Penalized Spline Estimation in the Partially Linear Model

by

Ashley D. Holland

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Applied and Interdisciplinary Mathematics)
in The University of Michigan
2012

Doctoral Committee:

Assistant Professor Matias D. Cattaneo, Co-Chair
Professor Virginia R. Young, Co-Chair
Associate Professor Mouli Banerjee
Associate Professor Volker W. Elling
Associate Professor Kristen S. Moore

To my husband, family, and friends who have all supported me throughout these years.

TABLE OF CONTENTS

DEDICATION ii

LIST OF TABLES v

LIST OF APPENDICES vi

ABSTRACT vii

CHAPTER

I. Introduction 1

1.1 The Partially Linear Model 1
1.2 Spline Estimation 3
1.3 Literature Review 6
1.3.1 Spline Estimation - Rates of Convergence 6
1.3.2 Spline Estimation - Asymptotic Distribution 9
1.3.3 Series Estimation in the Partially Linear Model 10
1.4 Overview of Results 11
1.4.1 Empirical and Fixed Mean Squared Error 12
1.4.2 Asymptotic Distribution and Standard Error Estimates for the Parametric Component 14
1.4.3 Asymptotic Distribution and Standard Errors for the Nonparametric Component 15

II. Rates of Convergence 17

2.1 Eigenvalues of the Design Matrix 17
2.2 Eigenvalues for the Penalization Matrix 18
2.2.1 Convergence of Eigenvalues 18
2.2.2 Decomposition of the Penalization Matrix 23
2.2.3 Asymptotic Representation of Eigenvalues 26
2.3 Conditional Mean Squared Error Expansion 26
2.3.1 Expansion 27
2.3.2 Simplication of the Variance Term 30
2.3.3 Simplification of the Penalization Bias Term 31
2.4 Rate of the Empirical Mean Squared Error 32

2.4.1 Without Derivatives 32
2.4.2 With Derivatives 37
2.5 Rate of the Fixed Mean Squared Error 42

III. Asymptotic Distribution of the Parametric and Nonparametric Components 44

3.1 Asymptotic Distribution of the Parametric Component 44
3.2 Asymptotic Distribution of the Nonparametric Component 52
3.2.1 Lower Bound on the Variance 52
3.2.2 Upper Bound on the Bias 55
3.2.3 Asymptotic Normality 59
3.3 Standard Errors 62
3.3.1 Parametric Component 62
3.3.2 Nonparametric Component 66

IV. Simulations Study 69

4.1 Description of Simulations 69
4.2 Results 69

V. Conclusion 72

APPENDICES 75

BIBLIOGRAPHY 99

LIST OF TABLES

Table

4.1 Parametric component - $\hat{\theta}$ 70
4.2 Nonparametric component - $\hat{g}(0.5, 0.5)$ 71

LIST OF APPENDICES

Appendix

A. Eigenvalues for the Penalization Matrix 76

B. Eigenvalues of the Design Matrix 85

ABSTRACT

Penalized Spline Estimation in the Partially Linear Model

by

Ashley D. Holland

Co-Chairs: Matias D. Cattaneo and Virginia R. Young

Penalized spline estimators have received considerable attention in recent years because of their good finite-sample performance, especially when the dimension of the regressors is large. In this project, we employ penalized B-splines in the context of the partially linear model to estimate the nonparametric component, when both the number of knots and the penalty factor vary with the sample size. We obtain mean-square convergence rates and establish asymptotic distributional approximations, with valid standard errors, for the resulting multivariate estimators of both the parametric and nonparametric components in this model. Our results extend and complement the recent theoretical work in the literature on penalized spline estimators by allowing for multivariate covariates, heteroskedasticity of unknown form, derivative estimation, and statistical inference in the semi-linear model, using weaker assumptions. The results from a simulation study are also reported.

CHAPTER I

Introduction

1.1 The Partially Linear Model

The partially linear model has a long tradition in statistics and econometrics (see, e.g., Ruppert, Wand & Carroll (2003) and Härdle, Müller, Sperlich & Werwatz (2004) for recent textbook discussions). In this model, for a dependent variable y and covariates $x \in \mathbb{R}^{d_x}$ and $z \in [0,1]^{d_z}$, the conditional mean function is assumed to satisfy

$$\mathbb{E}[y|x,z] = x'\theta + g(z),$$

where both the finite-dimensional parameter θ and the infinite-dimensional parameter $g(\cdot)$ are of potential interest. This is a very popular model in empirical work because it provides a parsimonious, yet flexible, approach to inference in different contexts. Typically, in this model the dimension of x is small while the dimension of z is large. In the program evaluation literature, for example, x is usually a treatment indicator and θ the scalar treatment effect of interest, while $g(\cdot)$ is a nonparametric nuisance function which is present to account for many possible confounding factors in a flexible way (see, e.g., Imbens & Wooldridge (2008) for a recent survey). The multivariate function $g(\cdot)$ and its derivatives are also parameters of interest in other cases, for instance in policy analysis (Stock (1989)).

Inference in the partially linear model is an important semiparametric problem. Large sample results are available for inference on θ and $g(\cdot)$ when the nonparametric component is estimated using kernel regression (Robinson (1988)), power series, or regression splines

(Donald & Newey (1994)). These results, however, rely on classical smoothing techniques which are usually quite sensitive to the specifics of their implementation in applications, a problem that is only exacerbated when the dimension of z is large. Partially motivated by the poor finite-sample performance of these classical smoothing techniques, a recent literature on penalized spline estimation has emerged and is receiving considerable attention. Originally proposed by O'Sullivan (1986), and later popularized by Eilers & Marx (1996), this alternative smoothing technique is nowadays commonly used in applications, being usually perceived as a superior alternative to other classical nonparametric estimators.

Motivated by their recent popularity, and with the explicit goal of increasing the finite-sample performance of the resulting statistical procedures, in this project we propose to employ multivariate penalized B-splines estimators, with n-varying knots and penalty (where n is the sample size), to estimate the nonparametric ingredient in the partially linear model. We investigate the large sample properties of the resulting estimators of θ and $g(\cdot)$ under quite general tuning parameter sequences, providing in particular asymptotic distributional approximations and consistent standard-error estimates. As an intermediate step, we also derive the mean-square convergence rate of penalized B-splines estimators of the regression function and its derivatives under general asymptotic sequences.

Despite the popularity of penalized spline smoothing, there is only a handful of papers analyzing its theoretical properties. Early work has obtained asymptotic results under fixed-knot asymptotics, where the number of knots is assumed to be fixed and the penalty factor converges to zero (see, e.g., Wand (1999); Aerts, Claeskens & Wand (2002); Yu & Ruppert (2002); and Wand & Ormerod (2008)), or under sequential asymptotics, Hall & Opsomer (2005). These asymptotics, however, are restrictive and may not always characterize appropriately the finite-sample behavior of the penalized splines. For this reason, recent work has focused on the asymptotic properties of penalized splines when both the knots and penalty vary with the sample size. Li & Ruppert (2008) studies univariate penalized splines when the number of knots is "large" and derive an asymptotic equivalence between kernel smoothing and penalized (smoothing) splines. Claeskens, Krivobokova & Opsomer (2009) study univariate penalized splines under quite general sequences of tuning parameters and show that these estimators are asymptotically equivalent in a mean-square-error sense to either regression splines or smoothing splines depending on the sequence of tuning parameters considered. Kauermann, Krivobokova & Fahrmeir (2009) extend some of the previous result to the context of univariate generalized spline smoothing. Krivobokova, Kneib & Claeskens (2010) propose asymptotically conservative confidence bands for univariate penalized spline estimators of the regression function. The present project substantially complements and extends some of the results in this emerging literature by allowing for multivariate covariates, heteroskedasticity of unknown form, derivative estimation and statistical inference on both the parametric and nonparametric components in the partially linear model.

In the rest of Chapter 1, we describe spline estimation, present main results in the literature on spline estimation and the partially linear model, and give an overview of our results. Chapter 2 presents the rate of the convergence of the penalized spline estimate of $g(\cdot)$. Chapter 3 gives the asymptotic distribution and standard errors for both the estimate of $\hat{\theta}$ and the penalized spline estimate of $g(\cdot)$. Chapter 4 discusses the results of a Monte Carlo study aimed to assess the finite-sample performance of these estimators, and Chapter 5 outlines the main contributions of this project.

1.2 Spline Estimation

To construct a B-spline basis $\{p_{jk}\}_{k=1}^{K^{1/d_z}}$ in direction j, $[0,1]$ is partitioned into $K^{1/d_z} - r + 1$ intervals

$$[t_{j,r}, t_{j,r+1}], [t_{j,r+1}, t_{j,r+2}], ..., [t_{j,K^{1/d_z}}, t_{j,K^{1/d_z}+1}],$$

with knots $t_{j,r} = 0 \leq \cdots \leq t_{j,K^{1/d_z}+1} = 1$, where r is the desired degree of the splines. A condition on the mesh ratio is assumed, for example

$$\max_{1\leq k\leq K^{1/d_z}} |h_{j,k+1} - h_{j,k}| = o(1/K^{1/d_z}), \quad h_j / \min_{1\leq k\leq K^{1/d_z}} h_{j,k} \leq M_j,$$

where $h_{j,k} \equiv t_{j,k} - t_{j,k-1}$, $h_j \equiv \max\limits_{1\leq k\leq K^{1/d_z}} h_{j,k}$, and $M_j > 0$ is a constant, in order to guarantee that $1/M_j \leq K^{1/d_z} h_j \leq M_j$ (Zhou, Shen & Wolfe (1998)). A weaker alternative is

$$t_{jk} - t_{j,k-1} \asymp 1/K_n^{1/d_z}$$

for all $k = 1, ..., K_n^{1/d_z}$, as in Huang (2003a), where $r_n \asymp \bar{r}_n$ indicates $r_n \geq c_1\bar{r}_n$ and $r_n \leq c_2\bar{r}_n$ for some $c_1 > 0$ and $c_2 < \infty$.

To manage boundary effects, an extra $2(r-1)$ knots are added with $t_{j,1} \leq \cdots \leq t_{j,r-1} \leq 0$ and $1 \leq t_{j,K^{1/d_z}+2} \leq \cdots \leq t_{j,K^{1/d_z}+r}$, creating an extended partition. The B-splines are then constructed using the well-known Cox-de Boor recursion relation (De Boor (2001)):

$$p_{j,k,1}(z_j) = \begin{cases} 1 & t_{j,k} \leq z < t_{j,k+1} \\ 0 & \text{otherwise} \end{cases}$$

$$p_{j,k,\ell}(z) = \frac{z - t_{j,k}}{t_{j,k+\ell-1} - t_{j,k}} p_{j,k,\ell-1}(z) + \frac{t_{j,k+\ell} - z}{t_{j,k+\ell} - t_{j,k+1}} p_{j,k+1,\ell-1}(z),$$

where $p_{j,k,\ell}$ is the kth spline of order ℓ in direction j, and the convention $0/0 = 0$ is used. The set $\{p_{j,k,r}\}_{k=1}^{K^{1/d_z}}$ spans the space

$$\mathcal{S}_{n,r} \equiv \{s(\cdot) \in \mathcal{C}^{r-2}[0,1] : s(z_j) \text{ is a polynomial of order } r \text{ on each subinterval } [t_{jk}, t_{j,k+1}]\}$$

(see for example Zhou et al. (1998), De Boor (2001), Schumaker (1981)). Normalized B-splines have the useful property

$$\sum_{k=1}^{K^{1/d_z}} p_{jk} = 1.$$

(Zhou et al. (1998)). Multivariate tensor product splines are formed using $p^K \equiv (p_1, ..., p_K)' = (p_{11}, ..., p_{1K^{1/d_z}})' \otimes \cdots \otimes (p_{d_z1}, ..., p_{d_zK^{1/d_z}})'$. Other references on B-splines include Stone (1994), De Boor (1976), and Eilers & Marx (1996).

The penalized spline estimate of $g(\cdot)$ minimizes the criterion function

$$S \equiv \sum_{i=1}^{n} (y_i - \hat{g}(z_i))^2 + \lambda_n \int_{[0,1]^d} \sum_{j_1,...,j_m=1}^{d} \left(\frac{\partial^m \hat{g}(z)}{\partial z_{j_1} \partial z_{j_2} \cdots \partial z_{j_m}} \right)^2 dz,$$

where λ_n is a smoothing parameter, $m = (m_1, ..., m_d)$, and $|m| = m_1 + ... + m_d$ (see Cox (1984) and Utreras (1988)). This method was first introduced by O'Sullivan (1986), with $d_z = 1$, $r = 4$, and $m = 2$ (see also Wand & Ormerod (2008)).

Let $\partial^m \hat{g}(z) = \dfrac{\partial^m \hat{g}(z)}{\partial z_{j_1} \partial z_{j_2} \cdots \partial z_{j_m}}$. Then noting that

$$\begin{aligned}
\int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^{d} \left(\partial^m \hat{g}(z)\right)^2 dz &= \sum_{j_1,\ldots,j_m=1}^{d} \int_{[0,1]^d} \left(\partial^m \left(\sum_{k=1}^{K} \beta_k p_k(z) \right) \right)^2 dz \\
&= \sum_{j_1,\ldots,j_m=1}^{d} \int_{[0,1]^d} \left(\sum_{k=1}^{K} \beta_k \partial^m p_k(z) \right)^2 dz \\
&= \sum_{k,\ell=1}^{K} \beta_k \beta_\ell \int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^{d} \partial^m p_k(z) \partial^m p_\ell(z) dz \\
&= \beta' D \beta
\end{aligned}$$

with $(D)_{k\ell} = \int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^{d} \partial^m p_k(z) \partial^m p_\ell(z) dz$, we rewrite the penalty term as

$$\lambda_n \int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^{d} \left(\frac{\partial^m \hat{g}(z)}{\partial z_{j_1} \partial z_{j_2} \cdots \partial z_{j_m}} \right)^2 dz = \lambda \beta' D \beta.$$

Then with the usual method of setting the derivative of S with respect to β equal to zero, we then find that

$$\hat{g}(z) = p^{K_n}(z)'(P'P + \lambda D)^{-1} P'Y, \;\; \hat{G} = P'(P'P + \lambda D)^{-1} P'Y$$

where P is the $n \times K$ matrix of spline basis functions evaluated at the observations $z_1, ..., z_n$, and $\hat{G} = (\hat{g}(z_1), ..., \hat{g}(z_n))'$. In the partially linear model, the estimate becomes

$$\hat{g}(z) = P'(P'P + \lambda D)^{-1} P'(Y - X\theta).$$

The standard assumption is that the penalty term $\beta' D \beta$ is bounded.

In the familiar framework of regression splines, the smoothing parameter is $\lambda_n = 0$. For asymptotic results, a rate condition such as $K_n^2/n \to 0$ (Newey (1997), Zhou et al. (1998), Zhou & Wolfe (2000), Claeskens et al. (2009)) or $K_n \log n/n \to 0$ (Huang (2003b), Huang (2003a)) is assumed. In contrast, smoothing spline estimation has $K_n = n$ and includes the penalization term to compensate for the resulting large variance (Cox (1984), Utreras (1988), Utreras (1979)). Choosing λ_n has been the topic of much research, and several methods including cross-validation and the information criterion have been considered (Wahba (1975), Wand (1999), Craven & Wahba (1978)), Utreras (1979), Li & Ruppert (2008).)

Penalized splines bridge the gap between regression splines and smoothing splines, in that the criterion function contains a penalty but $K_n = n$ is not required.

Reference books for splines include Wahba (1990), Green & Silverman (1994), and Eubank (1999). Also, see Ruppert et al. (2003) for applications of spline regression.

1.3 Literature Review

We define the following norms:

$$\|g\|_{2,a}^2 = \sup_{|q|\le a} \|\partial^q g\|_2^2 = \sup_{|q|\le a} \int_{[0,1]^d} (\partial^q g(z))^2 dF(z),$$

$$\|g\|_{2,a,n}^2 = \sup_{|q|\le a} \|\partial^q g\|_{2,n}^2 = \sup_{|q|\le a} \frac{1}{n}\sum_{i=1}^{n} (\partial^q g(Z_i))^2,$$

$$\|g\|_{\infty,a}^2 = \sup_{|q|\le a} \|\partial^q g\|_{\infty}^2 = \sup_{|q|\le a} \sup_{z\in[0,1]^d} |\partial^q g(z)|^2.$$

Main asymptotic results available in the literature for regression, smooothing, and penalized spline estimation include the following:

1.3.1 Spline Estimation - Rates of Convergence

- Regression Splines

  A paper that considers asymptotics of series estimators is Newey (1997), which gives mean-square and uniform rates of convergence for multivariate series estimators. Specifically, he gave the following results:

Newey (1997): If $K_n^2/n \to 0$,

$$\|\hat{g} - g\|_{2,0}^2 = \mathrm{O}_p\left(\frac{K_n}{n} + K_n^{-2p/d_z}\right)$$
$$\|\hat{g} - g\|_\infty \equiv \sup_{z \in [0,1]^{d_z}} |\hat{g}(z) - g(z)| = \mathrm{O}_p\left(\sqrt{K_n}\left(\sqrt{\frac{K_n}{n}} + K_n^{-p/d_z}\right)\right),$$

for $g(\cdot) \in \mathcal{C}^p[0,1]^{d_z}$ and (implicitly) $r - 2 \geq p$.

Huang (2003a) presented the same mean-square rate of convergence, specific to polynomial spline estimation, using a projection argument. The conditions were weaker, and he assumed

Huang (2003):
$$\frac{K_n \log n}{n} \to 0$$

instead of $K_n^2/n \to 0$, by using an argument with Bernstein's inequality (see Huang (2003b)). Huang claimed that this rate was minimal, and it is generally considered to be so. The results improved on Huang (1998).

See also Li & Ruppert (2008), Kohler & Krzyzak (2001), and Nychka (1995) for similar results. Hall & Opsomer (2005) also obtains mean-square and consistency results, using a white-noise model, and Li & Ruppert (2008) consider an equivalent kernel representation for degree zero and one B-splines with first- or second-degree order penalties.

- Smoothing Splines

  Using an argument based on projections and Green's functions, Cox (1984) presented a mean-square rate of convergence for multivariate smoothing spline estimates, as follows:

Cox (1984):
$$\mathbb{E}\|\hat{g} - g\|_{2,0}^2 = \mathrm{O}\left(\frac{\lambda_n}{n} + \frac{n^{(d_z - 2m)/2m}}{\lambda_n^{d_z/2m}}\right),$$

where m is the order of the derivative used in the penalty, as above. The argument relies on the order of the eigenvalues of a differential equation discussed therein, as in many other smoothing splines papers (Speckman (1985), Utreras (1988), Utreras (1979)). The order of the kth eigenvalue is shown to be k^{2m/d_z}, $k = 1, ..., K_n$, from Agmon (2010). See also Póo (1999), which uses results from Speckman (1985) for a slightly different treatment of the rates of convergence.

Also, Stone showed that under some conditions, the smoothing spline estimator achieves the optimal rate (Speckman (1985)).

- Penalized Splines

Claeskens et al. (2009) considered the mean-square rate of convergence of penalized spline estimators and presented a simple condition determining the form of this rate. Like in Cox (1984) and other smoothing spline papers discussed above, their argument relies on the eigenvalues of a differential equation, which they use to decompose the penalization matrix. Specifically, the authors show that for a constant $\tilde{c}_1$,

Claeskens, Krivobokova, Opsomer (2009):
If $K_m \equiv (K-m)(\lambda_n \tilde{c}_1)^{1/2m} n^{-1/2m} < 1$ for sufficiently large n and $g \in \mathcal{C}^p([a,b])$, then

$$\|\hat{g} - g\|_{2,0,n}^2 = O\left(\frac{K_n}{n} + \frac{\lambda^2 K_n^{2m}}{n^2} + K_n^{-2p}\right),$$

and if $K_m \geq 1$ for sufficiently large n and $g \in \mathcal{W}^{m,2}[a,b]$, then

$$\|\hat{g} - g\|_{2,0,n}^2 = O\left(\frac{n^{1/2m-1}}{\lambda_n^{1/2m}} + \frac{\lambda_n}{n} + K_n^{-2m}\right).$$

In both expressions the first term is the rate of the variance, the second term is the rate of the squared bias resulting from the penalization, and the third term is the rate of the squared approximation bias. For $K_m < 1$ and $d_z = 1$, the first and third terms match those in Newey (1997); and for $K_m \geq 1$, the first and second terms match those in Cox (1984). The assumption on the number of knots is that

$$K_n^2/n \to 0$$

for the random design case.

1.3.2 Spline Estimation - Asymptotic Distribution

There are also current results for the asymptotic distribution of the (properly standardized) $\hat{g}(\cdot)$, along with its derivatives, using the multi-index model. These results are of course necessary for hypothesis testing and inference.

- Regression Splines

  In Newey (1997), an asymptotic normality result, along with the standard errors, is given:

  Newey (1997): If $\sqrt{n}K_n^{-p/d_z} \to 0$, then

  $$\sqrt{n}V^{-1/2}(\hat{\theta}-\theta_0) \to_d \mathcal{N}(0,1), \quad \sqrt{n}\hat{V}^{-1/2}(\hat{\theta}-\theta_0) \to_d \mathcal{N}(0,1),$$

  where $V = A'Q^{-1}\Sigma Q^{-1}A$, $\hat{V} = \hat{A}'\hat{Q}^{-1}\hat{\Sigma}\hat{Q}^{-1}\hat{A}$, $A = (D(p_1),...,D(p_K))'$ (see Assumption 5 in that paper), $Q = \mathbb{E}[p^{K_n}(z)p^{K_n}(z)']$, $\Sigma = \mathbb{E}[p^{K_n}(z)p^{K_n}(z)'\sigma(z)^2]$ with $\sigma(z)^2 = \mathbb{E}\varepsilon(z)^2$, $\hat{A} = \frac{\partial a(\beta' p^K)}{\partial \beta}|_{\beta=\hat{\beta}}$, $\hat{Q} = \frac{P'P}{n}$, and $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} p^K(z_i)p^K(z_i)'(y_i - \hat{g}(z_i))^2$.

- Smoothing Splines

  For smoothing splines, Póo (1999) gives an asymptotic normality result, with a condition on the rate of λ_n.

Póo (1999): If $\lambda_n \asymp n^{-2m/(2m+1)}$, then

$$\frac{\hat{g}(z) - g(z)}{\sigma(z)} \to_d \mathcal{N}(0,1),$$

where $\sigma^2(z) = \mathbb{E}(\hat{g}(z) - g(z))^2$.

Also, some authors have considered methods for controlling the penalization term, other than the common method of using the eigenvalues of a differential equation. For example, Eilers & Marx (1996) consider a discretization for the penalty term; that is, they propose to use finite differences to approximate the integrated second derivative penalty. Similarly, Schwetlick & Kunert (1993) decouple the order of the derivative in the penalization and the order of the spline.

1.3.3 Series Estimation in the Partially Linear Model

A main paper on series estimation in the partially linear model is Donald & Newey (1994), which presents rate of convergence and asymptotic normality results. Specifically, the authors define $e_g(K_n)$ and $e_h(K_n)$ such that there are π and η with

$$\sup_{n\geq 1}\left[\sum_{i=1}^{n}\mathbb{E}(g(z_i) - p^K(z_i)'\pi)^2/n\right]^{1/2} \leq e_g(K_n)$$

$$\max_j \sup_{n\geq 1}\left[\sum_{i=1}^{n}\mathbb{E}(h_j(z_i) - p^K(z_i)'\eta)^2/n\right]^{1/2} \leq e_h(K_n)$$

where $h_j(z_i) = \mathbb{E}[x_{ji}|z_i]$ with x_{ji} equal to the ith observation of the jth regressor, as above. Thus, $e_g(K_n)$ and $e_h(K_n)$ describe how well an element of $\mathcal{S}_{n,r}$ can approximate $\hat{g}$. The authors showed that

Donald and Newey (1994) - Rate of Convergence:

$$\hat{\beta}-\beta = O_p(n^{-1/2})+O_p(e_g(k)e_h(K_n))+O_p(e_g(K_n)n^{-1/2})+O_p(e_h(K_n)n^{-1/2})+O_p(K_n^{1/2}n^{-1}),$$

under the assumption that

$$K_n/n \to 0.$$

They also showed that if $\sqrt{n}K^{-(p_g+p_h)/d_z} \to 0$, then

Donald and Newey (1994) - Asymptotic Normality:

$$(\bar{A}_n^{-1}\bar{B}_n\bar{A}_n)^{-1/2}\sqrt{n}(\hat{\beta}-\beta) \to_d \mathcal{N}(0, I),$$

where $\bar{A}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbb{E}u_iu_i'$ and $\bar{B}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\mathbb{E}\varepsilon_i^2 u_iu_i'$ are uniformly positive definite, with $u_i = x_i - h(z_i)$, as above.

