{N_s}\eta_i(\mathbf{s}_j)\right|,$$

$$R_1^\epsilon = \max_{1\le k\le p}\left|(nN_s)^{-1/2}\sum_{i=1}^n X_{ik}\sum_{j=1}^{N_s}\epsilon_{ij}\right|,$$

and

$$R_2^\eta = \max_{1\le k\le p}\left\|(nN_s)^{-1/2}\sum_{i=1}^n X_{ik}\sum_{j=1}^{N_s}\eta_i(\mathbf{s}_j)\mathbf{B}(\mathbf{s}_j)\right\|_\infty,$$

$$R_2^\epsilon = \max_{1\le k\le p}\left\|(nN_s)^{-1/2}\sum_{i=1}^n X_{ik}\sum_{j=1}^{N_s}\epsilon_{ij}\mathbf{B}(\mathbf{s}_j)\right\|_\infty.$$

Let C be a positive real number. It is shown in Li et al. (2021) that

$$R_1^\eta \leq C\sqrt{N_s \log(p)},\ \ R_2^\eta \leq C\sqrt{N_s \log(pJ_n)},$$

and

$$R_1^\epsilon \leq C\sigma\sqrt{\log(p)},\ \ R_2^\epsilon \leq C\sigma\sqrt{\log(pJ_n)}.$$

So, we get

$$\begin{aligned} ||\mathbf{Z}_{\mathcal{I}}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 &\leq (nN_s d')\left((C\sqrt{N_s\log(pJ_n)})^2 + (C\sigma\sqrt{\log(pJ_n)})^2\right) \\ &= (nN_s d' C^2)\log(pJ_n)(N_s+\sigma^2), \end{aligned}$$

and then,

$$||\mathbf{Z}_{\mathcal{I}}^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 = O_p\left(nN_s d'(N_s+\sigma^2)\log(pJ_n)\right).$$

Putting everything together, we have

$$||\boldsymbol{\eta}^* + \boldsymbol{\epsilon}^*||^2 = O_p\left(d'(N_s+\sigma^2)c^{-1}\log(pJ_n)\right),$$

$$||\boldsymbol{\xi}^*||^2 = O_p\left(d'(N_s+\sigma^2)c^{-1}\log(pJ_n)\right) + O_p\left(nN_s(d')^2|\Delta|^{2(d+1)}\right),$$

and therefore,

$$\begin{aligned} ||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 &= O_p\left(\frac{d'(N_s+\sigma^2)\log(pJ_n)}{nN_s c^2}\right) + O_p\left(\frac{nN_s(d')^2|\Delta|^{2(d+1)}}{nN_s c}\right) \\ &\quad + O\left(\frac{d'(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)}{(nN_s c)^2}\right). \end{aligned}$$

Since c is bounded by a positive constant and $d' = O(|A|)$, we get

$$||\tilde{\boldsymbol{\theta}}_{A'} - \boldsymbol{\theta}_{0,A'}||^2 = O_p\left(n^{-1}\log(pJ_n)\right) + O_p\left(|\Delta|^{2(d+1)}\right) + O\left((nN_s)^{-2}(\tilde{\rho}_{n_1}^2+\tilde{\rho}_{n_2}^2)\right),$$

which completes the proof of Part 2.

Recall from Assumption 6 that there exists $c_\alpha, c_\beta > 0$ such that

$$\min_{k\in A_c} |\alpha_{0k}| \geq c_\alpha, \ \min_{k\in A_v} ||\beta_{0k}||_{L_2(\Omega)} \geq c_\beta.$$