Another important paper on series estimation in the partially linear model is Cattaneo, Jansson & Newey (2010), in which many regressors are allowed, that is, $d_z = O(n)$. We refer to this paper in our proof of the asymptotic normality of $\hat{\beta}$ in Chapter 3.

1.4 Overview of Results

Let $(y_i, x_i', z_i')'$, $i = 1, ..., n$ be a random sample of the random vector $(y, x', z')'$, where $y \in \mathbb{R}$ is a dependent variables and $x \in \mathbb{R}^{d_x \times 1}$ and $z \in \mathbb{R}^{d_z+1}$ are explanatory variables. As discussed above, the partially linear model is given by

$$y_i = x_i'\theta + g(z_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i|x_i, z_i] = 0, \quad \sigma_\varepsilon^2(x_i, z_i) = \mathbb{E}[\varepsilon_i^2|x_i, z_i],$$

where $v_i = x_i - h(z_i)$ with $h(z_i) = \mathbb{E}[x_i|z_i]$ and $\sigma_v^2(z_i) = \mathbb{E}[v_i^2|z_i]$. A series estimator of β is obtained by regressing y_i on x_i and approximating functions of z_i.

For this project, we consider asymptotics for both the nonparametric and parametric components of the partially linear model. Specifically, in Chapter 1, we give mean-square rates of convergence for $\partial^\ell \hat{g}$ in the fixed norm $\|\hat{g} - g\|_{2,\ell}^2$ and the empirical norm $\|\hat{g} - g\|_{2,\ell,n}^2$. Section 1 handles inversion of $P'P/n$ by showing that its eigenvalues are bounded above and below by positive constants with probability approaching one under the rate condition $K_n \log n/n \to 0$. Section 2 presents an asymptotic expression for the eigenvalues of $(P'P/n)^{-1/2}D(P'P/n)^{-1/2}$, using theory from the field of partial differential equations.

Section 3 gives an expression for the conditional mean squared error of $\partial^\ell \hat{g}$, with $\ell = (\ell_1, ..., \ell_d)$, $\ell_j \leq r-2$, in terms of these eigenvalues, and Section 4 uses this expression to find the rates of convergence.

Chapter 2 considers the asymptotic distribution of both the parametric and nonparametric components. Section 1 gives the distribution of $\hat{\theta}$. Section 2 presents the distribution of $\partial^\ell \hat{g}(\cdot)$, and to that end, gives a lower bound on its pointwise variance and an upper bound on its pointwise bias. Also, Section 3 gives the standard error estimates for the parametric and nonparametric components.

1.4.1 Empirical and Fixed Mean Squared Error

Define

$$\mathcal{C}_n(m) = \left(\frac{\varsigma\lambda_n}{n}\right)^{d_z/2m}(K_n - m), \ \ \varsigma = \left(\frac{(\Gamma(\frac{1}{2m})/2\pi m)^{d_z}}{\Gamma(1+\frac{d_z}{2m})}\int\limits_{[0,1]^d} f(x)^{d_z/2m}dx\right)^{-2m/d_z}$$

Our first theorem present rates for the empirical and fixed mean squared error. As given above, the fixed mean squared error is the average of $(\hat{g}(z) - g(z))^2$ over the population, and the empirical mean squared error is the average of $(\hat{g}(z) - g(z))^2$ over the observations. Our assumptions are as follows:

(A1): $(y_1, x_1, z_1), ..., (y_n, x_n, z_n)$ are i.i.d.

(A2): $\sigma_\varepsilon(x,z)^2$ and $f(z)$ (the density of z) are bounded above and below away from zero, uniformly in z.

These assumptions are standard in the literature and are difficult to relax without affecting the rates of convergence.

(A3): $\dfrac{K_n \log n}{n} \to 0$.

This assumption is weaker than that common in the literature for regression splines and penalized splines, e.g. Newey (1997), Zhou et al. (1998), and Claeskens et al. (2009); and

Huang (2003a) claims this assumption is minimal. We use it to bound the eigenvalues of the design matrix away from zero and to obtain an asymptotic expression for the eigenvalues used in the decomposition of the penalization matrix.

(A4): For all k and j, $t_{j,k+1} - t_{jk} \asymp 1/K_n^{1/d_z}$.

We also choose $m \leq r_g$ and $m > d_z/4$.

Theorem 1: Under (A1)-(A4), for $g(\cdot) \in \mathcal{C}^p[0,1]^{d_z}$ and $r_g = \min\{p, r-2\}$, if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then

$$\|\hat{g} - g\|_{2,\ell,n}^2 = O_p\left(K_n^{2\ell/d_z}\left(\frac{K_n}{n} + \frac{\lambda_n^2}{n^2}K_n^{2m/d_z} + K_n^{-2r_g/d_z}\right)\right),$$

$$\|\hat{g} - g\|_{2,\ell}^2 = O_p\left(K_n^{2\ell/d_z}\left(\frac{K_n}{n} + \frac{\lambda_n^2}{n^2}K_n^{2m/d_z} + K_n^{-2r_g/d_z}\right)\right),$$

and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then

$$\|\hat{g} - g\|_{2,\ell,n}^2 = O_p\left(K_n^{2\ell/d_z}\left(\frac{n^{(d_z-2m)/2m}}{\lambda_n^{d_z/2m}} + \frac{\lambda_n}{n} + K_n^{-2r_g/d_z}\right)\right),$$

$$\|\hat{g} - g\|_{2,\ell}^2 = O_p\left(K_n^{2\ell/d_z}\left(\frac{n^{(d_z-2m)/2m}}{\lambda_n^{d_z/2m}} + \frac{\lambda_n}{n} + K_n^{-2r_g/d_z}\right)\right),$$

These results agree with the literature. In particular, for $d_z = 1$ and $\ell = 1$, we recover the result in Claeskens et al. (2009). (Note that Claeskens et al. (2009) assumed a different functional space for $g(\cdot)$ in the case $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, resulting in K_n^{-2m} instead of K_n^{-2p} in their result.) Also, when $\lambda_n = 0$, the rate when $\mathcal{C}_n(m) < 1$ for all sufficiently large n matches the result in Newey (1997). For $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, the first two terms are the same as in Cox (1984). The third term is not present in that result since $K_n = n$.

The first term in each expression is the rate of the variance. For $\mathcal{C}_n(m) < 1$, the variance grows with K_n and declines with n, since more observations results in a smoother estimate, and a larger K_n allows for a more jagged estimate. For $\mathcal{C}_n(m) \geq 1$, the variance decreases with λ_n, since a larger penalization forces a smoother estimate. The second term in each expression is the rate of the squared bias resulting from the penalization. In both cases, the squared bias decreases with n and increases with λ_n, since a larger penalization produces less fidelity to the data. Finally, the third term in each expression is the smoothing (approximation) bias. As the number of basis functions used increases, the approximation over $\mathcal{S}_{n,r}$ improves, causes this bias to decrease.

1.4.2 Asymptotic Distribution and Standard Error Estimates for the Parametric Component

We present an asymptotic linear representation of $\hat{\theta}$, along with an asymptotic normality result, under weak conditions on the tuning parameter sequences. We also give simple, plug-in standard error estimates, which are robust to heteroskedasticity of unknown form.

Define $T = I - P(P'P + \lambda_n D)^{-1}P'$; $V_n = \Gamma_n^{-1}\Omega_n\Gamma_n^{-1}$ with $\Gamma_n = XT'TX$, $\Omega_n = XT'T\Sigma TT'X$, and $\Sigma = \text{diag}(\sigma_\varepsilon^2(x_1, z_1), ..., \sigma_\varepsilon^2(x_n, z_n))$; $\hat{V}_n = \Gamma_n^{-1}\hat{\Omega}_n\Gamma_n^{-1}$, with $\hat{\Omega}_n = XT'T\hat{\Sigma}TT'X$ and $\hat{\Sigma} = \text{diag}(\hat{\varepsilon}_1^2, ..., \hat{\varepsilon}_n^2)$; and $\Gamma = \mathbb{E}[\nu_i\nu_i']$, $\Omega = \mathbb{E}[\nu_i\nu_i'\varepsilon_i^2]$. We need the following assumptions:

(A5): $\mathbb{E}[\|v_i\|^4|z_i]$ and $\mathbb{E}[\epsilon_i^4|z_i]$ are bounded above.

Let p_h be the minimum (over j) number of continuous derivatives of h^j, and define $r_h = \min\{r-2, p_h\}$.

(A6): (a) $\sqrt{n}K_n^{-(r_g+r_h)/d_z} \to 0$, (b) if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then $\sqrt{n}\lambda_n K_n^{m/d_z}/n \to 0$, and (c) if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then $\sqrt{n}\sqrt{\lambda_n/n} = \lambda_n \to 0$.

Note that for $\mathcal{C}_n(m) < 1$, (A6) and (A3) easily ensure that $\|\hat{g} - g\|_{2,0,n} \to_p 0$, since $K_n \to \infty$. For $\mathcal{C}_n(m) \geq 1$, since $K_n \gtrsim ((\lambda_n/n)^{-d_z/2m})$,

$$\begin{aligned}
\frac{K_n \log n}{n} &\gtrsim \frac{(\lambda_n/n)^{-d_z/2m}\log n}{n} \\
&= \frac{n^{d_z/2m}\log n}{\lambda_n^{d_z/2m} n} \\
&= \frac{n^{(d_z-2m)/2m}\log n}{\lambda_n^{d/2m}}.
\end{aligned}$$

So by (A3), $n^{(d_z-2m)/2m}\log n/\lambda_n^{d/2m} \to 0$, and again, $\|\hat{g}-g\|_{2,0,n} \to_p 0$.

Theorem 2: Under (A1), (A3), (A5), and (A6),

$$\text{(a)} \quad V_n^{-1/2}\sqrt{n}(\hat{\theta}-\theta) = V_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_i\varepsilon_i + o_p(1) \to_p \mathcal{N}(0,1),$$

$$V_n = \Gamma^{-1}\Omega\Gamma^{-1} + o_p(1),\ \ \Gamma = \mathbb{E}[\nu_i\nu_i'],\ \ \Omega = \mathbb{E}[\nu_i\nu_i'\varepsilon_i^2]$$

$$\text{(b)} \quad \Gamma_n = \Gamma + o_p(1), \quad \hat{\Omega}_n = \Omega + o_p(1).$$

The main differences between this result and a similar result in Cattaneo et al. (2010) is that a penalization is allowed but $K_n = O(n)$ is not allowed.

1.4.3 Asymptotic Distribution and Standard Errors for the Nonparametric Component

The asymptotic distribution for the nonparametric component (and its derivatives) are also given. The method of proof involves noting that

$$\begin{aligned}
\partial^\ell\hat{g}(z) - \partial^\ell g(z) &= \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'(Y-X\hat{\theta}) - \partial^\ell g(z) \\
&= \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'(Y-X\theta) - \\
&\quad\ \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta) - \partial^\ell g(z) \\
&= \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'(Y-G-X\theta) + \\
&\quad\ [\partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'G - \partial^\ell g(z)] - \\
&\quad\ \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta).
\end{aligned}$$

The first term approaches a normal distribution, and the second and third terms are bias terms that approach zero (in probability), under some assumptions.

Define $W_{\ell,n} = \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'\Sigma P(P'P+\lambda D)^{-1}\partial^\ell p^K(z)$ and $\hat{W}_{\ell,n} = \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'\hat{\Sigma}P(P'P+\lambda D)^{-1}\partial^\ell p^K(z)$, with Σ and $\hat{\Sigma}$ defined as above.

(A7): (a) If $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then

$$\frac{\lambda K_n^{m/d_z}/n + 1\{|\ell| = 0\} K_n^{-r_g/d_z}}{\sqrt{K_n/n}} + 1\{|\ell| > 0\}\sqrt{n}K_n^{-(r_g-|\ell|)/d_z} \to 0,$$

(b) if $1 \leq \mathcal{C}_n(m) < \infty$, then

$$\frac{\sqrt{\lambda_n/n} + 1\{|\ell| = 0\} K_n^{-r_g/d_z}}{\sqrt{n^{(d_z-2m)/2m}/\lambda_n^{d_z/2m}}} + 1\{|\ell| > 0\}\sqrt{n}K_n^{-(r_g-|\ell|)/d_z} \to 0.$$

(c) and if $\mathcal{C}_n(m) = \infty$, then

$$\frac{\lambda K_n^{2m/d_z}}{n}\left(\frac{\sqrt{\lambda_n/n}\sqrt{\lambda_n K_n^{2m/d_z}/n} + 1\{|\ell| = 0\} K_n^{-r_g/d_z}}{\sqrt{n^{(d_z-2m)/2m}/\lambda_n^{d_z/2m}}} + 1\{|\ell| > 0\}\sqrt{n}K_n^{-(r_g-|\ell|)/d_z}\right) \to 0,$$

$$\frac{\lambda_n K_n^{2m/d_z}}{n}\frac{n^{(d_z-2m)/2m}}{\lambda_n^{d_z/2m}} \to 0.$$

These assumptions are needed to guarantee that the bias (divided by the variance) disappears asymptotically.

(A8): $K_n^{2|\ell|/d_z} K_n/n \to 0$.

Note that for $|\ell| = 0$, this condition is already satisfied since $K_n \log n/n \to 0$.

(A9): $\sup_{z\in[0,1]^d} |\hat{g}(z) - g(z)| \leq O_p(1)$.

A discussion of this condition is included in Chapter 3.

Theorem 3: Under (A1)-(A4) and (A7)-(A9),

$$\text{(a)} \quad \frac{\partial^\ell \hat{g}(z) - \partial^\ell g(z)}{\sqrt{W_{\ell,n}(z)}} \to_d \mathcal{N}(0,1),$$

$$\text{(b)} \quad \hat{W}_{\ell,n}(z) = W_{\ell,n}(z) + o_p(1).$$

Because of the triangular array structure, the Lindeberg-Feller CLT is used in the proof.

CHAPTER II

Rates of Convergence

For ease of notation, we let $d \equiv d_z$, $K \equiv K_n$, and $\lambda \equiv \lambda_n$ throughout Chapters 2 and 3.

2.1 Eigenvalues of the Design Matrix

We first consider the eigenvalues of the design matrix $P'P/n$, in order to ensure invertibility, required for the mean squared error expansion. We multiply each basis function p_k by $\sqrt{K}$, as a normalization (see Newey (1997)). Our only noteworthy assumption is that $K \log n/n \to 0$, generally considered to be minimal (see Huang (2003b)). We show that these eigenvalues are bounded above and below asymptotically.

Let $\{\Delta_k\}_{k=1}^{K}$ be the set of hyper-intervals $(t_{1,k_1}, t_{1,k_1+1}] \times \cdots \times (t_{d,k_d}, t_{d,k_d+1}]$, $k_1, ..., k_d = 1, ..., K^{1/d}$, and let $\{\Delta_k^j\}_{k=1}^{K^{1/d}}$ be the set of intervals $(t_{j,k}, t_{j,k+1}]$, $k = 1, ..., K^{1/d}$, $j = 1, ..., d$.

Lemma 2.1: If $K \log n/n \to 0$,

$$c_1 + o_p(1) \leq \tilde{\lambda}_{\min} \leq \tilde{\lambda}_{\max} \leq c_2 + o_p(1),$$

for some constants $c_1, c_2 > 0$.

Proof: Given in the Appendix.

2.2 Eigenvalues for the Penalization Matrix

We now consider the eigenvalues of the penalization matrix D, where the $k\ell$ element of D is

$$D_{k\ell} = \int\limits_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^{d} \frac{\partial^m p_k}{\partial z_{j_1} \cdots \partial z_{j_m}} \frac{\partial^m p_\ell}{\partial z_{j_1} \cdots \partial z_{j_m}} dz.$$

as given previously. Let $\mu_k^{n,K}$ be the kth such eigenvalue, where $\mu_1^{n,K} \leq \mu_2^{n,K} \leq \cdots \leq \mu_K^{n,K}$.

2.2.1 Convergence of Eigenvalues

We show that $\mu_k^{n,K}$ approaches the kth eigenvalue μ_k of a well-known differential equation as $n, K \to \infty$.

Define $\tilde{a}(u,v)$, $b(u,v)$, $b_n(u,v)$, and $a(u,v)$ to be the bilinear forms

$$\tilde{a}(u,v) \equiv \int\limits_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^{d} \frac{\partial^m u}{\partial x_{j_1} \cdots \partial x_{j_m}} \frac{\partial^m v}{\partial x_{j_1} \cdots \partial x_{j_m}} dx,$$

$$b(u,v) \equiv \int\limits_{[0,1]^d} u(x)v(x)f(x)dx,$$

$$b_n(u,v) \equiv \frac{1}{n}\sum_{i=1}^{n} u(Z_i)v(Z_i),$$

$$a(u,v) = \tilde{a}(u,v) + b(u,v).$$

Define $H \equiv W^{m,2} = \{g \in L^2([0,1]^d) : \text{for all} \, |\alpha| \leq m, \partial^\alpha g \in L^2([0,1]^d)\}$, where $\alpha = (\alpha_1, ..., \alpha_d)$ and $|\alpha| = \alpha_1 + \cdots + \alpha_d$, as usual (see Adams & Fournier (2003) for a discussion of Sobolev spaces). We consider the eigenvalues of the equation

$$a(u,v) = \mu b(u,v) \, \text{for all} \, v \in H, \, \text{for some} \, u \in H. \tag{2.1}$$

Note that these eigenvalues are the eigenvalues of the equation $\tilde{a}(u,v) = \mu b(u,v)$, but with a value of 1 added to each. Since all of the eigenvalues of the latter equation are non-negative (as shown in the Appendix), all μ satisfying (2.1) are positive. Thus, a is positive definite on $H \times H$ and is therefore invertible, that is,

$$\inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |a(u,v)| \geq C > 0$$

and similarly $\inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |b(u,v)| \geq C > 0$. These inequalities also hold on $\mathcal{S}_{n,r}$ since $\mathcal{S}_{n,r} \subset H$. Also, since a and b are both integrals on finite domains, $|a(u,v)| \leq C\|u\|_H\|v\|_H$ and $|b(u,v)| \leq C\|u\|_H\|v\|_H$.

Define the operator T by $a(Tu,v) = \mu b(u,v)$ for all $v \in H$. (Note that T exists by the Riesz Representation Theorem, as discussed in Fix (1972)). Since a and b are both defined on $H \times H$, T is compact (Aziz & Babuska (1972), p. 305 and 319). Observe that μ satisfies $a(u,v) = \mu b(u,v)$ if and only if $1/\mu$ is an eigenvalue of T, so by the compactness of T, the eigenvalues form a countable set with accumulation only at infinity. This is confirmed in the Appendix.

Consider the equations

$$a(u^K, v^K) = \mu^K b(u_K, v^K) \text{ for all } v^K \in \mathcal{S}_{n,r}, \text{ for some } u^K \in \mathcal{S}_{n,r} \tag{2.2}$$

$$a(u^K, v^K) = \mu^{K,n} b_n(u^K, v^K) \text{ for all } v^K \in \mathcal{S}_{n,r}, \text{ for some } u^K \in \mathcal{S}_{n,r}, \tag{2.3}$$

We define T^K similarly to T, and define

$$\begin{aligned}
&d(\mu) \equiv \inf\nolimits_{\|u\|_H=1} \sup\nolimits_{\|v\|_H=1} |a(u,v) - \mu b(u,v)|,\\
&d_K(\mu) \equiv \inf\nolimits_{\|u^K\|_H=1} \sup\nolimits_{\|v^K\|_H=1} |a(u^K,v^K) - \mu^K b(u^K,v^K)|,\\
&d_{n,K}(\mu) \equiv \inf\nolimits_{\|u^K\|_H=1} \sup\nolimits_{\|v^K\|_H=1} |a(u^K,v^K) - \mu^{n,K} b_n(u^K,v^K)|.
\end{aligned}$$

We let $\mu_1 \leq \cdots \leq \mu_K$ and $\mu_1^K \leq \cdots \leq \mu_K^K$ as with $\mu_1^{n,K}, ..., \mu_K^{n,K}$.

Lemma 2.2.1: Under (A3), for all $k \geq 1$, $|\mu_k - \mu_k^{n,K}| \to_p 0$.

Proof: For conciseness, we drop the k subscript. We present an argument similar to the proof of Theorem 10.5.1 in Aziz & Babuska (1972).

We first show that if $d(\mu_*) = 0$, then given small $\epsilon > 0$ and a value $\mu^{n,K}$ such $|\mu_* - \mu^{n,K}| = \epsilon$ for large n and K, we have $d_{n,K}(\mu^{n,K}) \geq C_\epsilon + o_p(1)$, for some constant $C_\epsilon > 0$ independent of n and K.

Let $d(\mu_*) = 0$, then since the eigenvalues μ that satisfy (2.1) are isolated (as shown in the Appendix), there is some $\rho > 0$ such that $|\mu_* - \mu| > \rho$ for all other μ. Then given $\epsilon > 0$

such that $\epsilon < \rho$, there is no μ such that $|\mu_* - \mu| = \epsilon$.

So given a value μ_0 such that $|\mu_0 - \mu_*| = \epsilon$, since

$$\sup_{\|u\|_H=1} \inf_{\|v\|_H=1} |a(u,v) - \mu_* b(u,v)| \leq \inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |a(u,v) - \mu_* b(u,v)| = 0,$$

we have

$$\begin{aligned} d(\mu_0) &= \inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |a(u,v) - \mu_0 b(u,v)| \\ &\geq \inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |(\mu_* - \mu_0) b(u,v)| - \sup_{\|u\|_H=1} \inf_{\|v\|_H=1} |a(u,v) - \mu_* b(u,v)| \\ &\geq |\mu_* - \mu_0| \inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |b(u,v)| \\ &\geq C_\epsilon. \end{aligned}$$

Given $u_0^K \in \mathcal{S}_{n,r}$, let w_0 be such that $a(w_0, v) = \mu_0 b(u_0^K, v)$ for all $v \in H$, and let w_0^K be such that $a(w_0^K, v^K) = \mu_0 b(u_0^K, v^K)$ for all $v^K \in \mathcal{S}_{n,r}$. Then for all $v \in H$, $a(u_0^K, v) - \mu_0 b(u_0^K, v) = a(u_0^K, v) - a(w_0, v) = a(u_0^K - w_0, v)$, so

$$\begin{aligned} \sup_{\|v\|_H=1} |a(u_0^K - w_0, v)| &= \sup_{\|v\|_H=1} |a(u_0^K, v) - \mu_0 b(u_0^K, v)| \\ &= \|u_0^K\|_H \sup_{\|v\|_H=1} |a(u_0^K/\|u_0^K\|_H, v) - \mu_0 b(u_0^K/\|u_0^K\|_H, v)| \\ &\geq \|u_0^K\|_H \inf_{\|u\|_H=1} \sup_{\|v\|_H=1} |a(u_0, v) - \mu_0 b(u_0, v)| \\ &\geq \|u_0^K\|_H C_\epsilon. \end{aligned}$$

Then since $\sup_{\|v\|_H=1} |a(u_0^K - w_0, v)| \leq \|u_0^K - w_0\|_H$, we have $\|u_0^K - w_0\|_H \geq \|u_0^K\|_H C_\epsilon$. As shown in Fix, $w_0 = \mu_0 T u_0^K$ and $w_0^K = \mu_0 T^K u_0^K$. So

$$\begin{aligned} \|w_0 - w_0^K\|_H &= \mu_0 \|(T - T^K) u_0^K\|_H \\ &\leq \mu_0 \|u_0^K\|_H \sup_{\|u^K\|=1} \|(T - T^K) u_0^K\|_H \\ &\leq \mu_0 \|u_0^K\|_H s_{n,K}, \end{aligned}$$

for some sequence $s_{n,K} \to 0$ as $n, K \to \infty$. As above, for all $v^K \in \mathcal{S}_{n,r}$, $a(u_0^K, v^K) -$

$\mu_0 b(u_0^K, v^K) = a(u_0^K - w_0^K, v^K)$. So by the invertibility of a,

$$\begin{aligned}
\sup_{\|v^K\|_H=1} |a(u_0^K, v^K) - \mu_0 b(u_0^K, v^K)| &= \sup_{\|v^K\|_H=1} |a(u_0^K - w_0^K, v^K)| \\
&\geq \|u_0^K - w_0^K\|_H \inf_{\|u^K\|_H=1} \sup_{\|v^K\|_H=1} |a(u_0^K, v^K)| \\
&\geq C\|u_0^K - w_0^K\|_H \\
&\geq C(\|u_0^K - w_0\|_H - \|w_0 - w_0^K\|_H) \\
&\geq C(C_\epsilon \|u_0^K\|_H - \mu_0 s_{n,K} \|u_0^K\|_H) \\
&\geq C_\epsilon(\|u_0^K\|_H (1 - \mu_0 s_{n,K}/C_\epsilon) \\
&\geq C_\epsilon,
\end{aligned}$$

for sufficiently large n and K.

Now as shown above, since $K \log n/n \to 0$, for all $u^K, v^K \in \mathcal{S}_{n,r}$ with $\|u^K\|_H = \|v^K\|_H = 1$,

$$|b(u^K, v^K) - b_n(u^K, v^K)| = \left| \mathbb{E}[u^K(z) v^K(z)] - \frac{1}{n} \sum_{i=1}^{n} u^K(Z_i) v^K(Z_i) \right| \leq \tilde{s}_{n,K},$$

for some sequence $\tilde{s}_{n,K} \to_p 0$ as $n, K \to \infty$, and thus

$$\begin{aligned}
\sup_{\|v^K\|_H=1} |a(u_0^K, v^K) - \mu_0 b_n(u_0^K, v^K)| &\geq \sup_{\|v^K\|_H=1} |a(u_0^K, v^K) - \mu_0 b(u_0^K, v^K)| - \\
&\quad \mu_0 \inf_{\|v^K\|_H=1} |b(u_0^K, v^K) - b_n(u_0^K, v^K)| \\
&\geq C_\epsilon - \mu_0 \tilde{s}_{n,K} \\
&\geq C_\epsilon + o_p(1).
\end{aligned}$$

So we have shown that given μ_0 such that $|\mu_0 - \mu_*| = \epsilon$ and $\mu_0^K \in \mathcal{S}_{n,r}$,

$$\sup_{\|v^K\|_H=1} |a(u_0^K, v^K) - \mu_0 b_n(u_0^K, v^K)| \geq C_\epsilon + o_p(1).$$

So since μ_0 and μ_0^K were arbitrary, we have $d_{n,K}(\mu^{n,K}) \geq C_\epsilon + o_p(1)$, for all $\mu^{n,K}$ such that $|\mu^{n,K} - \mu_*| = \epsilon$, as desired.

Using this result, we now show that with probability approaching one (wpa1), for sufficiently large n and K, there exists $\mu_*^{n,K}$ such that $|\mu_* - \mu_*^{n,K}| < \epsilon$ and $d_{n,K}(\mu_*^{n,K}) = 0$.

Suppose that for all n and K, there is no zero of $d_{n,K}(\mu^{n,K})$ such that $|\mu^{n,K} - \mu_*| < \epsilon$.

Then $(d_{n,K}(\mu^{n,K}))^{-1}$ is subharmonic and attains its maximum on the circumference $|\mu^{n,K} - \mu_*| = \epsilon$ (see Aziz). Thus, $(d_{n,K}(\mu^{n,K}))^{-1} < 1/(C_\epsilon + o_p(1)) < \infty$, that is, $d_{n,K}(\mu^{n,K}) > C_\epsilon + o_p(1)$, for all $\mu^{n,K}$ such that $|\mu^{n,K} - \mu_*| < \epsilon$ and sufficiently large n and K. Then for all μ^K such that $|\mu^K - \mu_*| < \epsilon$,

$$\begin{aligned} d_K(\mu^K) &= \inf_{\|u^K\|_H=1} \sup_{\|v_K\|_H=1} |a(u^K, v^K) - \mu^K b(u^K, v^K)| \\ &\geq \inf_{\|u^K\|_H=1} \sup_{\|v_K\|_H=1} |(a(u^K, v^K) - \mu^K b_n(u^K, v^K))| - \\ &\quad \mu^{n,K} \sup_{\|u^K\|=1} \inf_{\|v^K\|_H=1} |b(u^K, v^K) - b_n(u^K, v^K)| \\ &\geq C_\epsilon + o_p(1). \end{aligned}$$

So since $|\mu_* - \mu_*| = 0 < \epsilon$,

$$\begin{aligned} \sup_{\|v^K\|_H=1} |a(u_*^K, v) - \mu_* b(u_*^K, v)| &\geq \|u_*^K\|_H \inf_{\|u^K\|_H=1} \sup_{\|v^K\|_H=1} |a(u^K, v) - \mu_* b(u^K, v)| \\ &\geq \|u_*^K\|_H (C_\epsilon + o_p(1)). \end{aligned}$$

Let u_* satisfy $a(u_*, v) = \mu_* b(u_*, v)$ for all $v \in H$, then as discussed above, $u_* = \mu_* T u_*$. So letting $u_*^K = \mu_* T^K u_*$, we have $\|u_* - u_*^K\|_H \leq \mu_* s_{n,K} \|u_*\|_H$. Then for all v^K,

$$\begin{aligned} |a(u_* - u_*^K, v^K) - \mu_* b(u_* - u_*^K, v^K)| &\leq |a(u_* - u_*^K, v^K)| + \mu_* |b(u_* - u_*^K, v^K)| \\ &\leq \|u_* - u_*^K\| \|v^K\|_H (1 + \mu_*) \\ &\leq \mu_* s_{n,K} \|u_*\|_H \|v^K\|_H (1 + \mu_*). \end{aligned}$$

Thus,

$$
\begin{aligned}
\sup_{\|v^K\|_H=1} |a(u_*, v^K) - \mu_* b(u_*, v^K)| &\geq \sup_{\|v^K\|_H=1} |a(u_*^K, v^K) - \mu_* b(u_*^K, v^K)| - \\
&\quad \inf_{\|v^K\|_H=1} |a(u_* - u_*^K, v^K) - \mu_* b(u_* - u_*^K, v^K)| \\
&\geq (C_\epsilon + o_p(1))\|u_*^K\|_H - \mu_* s_{n,K}\|u_*\|_H\|(1+\mu_*) \\
&\geq (C_\epsilon + o_p(1))(\|u_*\|_H - \|u_*^K - \\
&\quad u_*\|_H) - \mu_* s_{n,K}\|u_*\|_H(1+\mu_*) \\
&\geq (C_\epsilon + o_p(1))(\|u_*\|_H - \\
&\quad \mu_* s_{n,K}\|u_*\|_H) - \mu_* s_{n,K}\|u_*\|_H(1+\mu_*) \\
&= \|u_*\|_H[(C_\epsilon + o_p(1))(1 - \mu_* s_{n,K}) - \\
&\quad \mu_* s_{n,K}(1+\mu_*)] \\
&\geq (C_\epsilon + o_p(1))\|u_*\|_H + o_p(1)
\end{aligned}
$$

So wpa1 μ_* does not satisfy 2.1, which is contrary to our assumption. So wpa1 it must be that for sufficiently large n and K, there exists $\mu_*^{n,K}$ satisfying (3) such that $|\mu_* - \mu_*^{n,K}| < \epsilon$.

Thus we have shown that if μ_* is an eigenvalue of 2.1, then given sufficiently small $\epsilon > 0$, wpa1 there exists $\mu_*^{n,K}$ such that $|\mu_* - \mu_*^{n,K}| < \epsilon$ and $\mu_*^{n,K}$ satisfies (3). The result follows.

2.2.2 Decomposition of the Penalization Matrix

We now show that $(P'P/n)^{-1/2}D(P'P)^{-1/2}$ has an eigenvalue decomposition with the eigenvalues $\mu_1^{n,K}, ..., \mu_K^{n,K}$ just discussed. This decomposition is crucial to the expression for the mean squared error, shown in a later section, and the eigenvalues determine the rate of this expression.

Lemma 2.2.2: $(P'P/n)^{-1/2}D(P'P/n)^{-1/2} = UMU'$ where M is the diagonal matrix of eigenvalues $\mu_1^{n,K}, ..., \mu_K^{n,K}$ and U is an orthogonal matrix of eigenvectors.

Proof: Let $\psi_1, ..., \psi_K$ be the (random) eigenfunctions corresponding to $\mu_1^{n,K}, ..., \mu_K^{n,K}$. We note that for any $j \neq k$,

$$\begin{aligned} 0 &= a(\psi_j, \psi_k) - a(\psi_k, \psi_j) \\ &= \mu_j^{n,K} b_n(\psi_j, \psi_k) - \mu_k^{n,K} b_n(\psi_k, \psi_j) \\ &= (\mu_j^{n,K} - \mu_k^{n,K}) b_n(\psi_j, \psi_k), \end{aligned}$$

so $b_n(\psi_j, \psi_k) = 0$. Thus, $\psi_1, ..., \psi_K$ are orthonormal, in the sense that

$$\frac{1}{n}\sum_{i=1}^{n} \psi_j(Z_i)\psi_k(Z_i) = 1\{j = k\}$$

(noting that ψ_k can be normalized if $\|\psi_k\|_H \neq 1$). (Orthogonality of eigenfunctions is a well-known property of Hermitian operators.) Suppose that $\sum_{k=1}^{K} c_k \psi_k(x) = 0$ for some constants $c_1, ..., c_K$. Then for any $k_0 \in [1, K]$, $\sum_{k=1}^{K} c_k \psi_{k_0}(x)\psi_k(z) = 0$ and thus

$$\begin{aligned} 0 &= \sum_{k=1}^{K} c_k \frac{1}{n} \sum_{i=1}^{n} \psi_{k_0}(Z_i)\psi_k(Z_i) \\ &= \sum_{k=1}^{K} c_k 1\{k = k_0\} \\ &= c_{k_0}. \end{aligned}$$

So $c_k = 0$, $k = 1, ..., K$. Thus, $\psi_1, ..., \psi_K$ are linearly independent and span $\mathcal{S}_{n,r}$. So we can write $p_q = \sum_{k=1}^{K} \alpha_{qk}\psi_k$, for $q = 1, ..., K$, for some constants α_{qk}, $k = 1, ..., K$. Then

$$\begin{aligned} (P'P/n)_{qr} &= \frac{1}{n}\sum_{i=1}^{n} p_q(Z_i)p_r(Z_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}[\sum_{k=1}^{K} \alpha_{qk}\psi_k(Z_i)][\sum_{k=1}^{K} \alpha_{rk}\psi_k(Z_i)] \\ &= \sum_{k_1,k_2=1}^{K} \alpha_{qk_1}\alpha_{rk_2} \frac{1}{n}\sum_{i=1}^{n} \psi_{k_1}(Z_i)\psi_{k_2}(Z_i) \\ &= \sum_{k_1,k_2=1}^{K} \alpha_{qk_1}\alpha_{rk_2} 1\{k_1 = k_2\} \\ &= \sum_{k=1}^{K} \alpha_{qk}\alpha_{rk}, \end{aligned}$$

So defining A such that $A_{jk} = \alpha_{jk}$, we have

$$\begin{aligned} P'P/n &= \begin{pmatrix} \sum_{k=1}^K \alpha_{1k}\alpha_{1k} & \cdots & \sum_{k=1}^K \alpha_{1k}\alpha_{Kk} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^K \alpha_{Kk}\alpha_{1k} & \cdots & \sum_{k=1}^K \alpha_{Kk}\alpha_{Kk} \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1K} \\ \vdots & \ddots & \vdots \\ \alpha_{K1} & \cdots & \alpha_{KK} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{K1} \\ \vdots & \ddots & \vdots \\ \alpha_{1K} & \cdots & \alpha_{KK} \end{pmatrix} \\ &= AA'. \end{aligned}$$

Also,

$$\begin{aligned} (D)_{qr} &= \int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^d \frac{\partial^m p_q(x)}{\partial x_{j_1}\cdots\partial x_{j_m}} \frac{\partial^m p_r(x)}{\partial x_{j_1}\cdots\partial x_{j_m}} dx \\ &= \int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^d \frac{\partial^m [\sum_{k=1}^K \alpha_{qk}\psi_k(x)]}{\partial x_{j_1}\cdots\partial x_{j_m}} \frac{\partial^m [\sum_{k=1}^K \alpha_{rk}\psi_k(x)]}{\partial x_{j_1}\cdots\partial x_{j_m}} dx \\ &= \sum_{k_1,k_2=1}^K \alpha_{qk_1}\alpha_{rk_2} \int_{[0,1]^d} \sum_{j_1,\ldots,j_m=1}^d \frac{\partial^m \psi_{k_1}(x)}{\partial x_{j_1}\cdots\partial x_{j_m}} \frac{\partial^m \psi_{k_2}(x)}{\partial x_{j_1}\cdots\partial x_{j_m}} dx \\ &= \sum_{k_1,k_2=1}^K \alpha_{qk_1}\alpha_{rk_2}\mu_{k_1}^{n,K} 1\{k_1 = k_2\} \\ &= \sum_{k=1}^K \alpha_{qk}\alpha_{rk}\mu_k^{n,K} \end{aligned}$$

Then

$$\begin{aligned} D &= \begin{pmatrix} \sum_{k=1}^K \mu_k^{n,K}\alpha_{1k}\alpha_{1k} & \cdots & \sum_{k=1}^K \mu_k^{n,K}\alpha_{1k}\alpha_{Kk} \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^K \mu_k^{n,K}\alpha_{Kk}\alpha_{1k} & \cdots & \sum_{k=1}^K \mu_k^{n,K}\alpha_{Kk}\alpha_{Kk} \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1K} \\ \vdots & \ddots & \vdots \\ \alpha_{K1} & \cdots & \alpha_{KK} \end{pmatrix} \begin{pmatrix} \mu_1^{n,K} & & 0 \\ & \ddots & \\ 0 & & \mu_K^{n,K} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{K1} \\ \vdots & \ddots & \vdots \\ \alpha_{1K} & \cdots & \alpha_{KK} \end{pmatrix} \\ &= AMA' \end{aligned}$$

So $(P'P/n)^{-1/2}D(P'P/n)^{1/2} = (AA')^{-1/2}(AMA')(AA')^{-1/2}$. Let $U = (AA')^{-1/2}A$, then $UU' = (AA')^{-1/2}AA'(AA')^{1/2} = I$. Since U is square, we also have $U'U = I$.

2.2.3 Asymptotic Representation of Eigenvalues

We now present an expression for the eigenvalues of the equation

$$a(u, v) = \mu b(u, v) \text{ for all } v \in H$$

where $u \in H$, using theory from the field of multivariate differential equations. We use these eigenvalues in Section 2 above to decompose the penalization matrix.

The expression we give is an asymptotic expression, meaning in this case that it is valid as $k \to \infty$, where k is the index on the eigenvalues. That is, the expression is true for large eigenvalues. A result without the $o_k(1)$ term in the final expression is known in the univariate case (Claeskens et al. (2009), Speckman (1985)), but the authors are unaware of an existing result (with or without the $o_k(1)$) in the multivariate case. This result is key for the mean-square and uniform rates of convergence, and we believe it will prove useful for many researchers in the future.

Lemma 2.2.3: For $k \geq 1$,

$$\mu_k = \left[\left(\frac{(\Gamma(\frac{1}{2m})/2\pi m)^d}{\Gamma(1 + \frac{d}{2m})} \int_{[0,1]^d} f(x)^{d/2m} dx \right)^{-2m/d} + o_k(1) \right] k^{2m/d},$$

where $o_k(1)$ represents a term that goes to 0 as $k \to \infty$.

Proof: The proof is given in the Appendix.

2.3 Conditional Mean Squared Error Expansion

Using the expression for the eigenvalues found above, we now present an expression for the empirical mean squared error of the nonparametric component, g, of the partially linear model. We let $G = (g(Z_1), ..., g(Z_n))'$ as above, and let 1_n be the indicator for the smallest eigenvalue of $P'P/n$ being greater than c_1 and the smallest eigenvalue of $I + \lambda(P'P/n)^{-1/2} D (P'P/n)^{-1/2}/n$ being greater than 1/2. Note that since $\mu_k \geq 0$ for all $k \geq 1$ (as discussed in the Appendix), $\mu_k^{n,K} = \mu_k + o_p(1) \geq o_p(1)$. So the eigenvalues of $I + \lambda(P'P/n)^{-1/2} D (P'P/n)^{-1/2}/n = I + \lambda UMU'/n$ are bounded below by $1 + o_p(1)$, since $\lambda/n \nrightarrow \infty$. Thus, $1_n \to_p 1$ as $n, K \to \infty$.

2.3.1 Expansion

Lemma 2.3.1: If $\sigma_\varepsilon(x,z)$ is bounded above and below away from zero, the conditional empirical mean squared error is

$$\begin{aligned}\frac{1}{n}\mathbb{E}[1_n(\hat{G}-G)'(\hat{G}-G)|\mathbf{X},\mathbf{Z}] \quad &\asymp \quad 1_n\left(\frac{1}{n}\sum_{i=1}^{K}\frac{1}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}+\frac{\lambda^2}{n}\sum_{i=1}^{K}\frac{\left(\frac{\mu_k^{n,K}}{n}\right)^2 b_i^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}+\right.\\ &\left.\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X},\mathbf{Z}]-g(Z_i))^2\right),\end{aligned}$$

where $\hat{g}_r$ is the spline estimate of g when $\lambda = 0$.

Proof: For ease of notation, assume throughout the proof that $1_n = 1$. We have

$$\begin{aligned}\mathbb{E}[(\hat{G}-G)'(\hat{G}-G)|\mathbf{X},\mathbf{Z}] \quad &\asymp \quad \mathbb{E}[(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])'(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])|\mathbf{X},\mathbf{Z}]+\\ &(\mathbb{E}[\hat{G}-\hat{G}_r|\mathbf{X},\mathbf{Z}])'(\mathbb{E}[\hat{G}-\hat{G}_r|\mathbf{X},\mathbf{Z}])+\\ &(\mathbb{E}[\hat{G}_r|\mathbf{X},\mathbf{Z}]-G)'(\mathbb{E}[\hat{G}_r|\mathbf{X},\mathbf{Z}]-G)\end{aligned}$$

Consider $\mathbb{E}[(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])'(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])|\mathbf{X},\mathbf{Z}]$. Let $B = P(P'P)^{-1/2}U$, then

$$\begin{aligned}\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}] \quad &= \quad P(P'P+\lambda D)^{-1}P'(Y-X\theta-G)\\ &= \quad P(P'P)^{-1/2}(I+\lambda UMU'/n)^{-1}(P'P)^{-1/2}P'(Y-X\theta-G)\\ &= \quad P(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}P'(Y-X\theta-G).\end{aligned}$$

So

$$\begin{aligned}
\mathbb{E}[(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])'(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])|\mathbf{X},\mathbf{Z}] &= \mathbb{E}[(Y-X\theta-G)'P(P'P)^{-1/2}U \times \\
&\quad (I+\lambda M/n)^{-1}U'(P'P)^{1/2}P'P \times \\
&\quad (P'P)^{1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2} \times \\
&\quad P'(Y-X\theta-G)|\mathbf{X},\mathbf{Z}] \\
&= \mathbb{E}[(Y-X\theta-g)'P(P'P)^{-1/2}U \times \\
&\quad (I+\lambda M/n)^{-2}U'(P'P)^{-1/2}P' \times \\
&\quad (Y-X\theta-G)|\mathbf{X},\mathbf{Z}]
\end{aligned}$$

Let $B = P(P'P)^{-1/2}U$ and $(a_1 \cdots a_K) \equiv (Y-X\theta-G)'B$, then

$$(Y-X\theta-G)'B(I+\lambda M/n)^{-2}B'(Y-X\theta-G) = \sum_{k=1}^{K} \frac{a_k^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}.$$

Also,

$$\begin{aligned}
\mathbb{E}[B'(Y-X\theta-G)(Y-X\theta-G)'B|\mathbf{X},\mathbf{Z}] &= \mathbb{E}[B'(Y-X\theta-G) \times \\
&\quad (Y-X\theta-G)'B|\mathbf{X},\mathbf{Z}] \\
&= B'\mathbb{E}[(Y-X\theta-G) \times \\
&\quad (Y-X\theta-G)'|\mathbf{X},\mathbf{Z}]B \\
&\leq B'(\sigma_\varepsilon^2 I)B \\
&= \sigma_\varepsilon^2 I,
\end{aligned}$$

and similarly $\mathbb{E}[B'(Y-X\theta-G)(Y-X\theta-G)'B|\mathbf{X},\mathbf{Z}] \geq \tilde{\sigma}_\varepsilon^2 I$. So since a_k^2 is the kth element along the diagonal of $B'(Y-X\theta-G)(Y-X\theta-G)'B$, we have $\tilde{\sigma}_\varepsilon^2 \leq \mathbb{E}[a_k^2|\mathbf{X},\mathbf{Z}] \leq \sigma_\varepsilon^2$. So

$$\mathbb{E}[(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])'(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])|\mathbf{X},\mathbf{Z}] \asymp \sum_{k=1}^{K} \frac{1}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}.$$

In the case of homoskedasticity,

$$\mathbb{E}[(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])'(\hat{G}-\mathbb{E}[\hat{G}|\mathbf{X},\mathbf{Z}])|\mathbf{X},\mathbf{Z}] = \sum_{k=1}^{K} \frac{\sigma_\varepsilon^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}.$$

Consider now $(\mathbb{E}[\hat{G}-\hat{G}_r|\mathbf{X},\mathbf{Z}])'(\mathbb{E}[\hat{G}-\hat{G}_r|\mathbf{X},\mathbf{Z}])$. Letting b_k be the kth component of $G'P(P'P)^{-1/2}U$, we have

$$\begin{aligned}
(\mathbb{E}[\hat{G}-\hat{G}_r|\mathbf{X},\mathbf{Z}])'(\mathbb{E}[\hat{G}-\hat{G}_r|\mathbf{X},\mathbf{Z}]) &= [P(P'P)^{-1}P'G - P(P'P+\lambda D)^{-1}P'G]' \times \\
&\quad [P(P'P)^{-1}P'G - P(P'P+\lambda D)^{-1}P'G] \\
&= [P(P'P)^{-1/2}U(I-(I+\lambda M/n)^{-1}) \times \\
&\quad U'(P'P)^{-1/2}P'G]' \times \\
&\quad [P(P'P)^{-1/2}U(I-(I+\lambda M/n)^{-1}) \times \\
&\quad U'(P'P)^{-1/2}P'G] \\
&= G'P(P'P)^{-1/2}U(I-(I+\lambda M/n)^{-1}) \times \\
&\quad U'(P'P)^{-1/2}P'P(P'P)^{-1/2}U \times \\
&\quad (I-(I+\lambda M/n)^{-1})U'(P'P)^{-1/2}P'G \\
&= G'P(P'P)^{-1/2}U(I-(I+\lambda M/n)^{-1})^2 \times \\
&\quad U'(P'P)^{-1/2}P'G \\
&= \sum_{k=1}^{K} b_k^2 \left(\frac{\frac{\lambda}{n}\mu_k^{n,K}}{1+\frac{\lambda}{n}\mu_k^{n,K}}\right)^2.
\end{aligned}$$

Finally, we see that $(\mathbb{E}[\hat{G}_r|\mathbf{X},\mathbf{Z}]-G)'(\mathbb{E}[\hat{G}_r|\mathbf{X},\mathbf{Z}]-G) = \sum_{i=1}^{n}(g(Z_i)-\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X},\mathbf{Z}])^2$. So

$$\begin{aligned}
\frac{1}{n}\mathbb{E}[1_n(\hat{G}-G)'(\hat{G}-G)|\mathbf{X},\mathbf{Z}] &\asymp 1_n \left(\frac{1}{n}\sum_{k=1}^{K}\frac{1}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2} + \frac{\lambda^2}{n}\sum_{k=1}^{K}\frac{\left(\frac{\mu_k^{n,K}}{n}\right)^2 b_k^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2} + \right. \\
&\quad \left. \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X},\mathbf{Z}]-g(Z_i))^2\right),
\end{aligned}$$

as desired.