So for all $k \in A_c$, if $|\alpha_{0k}| \neq 0$ but $|\tilde{\alpha}_k| = 0$, then $|\alpha_{0k} - \tilde{\alpha}_k| \geq c_\alpha$. By Lemma A6, for all $k \in A_v$, if $||\boldsymbol{\gamma}_{0k}|| \neq 0$, but $||\tilde{\boldsymbol{\gamma}}_k|| = 0$, then $||\boldsymbol{\gamma}_{0k} - \tilde{\boldsymbol{\gamma}}_k|| \geq c_1 c_\beta$. However, this contradicts Part 2 when

$$n^{-1}\log(pJ_n) \xrightarrow{n\to\infty} 0, \text{ and } \frac{(\tilde{\rho}_{n_1}^2 + \tilde{\rho}_{n_2}^2)}{(nN_s)^2} \xrightarrow{n,N_s\to\infty}.$$

Since $N_s^{-1/2} < |\Delta| < n^{\frac{-1}{2(d+1)}}$, $|\Delta|^{2(d+1)} \longrightarrow 0$, as $n, N_s \to \infty$.

Therefore, with probability converging to one, all of the nonzero parameters α_{0k}, $k \in A_c^*$ and $\beta_{0k}(\cdot)$, $k \in A_v$, are selected. This completes the proof of part 3. □

As defined in Li et al. (2021), let

$$\hat{A}_c^* = \{k : |\hat{\alpha}_k| \neq 0, 1 \leq k \leq p\}.$$

Then, $A_c = A_c^* \backslash A_v$ and $\hat{A}_c = \hat{A}_c^* \backslash \hat{A}_v$.

The following lemma from Li et al. (2021) is used to help prove Theorem 3.1.

Lemma A8. *Suppose that the assumptions in Sections 2.2 and 3.1 hold. Then, as $n \to \infty$ and $N_s \to \infty$,*

$$\mathrm{P}\left(\hat{A}_c^* = A_c^*\right) \longrightarrow 1, \ \mathrm{P}\left(\hat{A}_v = A_v\right) \longrightarrow 1.$$

Proof. By the Karush-Kuhn-Tucker (KKT) conditions in Boyd et al. (2004), the unique minimizer, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\gamma}}^T)^T$ of

$$L_n(\boldsymbol{\alpha}, \boldsymbol{\gamma}; \rho_{n1}, \rho_{n2}) = \sum_{i=1}^{n} \sum_{j=1}^{N_s} \left[Y_i(\mathbf{s}_j) - \sum_{k=1}^{p} X_{ik}\alpha_k - \sum_{k=1}^{p} X_{ik}\mathbf{B}^T(\mathbf{s}_j)\boldsymbol{\gamma}_k \right]^2 + \rho_{n1} \sum_{k=1}^{p} w_{n,k}^c(|\alpha_k|) + \rho_{n2} \sum_{k=1}^{p} w_{n,k}^v(||\boldsymbol{\gamma}_k||)$$

satisfies the following conditions:

(1.1) $(\mathbf{X}_k \otimes \mathbf{1}_{N_s})^T[\mathbf{Y} - (\mathbf{X}_k \otimes \mathbf{1}_{N_s})\hat{\boldsymbol{\alpha}} - (\mathbf{X}_k \otimes \mathbf{B})\hat{\boldsymbol{\gamma}}] = \rho_{n1} w_{n,k}^c \frac{\alpha_k}{|\alpha_k|}, \forall k \in A_c^*,$

(1.2) $(\mathbf{X}_k \otimes \mathbf{B})^T[\mathbf{Y} - (\mathbf{X}_k \otimes \mathbf{1}_{N_s})\hat{\boldsymbol{\alpha}} - (\mathbf{X}_k \otimes \mathbf{B})\hat{\boldsymbol{\gamma}}] = \rho_{n2} w_{n,k}^v \frac{\boldsymbol{\gamma}_k}{||\boldsymbol{\gamma}_k||}, \forall k \in A_v,$

(2) $\left|(\mathbf{X}_k \otimes \mathbf{1}_{N_s})^T[\mathbf{Y} - (\mathbf{X}_k \otimes \mathbf{1}_{N_s})\hat{\boldsymbol{\alpha}} - (\mathbf{X}_k \otimes \mathbf{B})\hat{\boldsymbol{\gamma}}]\right| \leq \rho_{n1} w_{n,k}^c, \forall k \notin A_c^*,$

(3) $\left|(\mathbf{X}_k \otimes \mathbf{B})^T[\mathbf{Y} - (\mathbf{X}_k \otimes \mathbf{1}_{N_s})\hat{\boldsymbol{\alpha}} - (\mathbf{X}_k \otimes \mathbf{B})\hat{\boldsymbol{\gamma}}]\right| \leq \rho_{n2} w_{n,k}^v, \forall k \notin A_v.$

Define $\bar{\boldsymbol{\theta}}_0 = (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T\mathbf{Y}$, which is a vector with length $|A_c^*| + |A_v|J_n$. Define two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, both with length $|A|$, such that

$$v_{1m} \frac{w_{n,m}^c \bar{\theta}_{0m}}{|\bar{\theta}_{0m}|} I(m \in A_c^*) + \mathbf{0}_{J_n} I\left(m \notin A_c^*\right), \forall m \in A,$$

$$v_{2m} \frac{w_{n,(m-|A_c^*|)}^v \bar{\boldsymbol{\theta}}_{0m}}{||\bar{\boldsymbol{\theta}}_{0m}||} I\left((m - |A_c^*|) \in A_v\right), \forall m \in A.$$

Define

$$\hat{\boldsymbol{\theta}}_{0,A} = (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\left(\mathbf{Z}_A^T\mathbf{Y} - \rho_{n_1}\mathbf{v}_1 - \rho_{n_2}\mathbf{v}_2\right),$$

and decompose it into two vectors defined as

$$\hat{\boldsymbol{\theta}}_{0,A_c^*} = \left(\hat{\boldsymbol{\theta}}_{0,m}, m \in A_c^*\right)^T, \text{ and } \hat{\boldsymbol{\theta}}_{0,A_v} = \left(\hat{\boldsymbol{\theta}}_{0,m}, m \in A_v\right)^T.$$

Let $\hat{A}^0 = \{1 \leq m \leq 2p : ||\hat{\boldsymbol{\theta}}_{0,m}|| > 0\} \subseteq A$. and define

$$\hat{\boldsymbol{\theta}}_0 = \left(\hat{\boldsymbol{\theta}}_{0,A_c^*}^T, \mathbf{0}_{p-|A_c^*|}^T, \hat{\boldsymbol{\theta}}_{0,A_v}^T, \mathbf{0}_{p-|A_v|}^T\right)^T.$$

The objective is to show that $\hat{\boldsymbol{\theta}}_0$ satisfies the KKT conditions and hence, is the unique minimizer of L_n.

Note that $\mathbf{Z}\hat{\boldsymbol{\theta}}_0$ and $\{\mathbf{Z}_m, m \in A\}$ are linearly independent, so conditions (1.1) and (1.2) will hold for $\hat{\boldsymbol{\theta}}_0$ if $A = \hat{A}^0$. Let condition

$$(1') \quad A \subseteq \hat{A}^0,$$

and if we show this holds, along with conditions (2) and (3) for $\hat{\boldsymbol{\theta}}_0$, then $\hat{\boldsymbol{\theta}}_0$ is the unique minimizer of L_n. This is equivalent to showing

$$\mathrm{P}\left(\hat{A}_c^* = A_c^*\right) \xrightarrow{n,N_s \to \infty} 1, \; \mathrm{P}\left(\hat{A}_v = A_v\right) \xrightarrow{n,N_s \to \infty} 1.$$

To prove condition (1'), we must have

$$||\boldsymbol{\theta}_{0m}|| - ||\hat{\boldsymbol{\theta}}_{0m}|| \leq ||\boldsymbol{\theta}_{0m} - \hat{\boldsymbol{\theta}}_{0m}|| < ||\boldsymbol{\theta}_{0m}||,$$

for all $m \in A$, as $n \to \infty$ and $N_s \to \infty$. It is shown in Li et al. (2021) that

$$\mathrm{P}\left(||\boldsymbol{\theta}_{0m} - \hat{\boldsymbol{\theta}}_{0m}|| \geq ||\boldsymbol{\theta}_{0m}||, \exists m \in A\right) \xrightarrow{n,N_s \to \infty} 0,$$

which implies that $||\hat{\boldsymbol{\theta}}_{0m}|| > 0$, for all $m \in A$. Thus, every $m \in A$ is also in $\hat{A}^0$, which proves (1').