2.3.2 Simplication of the Variance Term

Now as shown in Section 2, $\mu_k^{n,K} = \mu_k + o_p(1) = (\varsigma + o_k(1))k^{2m/d} + o_p(1)$, where $o_k(1)$ denotes a term that approaches zero as $k \to \infty$. Then for some sequences δ_k and $\delta_{n,K}$ such that $\delta_k \to 0$ as $k \to \infty$ and $\delta_{n,K} \to_p 0$ as $n, K \to \infty$, we have $\mu_k^{n,K} + o_p(1) = (\varsigma + \delta_k)k^{2m/d} + \delta_{n,K}$. We now show that $\mu_k + o_p(1)$ in the above expression can be replaced with its leading term. Lemma 2.3.2 deals with the first term of the mean-square error, and Lemma 2.3.3 deals with the second.

Lemma 2.3.2: The first term in the conditional mean squared error expansion can be rewritten as

$$\sum_{k=1}^{K} \frac{1}{(1 + \frac{\lambda}{n}\mu_k^{n,K})^2} = \sum_{k=1}^{K} \frac{1 + \epsilon_{n,K,\lambda,k}}{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2} \asymp \sum_{k=1}^{K} \frac{1 + \epsilon_{n,K,\lambda,k}}{(1 + \frac{\varsigma\lambda}{n}k^{2m/d})^2},$$

where $\epsilon_{n,K,\lambda,k} \to_p 0$ as $n, K \to \infty$.

Proof: We have

$$\begin{aligned}
\sum_{k=1}^{K} \frac{1}{(1 + \frac{\lambda}{n}\mu_k^{n,K})^2} &= \sum_{k=1}^{K} \frac{1}{\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d} + \delta_{n,K}\right)^2} \\
&= \sum_{k=1}^{K} \frac{1}{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2} \frac{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2}{\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d} + \delta_{n,K})\right)^2} \\
&= \sum_{k=1}^{K} \frac{1}{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2} \times \\
&\quad \frac{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2}{\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d}\right)^2 + 2\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d}\right)\frac{\lambda}{n}\delta_{n,K} + \left(\frac{\lambda}{n}\delta_{n,K}\right)^2} \\
&= \sum_{k=1}^{K} \frac{1}{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2} [1- \\
&\quad \left.\frac{2\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d}\right)\frac{\lambda}{n}\delta_{n,K} + \left(\frac{\lambda}{n}\delta_{n,K}\right)^2}{\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d}\right)^2 + 2\left(1 + \frac{\lambda}{n}((\varsigma + \delta_k)k^{2m/d}\right)\frac{\lambda}{n}\delta_{n,K} + \left(\frac{\lambda}{n}\delta_{n,K}\right)^2}\right] \\
&= \sum_{k=1}^{K} \frac{1 + \epsilon_{n,K,\lambda,k}}{(1 + (\varsigma + \delta_k)\frac{\lambda}{n}k^{2m/d})^2},
\end{aligned}$$

where

$$\epsilon_{n,K,\lambda,k} = -\frac{2\left(1+\frac{\lambda}{n}((\varsigma+\delta_k)k^{2m/d}\right)\frac{\lambda}{n}\delta_{n,K}+\left(\frac{\lambda}{n}\delta_{n,K}\right)^2}{\left(1+\frac{\lambda}{n}((\varsigma+\delta_k)k^{2m/d}\right)^2+2\left(1+\frac{\lambda}{n}((\varsigma+\delta_k)k^{2m/d}\right)\frac{\lambda}{n}\delta_{n,K}+\left(\frac{\lambda}{n}\delta_{n,K}\right)^2}.$$

Note that $\epsilon_{n,K,\lambda,k} \to_p 0$ for all $k = 1, ..., K$, as $n, K \to \infty$, since

$$1 + (\varsigma + \delta_k)\lambda k^{2m/d}/n \gtrsim \lambda\delta_{n,K}/n.$$

Similarly, to deal with δ_k,

$$\begin{aligned}
\sum_{k=1}^{K} \frac{1+\epsilon_{n,K,\lambda,k}}{(1+(\varsigma+\delta_k)\frac{\lambda}{n}k^{2m/d})^2} &= \sum_{k=1}^{K} \frac{1+\epsilon_{n,K,\lambda,k}}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2}\frac{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2}{(1+(\varsigma+\delta_k)\frac{\lambda}{n}k^{2m/d})^2} \\
&= \sum_{k=1}^{K} \frac{1+\epsilon_{n,K,\lambda,k}}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2}\Bigg[1- \\
&\quad \left.\frac{2(1+\frac{\varsigma\lambda}{n}k^{2m/d})\delta_k\frac{\lambda}{n}k^{2m/d}+(\delta_k\frac{\lambda}{n}k^{2m/d})^2}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2+2(1+\frac{\varsigma\lambda}{n}k^{2m/d})\delta_k\frac{\lambda}{n}k^{2m/d}+(\delta_k\frac{\lambda}{n}k^{2m/d})^2}\right] \\
&\asymp \sum_{k=1}^{K} \frac{1+\epsilon_{n,K,\lambda,k}}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2},
\end{aligned}$$

since

$$\epsilon_k \equiv \frac{2(1+\varsigma\frac{\lambda}{n}k^{2m/d})\delta_k\frac{\lambda}{n}k^{2m/d}+(\delta_k\frac{\lambda}{n}k^{2m/d})^2}{(1+\varsigma\frac{\lambda}{n}k^{2m/d})^2+2(1+\varsigma\frac{\lambda}{n}k^{2m/d})\delta_k\frac{\lambda}{n}k^{2m/d}+(\delta_k\frac{\lambda}{n}k^{2m/d})^2} \to 0$$

as $k \to \infty$ (since $1 + \varsigma\lambda k^{2m/d}/n > \delta_k\lambda k^{2m/d}/n$ for large k).

2.3.3 Simplification of the Penalization Bias Term

Lemma 2.3.3: The second term in the conditional mean squared error expansion can be rewritten as

$$\begin{aligned}
\sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)^2 b_k^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2} &= \sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{\varsigma+\delta_k}{n}k^{2m/d}\right)b_k^2(1+\epsilon'_{n,K,\lambda,k})}{(1+(\varsigma+\delta_k)\frac{\lambda}{n}k^{2m/d})^2} \\
&\asymp \sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{\varsigma}{n}k^{2m/d}\right)b_k^2(1+\epsilon'_{n,K,\lambda,k})}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2},
\end{aligned}$$

where $\epsilon'_{n,K,\lambda,k} \to_p 0$ as $n, K \to \infty$.

Proof: We have

$$\sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)^2 b_k^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2} = \sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{(\varsigma+\delta_k)k^{2m/d}+\delta_{n,K}}{n}\right) b_k^2(1+\epsilon_{n,K,\lambda,k})}{(1+(\varsigma+\delta_k)\frac{\lambda}{n}k^{2m/d})^2}$$
$$= \sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{\varsigma+\delta_k}{n}k^{2m/d}\right) b_k^2(1+\epsilon'_{n,K,\lambda,k})}{(1+(\varsigma+\delta_k)\frac{\lambda}{n}k^{2m/d})^2},$$

where $\epsilon'_{n,K,\lambda,k} = \epsilon_{n,K,\lambda,k} + \dfrac{\delta_{n,K}/n}{(\varsigma+\delta_k)k^{2m/d}/n}(1+\epsilon_{n,K,\lambda,k}) \to_p 0$, since $\epsilon_{n,K,\lambda,k} \to_p 0$ and $\delta_{n,K} \to_p 0$ as $n, K \to \infty$, and the first equality follow from Lemma 2.3.2. Also as in Lemma 2.3.2,

$$\sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{\varsigma+\delta_k}{n}k^{2m/d}\right) b_k^2(1+\epsilon'_{n,K,\lambda,k})}{(1+(\varsigma+\delta_k)\frac{\lambda}{n}k^{2m/d})^2} = \sum_{k=1}^{K} \frac{\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{\varsigma+\delta_k}{n}k^{2m/d}\right) b_k^2(1+\epsilon'_{n,K,\lambda,k})(1+\epsilon_k)}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2}$$
$$\asymp \sum_{k=1}^{K} \frac{\left(\frac{\varsigma}{n}k^{2m/d}\right)^2 b_k^2(1+\epsilon'_{n,K,\lambda,k})}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2},$$

since $\delta_k < \varsigma$ for large k.

2.4 Rate of the Empirical Mean Squared Error

2.4.1 Without Derivatives

We now find the empirical mean squared rate of convergence of $\hat{g}$, given the expressions found in Section 3. We first note that $\mathcal{C}_n(m) = (\varsigma\lambda/n)^{d/2m}(K-m) \to \infty$ if and only if $(\varsigma\lambda/n)^{d/2m}K \to \infty$. This case is equivalent to smoothing splines if $K = O(n)$, and we find that the rate of the mean squared error is equivalent to that for smoothing splines. When $\mathcal{C}_n(m) \to \infty$ but $K < O(n)$, the framework could be considered "almost" smoothing splines, and the mean-square rate has an additional approximation bias term of order $K^{-r_g/d}$. If K is chosen sufficiently large, then this bias will be dominated by the bias resulting from the penalization, showing that the estimation procedure is asymptotically equivalent to smoothing spline estimation.

Similarly, $\mathcal{C}_n(m) \to 0$ if and only if $(\varsigma\lambda/n)^{d/2m}K \to 0$, since $K - m = O(K)$. This case is equivalent to regression splines if $\lambda = 0$. If $\mathcal{C}_n(m) \to 0$ but $\lambda > 0$, there is an additional bias term of order $\lambda K^{m/d}/n$ resulting from the penalization (producing less fidelity to the data). If λ is chosen sufficiently small so that the this penalization bias is smaller order than the approximation bias, then this framework is asymptotically equivalent to regression splines.

Finally, if $\mathcal{C}_n(m) \to c$ for some constant c, then since $\mathcal{C}_n(m) = (\varsigma\lambda/n)^{d/2m}K - (\varsigma\lambda/n)^{d/2m}m$ and $K \to \infty$, it must be that $(\varsigma\lambda/n)^{d/2m}K \to c$ and $(\varsigma\lambda/n)^{d/2m}m \to 0$; and if $(\varsigma\lambda/n)^{d/2m}K \to c$ for some constant c, then since $K \to \infty$, we have $(\varsigma\lambda/n)^{d/2m}m \to 0$, so $\mathcal{C}_n(m) \to c$. So $\lim_{n\to\infty}(\varsigma\lambda/n)^{d/2m}K = \lim_{n\to\infty}\mathcal{C}_n(m)$. This case is neither asymptotically regression spline estimation nor smoothing spline estimation, and most researchers in penalized spline estimation are particularly interested in this case. The situation $c = 1$ can be considered the "knife-edge" case, in that this is when the form for the rates of convergence switches between that for regression splines and that for smoothing splines.

For ease of notation, define $\eta_k \equiv \varsigma k^{2m/d}$. We first present a proof in the case $|\ell| = 0$.

Theorem 1 - empirical norm, $|\ell| = \mathbf{0}$ (restated): If $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then

$$\|\hat{g} - g\|^2_{2,0,n} = O_p\left(\frac{K}{n} + \frac{\lambda^2}{n^2}K^{2m/d} + K^{-2r_g/d}\right),$$

and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then

$$\|\hat{g} - g\|^2_{2,0,n} = O_p\left(\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}} + \frac{\lambda}{n} + K^{-2r_g/d}\right),$$

Proof: From Lemmas 2.3.1, 2.3.2, and 2.3.3,

$$\begin{aligned}\frac{1}{n}\mathbb{E}[1_n(\hat{G}-G)'(\hat{G}-G)|\mathbf{X},\mathbf{Z}] &\asymp 1_n\left[\frac{1}{n}\sum_{k=1}^{K}\frac{1}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}+\frac{\lambda^2}{n}\sum_{k=1}^{K}\frac{\left(\frac{\mu_k^{n,K}}{n}\right)^2 b_k^2}{(1+\frac{\lambda}{n}\mu_k^{n,K})^2}+\right.\\ &\qquad\left.\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X},\mathbf{Z}]-g(Z_i))^2\right]\\ &\asymp 1_n\left[\frac{1}{n}\sum_{k=1}^{K}\frac{1+\epsilon_{n,K,\lambda,k}}{(1+\frac{\lambda}{n}\eta_k)^2}+\right.\\ &\qquad\frac{\lambda^2}{n}\sum_{k=1}^{K}\frac{(\frac{\mu_k^{n,K}}{n})(\frac{\eta_k}{n})b_k^2(1+\epsilon'_{n,K,\lambda,k})}{(1+\frac{\lambda}{n}\eta_k)^2}+\\ &\qquad\left.\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X},\mathbf{Z}]-g(Z_i))^2\right].\end{aligned}$$

Define $1_{n,K}=1\{1_n=1,|\epsilon_{n,K,\lambda,k}|<0.1,|\epsilon'_{n,K,\lambda,k}|<0.1\}\to_p 1$ as $n,K\to\infty$. Then

$$\begin{aligned}\frac{1}{n}\mathbb{E}[1_{n,K}(\hat{G}-G)'(\hat{G}-G)|\mathbf{X},\mathbf{Z}] &\asymp 1_{n,K}\left(\frac{1}{n}\sum_{k=1}^{K}\frac{1}{(1+\frac{\lambda}{n}\eta_k)^2}+\frac{\lambda^2}{n}\sum_{k=1}^{K}\frac{(\frac{\mu_k^{n,K}}{n})(\frac{\eta_k}{n})b_k^2}{(1+\frac{\lambda}{n}\eta_k)^2}+\right.\\ &\qquad\left.\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X},\mathbf{Z}]-g(Z_i))^2\right).\end{aligned}$$

For ease of notation, assume throughout the rest of the proof that $1_{n,K}=1$. Consider first $\mathcal{C}_n(m)<1$ for all sufficiently large n. Since $\lambda\eta_k/n>0$ for $k=1,...,K$, $\frac{1}{n}\sum_{k=1}^{K}\frac{1}{(1+\frac{\lambda}{n}\eta_k)^2}\leq\frac{K}{n}$.

Consider the second term. Since $1+\lambda\eta_k/n\geq 1$,

$$\frac{\lambda^2}{n}\sum_{k=1}^{K}\frac{(\frac{\mu_k^{n,K}}{n})(\frac{\eta_k}{n})b_k^2}{(1+\frac{\lambda}{n}\eta_k)^2}\leq\frac{\lambda^2}{n}\sum_{k=1}^{K}\left(\frac{\mu_k^{n,K}}{n}\right)\left(\frac{\eta_k}{n}\right)b_k^2\asymp\frac{\lambda^2}{n^2}K^{2m/d}\sum_{k=1}^{K}\frac{\mu_k^{n,K}}{n}b_k^2.$$

Now

$$\sum_{k=1}^{K} \frac{\mu_k^{n,K}}{n} b_k^2 = g'P(P'P)^{-1/2}UMU'(P'P)^{-1/2}P'g/n = g'P(P'P)^{-1}D(P'P)^{-1}P'g.$$

Define β_K such that $s_g = \beta_K' p^K$, where $s_g = \inf_{s \in \mathcal{S}_{n,r}} \sup_{z \in [0,1]^d} |s(z) - g(z)|$, and $S_g = (s_g(Z_1), ..., s_g(Z_n))$. Then adding and subtracting s_g from g, we have

$$\begin{aligned}
\sum_{k=1}^{K} \frac{\mu_k^{n,K}}{n} b_k^2 &= (G - S_g)'P(P'P)^{-1}D(P'P)^{-1}P'(G - S_g) + \\
&\quad 2S_g'P(P'P)^{-1}D(P'P)^{-1}P'(G - S_g) + S_g'P(P'P)^{-1}D(P'P)^{-1}P'S_g \\
&\asymp (G - S_g)'P(P'P)^{-1}D(P'P)^{-1}P'(G - S_g) + \\
&\quad S_g'P(P'P)^{-1}D(P'P)^{-1}P'S_g \\
&= \beta_K' D \beta_K + (G - S_g)'P(P'P)^{-1}D(P'P)^{-1}P'(G - S_g).
\end{aligned}$$

using $S_g' = \beta_K' P'$. Since the number of observations in any hyper-interval is $\asymp n/K$, for all $k = 1, ..., K$,
$\sum_{i=1}^{n} p_k(Z_i) \lesssim \sqrt{K} \cdot \frac{n}{K} = \frac{n}{\sqrt{K}}$. So

$$\begin{aligned}
(G - S_g)'P(P'P)^{-1}D(P'P)^{-1}P'(G - S_g) &= (G - S_g)'P(P'P/n)^{-1/2}(P'P/n)^{-1/2} \times \\
&\quad D(P'P/n)^{-1/2}(P'P/n)^{-1/2}P'(G - S_g)/n^2 \\
&\leq \lambda_{\min}^{-1}(P'P/n)(G - S_g)'PUMU'P' \times \\
&\quad (G - S_g)/n^2 \\
&\asymp \frac{1}{n^2}\lambda_{\max}(M) \times \\
&\quad \sum_{k=1}^{K} \left(\sum_{i=1}^{n} (g(Z_i) - s_g(Z_i)) p_k(Z_i) \right)^2 \\
&\lesssim \frac{1}{n^2} K^{2m/d} \sup_{z \in [0,1]^d} (g(z) - s_g(z))^2 \cdot K \cdot \frac{n^2}{K} \\
&\lesssim K^{2(m - r_g)/d} \\
&= O(1)
\end{aligned}$$

since $m \leq r_g$. Also, $\beta_K' D \beta_K$ is the penalty term in our criterion function, which we assumed was bounded. So $\sum_{k=1}^{K} \frac{\mu_k^{n,K}}{n} b_k^2$ is bounded, and thus

$$\frac{\lambda^2}{n} \sum_{k=1}^{K} \frac{(\frac{\mu_k^{n,K}}{n})(\frac{\eta_k}{n}) b_k^2}{(1 + \frac{\lambda}{n}\mu_k^{n,K})^2} \lesssim \frac{\lambda^2}{n^2} K^{2m/d}.$$

Finally, consider the last term, which is $\frac{1}{n}\sum_{i=1}^{n} (\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X}, \mathbf{Z}] - g(Z_i))^2$. We have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} (\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X}, \mathbf{Z}] - g(Z_i))^2 &= \frac{1}{n}\sum_{i=1}^{n} (\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X}, \mathbf{Z}] - s_g(Z_i))^2 + \\ &\quad \frac{1}{n}\sum_{i=1}^{n} (g(Z_i) - s_g(Z_i))^2. \end{aligned}$$

Note that since $s_g \in \mathcal{S}_{n,r}$, we have $s_g = p_K'(P'P)^{-1}P'S_g$. Then as above,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} (\mathbb{E}[\hat{g}_r(Z_i)|\mathbf{X}, \mathbf{Z}] - s_g(Z_i))^2 &= (G - S_g)'P(P'P)^{-1}P'P(P'P)^{-1}P'(G - S_g)/n \\ &\leq \lambda_{\min}^{-1}(P'P/n)(G - S_g)'PP'(G - S_g)/n^2 \\ &\asymp \frac{1}{n^2}\sum_{k=1}^{K}\left(\sum_{i=1}^{n}(g(Z_i) - s_g(Z_i))^2 p_k(Z_i)\right)^2 \\ &\asymp K^{-2r_g/d}. \end{aligned}$$

Also, $\frac{1}{n}\sum_{k=1}^{n}(g(Z_i) - s_g(Z_i))^2 \leq \sup_{z \in [0,1]^d} |g(z) - s_g(z)|^2 \lesssim K^{-2r_g/d}$, by Schumaker (1981) (see also Newey (1997)). So

$$\frac{1}{n}\mathbb{E}[1_{n,K}(\hat{G} - G)'(\hat{G} - G)|\mathbf{X}, \mathbf{Z}] \lesssim \frac{K}{n} + \frac{\lambda^2}{n^2}K^{2m/d} + K^{-2r_g/d}.$$

Then by Markov's inequality and the fact that $1_{n,K} \to_p 1$,

$$\|\hat{g} - g\|_{0,2,n}^2 = O_p\left(\frac{K}{n} + \frac{\lambda^2}{n^2}K^{2m/d} + K^{-2r_g/d}\right).$$

Consider now $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n. Letting r_m be the remainder term from the Euler-Maclaurin formula,

$$\begin{aligned}
\sum_{k=1}^{K} \frac{1}{(1+\frac{\lambda}{n}\eta_k)^2} &= \sum_{k=1}^{K} \frac{1}{(1+\frac{\varsigma\lambda}{n}k^{2m/d})^2} \\
&= \int_0^K \frac{dx}{(1+\frac{\varsigma\lambda}{n}x^{2m/d})^2} + r_m \\
&= \left(\frac{\varsigma\lambda}{n}\right)^{-d/2m} \int_0^{(\varsigma\lambda/n)^{d/2m}K} \frac{du}{(1+u^{2m/d})^2} + r_m \\
&\lesssim \left(\frac{\lambda}{n}\right)^{-d/2m},
\end{aligned}$$

since the integral is finite for $m > d/4$, even if $\mathcal{C}_n(m) = \infty$, and where we use the substitution $u = (\varsigma\lambda/n)^{d/2m}x$.

Consider now the second term. Since $x(1+x)^{-2} \leq \frac{1}{4}$ for $x \geq 1$,

$$\frac{\lambda}{n}\sum_{k=1}^{K} \frac{(\frac{\mu_k^{n,K}}{n})(\frac{\lambda\eta_k}{n})b_k^2}{(1+\frac{\lambda}{n}\eta_k)^2} \leq \frac{\lambda}{4n}\sum_{k=1}^{K} \frac{\mu_k^{n,K}}{n}b_k^2 \lesssim \frac{\lambda}{n}.$$

So using Markov's inequality and the fact that $1_{n,K} \to_p 1$,

$$\|\hat{g} - g\|_{0,2,n}^2 = O_p\left(\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}} + \frac{\lambda}{n} + K^{-2r_g/d}\right),$$

as desired.

2.4.2 With Derivatives

We now consider the mean squared rate of convergence in estimating the derivatives of g. Let $\ell = (\ell_1, ..., \ell_d)$ be a vector of nonnegative integers, and let $\partial^{\ell}(z) \equiv \partial^{(\ell)}h(z)/\partial x_1^{\ell_1} \cdots \partial x_d^{\ell_d}$ with $|\ell| = \sum_{j=1}^{d} \ell_j$. We consider derivatives up to order $r-2$ in any one direction, where r is again the order of the B-splines. This includes the popular case of cubic B-splines with a first-order derivative in any direction.

We first present a lemma giving the best L_∞ approximation rate to derivatives of g over $\mathcal{S}_{n,r}$. This sort of result is available in the literature in the univariate case (see Zhou & Wolfe (2000) and Newey (1997)), but we are unaware of a similar result in the multivariate case.

Lemma 2.4.2: Given $\ell = (\ell_1, ..., \ell_d)$, there exists $\bar{s}_g \in \mathcal{S}_{n,r}$ such that

$$\sup_{z \in [0,1]^d} |\partial^\ell g(z) - \partial^\ell \bar{s}_g(z)| = O(K^{-(r_g - |\ell|)/d}),$$

where $\partial^0 g(z) \equiv g(z)$, as usual.

Proof: Let $\Sigma_{\partial^\ell g}$ be defined by $\partial^\ell \Sigma_{\partial^\ell g} = s_{\partial^\ell g}$, where $s_{\partial^\ell g}$ is the best L_∞ approximation to $\partial^\ell g$, as above. Define

$$\int\limits_{\ell, z, t_{k_1,...,k_d}} h(\zeta) d\zeta \equiv \underbrace{\int\limits_{t_{1k_1}}^{z_1} \cdots \int\limits_{t_{1k_1}}^{z_1}}_{\ell_1} \cdots \underbrace{\int\limits_{t_{dk_d}}^{z_d} \cdots \int\limits_{t_{dk_d}}^{z_d}}_{\ell_d} \partial^\ell h(\zeta_1, ..., \zeta_d) \underbrace{d\zeta_1 \cdots d\zeta_1}_{\ell_1} \cdots \underbrace{d\zeta_d \cdots d\zeta_d}_{\ell_d}$$

for a function h and $f_{\ell,h}(z, t_{k_1,...,k_d}) = \int\limits_{\ell, x, t_{k_1,...,k_d}} \partial^\ell h(\zeta) d\zeta - h(z)$. Then over each hyper-interval $[t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{dk_d}, t_{d,k_d+1})$, $k_1, ..., k_d = 1, ..., K^{1/d}$, define

$$\begin{aligned} \bar{s}_g(z) &= \left[\int\limits_{\ell, z, t_{k_1,...,k_d}} s_{\partial^\ell g}(\zeta) d\zeta - f_{\ell, s_g}(z, t_{k_1,...,k_d}) \right] \times \\ &\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{dk_d}, t_{d,k_d+1})\} \\ &= [\Sigma_{\partial^\ell g}(z) + f_{\ell, \Sigma_{\partial^\ell g}}(z, t_{k_1}, ..., t_{k_d}) - f_{\ell, s_g}(z, t_{k_1,...,k_d})] \times \\ &\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{dk_d}, t_{d,k_d+1})\} \end{aligned}$$

Since each term in both $f_{\ell, \Sigma_{\partial^\ell g}}(z, t_{k_1,...,k_d})$ and $f_{\ell, s_g}(z, t_{k_1,...,k_d})$ is a function of at most $d-1$ elements of

$z = (z_1, ..., z_d)$, $\partial^\ell f_{\ell, \Sigma_{\partial^\ell g}} = 0$ and $\partial^\ell f_{\ell, s_g} = 0$, so $\partial^\ell \bar{s}_g = s_{\partial^\ell g}$. Thus,

$$\sup_{z \in [0,1]^d} |\partial^\ell g(z) - \partial^\ell \bar{s}(z)| = O(K^{-(r_g - \ell)/d}),$$

since the modulus of smoothness of $\partial^\ell g$ is $p - \ell$ (Newey (1997)).

Also, for all $z \in [0,1]^d$,

$$
\begin{aligned}
|g(z) - \bar{s}_g(z)| &= \Bigg| \left[\int\limits_{\ell,z,t_{k_1,\ldots,k_d}} \partial^\ell g(\zeta) d\zeta - f_{\ell,g}(z, t_{k_1,\ldots,k_d}) \right] \times \\
&\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{1k_1}, t_{1,k_1+1})\} - \\
&\quad \left[\int\limits_{\ell,z,t_{k_1,\ldots,k_d}} s_{\partial^\ell g}(\zeta) d\zeta - f_{\ell,s_g}(x, t_{k_1,\ldots,k_d}) \right] \times \\
&\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{1k_1}, t_{1,k_1+1})\} \Bigg| \\
&\leq \left[\int\limits_{\ell,z,t_{k_1,\ldots,k_d}} |\partial^\ell g(\zeta) - s_{\partial^\ell g}(\zeta)| d\zeta \right] \times \\
&\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{1k_1}, t_{1,k_1+1})\} + \\
&\quad |f_{\ell,g}(z, t_{k_1,\ldots,k_d}) - f_{\ell,s_g}(x, t_{k_1,\ldots,k_d})| \times \\
&\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{1k_1}, t_{1,k_1+1})\} \\
&\lesssim \sup_{z \in [0,1]^d} |\partial^\ell g(z) - s_{\partial^\ell g}(z)| \cdot (z_1 - t_{1,k_1})^{\ell_1} \cdots (z_d - t_{d,k_d})^{\ell_d} \times \\
&\quad 1\{z \in [t_{1k_1}, t_{1,k_1+1}) \times \cdots \times [t_{1k_1}, t_{1,k_1+1})\} + \\
&\quad \sup_{z \in [0,1]^d} |g(z) - s_g(z)| \\
&= O(K^{-(r_g - |\ell|)/d}) \cdot (K^{1/d})^{\ell_1 + \ldots + \ell_d} + O(K^{-r_g/d}) \\
&= O(K^{-r_g/d}),
\end{aligned}
$$

where the third-to-last line follows since each term in $f_{\ell,g}(z, t_{k_1,\ldots,k_d}) - f_{\ell,s_g}(z, t_{k_1,\ldots,k_d})$ is of the form $g(a_1, ..., a_d) - s_g(a_1, ..., a_d)$ where each a_j equals z_j or t_{jk_j}.

We now present the proof for the rate of a convergence for an estimate of $\partial^\ell \hat{g}$, in the empirical norm. In the proof, we find the rate of $\|\hat{\beta} - \bar{\beta}\|_2^2$ (where $\bar{s}_g = p^{K\prime}\bar{\beta}$) from the rate of $\|\hat{g} - g\|_{2,0,n}^2$, which is key to the result. This shows, as is intuitive, that the rates are governed by the rate of approximation of the estimated coefficients on the spline basis functions compared to the best L_∞ coefficients.

Theorem 1 - empirical norm, $\ell > 0$ (restated): If $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then

$$\|\hat{g}-g\|_{2,\ell,n}^2 = O_p\left(K^{2\ell/d}\left(\frac{K}{n}+\frac{\lambda^2}{n^2}K^{2m/d}+K^{-2r_g/d}\right)\right),$$

and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then

$$\|\hat{g}-g\|_{2,\ell,n}^2 = O_p\left(K^{2\ell/d}\left(\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}}+\frac{\lambda}{n}+K^{-2r_g/d}\right)\right),$$

Proof: Let $1_{n,K}=1$ throughout the proof, and define $D_j^{(\ell_j)} = M_{j,1}'M_{j,2}'\cdots M_{j,\ell_j}'$ with

$$M_{j,\eta} = (r-\eta)\begin{pmatrix} \frac{-1}{t_{j,r}-t_{j,\eta}} & 0 & 0 & \cdots & 0 \\ \frac{1}{t_{j,r}-t_{j,\eta}} & \frac{-1}{t_{j,r+1}-t_{j,1+\eta}} & 0 & \cdots & 0 \\ 0 & \frac{1}{t_{j,r+1}-t_{j,1+\eta}} & \frac{-1}{t_{j,r+2}-t_{j,2+\eta}} & \cdots & 0 \\ 0 & 0 & \frac{1}{t_{j,r+2}-t_{j,2+\eta}} & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \frac{-1}{t_{j,K^{1/d}+r-1-\eta}-t_{j,K^{1/d}-1}} \\ 0 & 0 & 0 & \cdots & \frac{1}{t_{j,K^{1/d}+r-1-\eta}-t_{j,K^{1/d}-1}} \end{pmatrix}$$

for $\eta = 1,...,\ell_j$, where t_{jk} is the kth knot in direction j. As shown in Zhou & Wolfe (2000) using De Boor (2001), when $d=1$,

$$\hat{g}^{(\ell)}(z) = (p^{K\prime}(z)\hat{\beta})^{(\ell)} = p_{-\ell}^{K\,\prime}(z)D^{(\ell)}\hat{\beta},$$

where $p_{-\ell}^K$ is the vector of spline basis functions of order $r-\ell$. Define $p_{j,-\ell_j}^K$ to be the vector of basis functions in direction j of degree $r-\ell_j$, and let $D_j^{(\ell_j)}$ be the matrix $D^{(\ell_j)}$ using the knots in direction j. Also, let $P_{-\ell}$ be the spline design matrix using splines of order $r-\ell_j$ in direction j and $D^\ell = D_1^{(\ell_1)}\otimes\cdots\otimes D_d^{(\ell_d)}$. Finally, let $\hat{\beta}-\bar{\beta} = (\hat{\alpha}_1-\bar{\alpha}_1)\otimes\cdots\otimes(\hat{\alpha}_d-\bar{\alpha}_d)$, where each vector $\hat{\alpha}_j-\bar{\alpha}_j$, $j=1,...,d$, is chosen appropriately, and let $\hat{\alpha}_{jk}-\bar{\alpha}_{jk}$ be the k

component of $\hat{\alpha}_j - \bar{\alpha}_j$. Then

$$\begin{aligned}
\partial^\ell \hat{g}(z) - \partial^\ell \bar{s}_g(z) &= \partial^\ell \left(\sum_{k=1}^{K} p_k(z)(\hat{\beta}_k - \bar{\beta}_k) \right) \\
&= \partial^\ell \left(\sum_{k_1,\ldots,k_d=1}^{K^{1/d}} p_{1k_1}(z_1) \cdots p_{dk_d}(z_d)(\hat{\alpha}_{1k_1} - \bar{\alpha}_{1k_1}) \cdots (\hat{\alpha}_{dk_d} - \bar{\alpha}_{dk_d}) \right) \\
&= \partial^\ell \left(\prod_{j=1}^{d} \sum_{k_j=1}^{K^{1/d}} p_{jk_j}(z_j)(\hat{\alpha}_{jk_j} - \bar{\alpha}_{jk_j}) \right) \\
&= \partial^\ell \left(\prod_{j=1}^{d} p_j^{K\prime}(z_j)(\hat{\alpha}_j - \bar{\alpha}_j) \right) \\
&= \prod_{j=1}^{d} p_{j,-\ell_j}^{K}{}'(z_j) D_j^{(\ell_j)}(\hat{\alpha}_j - \bar{\alpha}_j) \\
&= \left[p_{1,\ell_1}^{K}{}'(z_j) \otimes \cdots \otimes p_{d,\ell_d}^{K}{}'(z_j) \right] \left[D_1^{(\ell_1)} \otimes \cdots \otimes D_d^{(\ell_1)} \right] \times \\
&\quad [(\hat{\alpha}_1 - \bar{\alpha}_1) \otimes \cdots \otimes (\hat{\alpha}_d - \bar{\alpha}_d)] \\
&= p_{-\ell}^{K}(z)' D^{(\ell)}(\hat{\beta} - \bar{\beta})
\end{aligned}$$

So $\partial^\ell \hat{G} - \partial^\ell \bar{S}_g = P_{-\ell} D^{(\ell)}(\hat{\beta} - \bar{\beta})$.

Now let $r_{n,K} = K/n + \lambda^2 K^{2m/d}/n^2 + K^{-2r_g/d}$ for $\mathcal{C}_n(m) < 1$ and $r_{n,K} = n^{(d-2m)/2m}/\lambda^{d/2m} + \lambda/n + K^{-2r_g/d}$ for $\mathcal{C}_n(m) \geq 1$. Since $\bar{s}_g$ achieves the optimal rate of approximation for g, using the results for $|\ell| = 0$,

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{g}(Z_i) - \bar{s}_g(Z_i))^2 \asymp \frac{1}{n}\sum_{i=1}^{n}(\hat{g}(Z_i) - g(Z_i))^2 + \frac{1}{n}\sum_{i=1}^{n}(\bar{s}_g(Z_i) - g(Z_i))^2 = O_p(r_{n,K}).$$

Also,

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{g}(Z_i) - \bar{s}_g(Z_i))^2 = (\beta - \bar{\beta})' P_{-\ell}' P_{-\ell}(\hat{\beta} - \bar{\beta})/n \gtrsim (\hat{\beta} - \bar{\beta})'(\hat{\beta} - \bar{\beta}).$$

So $\|\hat{\beta} - \bar{\beta}\|_2^2 \equiv (\hat{\beta} - \bar{\beta})'(\hat{\beta} - \bar{\beta}) = O_p(r_{n,K})$. Then using the structure of $D^{(\ell)}$ and the fact that its maximum element is $\|D^{(\ell)}\|_\infty = O(K^{\ell/d})$,

$$\begin{aligned}(\hat{G}^{(\ell)} - \bar{S}_g^{(\ell)})'(\hat{G}^{(\ell)} - \bar{S}_g^{(\ell)})/n &= (\hat{\beta} - \bar{\beta})' D^{(\ell)\prime} P' P D^{(\ell)} (\hat{\beta} - \bar{\beta})/n \\ &\leq \lambda_{\max}(P'P/n)(\hat{\beta} - \bar{\beta})' D^{(\ell)\prime} D^{(\ell)} (\hat{\beta} - \bar{\beta}) \\ &\lesssim \|D^{(\ell)}\|_\infty^2 \|\hat{\beta} - \bar{\beta}\|^2 \\ &= O_p(K^{2\ell/d} r_{n,K}).\end{aligned}$$

Thus,

$$\|\hat{g} - g\|_{2,\ell,n}^2 \asymp \|\hat{g} - \bar{s}_g\|_{2,\ell,n}^2 + \|\bar{s}_g - g\|_{2,\ell,n}^2 = O_p(K^{2\ell/d} r_{n,K}),$$

as desired.

2.5 Rate of the Fixed Mean Squared Error

We now consider the rate of convergence in the fixed norm, in which the average is taken over the full distribution instead of the specific data set. This norm is more common and is used in particular in Newey (1997). We first find the rate of $\|D^{(\ell)}(\hat{\beta} - \bar{\beta})\|_2^2$ from the rate in the empirical norm, which allows the rate in the fixed norm to follow easily.

Theorem 1 - fixed norm (restated): If $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then

$$\|\hat{g} - g\|_{2,\ell}^2 = O_p\left(K^{2\ell/d}\left(\frac{K}{n} + \frac{\lambda^2}{n^2} K^{2m/d} + K^{-2r_g/d}\right)\right),$$

and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then

$$\|\hat{g} - g\|_{2,\ell}^2 = O_p\left(K^{2\ell/d}\left(\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}} + \frac{\lambda}{n} + K^{-2r_g/d}\right)\right),$$

Proof: For notational convenience, define $D^{(0)} \equiv I$. Similarly to above, since $\partial^\ell \bar{s}_g$ achieves the optimal rate of approximation for $\partial^\ell g$, using the results for the empirical norm,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(\partial^\ell \hat{g}(Z_i) - \partial^\ell \bar{s}_g(Z_i))^2 &\asymp \frac{1}{n}\sum_{i=1}^{n}(\partial^\ell \hat{g}(Z_i) - \partial^\ell g(Z_i))^2 + \\
&\quad \frac{1}{n}\sum_{i=1}^{n}(\partial^\ell \bar{s}_g(Z_i) - \partial^\ell g(Z_i))^2 \\
&= O_p(K^{2\ell/d} r_{n,K}).
\end{aligned}$$

Also,

$$\frac{1}{n}\sum_{i=1}^{n}(\partial^\ell \hat{g}(Z_i) - \partial^\ell \bar{s}_g(Z_i))^2 = (\beta - \bar{\beta})' D^{\ell\prime} P_{-\ell}' P_{-\ell} D^\ell (\hat{\beta} - \bar{\beta})/n \gtrsim (\hat{\beta} - \bar{\beta})' D^{\ell\prime} D^\ell (\hat{\beta} - \bar{\beta}).$$

So $\|D^{(\ell)}(\hat{\beta} - \bar{\beta})\|^2 = O_p(K^{2\ell/d} r_{n,K})$. So

$$\begin{aligned}
\|\hat{g} - g\|_{\ell,2}^2 &= \int (\partial^\ell \hat{g}(z) - \partial^\ell g(z))^2 dF_0(z) \\
&= \int (p^K(z)' D^{(\ell)}(\hat{\beta} - \bar{\beta}) + p^K(z)' D^{(\ell)} \bar{\beta} - \partial^\ell g(z))^2 dF_0(z) \\
&\asymp (\hat{\beta} - \bar{\beta})' D^{(\ell)\prime} \int p^K(z) p^K(z)' dF_0(z) D^{(\ell)}(\hat{\beta} - \bar{\beta}) + \\
&\quad \int (\partial^\ell \bar{s}_g(z) - \partial^\ell g(z))^2 dF_0(z) \\
&\leq \lambda_{\max}(\mathbb{E}[p^K(z) p^K(z)']) \|D^{(\ell)}(\hat{\beta} - \bar{\beta})\|_2^2 + \sup_{z \in [0,1]^d} (\partial^\ell \bar{s}_g(z) - \partial^\ell g(z))^2 \\
&= O_p(K^{2\ell/d} r_{n,K}),
\end{aligned}$$

giving the result.

CHAPTER III

Asymptotic Distribution of the Parametric and Nonparametric Components

3.1 Asymptotic Distribution of the Parametric Component

We now consider the asymptotic distribution of $\hat{\theta}$. Define $Q = P'P/n$, $R = P(P'P + \lambda D)^{-1}P'$, $R_r = P(P'P)^{-1}P'$, $T = I - R$, and $T_r = I - R_r$. Also, as above, let $h^j(z) = \mathbb{E}[x_j|z]$, where x_j is the jth component of x, and $v^j(x,z) = x - h^j(z)$.

We use the assumptions that (a) $\sqrt{n}K^{-(r_g+r_h)/d} \to 0$, (b) if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then $\sqrt{n}\lambda K^{m/d}/n \to 0$, and (c) if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then $\sqrt{n}\sqrt{\lambda/n} \to 0$.

Lemma 3.1-1: $X'TT'X/n = \Gamma + o_p(1)$, where $\Gamma = \mathbb{E}[v_i v_i'|z_i]$.

Proof: Let $1_{n,K} = 1$. First, we see that

$$\frac{1}{n}X'TX = \frac{1}{n}H'TH + \frac{1}{n}H'TV + \frac{1}{n}V'TH + \frac{1}{n}V'V - \frac{1}{n}V'RV.$$

Noting as above that $\hat{h}_r^j \equiv p^{K'}(P'P)^{-1}P'H^j$ and $\hat{h}^j \equiv p^{K'}(P'P + \lambda D)^{-1}P'H^j$ can be considered the conditional expectation of regression and penalized spline estimates of h^j, respectively, we have

$$
\begin{aligned}
{H^j}'(T-T_r)H^j/n &= {H^j}'(P(P'P)^{-1}P' - P(P'P+\lambda D)^{-1}P')H^j/n \\
&= {H^j}'\mathbb{E}[\hat{H}_r^j - \hat{H}^j|\mathbf{Z}]/n \\
&= \frac{1}{n}\sum_{i=1}^n h^j(Z_i)\mathbb{E}[\hat{h}_r^j(Z_i) - \hat{h}_j(Z_i)|\mathbf{Z}] \\
&\le \frac{1}{n}\sup_{z\in[0,1]^d} h^j(z)\sum_{i=1}^n \mathbb{E}[\hat{h}_r^j(Z_i) - \hat{h}_j(Z_i)|\mathbf{Z}] \\
&\lesssim \left(\frac{1}{n}\sum_{i=1}^n (\mathbb{E}[\hat{h}_r^j(Z_i) - \hat{h}_j(Z_i)|\mathbf{Z}])^2\right)^{1/2} \\
&= \Big((\mathbb{E}[\hat{H}_r^j - \hat{H}^j|\mathbf{Z}])'(\mathbb{E}[\hat{H}_r^j - \hat{H}^j|\mathbf{Z}])/n\Big)^{1/2},
\end{aligned}
$$

where $H^j = (h^j(Z_1)\cdots h^j(Z_n))'$ and similarly for $\hat{H}^j$ and $\hat{H}_r^j$. Then ${H^j}'(T-T_r)H^j/n = \mathcal{O}_p(\lambda K^{m/d}/n)$ if $\mathcal{C}_n(m) < 1$ for all sufficiently large n and ${H^j}'(T-T_r)H^j/n = \mathcal{O}_p(\sqrt{\lambda/n})$ if $\mathcal{C}_n(m) \ge 1$ for all sufficiently large n, using Theorem 1 and the fact that $1_{n,K} \to_p 1$. Since $P(P'P)^{-1}P'H^j$ is the projection of the vector H^j onto $\mathcal{S}_{n,r}$, we have $T_r H^j = H^j - S_{h^j}$, where S_{h^j} is the vector of best L_∞ approximations to the elements of H^j. So since T_r is idempotent,

$$
\begin{aligned}
{H^j}'T_r H^j/n &= {H^j}'T_r T_r' H^j/n \\
&= \frac{1}{n}\sum_{i=1}^n (h^j(Z_i) - s_{h^j}(Z_i))^2 \\
&= O(K^{-2r_h/d}).
\end{aligned}
$$

So $H'TH/n = o_p(1)$. Also, since $\mathbb{E}[V^j V^{j'}|\mathbf{Z}]$ is a diagonal matrix with bounded elements,

$$
\begin{aligned}
\mathbb{E}[V^{j'} R V^j]/n &= \mathbb{E}[\mathrm{Tr}(R\mathbb{E}[V^j V^{j'}|\mathbf{Z}])]/n \\
&\lesssim \mathbb{E}[\mathrm{Tr}(R)]/n \\
&= \mathbb{E}[\mathrm{Tr}(P(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}P')]/n \\
&= \mathbb{E}[\mathrm{Tr}((I+\lambda M/n)^{-1})]/n \\
&= \frac{1}{n}\sum_{k=1}^{K} \frac{1}{1+\frac{\lambda}{n}\mu_k^{n,K}} \\
&\asymp \frac{1}{n}\sum_{k=1}^{K} \frac{1}{1+\frac{\lambda}{n}\eta_k} \\
&\leq K/n.
\end{aligned}
$$

So by Markov's inequality and the fact that $1_{n,K} \to_p 1$, $V'RV/n = \mathcal{O}_p(K/n)$. Finally, since $V'V/n \to_p \mathbb{E}[v_i v_i'] = O_p(1)$, we have $H'TV/n = V'TH = o_p(1)$. So $X'TX/n = \mathbb{E}[v_i v_i'] + o_p(1)$.

Now

$$
\begin{aligned}
X'TT'X &= X'TX + X'T(T-T_r)'X + X'T(T_r - I)'X \\
&= X'TX + X'T_r(T-T_r)'X + X'(T-T_r)(T-T_r)'X + X'T(T_r-I)'X
\end{aligned}
$$

Consider the last term. We have

$$
\begin{aligned}
T(T_r - I)' &= [I - P(P'P+\lambda D)^{-1}P'][I - P(P'P)^{-1}P' - I] \\
&= -[I - P(P'P+\lambda D)^{-1}P']P(P'P)^{-1}P' \\
&= -[P(P'P)^{-1}P - P(P'P+\lambda D)^{-1}P'] \\
&= -[(I - P(P'P+\lambda D)^{-1}P') - (I - P(P'P)^{-1}P)] \\
&= T_r - T,
\end{aligned}
$$

so

$$
\begin{aligned}
X'T(T_r - I)X/n &= X'(T_r - T)X/n \\
&= H'(T_r - T)H/n + H'(T_r - T)V/n + V'(T_r - T)H/n + V'(T_r - T)V/n \\
&= H'(T_r - T)H/n + 2H'(T_r - T)V/n + V'RV/n - V'R_rV/n \\
&= o_p(1),
\end{aligned}
$$

by the above arguments (note that $V'R_rV/n$ is equal to $V'RV/n$ with $\lambda = 0$).

Consider the third term. We have

$$
\begin{aligned}
X'(T - T_r)(T - T_r)'X &= X'[P(P'P)^{-1/2}\left(I - (I + \lambda M/n)^{-1}\right)(P'P)^{-1/2}P'] \times \\
&\quad [P(P'P)^{-1/2}\left(I - (I + \lambda M/n)^{-1}\right)(P'P)^{-1/2}P']X \\
&= X'P(P'P)^{-1/2}\left(I - (I + \lambda M/n)^{-1}\right)^2(P'P)^{-1/2}P'.
\end{aligned}
$$

Since the kth diagonal element of the (diagonal) matrix $|I - (I + \lambda M/n)^{-1}|$ is $\mu_k^{n,K}/(1 - \mu_k^{n,K}) < 1$, the elements of $\left(I - (I + \lambda M/n)^{-1}\right)^2$ are less than the (absolute value of the) elements of $I - (I + \lambda M/n)^{-1}$. So since $X'(T - T_r)X/n = o_p(1)$, we also have $X'(T - T_r)(T - T_r)'X/n = o_p(1)$.

Consider the second term, and note that since ${H^j}'T_rH^j/n$ and ${H^j}'(T - T_r)(T - T_r)'H^j/n$ are $o_p(1)$, we have ${H^j}'T_r(T - T_r)H^j/n = o_p(1)$, and similarly ${H^j}'T_r(T - T_r)'V^j/n = o_p(1)$. Also, ${V^j}'T_r(T - T_r)V^j = {V^j}'(T - T_r)V^j - {V^j}'R_r(T - T_r)V^j$. Since ${V^j}'R_rV^j/n = {V^j}'R_rR_r'V^j/n = o_p(1)$ and ${V^j}'(T - T_r)(T - T_r)'V^j/n = o_p(1)$, we have ${V^j}'R_r(T - T_r)V^j/n = o_p(1)$. Then since ${V^j}'(T - T_r)V^j/n = o_p(1)$ as shown above, ${V^j}'T_r(T - T_r)V^j/n = o_p(1)$. So $X'T_r(T - T_r)X/n = o_p(1)$.