Further, it is shown in Li et al. (2021) that

$$\mathrm{P}\left(|(\mathbf{X}_k \otimes \mathbf{1}_{N_s})^T(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\theta}})| > \rho_{n1} w_{n,k}^c, \exists k \notin A_c^*\right) \xrightarrow{n,N_s \to \infty} 0,$$

$$\mathrm{P}\left(||(\mathbf{X}_k \otimes \mathbf{B})^T(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\theta}})|| > \rho_{n2} w_{n,k}^v, \exists k \notin A_v\right) \xrightarrow{n,N_s \to \infty} 0,$$

which prove conditions (2) and (3).

Therefore, conditions (1'), (2), and (3) hold, which implies

$$\mathrm{P}\left(\hat{A}_c^* = A_c^*\right) \xrightarrow{n,N_s \to \infty} 1, \; \mathrm{P}\left(\hat{A}_v = A_v\right) \xrightarrow{n,N_s \to \infty} 1.$$

□

Proof of Theorem 3.1. Recall that $A_c = A_c^* \backslash A_v$ and $\hat{A}_c = \hat{A}_c^* \backslash \hat{A}_v$. From Lemma A8, we had

$$\mathrm{P}\left(\hat{A}_c^* = A_c^*\right) \xrightarrow{n, N_s \to \infty} 1, \ \mathrm{P}\left(\hat{A}_v = A_v\right) \xrightarrow{n, N_s \to \infty} 1.$$

Since we have $\mathrm{P}\left(\hat{A}_c = A_c\right) = \mathrm{P}\left(\hat{A}_c^* \backslash \hat{A}_v = A_c^* \backslash A_v\right)$, and $\mathrm{P}\left(\hat{A}_v = A_v\right) \xrightarrow{n, N_s \to \infty} 1$,

$$\mathrm{P}\left(\hat{A}_c^* \backslash \hat{A}_v = A_c^* \backslash A_v\right) = \mathrm{P}\left(\hat{A}_c^* = A_c^*\right) \xrightarrow{n, N_s \to \infty} 1.$$

Therefore, as $n, N_s \to \infty$,

$$\mathrm{P}\left(\hat{A}_c = A_c\right) \longrightarrow 1, \ \text{and } \mathrm{P}\left(\hat{A}_v = A_v\right) \longrightarrow 1,$$

which completes the proof. □

Proof of Theorem 3.2. Denote π_1 as the minimum eigenvalue of $\mathbf{C}_A$. We have

$$\hat{\boldsymbol{\theta}}_A = \hat{\boldsymbol{\theta}}_{0,A} = (\mathbf{Z}_A^T \mathbf{Z}_A)^{-1} \left(\mathbf{Z}_A^T \mathbf{Y} - \rho_{n_1} \mathbf{v}_1 - \rho_{n_2} \mathbf{v}_2\right),$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors defined in the proof of Lemma A8. Then,