Thus,

$$
\frac{1}{n}X'TT'X = \mathbb{E}[v_iv_i'] + o_p(1).
$$

Lemma 3.1-2: $\dfrac{1}{\sqrt{n}}X'TT'(Y-X\theta) = \dfrac{1}{\sqrt{n}}V'\varepsilon + o_p(1)$.

Proof: Again, let $1_{n,K}=1$. We have

$$\frac{1}{\sqrt{n}}X'T(Y-X\theta) = \frac{1}{\sqrt{n}}H'T\varepsilon + \frac{1}{\sqrt{n}}V'\varepsilon - \frac{1}{\sqrt{n}}V'R\varepsilon + \frac{1}{\sqrt{n}}H'TG + \frac{1}{\sqrt{n}}V'TG.$$

Consider $H'T\varepsilon/\sqrt{n}$. We have

$$\begin{aligned}
\mathbb{E}[(H^{j\prime}(T-T_r)\varepsilon/\sqrt{n})^2|\mathbf{X},\mathbf{Z}] &= H^{j\prime}(T-T_r)\mathbb{E}[\varepsilon\varepsilon'|\mathbf{X},\mathbf{Z}](T-T_r)'H^j/n \\
&\lesssim H^{j\prime}[P(P'P)^{-1}P' - P(P'P+\lambda D)^{-1}P']' \times \\
&\quad [P(P'P)^{-1}P' - P(P'P+\lambda D)^{-1}P']H^j/n \\
&= (\mathbb{E}[\hat{H}^j_r - \hat{H}^j|\mathbf{Z}])'(\mathbb{E}[\hat{H}^j_r - \hat{H}^j|\mathbf{Z}])/n,
\end{aligned}$$

by Chebyshev's inequality. So if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then $H^{j\prime}(T-T_r)\varepsilon/\sqrt{n} = O_p(\lambda K^{m/d}/n)$, and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then $H^{j\prime}(T-T_r)\varepsilon/\sqrt{n} = O_p(\sqrt{\lambda/n})$. Then since

$$\begin{aligned}
\mathbb{E}[(H^{j\prime}T_r\varepsilon/\sqrt{n})^2|\mathbf{X},\mathbf{Z}] &= \mathrm{Tr}(H^{j\prime}T_r\mathbb{E}[\varepsilon\varepsilon'|\mathbf{X},\mathbf{Z}]T_rH^j)/n \\
&\lesssim \mathrm{Tr}(H^{j\prime}T_rH^j)/n \\
&= O_p(K^{-r_h/d}),
\end{aligned}$$

we have $H'T\varepsilon/\sqrt{n} = o_p(1)$.

Similarly, consider $V'TG/\sqrt{n}$. We have

$$\begin{aligned}
\mathbb{E}[(V^{j\prime}(T-T_r)G/\sqrt{n})^2|\mathbf{X},\mathbf{Z}] &= G'(T-T_r)'\mathbb{E}[V^{j\prime}V^j|\mathbf{X},\mathbf{Z}](T-T_r)G/n \\
&\lesssim G'[P(P'P)^{-1}P' - P(P'P+\lambda D)^{-1}P']' \times \\
&\quad [P(P'P)^{-1}P' - P(P'P+\lambda D)^{-1}P']G/n \\
&= (\mathbb{E}[\hat{G}_r - \hat{G}|\mathbf{X},\mathbf{Z}])'(\mathbb{E}[\hat{G}_r - \hat{G}|\mathbf{X},\mathbf{Z}])/n
\end{aligned}$$

So if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then $V^{j\prime}(T-T_r)G/\sqrt{n} = O_p(\lambda K^{m/d}/n)$, and if

$\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then $V^{j\prime}(T - T_r)G/\sqrt{n} = O_p(\sqrt{\lambda/n})$. Also,

$$\begin{aligned}
\mathbb{E}[(V^{j\prime} T_r G/\sqrt{n})^2|\mathbf{Z}] &= \mathrm{Tr}(G' T_r \mathbb{E}[V^j V^{j\prime}|\mathbf{Z}] T_r G)/n \\
&\lesssim \mathrm{Tr}(G' T_r G)/n \\
&= O_p(K^{-2r_g/d}).
\end{aligned}$$

So $V' T G/\sqrt{n} = o_p(1)$.

Next, consider $H'(T - T_r)G/\sqrt{n}$. We have

$$\begin{aligned}
H^{j\prime}(T - T_r)G/\sqrt{n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^j(Z_i) \mathbb{E}[\hat{g}_r(Z_i) - \hat{g}(Z_i)|\mathbf{X}, \mathbf{Z}] \\
&\lesssim \sqrt{n} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} (\mathbb{E}[\hat{g}_r(Z_i) - \hat{g}(Z_i)|\mathbf{X}, \mathbf{Z}])^2 \right)^{1/2} \\
&\leq \sqrt{n} \left(\mathbb{E}[\hat{G}_r - \hat{G}|\mathbf{X}, \mathbf{Z}])'(\mathbb{E}[\hat{G}_r - \hat{G}|\mathbf{X}, \mathbf{Z}]/n \right)^{1/2}
\end{aligned}$$

So if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then $H^{j\prime}(T - T_r)G/\sqrt{n} = O_p(\sqrt{n}\lambda K^{m/d}/n)$, and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then $H^j(T - T_r)G/\sqrt{n} = O_p(\sqrt{n}\sqrt{\lambda/n})$. Also,

$$\begin{aligned}
(H^{j\prime} T_r G/\sqrt{n})^2 &= G' T_r H^j H^{j\prime} T_r G/n \\
&= \mathrm{Tr}(T_r H^j H^{j\prime} T_r G G')/n \\
&\leq \sqrt{n}\sqrt{H^{j\prime} T_r H^j/n}\sqrt{G' T_r G/n} \\
&= O(\sqrt{n} K^{-(r_g + r_h)/d})
\end{aligned}$$

So $H' T G/\sqrt{n} = o_p(1)$.

Finally, consider $V'R\varepsilon/\sqrt{n}$. Similarly to above,

$$\begin{aligned}
\mathbb{E}(V^{j\prime}R\varepsilon/\sqrt{n})^2 &= \mathbb{E}[V^{j\prime}R\mathbb{E}[\varepsilon\varepsilon'|\mathbf{X},\mathbf{Z}]R'V^j]/n \\
&\lesssim \mathbb{E}[V^{j\prime}R'RV^j]/n \\
&= \mathbb{E}[V^{j\prime}P(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}P'P(P'P)^{-1/2}U \times \\
&\quad (I+\lambda M/n)^{-1})U'P(P'P)^{-1/2}V^j]/n \\
&\lesssim \lambda_{\min}^{-2}(I+\lambda M/n)\mathbb{E}[V^{j\prime}RV^j]/n \\
&\lesssim K/n
\end{aligned}$$

so $V'R\varepsilon/\sqrt{n} = o_p(1)$. Thus,

$$\frac{1}{\sqrt{n}}X'T(Y-X\theta) = \frac{1}{\sqrt{n}}V'\varepsilon + o_p(1).$$

Now

$$\begin{aligned}
\frac{1}{\sqrt{n}}X'TT'(Y-X\theta) &= \frac{1}{\sqrt{n}}X'T(Y-Y\theta) + \frac{1}{\sqrt{n}}X'T_r(T-T_r)'(Y-X\theta) + \\
&\quad \frac{1}{\sqrt{n}}X'(T-T_r)(T-T_r)'(Y-X\theta) + \frac{1}{\sqrt{n}}X'T(T_r-I)'(Y-X\theta).
\end{aligned}$$

Consider the last term. Since $T(T_r - I) = T_r - T$,

$$\begin{aligned}
\frac{1}{\sqrt{n}}X'T(T_r-I)'(Y-X\theta) &= \frac{1}{\sqrt{n}}H'(T_r-T)\varepsilon + \frac{1}{\sqrt{n}}V'R\varepsilon - \frac{1}{\sqrt{n}}V'R_r\varepsilon + \\
&\quad \frac{1}{\sqrt{n}}H'(T_r-T)G + \frac{1}{\sqrt{n}}V'(T_r-T)G \\
&= o_p(1),
\end{aligned}$$

as shown above.

Consider the third term. As discussed above, the diagonal elements of $T_r - T$ have absolute value less than one, so the elements of $(T_r - T)(T_r - T)'$ are less than the (absolute value of the) elements of $T_r - T$. So $X'(T - T_r)(T - T_r)'(Y - X\theta)/\sqrt{n} = o_p(1)$.

Finally, consider the second term. Since $H'(T_r - T)(T_r - T)'\varepsilon/\sqrt{n} = o_p(1)$ and $H'T_rT_r'\varepsilon/\sqrt{n} = o_p(1)$, we have $H'T_r(T_r - T)'\varepsilon/\sqrt{n} = o_p(1)$, and similarly, $H'T_r(T_r - T)G/\sqrt{n} = o_p(1)$ and $V'T_r(T_r - T)G/\sqrt{n} = o_p(1)$. Also, $V^{j\prime}T_r(T - T_r)\varepsilon/\sqrt{n} = V^{j\prime}(T - T_r)\varepsilon/\sqrt{n} - V^{j\prime}R_r(T - T_r)\varepsilon/\sqrt{n}$. Since $V^{j\prime}R_r\varepsilon/\sqrt{n} = o_p(1)$ and $V^{j\prime}(T - T_r)(T - T_r)'\varepsilon/\sqrt{n} = o_p(1)$ as shown above, we have $V^{j\prime}R_r(T - T_r)\varepsilon/\sqrt{n} = o_p(1)$. Then since $V^{j\prime}(T - T_r)\varepsilon/\sqrt{n} = o_p(1)$, we have $V^{j\prime}T_r(T - T_r)\varepsilon/\sqrt{n} = o_p(1)$. So $X'T_r(T - T_r)\varepsilon/\sqrt{n} = o_p(1)$.

Thus, $X'TT'\varepsilon/\sqrt{n} = X'T\varepsilon/\sqrt{n} + o_p(1)$, and therefore

$$\frac{1}{\sqrt{n}}X'TT'(Y - X\theta) = \frac{1}{\sqrt{n}}V'\varepsilon + o_p(1).$$

Theorem 2, part (a) (restated):

$$V_n^{-1/2}\sqrt{n}(\hat{\theta} - \theta) = V_n^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_i\varepsilon_i + o_p(1) \to_p \mathcal{N}(0,1),$$

$$V_n = \Gamma^{-1}\Omega\Gamma^{-1} + o_p(1),\ \ \Gamma = \mathbb{E}[\nu_i\nu_i'],\ \ \Omega = \mathbb{E}[\nu_i\nu_i'\varepsilon_i^2]$$

Proof: We have $\hat{\theta} = (X'TT'X)^{-1}X'TT'Y$. Note that $\mathbb{E}[v_i\varepsilon_i] = \mathbb{E}[v_i\mathbb{E}[\varepsilon_i|\mathbf{X},\mathbf{Z}]] = 0$. So by independence across observations,

$$\begin{aligned}\Omega_n &\equiv \mathbb{V}(V'\varepsilon/\sqrt{n}|\mathbf{X},\mathbf{Z}) \\ &= \frac{1}{n}\mathbb{V}\left(\sum_{i=1}^{n}v_i\varepsilon_i|x_i,z_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[v_iv_i'\varepsilon_i^2|x_i,z_i].\end{aligned}$$

So by the CLT, $\Omega_n^{-1/2}V'\varepsilon/\sqrt{n} \to_d \mathcal{N}(0,1)$. Then since $X'TT'X/n = \Gamma + o_p(1)$, $X'TT'(Y - X\theta)/\sqrt{n} = V'\varepsilon/\sqrt{n} = o_p(1)$, $\sqrt{n}(\hat{\theta} - \theta) = (X'TT'X/n)^{-1}X'TT'(Y - X\theta)/\sqrt{n}$, and $\Omega_n \to_p \Omega \equiv \mathbb{E}[v_iv_i'\varepsilon_i^2]$, we have

$$V_n^{-1/2}\sqrt{n}(\hat{\theta} - \theta) \to_p \mathcal{N}(0,1),$$

where $V_n \equiv \Gamma^{-1}\Omega\Gamma^{-1} + o_p(1)$, using Slutzky.

3.2 Asymptotic Distribution of the Nonparametric Component

3.2.1 Lower Bound on the Variance

We now consider the asymptotic distribution of $\partial^\ell g$. First, we present a lemma giving a lower bound on the pointwise conditional variance of $\partial^\ell g$, in order to give conditions for the bias (divided by the variance) to vanish.

Lemma 3.2.1: Defining $W_{\ell,n}(z) \equiv \mathbb{V}(\partial^\ell p^K(z)'(P'P + \lambda D)^{-1}P'\varepsilon|\mathbf{X},\mathbf{Z})$, if $\mathcal{C}_n(m) < 1$ for all sufficiently large n, then

$$W_{\ell,n}(z) \gtrsim_p K^{2|\ell|/d}\frac{K}{n};$$

if $\mathcal{C}_n(m) \geq 1$ and $\mathcal{C}_n(m)$ is bounded above for all sufficiently large n, then

$$W_{\ell,n}(z) \gtrsim_p K^{2|\ell|/d}\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}};$$

and if $\mathcal{C}_n(m)$ is unbounded for large n, then

$$W_{\ell,n}(z) \gtrsim_p K^{2|\ell|/d}\left(\frac{n}{\lambda K^{2m/d}}\right)^2\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}}.$$

Proof: First, consider the structure of $D_j^{(\ell_j)}$. We note that since each $M_{j\eta}$ is a lower triangular matrix (with the last column missing), $D_j^{(\ell_j)}$ is an upper triangular band matrix (with the last ℓ_j rows missing). Also, the kth diagonal entry of $D_j^{(\ell_j)}$ is $\prod_{L=1}^{\ell_j}\frac{L-r}{t_{j,k}-t_{j,k-r+L}}$. Letting $\left(D_j^{(\ell_j)}\right)_k$ be the kth column of $D_j^{(\ell_j)}$ and $\left(D_j^{(\ell_j)}\right)_{k\kappa}$ be the $k\kappa$ element of $D_j^{(\ell_j)}$. Then since $D_j^{(\ell_j)}$ is upper triangular, we see that

$$p_j^K(z_j)'\left(D_j^{(\ell_j)}\right)_{k_{z_j}} = p_{jk_{z_j}}(z_j)\left(D_j^{(\ell_j)}\right)_{k_{z_j}k_{z_j}} \asymp K^{1/2d}\prod_{L=1}^{\ell_j}\frac{L-r}{t_{j,k_{z_j}}-t_{j,k_{z_j}-r+L}}.$$

So

$$
\begin{aligned}
p_{j,-\ell_j}^K(z_j)' D_j^{(\ell_j)} D_j^{(\ell_j)\prime} p_{j,-\ell_j}^K(z) &= \sum_{k=1}^{K^{1/d}-\ell_j} \left(p_j^K(z_j)' \left(D_j^{(\ell_j)} \right)_k \right)^2 \\
&\geq \left(p_j^K(z_j)' \left(D_j^{(\ell_j)} \right)_{k_{z_j}} \right)^2 \\
&\asymp K^{1/d} \left(\prod_{L=1}^{\ell_j} \frac{1}{t_{j,k_{z_j}} - t_{j,k_{z_j}-r+L}} \right)^2 \\
&\asymp K^{(2\ell_j+1)/d}.
\end{aligned}
$$

(This result is also shown in Zhou & Wolfe (2000) but with a different method.) Then since $\partial^\ell p^K(z)'(P'P + \lambda D)^{-1} P'\varepsilon = p_{-\ell}^K(z)' D^{(\ell)} (P'P + \lambda D)^{-1} P'\varepsilon$ (as shown in the proof for the rate of $\|\hat{g} - g\|_{2,\ell,n}$, $\ell > 0$), we have

$$
\begin{aligned}
p_{j,-\ell}^K(z)' D^{(\ell)} D^{(\ell)\prime} p_{-\ell}^K(z) &= [p_{1,-\ell_1}^K(z)' \otimes \cdots \otimes p_{d,-\ell_d}^K(z)'][D_1^{(\ell_1)} \otimes \cdots \otimes D_d^{(\ell_d)}] \times \\
&\quad [D_1^{(\ell_1)\prime} \otimes \cdots \otimes D_d^{(\ell_d)\prime}][p_{1,-\ell_1}^K(z) \otimes \cdots \otimes p_{d,-\ell_d}^K(z)] \\
&= [p_{1,-\ell_1}^K(z)' D_1^{(\ell_1)} D_1^{(\ell_1)\prime} p_{1,-\ell_1}^K(z)] \otimes \cdots \otimes \\
&\quad [p_{d,-\ell_d}^K(z)' D_d^{(\ell_d)} D_d^{(\ell_d)\prime} p_{d,-\ell_d}^K(z)] \\
&= \prod_{j=1}^d p_{j,-\ell_j}^K(z)' D_j^{(\ell_j)} D_j^{(\ell_j)\prime} p_{j,-\ell_j}^K(z) \\
&\asymp K^{1+(2|\ell|/d)}.
\end{aligned}
$$

Also,

$$
\lambda_{\max}(I + \lambda M/n) = 1 + \lambda \mu_K^{n,K}/n = 1 + \lambda[(\varsigma + \delta_k) K^{2m/d} + \delta_{n,K}]/n = O_p(1 + \lambda K^{2m/d}/n),
$$

so $\lambda_{\max}^{-1}(I+\lambda M/n) \gtrsim_p 1/(1+\lambda K^{2m/d}/n)$. Then since $\sigma_\varepsilon(x,z)$ is bounded below,

$$\begin{aligned}
W_{\ell,n}(z) &\equiv \mathbb{V}(\partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'\varepsilon|\mathbf{X},\mathbf{Z}) \\
&= p_{-\ell}^K(z)'D^{(\ell)}(P'P+\lambda D)^{-1}P'\mathbb{E}[\varepsilon\varepsilon'|\mathbf{X},\mathbf{Z}]P(P'P+\lambda D)^{-1}D^{(\ell)\prime}p_{-\ell}^K(z) \\
&\gtrsim p_{-\ell}^K(z)'D^{(\ell)}(P'P)^{-1/2}U(I+\lambda M/n)^{-2}U'(P'P)^{-1/2}D^{(\ell)\prime}p_{-\ell}^K(z) \\
&\geq \lambda_{\max}^{-2}(I+\lambda M/n)p_{-\ell}^K(z)'D^{(\ell)}D^{(\ell)\prime}p_{-\ell}^K(z)/n \\
&\geq O_p((1+\lambda K^{2m/d}/n)^{-2}K^{1+(2|\ell|/d)}/n).
\end{aligned}$$

If $\lim_{n\to\infty}\mathcal{C}_n(m)<\infty$, then $(1+\lambda K^{2m/d}/n)^{-2}=O(1)$, so

$$W_{\ell,n}(z) \gtrsim_p K^{2|\ell|/d}\frac{K}{n}.$$

If $\lim_{n\to\infty}\mathcal{C}_n(m)\geq 1$ for all sufficiently large n, then $K\geq O((\lambda/n)^{-d/2m})$, and thus if $1\leq \lim_{n\to\infty}\mathcal{C}_n(m)<\infty$, then

$$W_{\ell,n}(z) \gtrsim_p K^{2|\ell|/d}\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}}.$$

On the other hand, if $\lim_{n\to\infty}\mathcal{C}_n(m)=\infty$, then $(1+\lambda K^{2m/d}/n)^{-2}=O((\lambda K^{2m/d}/n)^{-2})$, so

$$W_{\ell,n}(z) \gtrsim_p K^{2|\ell|/d}\left(\frac{n}{\lambda K^{2m/d}}\right)^2\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}},$$

as desired.

3.2.2 Upper Bound on the Bias

We now note that

$$\begin{aligned}
1_{n,K}\left(\partial^\ell \hat{g}(z) - \partial^\ell g(z)\right) &= 1_{n,K}\left(\partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'(Y - X\hat{\theta}) - \partial^\ell g(z)\right) \\
&= 1_{n,K}\Big(\partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'(Y - X\theta) - \\
&\quad \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta) - \partial^\ell g(z)\Big) \\
&= 1_{n,K}\Big(\partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'\varepsilon + \\
&\quad \left[\partial^\ell p^K(z)'(P'P+\lambda D)^{-}P'G - \partial^\ell g(z)\right] - \\
&\quad \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta)\Big).
\end{aligned}$$

In the next three lemmas, we consider the second term in this expression. To that end, we also give an upper bound on the pointwise conditional bias of $\partial^\ell g$, which is a useful result in its own right.

Lemma 3.2.2-1: For all $z \in [0,1]^d$, if $\mathcal{C}_n(m) < 1$ for all sufficiently large n,

$$\mathbb{E}[\partial^\ell \hat{g}(z) - \partial^\ell \hat{g}_r(z) | \mathbf{X}, \mathbf{Z}] \lesssim_p K^{|\ell|/d}\frac{\lambda}{n}K^{m/d},$$

and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, then

$$\mathbb{E}[\partial^\ell \hat{g}(z) - \partial^\ell \hat{g}_r(z) | \mathbf{X}, \mathbf{Z}] \lesssim_p K^{|\ell|/d}\sqrt{\frac{\lambda}{n}}.$$

Proof: Assume throughout that $1_{n,K} = 1$.

We have

$$\begin{aligned}
\mathbb{E}[\partial^\ell \hat{g}(z) - \partial^\ell \hat{g}_r(z) | \mathbf{X}, \mathbf{Z}] &= -\lambda p^K(z)'D^{(\ell)}(P'P+\lambda D)^{-1}D(P'P)^{-1}P'G \\
&= -\lambda p^K(z)'D^{(\ell)}(P'P+\lambda D)^{-1}D(P'P)^{-1}P'\bar{S}_g - \\
&\quad \lambda p^K(z)'D^{(\ell)}(P'P+\lambda D)^{-1}D(P'P)^{-1}P'(G - \bar{S}_g).
\end{aligned}$$

For the first term,

$$
\begin{aligned}
\lambda p^K(z)'(P'P+\lambda D)^{-1}D(P'P)^{-1}P'\bar{S}_g &= \lambda p^K(z)'(P'P)^{-1/2}(I+\lambda M/n)^{-1}(P'P)^{-1/2}D \times \\
&\quad (P'P)^{-1/2}P'\bar{S}_g \\
&= \lambda p^K(z)'(P'P)^{-1/2}(I+\lambda M/n)^{-1}(M/n) \times \\
&\quad (P'P)^{-1/2}P'\bar{S}_g \\
&\lesssim \lambda p^K(z)'(I+\lambda M/n)^{-1}(M/n)P'\bar{S}_g/n \\
&= \frac{\lambda}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\frac{\mu_k^{n,K}}{n}}{1+\frac{\lambda}{n}\mu_k^{n,K}}p_k(z)p_k(z_i)\bar{S}_g(z_i) \\
&\lesssim \frac{\lambda}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\frac{\mu_k^{n,K}}{n}}{1+\frac{\lambda}{n}\mu_k^{n,K}}p_k(z)p_k(z_i).
\end{aligned}
$$

If $\mathcal{C}_n(m) < 1$, since $1_{n,K} = 1$ and $1 + \lambda\eta_k/n \geq 1$

$$
\begin{aligned}
\frac{\lambda}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\frac{\mu_k^{n,K}}{n}}{1+\frac{\lambda}{n}\mu_k^{n,K}}p_k(z)p_k(z_i) &\lesssim \frac{\lambda}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\eta_k}{n}p_k(z)p_k(z_i) \\
&\lesssim \frac{\lambda}{n}K^{m/d}\frac{1}{n}\sum_{i=1}^{n}\sum_{k\in S_z}p_k(z)p_k(z_i) \\
&\lesssim \frac{\lambda}{n}K^{m/d}\sqrt{K}\frac{1}{n}\sum_{i=1}^{n}p_k(z_i) \\
&= O_p(\lambda K^{m/d}/n),
\end{aligned}
$$

since the number of observations for which p_k is nonzero is $O_p(n/K)$ as shown above, and where S_z is the set of all k such that $p_k(z) \geq 0$.