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_{0,A} &= (\mathbf{Z}_A^T \mathbf{Z}_A)^{-1} \left(\mathbf{Z}_A^T \mathbf{Y} - \rho_{n_1} \mathbf{v}_1 - \rho_{n_2} \mathbf{v}_2\right) \\
&= (\mathbf{Z}_A^T \mathbf{Z}_A)^{-1} \left(\mathbf{Z}_A^T \left((\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}) + \mathbf{Z}_A \boldsymbol{\theta}_{0,A}\right) - \rho_{n_1} \mathbf{v}_1 - \rho_{n_2} \mathbf{v}_2\right) \\
&= (\mathbf{Z}_A^T \mathbf{Z}_A)^{-1} \left(\mathbf{Z}_A^T \mathbf{Z}_A \boldsymbol{\theta}_{0,A} + \mathbf{Z}_A^T (\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}) - \rho_{n_1} \mathbf{v}_1 - \rho_{n_2} \mathbf{v}_2\right) \\
&= \boldsymbol{\theta}_{0,A} + (\mathbf{Z}_A^T \mathbf{Z}_A)^{-1} \left(\mathbf{Z}_A^T (\boldsymbol{\eta} + \boldsymbol{\epsilon} + \boldsymbol{\zeta}) - \rho_{n_1} \mathbf{v}_1 - \rho_{n_2} \mathbf{v}_2\right),
\end{aligned}$$

and so,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{0,A} - \boldsymbol{\theta}_{0,A} &= (\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\left(\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon}+\boldsymbol{\zeta}) - \rho_{n_1}\mathbf{v}_1 - \rho_{n_2}\mathbf{v}_2\right)\\ &= (nN_s)^{-1}\mathbf{C}_A^{-1}\left(\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon}+\boldsymbol{\zeta}) - \rho_{n_1}\mathbf{v}_1 - \rho_{n_2}\mathbf{v}_2\right).\end{aligned}$$

Define $\boldsymbol{\xi}_*$, $\boldsymbol{\eta}_*$, and $\boldsymbol{\epsilon}_*$ as the projections of $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, and $\boldsymbol{\epsilon}_*$ onto the column space of $\mathbf{Z}_A$, respectively. So,

$$||\boldsymbol{\eta}_* + \boldsymbol{\epsilon}_*||^2 = ||\mathbf{Z}_A(\mathbf{Z}_A^T\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 = ||(\mathbf{Z}_A^T\mathbf{Z}_A)^{-1/2}\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2.$$

Then,

$$||\boldsymbol{\eta}_* + \boldsymbol{\epsilon}_*||^2 = ||(nN_s)^{-1/2}\mathbf{C}_A^{-1/2}\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}||(nN_s)^{-1/2}\mathbf{C}_A^{-1/2}\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 &\leq (nN_s)^{-1}||\mathbf{C}_A^{-1/2}||^2||\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2\\ &\leq (nN_s\pi_1)^{-1}||\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2.\end{aligned}$$

By the triangle inequality,

$$||\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 \leq \left(||\mathbf{Z}_A^T\boldsymbol{\eta}|| + ||\mathbf{Z}_A^T\boldsymbol{\epsilon}||\right)^2 \leq 2||\mathbf{Z}_A^T\boldsymbol{\eta}||^2 + 2||\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2.$$

Multiplying and dividing by $n^2N_s^2$, we get

$$||\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 \leq 2(nN_s)^2\left(\frac{|\mathbf{Z}_A^T\boldsymbol{\eta}||^2}{(nN_s)^2} + \frac{|\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2}{(nN_s)^2}\right).$$

By Lemmas A4 and A5,

$$2(nN_s)^2\left(\frac{|\mathbf{Z}_A^T\boldsymbol{\eta}||^2}{(nN_s)^2} + \frac{|\mathbf{Z}_A^T\boldsymbol{\epsilon}||^2}{(nN_s)^2}\right) \leq 2(nN_s)^2\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2}\right),$$

and thus,

$$||\mathbf{Z}_A^T(\boldsymbol{\eta}+\boldsymbol{\epsilon})||^2 = O_p\left(nN_s^2 + nN_s|\Delta|^{-2}\right).$$

We also have

$$||\boldsymbol{\zeta}||^2 = O_p\left(nN_s|\Delta|^{2(d+1)}\right).$$

Now, by the triangle inequality,

$$||\boldsymbol{\xi}_*||^2 = ||(\boldsymbol{\eta}_* + \boldsymbol{\epsilon}_*) + \boldsymbol{\zeta}||^2 \leq (||\boldsymbol{\eta}_* + \boldsymbol{\epsilon}_*|| + ||\boldsymbol{\zeta}||)^2 \leq 2||\boldsymbol{\eta}_* + \boldsymbol{\epsilon}_*||^2 + 2||\boldsymbol{\zeta}||^2,$$

and therefore, since π_1^{-1} is bounded by a positive real number,

$$\begin{aligned}||\boldsymbol{\xi}_*||^2 &= O_p\left((nN_s\pi_1)^{-1}(nN_s^2 + nN_s|\Delta|^{-2})\right) + O_p\left(nN_s|\Delta|^{2(d+1)}\right)\\ &= O_p\left(N_s + |\Delta|^{-2} + nN_s|\Delta|^{2(d+1)}\right).\end{aligned}$$

In a similar way as in Part 2 of the proof of Theorem A2,

$$||\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_{0,A}||^2 \leq \frac{8||\boldsymbol{\xi}_*||^2}{nN_s\pi_1} + \frac{4(\rho_{n_1}^2|A_c| + \rho_{n_2}^2|A_v|)}{n^2N_s^2\pi_1^2},$$

and so,

$$\begin{aligned}||\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_{0,A}||^2 = O_p\left(\frac{N_s + |\Delta|^{-2}}{nN_s}\right) + O_p\left(\frac{nN_s|\Delta|^{2(d+1)}}{nN_s}\right)\\ + O\left(\frac{\rho_{n_1}^2|A_c| + \rho_{n_2}^2|A_v|}{n^2N_s^2}\right).\end{aligned}$$

Thus,

$$||\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_{0,A}||^2 = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)} + \frac{\rho_{n_1}^2 + \rho_{n_2}^2}{n^2N_s^2}\right).$$

Since

$$(\hat{\boldsymbol{\alpha}}_{A_c} - \boldsymbol{\alpha}_{0,A_c}) = \left[\mathbf{I}_{|A_c|} \quad \mathbf{0}_{|A_c|\times|A_v|J_n}\right](\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_{0,A}),$$

we get

$$\sum_{k\in A_c}(\hat{\alpha}_k - \alpha_{0k})^2 = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)} + \frac{\rho_{n_1}^2 + \rho_{n_2}^2}{n^2N_s^2}\right),$$

which completes the proof of (a). Since

$$(\hat{\boldsymbol{\gamma}}_{A_v} - \boldsymbol{\gamma}_{0,A_v}) = \left[\mathbf{0}_{|A_v|J_n \times |A_c|} \quad \mathbf{I}_{|A_v|J_n}\right] (\hat{\boldsymbol{\theta}}_A - \boldsymbol{\theta}_{0,A}),$$

and for all $\mathbf{s} \in \Omega$,

$$\hat{\boldsymbol{\beta}}_{A_v}(\mathbf{s}) - \boldsymbol{\beta}_{0,A_v}(\mathbf{s}) = \left[\boldsymbol{I}_{|A_v|} \otimes \mathbf{B}^T(\mathbf{s})\right] (\hat{\boldsymbol{\gamma}}_{A_v} - \boldsymbol{\gamma}_{0,A_v}),$$

we get

$$\sum_{k \in A_v} ||\hat{\beta}_k - \beta_{0k}||^2_{L_2(\Omega)} = O_p\left(\frac{1}{n} + \frac{1}{nN_s|\Delta|^2} + |\Delta|^{2(d+1)} + \frac{\rho_{n_1}^2 + \rho_{n_2}^2}{n^2 N_s^2}\right),$$

which completes the proof of (b).

$\square$

VITA AUCTORIS

Noah Fuerth was born in 1999 in Windsor, Ontario, and grew up in Essex, Ontario. He attended Cardinal Carter Catholic Secondary School in Leamington, Ontario and graduated in 2017. Noah graduated with a B.Sc. in Mathematics and Statistics from the University of Windsor in 2021, and is now pursuing a M.Sc. in Statistics at the University of Windsor, and hopes to graduate in May 2023.