If $\mathcal{C}_n(m) \geq 1$, $\lambda\eta_k/n \to c \geq 1$, so $\sqrt{\lambda\eta_k/n}/(1+\lambda\eta_k/n) \leq 1/2$. Then

$$
\begin{aligned}
\frac{\lambda}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\frac{\mu_k^{n,K}}{n}}{1+\frac{\lambda}{n}\mu_k^{n,K}}p_k(z)p_k(z_i) &= \sqrt{\frac{\lambda}{n}}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\sqrt{\frac{\lambda\eta_k}{n}}}{1+\frac{\lambda}{n}\mu_k^{n,K}}p_k(z)p_k(z_i) \\
&\lesssim \sqrt{\frac{\lambda}{n}}\frac{1}{n}\sum_{k=1}^{K}p_k(z)p_k(z_i) \\
&= O_p(\sqrt{\lambda/n}).
\end{aligned}
$$

So using the structure of $D^{(\ell)}$ as in the proof of Theorem 1 for $|\ell| > 0$ and the fact that $\sup_{z\in[0,1]^d} |g(z) - s_g(z)|$, we obtain the rates $K^{|\ell|/d}\lambda K^{2m}/n$, $K^{|\ell|/d}\sqrt{\lambda/n}$, and $K^{|\ell|/d}\sqrt{\lambda K^{2m/d}/n}\sqrt{\lambda/n}$ for $\lambda p^K(z)'D^{(\ell)}(P'P+\lambda D)^{-1}D(P'P)^{-1}P'G$.

Lemma 3.2.2-2: For all $z \in [0,1]^d$,

$$|\partial^\ell \mathbb{E}[\hat{g}_r(z)|\mathbf{X},\mathbf{Z}] - \partial^\ell g(z)| \lesssim K^{1\{|\ell|>0\}/2} K^{-(r_g - |\ell|)/d}.$$

Proof: From Huang (2003b) (see Lemma 5.1), since $\sup_{z\in[0,1]^d} |\bar{s}_g(z) - g(z)| \asymp \inf_{s\in\mathcal{S}_{n,r}} \sup_{z\in[0,1]^d} |s(z) - g(z)| \lesssim K^{-r_g/d}$, we have

$$\sup_{z\in[0,1]^d} |\mathbb{E}[\hat{g}_r(z)|\mathbf{X},\mathbf{Z}] - \bar{s}_g(z)| \lesssim \sup_{z\in[0,1]^d} |g(z) - \bar{s}_g(z)| \lesssim K^{-r_g/d},$$

which gives the result for $|\ell| = 0$. For $|\ell| > 0$, for all $z \in [0,1]^d$,

$$\begin{aligned}
(\partial^\ell \mathbb{E}[\hat{g}_r(z)|\mathbf{X},\mathbf{Z}] - \partial^\ell \bar{s}_g(z))^2 &= (\partial^\ell p^K(z)'(P'P)^{-1/2}P'(G-\bar{S}_g))^2 \\
&= (G-\bar{S}_g)'P(P'P)^{-1}D^{(\ell)\prime}p^K_{-\ell}(z)p^K_{-\ell}(z)'D^{(\ell)} \times \\
&\quad (P'P)^{-1}P'(G-\bar{S}_g) \\
&\leq \lambda_{\max}(D^{(\ell)\prime}p^K_{-\ell}(z)p^K_{-\ell}(z)'D^{(\ell)})\lambda_{\max}((P'P)^{-1}) \times \\
&\quad (G-\bar{S}_g)'P(P'P)^{-1}P'(G-\bar{S}_g) \\
&\lesssim K^{1+(2|\ell|/d)}(G-\bar{S}_g)'P(P'P)^{-1}P'P(P'P)^{-1}P'(G-\bar{S}_g)/n \\
&\asymp K^{1+(2|\ell|/d)}\frac{1}{n}\sum_{i=1}^n(\mathbb{E}[g_r(z_i)|\mathbf{X},\mathbf{Z}] - \bar{s}_g(z_i))^2 \\
&\lesssim K^{1+(2|\ell|/d)}K^{-2r_g/d}.
\end{aligned}$$

Then since $\sup_{z\in[0,1]^d} |\partial^\ell \bar{s}_g(z) - \partial^\ell g(z)| \lesssim K^{-(r_g-|\ell|)/d}$, we have

$$|\partial^\ell \mathbb{E}[\hat{g}_r(z)|\mathbf{X},\mathbf{Z}] - \partial^\ell g(z)| \lesssim \sqrt{K}K^{-(r_g-|\ell|)/d},$$

giving the result for $|\ell| > 0$.

Lemma 3.2.2-3: For all $z \in [0,1]^d$, if $\mathcal{C}_n(m) \leq 1$ for all sufficiently large n,

$$\partial^\ell p^K(z)'(P'P + \lambda D)^- P'G - \partial^\ell g(z) \lesssim_p K^{|\ell|/d} \left(\frac{\lambda K^{m/d}}{n} + K^{-r_g/d + 1\{|\ell|>0\}/2} \right),$$

and if $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n,

$$\partial^\ell p^K(z)'(P'P + \lambda D)^- P'G - \partial^\ell g(z) \lesssim_p K^{|\ell|/d} \left(\sqrt{\frac{\lambda}{n}} + K^{-r_g/d + 1\{|\ell|>0\}/2} \right),$$

Proof: Combine Lemmas 3.2.2-1 and 3.2.2-2.

We now consider the third term of the above expansion, which is $\partial^\ell p_K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta)$. In order to bound this term, we also give the a bound on the rate of the uniform error of $\partial^\ell \hat{g}$. This bound is not sharp but will suffice for our purposes. We expect that the $\sqrt{K}$ coming from the basis could be replaced with $\log n$, most likely by using Bernstein's inequality along with truncation.

Lemma 3.2.2-4: $\sup_{z \in [0,1]^d} |\partial^\ell \hat{g}(z) - \partial^\ell \bar{s}_g(z)| = O_p(K^{|\ell|/d}\sqrt{K r_{n,K}})$, and thus

$$\partial^\ell p_K(z)'(P'P + \lambda D)^- P'X' = O_p(K^{|\ell|/d}\sqrt{K r_{n,K}} + 1).$$

Proof: As shown previously, for all $|\ell| \leq r - 2$, under the assumptions of Theorem 1, $\|D^{(\ell)}(\hat{\beta} - \bar{\beta})\|_2^2 = O_p(K^{2|\ell|/d} r_{n,K})$. Then since $\sup_{z \in [0,1]^d} |p_K(z)| \leq \sqrt{K}$,

$$\begin{aligned} \sup_{z \in [0,1]^d} |\partial^\ell \hat{g}(z) - \partial^\ell \bar{s}_g(z)| &= \sup_{z \in [0,1]^d} |p^K_{-\ell}(z)' D^{(\ell)}(\hat{\beta} - \bar{\beta})| \\ &\leq \sqrt{K} \|D^{(\ell)}(\hat{\beta} - \bar{\beta})\|_2 \\ &= O_p\left(\sqrt{K \cdot K^{2|\ell|/d} r_{n,K}}\right), \end{aligned}$$

and thus

$$\begin{aligned}\sup_{z\in[0,1]^d}|\partial^\ell\hat{g}(z)-\partial^\ell\bar{s}_g(z)| &\asymp \sup_{z\in[0,1]^d}|\partial^\ell\hat{g}(z)-\partial^\ell\bar{s}_g(z)|+\sup_{z\in[0,1]^d}|\partial^\ell g(z)-\partial^\ell\bar{s}_g(z)|\\ &= O_p(K^{|\ell|/d}\sqrt{Kr_{n,K}}),\end{aligned}$$

which gives the first result.

Now note that $\partial^\ell p_K(z)'(P'P+\lambda D)^{-1}P'X$ can be considered the penalized spline approximation $\partial^\ell\hat{h}(z)$ to $\partial^\ell h(z)=\partial^\ell E[x|z]$. Then since $\partial^\ell E[x|z]$ is bounded,

$$\partial^\ell p_K(z)'(P'P+\lambda D)^{-}P'X'=(\hat{h}^\ell(z)-h^\ell(z))+h^\ell(z)=O_p(K^{|\ell|/d}\sqrt{Kr_{n,K}}+1).$$

which is the second result.

3.2.3 Asymptotic Normality

We now give a proof for the asymptotic normality of $\partial^\ell\hat{g}$, for $\ell_j\leq r-2$ (for each j). We assume that if $\mathcal{C}_n(m)<1$ for all sufficiently large n, then

$$\frac{\lambda K_n^{m/d_z}/n+1\{|\ell|=0\}K_n^{-r_g/d_z}}{\sqrt{K_n/n}}+1\{|\ell|>0\}\sqrt{n}K_n^{-(r_g-|\ell|)/d_z}\to 0,$$

if $\mathcal{C}_n(m)\geq 1$ and $\mathcal{C}_n(m)$ is bounded above for all sufficiently large n.

$$\frac{\sqrt{\lambda_n/n}+1\{|\ell|=0\}K_n^{-r_g/d_z}}{\sqrt{n^{(d_z-2m)/2m}/\lambda_n^{d_z/2m}}}+1\{|\ell|>0\}\sqrt{n}K_n^{-(r_g-|\ell|)/d_z}\to 0.$$

For $\mathcal{C}_n(m)$ unbounded for large n, we assume that

$$\frac{\lambda K_n^{2m/d_z}}{n}\left(\frac{\sqrt{\lambda_n/n}\sqrt{\lambda_n K_n^{2m/d_z}/n}+1\{|\ell|=0\}K_n^{-r_g/d_z}}{\sqrt{n^{(d_z-2m)/2m}/\lambda_n^{d_z/2m}}}+1\{|\ell|>0\}\sqrt{n}K_n^{-(r_g-|\ell|)/d_z}\right)\to 0,$$

The added factor of $\lambda K^{2m/d}/n$ is needed since $\lim_{n\to\infty}\mathcal{C}_n(m)=\infty$ allows K to go to infinity very quickly, causing the variance to vanish quickly. This can then lead to a degenerate asymptotic distribution without an extra assumption. We also assume that for $\mathcal{C}_n(m)$

unbounded,

$$\frac{\lambda_n K_n^{2m/d_z}}{n} \frac{n^{(d_z-2m)/2m}}{\lambda_n^{d_z/2m}} \to 0.$$

These assumptions are needed to guarantee that the bias (divided by the variance) disappears asymptotically.

Theorem 3, part (a) (restated): Using $W_{\ell,n} = \mathbb{V}(\partial^\ell p_K(z)'(P'P + \lambda D)^{-1} P'\varepsilon|\mathbf{X}, \mathbf{Z}])$,

$$\frac{\partial^\ell \hat{g}(z) - \partial^\ell g(z)}{\sqrt{W_{\ell,n}(z)}} \to_d \mathcal{N}(0,1),$$

Proof: Let $1_{n,K} = 1$ throughout. Using the above lemmas, if $\lim_{n\to\infty} \mathcal{C}_n(m) < 1$ for all sufficiently large n,

$$\begin{aligned}
\frac{\partial^\ell p_K(z)'(P'P + \lambda D)^{-1} P'G - \partial^\ell g(z)}{\sqrt{W_{\ell,n}(z)}} &= O_p\left(\frac{K^{|\ell|/d}\left(\frac{\lambda}{n}K^{m/d} + K^{-r_g/d+1\{|\ell|>0\}/2}\right)}{K^{|\ell|/d}\sqrt{\frac{K}{n}}}\right) \\
&= O_p\left(\frac{\lambda K^{m/d}/n + K^{-r_g/d+1\{|\ell|>0\}/2}}{\sqrt{K/n}}\right) \\
&= o_p(1).
\end{aligned}$$

If $1 \leq \lim_{n\to\infty} \mathcal{C}_n(m) < \infty$,

$$\begin{aligned}
\frac{\partial^\ell p_K(z)'(P'P + \lambda D)^{-1} P'G - \partial^\ell g(z)}{\sqrt{W_{\ell,n}(z)}} &= O_p\left(\frac{K^{|\ell|/d}\left(\sqrt{\frac{\lambda}{n}} + K^{-r_g/d+1\{|\ell|>0\}/2}\right)}{K^{|\ell|/d}\sqrt{\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}}}}\right) \\
&= O_p\left(\frac{\sqrt{\lambda/n} + K^{-r_g/d+1\{|\ell|>0\}/2}}{n^{(d-2m)/2m}/\lambda^{d/2m}}\right) \\
&= o_p(1),
\end{aligned}$$

and if $\lim_{n\to\infty} \mathcal{C}_n(m) = \infty$,

$$\begin{aligned}
\frac{\partial^\ell p_K(z)'(P'P+\lambda D)^{-1}P'G - \partial^\ell g(z)}{\sqrt{W_{\ell,n}(z)}} &= O_p\left(\frac{K^{|\ell|/d}\left(\sqrt{\frac{\lambda}{n}} + K^{-r_g/d+1\{|\ell|>0\}/2}\right)}{K^{|\ell|/d}\left(\frac{\lambda K^{2m/d}}{n}\right)^{-1}\sqrt{\frac{n^{(d-2m)/2m}}{\lambda^{d/2m}}}}\right) \\
&= O_p\left(\frac{\lambda K^{2m/d}}{n}\left(\frac{\sqrt{\lambda/n} + K^{-r_g/d+1\{|\ell|>0\}/2}}{n^{(d-2m)/2m}/\lambda^{d/2m}}\right)\right) \\
&= o_p(1).
\end{aligned}$$

Similarly, since $\hat{\theta} - \theta = O_p(n^{-1/2})$, if $\lim_{n\to\infty} \mathcal{C}_n(m) < \infty$,

$$\begin{aligned}
\frac{\partial^\ell p_K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta)}{\sqrt{W_{\ell,n}(z)}} &= O_p\left(\frac{(K^{|\ell|/d}\sqrt{Kr_{n,K}} + C)O_p(n^{-1/2})}{K^{|\ell|/d}\sqrt{K/n}}\right) \\
&= O_p(\sqrt{r_{n,K}} + 1/K^{|\ell|/d+1/2}) \\
&= o_p(1),
\end{aligned}$$

and if $\lim_{n\to\infty} \mathcal{C}_n(m) = \infty$,

$$\begin{aligned}
\frac{\partial^\ell p_K(z)'(P'P+\lambda D)^{-1}P'X(\hat{\theta}-\theta)}{\sqrt{W_{\ell,n}(z)}} &= O_p\left(\frac{(K^{|\ell|/d}\sqrt{Kr_{n,K}} + C)O_p(n^{-1/2})}{K^{|\ell|/d}(n/\lambda K^{2m/d})\sqrt{K/n}}\right) \\
&= O_p\left((\lambda K^{2m/d}/n)(\sqrt{r_{n,K}} + 1/K^{|\ell|/d+1/2})\right) \\
&= o_p(1).
\end{aligned}$$

where $r_{n,K} = K/n + \lambda K^{m/d}/n + K^{-r_g/d}$ for $\mathcal{C}_n(m) < 1$ and $r_{n,K} = n^{(d-2m)/4m}/\lambda^{d/4m} + \sqrt{\lambda/n} + K^{-r_g/d}$ for $\mathcal{C}_n(m) \geq 1$.

It remains to show that $\partial^\ell p_K(z)'(P'P+\lambda D)^{-1}P'\varepsilon/\sqrt{W_{\ell,n}(z)} \to_d \mathcal{N}(0,1)$. We have

$$\begin{aligned} W_{\ell,n}(z) &\equiv \mathbb{V}(\partial^\ell p^K(z)'(P'P+\lambda D)^{-1}P'\varepsilon|\mathbf{X},\mathbf{Z}) \\ &= p^K_{-\ell}(z)'D^{(\ell)}(P'P+\lambda D)^{-1}P'\mathbb{E}[\varepsilon\varepsilon'|\mathbf{X},\mathbf{Z}]P(P'P+\lambda D)^{-1}D^{(\ell)'}p^K_{-\ell}(z) \\ &\gtrsim p^K_{-\ell}(z)'D^{(\ell)}(P'P+\lambda D)^{-1}P'P(P'P+\lambda D)^{-1}D^{(\ell)'}p^K_{-\ell}(z) \end{aligned}$$

Also, $\partial^\ell p_K(z)'(P'P+\lambda D)P'\varepsilon = \sum_{i=1}^n d_i\varepsilon_i$, where $d_i = \partial^\ell p^K(z)'(P'P+\lambda D)^{-1}p^K(z_i)$. Since $p^K(z)'p^K(z) \le K$ and $1_{n,K}=1$,

$$\begin{aligned} d_i^2 &= p^K_{-\ell}(z)'D^{(\ell)}(P'P+\lambda D)^{-1}p^K(z_i)p^K(z_i)'(P'P+\lambda D)^{-1}D^{(\ell)'}p^K_{-\ell}(z) \\ &\le \lambda_{\max}(p^K(z_i)p^K(z_i)')p^K_{-\ell}(z)'D^{(\ell)}(P'P+\lambda D)^{-1}(P'P)^{-1}P'P(P'P+\lambda D)^{-1}D^{(\ell)'}p^K_{-\ell}(z) \\ &\lesssim Kp^K_{-\ell}(z)'D^{(\ell)}(P'P+\lambda D)^{-1}P'P(P'P+\lambda D)^{-1}D^{(\ell)'}p^K_{-\ell}(z)/n \\ &\lesssim KW_{\ell,n}/n \\ &= o(W_{\ell,n}). \end{aligned}$$

So since $\sum_{i=1}^n d_i^2 \asymp W_{\ell,n}(z)$, we have $\max_{1\le i\le n} d_i^2 = o\left(\sum_{i=1}^n d_i^2\right) = o(W_{\ell,n}(z))$, and by the Lindeberg-Feller CLT,

$$\frac{1_{n,K}(\partial^\ell\hat{g}(z)-\partial^\ell g(z))}{\sqrt{W_{\ell,n}(z)}} \to_d \mathcal{N}(0,1).$$

Then since $\frac{(1_{n,K}-1)(\partial^\ell\hat{g}(z)-\partial^\ell g(z))}{\sqrt{W_{\ell,n}(z)}} \to_p 0$, we have

$$\frac{\partial^\ell\hat{g}(z)-\partial^\ell g(z)}{\sqrt{W_{\ell,n}(z)}} \to_d \mathcal{N}(0,1),$$

as desired.

3.3 Standard Errors

3.3.1 Parametric Component

We now consider the standard errors for the parametric component given in the statement of Theorem 2.

Theorem 2, part (b) (restated): $\hat{\Gamma} = \Gamma + o_p(1)$ and $\hat{\Omega} = \Omega + o_p(1)$

Proof - Theorem 2, part (b): We again assume that $1_{n,K} = 1$ and note that as shown above, $X'TT'X/n = \mathbb{E}[v_i v_i'] + o_p(1)$, that is, $\hat{\Gamma} = \Gamma + o_p(1)$.

So we now consider $\hat{\Omega} = X'TT'\hat{\Sigma}T'TX/n$. We have

$$\begin{aligned}
\sum_{j=1}^{n} R_{ij}^2 &= \sum_{j=1}^{n} p^K(Z_i)'(P'P+\lambda D)^{-1}p^K(Z_j)p^K(Z_j)'(P'P+\lambda D)^{-1}p^K(Z_i) \\
&= p^K(Z_i)'(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2} \times \\
&\quad \sum_{j=1}^{n} p^K(Z_j)p^K(Z_j)'(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}p^K(Z_i) \\
&= p^K(Z_i)'(P'P)^{-1/2}U(I+\lambda M/n)^{-2}U'(P'P)^{-1/2}p^K(Z_i) \\
&\leq \lambda_{\min}^{-1}(P'P/n)\lambda_{\min}^{-2}(I+\lambda M/n)p^K(Z_i)'p^K(Z_i)/n \\
&\lesssim K/n
\end{aligned}$$

Let N_δ be the number of observations lying in a hyper-interval δ, then

$$\mathbb{E}N_\delta = \frac{n\int_\delta f(z)dz}{\int_{[0,1]^d} f(z)dz} \lesssim \frac{n}{K}$$

(where f is the density of z). So by Markov's inequality, $N_\delta = O_p(n/K)$ for all δ, and thus $\sum_{i=1}^{n} p_k(Z_i) \lesssim \sqrt{K}O_p(n/K) = O_p(n/\sqrt{K})$. (To be precise, given $\varepsilon > 0$, let $M = \varepsilon n/K$. Then

$$\begin{aligned}
\mathrm{P}(N_\delta/(n/K) > \varepsilon) &= \mathrm{P}(N_\delta > \varepsilon n/K) \\
&\leq \mathbb{E}N_\delta/(\varepsilon n/K) \\
&\lesssim (n/K)/(\varepsilon n/K) \\
&= \varepsilon.
\end{aligned}$$

So $N_\delta/(n/K)$ is tight. Then letting $Y_{n,\delta} \equiv N_\delta/(n/K)$, we have $N_\delta = Y_{n,\delta}n/K$ with $Y_{n,\delta}$ tight. So by definition, $N_\delta = O_p(n/K)$.)

So

$$\begin{aligned}\sum_{j=1}^n R_{ij} &= p^K(Z_i)'(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}\sum_{j=1}^n p^K(Z_j) \\ &\leq \frac{1}{n}\lambda_{\min}^{-1}(P'P/n)\lambda_{\min}^{-1}(I+\lambda M/n)p^K(Z_i)'\sum_{j=1}^n p^K(Z_j) \\ &= \frac{1}{n}O_p(n/\sqrt{K})\sum_{k=1}^K p_k(Z_i) \\ &= O_p(1).\end{aligned}$$

Also,

$$\begin{aligned}R_{ii} &= p^K(Z_i)'(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}p^K(Z_i) \\ &\geq \lambda_{\max}^{-1}(P'P/n)\lambda_{\max}^{-1}(I+\lambda M/n)p^K(Z_i)'p^K(Z_i)/n \\ &\geq 0,\end{aligned}$$

so $|R_{ii}| = R_{ii}$, and

$$R_{ii} = p^K(Z_i)'(P'P)^{-1/2}U(I+\lambda M/n)^{-1}U'(P'P)^{-1/2}p^K(Z_i) \lesssim p^K(Z_i)'p^K(Z_i)/n = K/n.$$

Then

$$\sum_{j=1}^n T_{ij}^2 = \sum_{j=1, j\neq i}^n R_{ij}^2 + (1-R_{ii})^2 \leq \sum_{j=1}^n R_{ij}^2 + 2R_{ii} + 1 \lesssim 1,$$

and

$$\left(\sum_{j=1}^n T_{ij}\right)^2 = \left(\sum_{j=1, j\neq i}^n R_{ij} + 1 - R_{ii}\right)^2 \lesssim \left(\sum_{j=1}^n R_{ij}\right)^2 + (1-R_{ii})^2 = O_p(1).$$

Now letting $\bar{\varepsilon} = T\varepsilon$ and similarly for $\bar{X}$ and $\bar{y}$,

$$\begin{aligned}
\hat{\varepsilon} &= Y - X\hat{\theta} - \hat{G}(Z) \\
&= Y - X\hat{\theta} - P(P'P + \lambda D)^{-1}P'(Y - X\hat{\theta}) \\
&= T(Y - X\hat{\theta}) \\
&= \bar{Y} - \bar{X}\hat{\theta} + \bar{G} - \bar{G} - \bar{X}\theta + \bar{X}\theta \\
&= \bar{\varepsilon} + \bar{X}(\theta - \hat{\theta}) + \bar{G}.
\end{aligned}$$

So $\hat{\varepsilon}_i = \bar{\varepsilon}_i + \bar{X}_i'(\theta - \hat{\theta}) + \bar{G}_i$, where X_i' is the ith row of X. Then

$$\hat{\varepsilon}_i^2 = \bar{\varepsilon}_i^2 + 2\bar{\varepsilon}_i(\hat{\varepsilon}_i - \bar{\varepsilon}_i) + (\hat{\varepsilon}_i - \bar{\varepsilon}_i)^2 = \bar{\varepsilon}_i^2 + 2\bar{\varepsilon}_i(\bar{X}_i'(\hat{\theta} - \theta) + \bar{g}(Z_i)) + (\bar{X}_i'(\hat{\theta} - \theta) + \bar{g}(Z_i))^2.$$

Consider $\bar{X}_i'(\hat{\theta} - \theta)$. Noting that $X_{\ell j}$ is the jth observation of the ℓth component of x, $\mathbb{E}[X_{\ell j}|\mathbf{Z}]$ is bounded above, so $X_{\ell j} = O_p(1)$ for all $\ell = 1, ..., d$ and $j = 1, ..., n$. So $\sum_{j=1}^{n} T_{ij}X_{\ell j} \leq O_p(1)\sum_{j=1}^{n} T_{ij} = O_p(1)$. Then since $\hat{\theta} - \theta = O_p(n^{-1/2})$,

$$\bar{X}_i'(\hat{\theta} - \theta) = T_i'X(\hat{\theta} - \theta) = \sum_{\ell=1}^{d}\left(\sum_{j=1}^{n} T_{ij}X_{\ell j}\right)(\hat{\theta}_\ell - \theta_\ell) = O_p(n^{-1/2}),$$

where T_i' is the ith row of T. Also,

$$\bar{G}_i = T_i'G = \sum_{j=1}^{n} T_{ij}g(Z_j) \leq \sup_{z\in[0,1]^d} g(z)\sum_{j=1}^{n} T_{ij} = O_p(1).$$

So $\bar{X}_i'(\hat{\theta} - \theta) + \bar{G}_i = O_p(1)$. Finally,

$$\begin{aligned}
\mathbb{E}[\bar{\varepsilon}_i^2|\mathbf{X}, \mathbf{Z}] &= \mathbb{E}[T_i'\varepsilon\varepsilon'T_i|\mathbf{X}, \mathbf{Z}] \\
&= T_i'\mathbb{E}[\varepsilon\varepsilon'|\mathbf{X}, \mathbf{Z}]T_i \\
&= \sum_{j=1}^{n} T_{ij}^2\mathbb{E}[\varepsilon_j^2|\mathbf{X}, \mathbf{Z}] \\
&\lesssim 1.
\end{aligned}$$

So using Markov's inequality, $\bar{\varepsilon}_i^2 = O_p(1)$ and thus $\hat{\varepsilon}_i^2 = O_p(1)$.

Now letting $\hat{\Omega}_r = X'T_r\hat{\Sigma}T_rX/n$, consider $\hat{\Omega} - \hat{\Omega}_r = (X'TT'\hat{\Sigma}T'TX - X'T_r\hat{\Sigma}T_rX)/n$, which can be written as a sum of fifteen term, each of the form $X'AA'\hat{\Sigma}AA'X/n$, where A is either T_r or $T - T_r$. As above, since the elements of the diagonal matrix $T - T_r$ have absolute value less than one, the elements of $(T - T_r)(T - T_r)'$ are less than (the absolute value of) the elements of $T - T_r$. Since $\hat{\varepsilon}_i^2 = O_p(1)$,

$$\begin{aligned} X^{j'}(T - T_r)\hat{\Sigma}(T - T_r)X^j/n &= (\hat{H}_r^j - \hat{H}^j)'\hat{\Sigma}(\hat{H}_r^j - \hat{H}^j)/n \\ &= O_p(1)\frac{1}{n}\sum_{i=1}^{n}(\hat{H}_r^j(Z_i) - \hat{H}^j(Z_i))^2 \\ &\leq O_p(r_{n,K}). \end{aligned}$$

So $X'(T - T_r)(T - T_r)'\hat{\Sigma}(T - T_r)'(T - T_r)X/n = o_p(1)$. Then since $\hat{\Omega}_r = O_p(1)$, it must be that each term in $\hat{\Omega} - \hat{\Omega}_r = (X'TT'\hat{\Sigma}T'TX - X'T_r\hat{\Sigma}T_rX)/n$ is $o_p(1)$, and thus $\hat{\Omega} = \hat{\Omega}_r + o_p(1)$. Then since $\hat{\Omega}_r = \Omega + o_p(1)$ as shown in Cattaneo et al. (2010), we have

$$\hat{\Omega} - \Omega = (\hat{\Omega} - \hat{\Omega}_r) + (\hat{\Omega}_r - \Omega) = o_p(1),$$

as desired.

3.3.2 Nonparametric Component

For part (b) of Theorem 3, we use the assumption that $K^{2|\ell|/d}K/n \to 0$. More importantly, we also assume that $\sup_{z\in[0,1]^d} |\hat{g}(z) - g(z)| = O_p(1)$. Considering the case $|\ell| = 0$, which is more familiar, as shown above, $\sup_{z\in[0,1]^d} |\mathbb{E}\hat{g}(z) - g(z)| = O_p(\lambda K^{m/d}/n + K^{-r_g/d}) = o_p(1)$ for $\mathcal{C}_n(m) < 1$ for all sufficiently large n and $\sup_{z\in[0,1]^d} |\mathbb{E}\hat{g}(z) - g(z)| = O_p(\sqrt{\lambda/n} + K^{-r_g/d}) = o_p(1)$ for $1 \leq \mathcal{C}_n(m) < \infty$. So for $\mathcal{C}_n(m) < 1$ for all sufficiently large n, $K^2/n = O(1)$ would ensure that $\sup_{z\in[0,1]^d} |\hat{g}(z) - g(z)| = O_p(1)$, and for $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, $Kn^{(d-2m)/2m}/\lambda^{d/2m} = O(1)$ would ensure that $\sup_{z\in[0,1]^d} |\hat{g}(z) - g(z)| = O_p(1)$. However, these assumptions are stronger than needed, since the bound on $\sup_{z\in[0,1]^d} |\hat{g}(z) - g(z)|$ given above is not tight. If $\sqrt{K}$ in this bound is replaced with $\log n$, as discussed previously, we would only need

$$\frac{K\log n}{n} = O(1)$$

for $\mathcal{C}_n(m) < 1$ for all sufficiently large n, and

$$\frac{n^{(d-2m)/2m}\log n}{\lambda^{d/2m}} = O(1)$$

for $\mathcal{C}_n(m) \geq 1$ for all sufficiently large n, so that $\sup_{z\in[0,1]^d} |\hat{g}(z) - g(z)| = O_p(1)$.

Theorem 3, part (b) (restated): For all $z \in [0,1]^d$, $\hat{W}_{\ell,n}(z) = W_{\ell,n}(z) + o_p(1)$

Proof - Theorem 3, part (b): Since $\mathbb{E}|\varepsilon_i| = \mathbb{E}[\mathbb{E}[|\varepsilon_i||\, X_i, Z_i]]$ is bounded above, we have $|\varepsilon_i| = O_p(1)$. Also,

$$X_i'(\hat{\theta} - \theta) = \sum_{j=1}^{d} X_{ji}(\hat{\theta}_j - \theta_j) = O_p(1)\sum_{j=1}^{d}(\hat{\theta}_j - \theta_j) = O_p(n^{-1/2}).$$

Then letting $\zeta^K(z) \equiv (P'P + \lambda D)^{-1} D^{\ell\prime} p^K_{-\ell}(z)$, since $\sup_{z \in [0,1]^d} |\hat{g}(z) - g(z)| \leq O_p(1)$, we have

$$\begin{aligned}
\hat{W}_{\ell,n}(z) - W_{\ell,n}(z) &= p^K(z)' D^{(\ell)} (P'P + \lambda D)^{-1} P'(\hat{\Sigma} - \Sigma) P (P'P + \lambda D)^{-1} {D^{(\ell)}}' p^K(z) \\
&= \zeta^K(z)' \sum_{i=1}^n p^K(z_i) p^K(z_i)' (\hat{\varepsilon}_i^2 - \varepsilon_i^2) \zeta^K(z) \\
&= \sum_{i=1}^n (\zeta^K(z)' p^K(z_i))^2 (2\varepsilon_i(\hat{\varepsilon}_i - \varepsilon_i) + (\hat{\varepsilon}_i - \varepsilon_i)^2) \\
&= \sum_{i=1}^n (\zeta^K(z) p^K(z_i))^2 (2\varepsilon_i(X_i'(\hat{\theta} - \theta) + \hat{g}(Z_i) - g(Z_i)) + \\
&\quad (X_i'(\hat{\theta} - \theta) + \hat{g}(Z_i) - g(Z_i))^2) \\
&\leq O_p(1) \sum_{i=1}^n (\zeta^K(z) p^K(Z_i))^2 (2|\varepsilon_i| + 1) \\
&= O_p(1) \zeta^K(z)' \sum_{i=1}^n p^K(z_i) p^K(z_i)' \zeta^K(z) \\
&= O_p(1) p^K(z)' D^{(\ell)} (P'P)^{-1/2} U (I + \lambda M/n)^{-2} U' (P'P)^{1/2} {D^{(\ell)}}' p^K(z) \\
&= O_p(1) \mathrm{Tr}(p^K(z)' {D^{(\ell)}}' D^{(\ell)} p^K(z))/n \\
&= O_p(K^{2|\ell|/d} K/n) \\
&= o_p(1),
\end{aligned}$$

as desired.

CHAPTER IV

Simulations Study

4.1 Description of Simulations

We have conducted a small simulations study, in order to illustrate the results given above, using Stata. We have also written a general ado file which will give the penalized B-spline estimate given a set of observations and a value of K^{1/d_z} and λ. Equally-spaced knots are used with an extended partition.

First, we used the bivariate (additive) function $g(z_1, z_2) = (z_1 + 2e^{-16z_1^2}) + (\sin(2z_2) + 2e^{-16z_2^2})$ with $z_1, z_2 \sim \mathcal{U}(0,1)$. We also used $h(z) = 0.1\sqrt{z_1^2 + z_2^2}$ and $\theta = 1$. The errors were normally distributed with $\varepsilon, \nu \sim \mathcal{N}(0,1)$. We generated 500 observations and did 1000 Monte-Carlo repetitions. Since g was additively separable, we used no interactions between the B-splines for z_1 and the B-splines for z_2, substantially improving run time. Since the knots were equally spaced, we found the elements of the penalization matrix D in Maple and entered them in the Stata code. With g additively separable, D has a block diagonal structure.

4.2 Results

Our results are given in the following tables. These results illustrate the asymptotic normality given in Theorems 2 and 3 and suggest that the pointwise confidence intervals using the standard error estimates given produce appropriate coverage rates along assumed sequences of λ_n and K_n. The first table illustrates that a larger K_n is needed for a larger λ_n to achieve the same coverage rate. The second table shows the familiar pattern that for

	0	1	2
4	**.942**	.976	.995
6	.931	.953	.973
8	.931	**.951**	.963
10	.938	.946	.954
12	.935	.946	.956
14	.936	.943	.954
16	.937	.945	.954
18	.935	.944	.952
20	.936	.941	.952
22	.933	.945	**.95**
24	.931	.944	.948
26	.926	.943	.948
28	.929	.944	.949
30	.925	.944	.947

Table 4.1: Parametric component - $\hat{\theta}$

each λ_n, as K_n increases, the coverage rate reaches approximately 95% and then decreases again.

	0	1	2
4	.011	.106	0
6	.713	.067	.208
8	.952	.934	.883
10	.957	**.955**	.945
12	**.96**	.95	**.95**
14	.952	.947	.94
16	.945	.942	.938
18	.94	.942	.938
20	.936	.94	.94
22	.935	.936	.937
24	.93	.931	.939
26	.926	.93	.935
28	.927	.928	.93
30	.928	.922	.927

Table 4.2: Nonparametric component - $\hat{g}(0.5, 0.5)$

CHAPTER V

Conclusion

In summary, we have presented a method for robust inference in the partially linear model under weak conditions, along with rates of convergence for the nonparametric component. The main contributions previously unavailable include the following:

Chapter 2:

- Rates of convergence of penalized spline estimators for $d_z > 1$

  These rates were available in the literature for regression and smoothing splines but not for penalized splines in the multivariate case

- Weaker conditions needed for consistency (and asymptotic normality)

  As discussed above, Huang (2003a) gives a minimal condition $K_n \log n/n \to 0$, as opposed to the condition $K_n^2/n \to 0$ used in Newey (1997) and implicitly in Claeskens et al. (2009) and Zhou et al. (1998). We use this condition to bound the eigenvalues of $P'P/n$ and convergence of the eigenvalues of $(P'P/n)^{-1/2}D(P'P/n)^{-1/2}$ to the eigenvalues of a well-known differential equation, as discussed above.

- An asymptotic expression for the eigenvalues used in the penalization matrix D and a decomposition of D

  This has been an open question in the literature and was at the heart of the results herein, since these eigenvalues are used to the determine the expression for the mean squared error and the resulting rate. We believe the asymptotic expression given will be useful to many researchers in the future considering penalized least squares estimators in various forms, not specific to spline estimation.

- An formula for ς (in the definition of $\mathcal{C}_n(m)$) for any density of z that is bounded above and below away from zero, even for $d_z = 1$

  Previously for $d_z = 1$, this expression was known only for densities that were regular (see Speckman (1981), equation (2.2), or Claeskens et al. (2009), Lemma A3 for a definition). It was not available in the literature even for regular densities when $d_z > 1$ to the author's knowledge.

- An expression for a best L_∞ approximation $\bar{s}_g$ to g such that $\partial^\ell \bar{s}_g$ is also a best L_∞ approximation to $\partial \bar{s}_g$

  This result is crucial for considering $\partial^\ell \hat{g}(z)$ instead of only $\hat{g}(z)$.

- Rates of convergence for $\partial^\ell \hat{g}(z)$ even for $d_z = 1$.

  This result was not previously available to the author's knowledge.

- Rates for the fixed mean squared error (in which the average is taken over the population as opposed to the observations), even for $d_z = 1$

  Previously, this rate was available only for the empirical mean squared error. The fixed mean squared error is more prevalent in the literature, so these results rate to the literature more clearly.

Chapter 3:

- The asymptotic distribution and standard error estimates for penalized spline estimators

  Asymptotic normality results were not previously available for penalized spline estimators even in the univariate case. They were also not available for smoothing spline estimators, to the author's knowledge.

- The asymptotic distribution and standard errors for the parametric component of the partially linear model

This result was available previously in Donald & Newey (1994) for regression splines but not for penalized splines.

- Improving the conditions for the asymptotic normality of $\hat{g}(z)$ for $\lambda_n = 0$

  As mentioned above, it was previously assumed in the literature that $\sqrt{n}K_n^{-r_g/d_z} \to 0$, which is stronger than the assumptions given in Chapter 3.

APPENDICES

APPENDIX A

Eigenvalues for the Penalization Matrix

Proof of Lemma 2.1: We give the proof in a series of lemmas, following the proofs in Huang (2003a) (see Lemma 1, Lemma 2, and Corollary 3).

Lemma 2.1-1: We have

$$P\left(\sup_{f,g\in\mathcal{S}_{n,r}} \frac{|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i) - \mathbb{E}[f(z)g(z)]|}{\|f\|\|g\|} > t\right) \leq$$
$$P\left(\sup_{\Delta}\sup_{f,g\in\mathcal{S}_{n,r}} \frac{|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i) - \mathbb{E}[f(z)g(z)I_\Delta(z)]|}{\|f\|_\Delta\|g\|_\Delta} > t\right) \lesssim$$
$$K\exp\left(-\frac{1}{C}\frac{(nt)^2}{nK + \frac{1}{3}\sqrt{K}A_K nt}\right),$$

where r_Δ is the number of nonzero splines on $\Delta \in \{\Delta_k\}_{k=1}^K$ and $A_{K,\Delta} = \sup\limits_{g\in\mathcal{S}_{n,r}} \dfrac{\|g\|_{\infty,\Delta}}{\|g\|_\Delta}$ with $\|g\|_\Delta^2 = \mathbb{E}[g^2(Z)I_\Delta(Z)]$ and $\|g\|_{\infty,\Delta} = \sup\limits_{z\in\Delta}|g(z)|$.

Proof: Given $\Delta = \Delta^1 \times \cdots \times \Delta^d \in \{\Delta\}_{k=1}^K$ with $\Delta^j \in \{\Delta_k^j\}_{k=1}^{K^{1/d}}$, $j = 1, ..., d$, let k_Δ and K_Δ be the smallest and largest values of k (with a possible reordering of the indices) such that $p_k \geq 0.1\sqrt{K}$ on Δ, and similarly for k_{Δ^j} and K_{Δ^j}. Then given $f, g \in \mathcal{S}_{n,r}$, for some constants $f, ..., f_K$, $g_1, ..., g_K$, $fI_\Delta = \sum\limits_{k=k_\Delta}^{K_\Delta} f_k p_k$ and $gI_\Delta = \sum\limits_{k=k_\Delta}^{K_\Delta} g_k p_k$. Then since the density of z is bounded away from zero,

$$\begin{aligned}
\mathbb{E} f^2 I_\Delta &\gtrsim \int_\Delta \left(\sum_{k=k_\Delta}^{K_\Delta} f_k p_k(z) \right)^2 dz \\
&= \prod_{j=1}^d \int_{\Delta^j} \left(\sum_{k_j=k_{\Delta^j}}^{K_{\Delta^j}} f_{jk_j} p_{jk_j}(z_j) \right)^2 dz_j \\
&= \prod_{j=1}^d \sum_{k_j=k_{\Delta^j}}^{K_{\Delta^j}} f_{jk} f_{j\ell} \int_{\Delta^j} p_{jk}(z_j) p_{j\ell}(z_j) dz_j,
\end{aligned}$$

where $f^1 \otimes \cdots \otimes f^d = f^K$ with $f^j = (f_{j1} \cdots f_{jK^{1/d}})'$ for $j = 1, ..., d$ and $f^K = (f_1 \cdots f_K)'$, and z_j is the jth component of z. If the degree $r - 1$ of the B-splines is zero, then for $k, \ell = k_\Delta, ..., K_\Delta$,

$$\int_{\Delta^j} p_{jk}(z_j) p_{j\ell}(z_j) dz_j \gtrsim \frac{1}{K^{1/d}} \cdot K^{1/d} 1\{k = \ell\} = 1\{k = \ell\},$$

since the length of Δ^j is $\asymp 1/K^{1/d}$. So $\mathbb{E} f I_\Delta^2 \gtrsim \prod_{j=1}^{d} \sum_{k=k_{\Delta_j}}^{K_{\Delta j}} f_{jk}^2 = \sum_{k=k_\Delta}^{K_\Delta} f_k^2$. If $r = 1$, then letting $p_{j,k,s}$ be the kth sth-degree B-spline in direction j, for $k, \ell = 1, ..., K^{1/d}$,

$$
\begin{aligned}
\int_{\Delta^j} p_{jk}(z_j) p_{j\ell}(z_j) dz_j &= \int_{\Delta^j} \left(\frac{z_j - t_{j,k}}{t_{j,k+1} - t_{j,k}} p_{j,k,0}(z_j) + \frac{t_{j,k+2} - z_j}{t_{j,k+2} - t_{j,k+1}} p_{j,k+1,0}(z_j) \right) \times \\
&\quad \left(\frac{z_j - t_{j,\ell}}{t_{j,\ell+1} - t_{j,\ell}} p_{j,\ell,0}(z_j) + \frac{t_{j,\ell+2} - z_j}{t_{j,\ell+2} - t_{j,\ell+1}} p_{j,\ell+1,0}(z_j) \right) dz_j \\
&= K^{1/d} \int_{\Delta^j} \left(\frac{z_j - t_{j,k}}{t_{j,k+1} - t_{j,k}} 1\{z_j \in [t_{j,k}, t_{j,k+1})\} + \right. \\
&\quad \left. \frac{t_{j,k+2} - z_j}{t_{j,k+2} - t_{j,k+1}} 1\{z_j \in [t_{j,k+1}, t_{j,k+2})\} \right) \times \\
&\quad \left(\frac{z_j - t_{j,\ell}}{t_{j,\ell+1} - t_{j,\ell}} 1\{z_j \in [t_{j,\ell}, t_{j,\ell+1})\} + \right. \\
&\quad \left. \frac{t_{j,\ell+2} - z_j}{t_{j,\ell+2} - t_{j,\ell+1}} 1\{z_j \in [t_{j,\ell+1}, t_{j,\ell+2})\} \right) dz_j \\
&\geq K^{1/d} \int_{\Delta^j} \left(\frac{z_j - t_{j,k}}{t_{j,k+1} - t_{j,k}} 1\{z_j \in [(t_{j,k} + t_{j,k+1})/2, t_{j,k+1})\} \right) \times \\
&\quad \left(\frac{z_j - t_{j,\ell}}{t_{j,\ell+1} - t_{j,\ell}} 1\{z_j \in [(t_{j,\ell} + t_{j,\ell+1})/2, t_{j,\ell+1})\} \right) dz_j \\
&\gtrsim K^{1/d} \int_{\Delta^j} 1\{z_j \in [(t_{j,k} + t_{j,k+1})/2, t_{j,k+1})\} dz_j \cdot 1\{k = \ell\} \\
&\gtrsim K^{1/d} \cdot \frac{1}{K^{1/d}} \cdot 1\{k = \ell\} \\
&= 1\{k = \ell\}
\end{aligned}
$$

So $\mathbb{E}f^2 I_\Delta \gtrsim \sum_{k=k_\Delta}^{K_\Delta} f_k^2$. If $r=2$, then

$$
\begin{aligned}
\int_{\Delta^j} p_{jk}(z_j)p_{j\ell}(z_j)dz_j &= \int_{\Delta^j}\left(\frac{z_j - t_{j,k}}{t_{j,k+2}-t_{j,k}}p_{j,k,1}(z_j) + \frac{t_{j,k+3}-z_j}{t_{j,k+3}-t_{j,k+1}}p_{j,k+1,1}(z_j)\right)\times \\
&\quad \left(\frac{z_j - t_{j,\ell}}{t_{j,\ell+2}-t_{j,\ell}}p_{j,\ell,1}(z_j) + \frac{t_{j,\ell+3}-z_j}{t_{j,\ell+3}-t_{j,\ell+1}}p_{j,\ell+1,1}(z_j)\right)dz_j \\
&= \int_{\Delta^j}\left[\frac{z_j - t_{j,k}}{t_{j,k+2}-t_{j,k}}\left(\frac{z_j - t_{j,k}}{t_{j,k+1}-t_{j,k}}p_{j,k,0}(z_j) + \right.\right. \\
&\quad \left.\frac{t_{j,k+2}-z_j}{t_{j,k+2}-t_{j,k+1}}p_{j,k+1,0}(z_j)\right) + \\
&\quad \frac{t_{j,k+3}-z_j}{t_{j,k+3}-t_{j,k+1}}\left(\frac{z_j - t_{j,k+1}}{t_{j,k+2}-t_{j,k+1}}p_{j,k+1,0}(z_j) + \right. \\
&\quad \left.\left.\frac{t_{k+3}-z_j}{t_{k+3}-t_{j,k+2}}p_{j,k+2,0}(z_j)\right)\right]\times \\
&\quad \left[\frac{z_j - t_{j,\ell}}{t_{j,\ell+2}-t_{j,\ell}}\left(\frac{z_j - t_{j,\ell}}{t_{j,\ell+1}-t_{j,\ell}}p_{j,\ell,0}(z_j) + \right.\right. \\
&\quad \left.\frac{t_{j,\ell+2}-z_j}{t_{j,\ell+2}-t_{j,\ell+1}}p_{j,\ell+1,0}(z_j)\right) + \\
&\quad \frac{t_{j,\ell+3}-z_j}{t_{j,\ell+3}-t_{j,\ell+1}}\left(\frac{z_j - t_{j,\ell+1}}{t_{j,\ell+2}-t_{j,\ell+1}}p_{j,\ell+1,0}(z_j) + \right. \\
&\quad \left.\left.\frac{t_{\ell+3}-z_j}{t_{\ell+3}-t_{j,\ell+2}}p_{j,\ell+2,0}(z_j)\right)\right]dz_j \\
&\gtrsim K^{1/d}\int_{\Delta^j} 1\{z_j \in [(t_{j,k}+t_{j,k+1})/2, t_{j,k+1})\}\cdot 1\{k=\ell\} \\
&\geq 1\{k=\ell\},
\end{aligned}
$$

and again

$$\mathbb{E}f^2 I_\Delta \gtrsim \sum_{k=k_\Delta}^{K_\Delta} f_k^2. \tag{A.1}$$

We can show the same result similarly for any degree of the spline basis functions.

Now for any $k, \ell = k_\Delta, ..., K_\Delta$, since $\sup_{z\in[0,1]^d}|p_k(z)| \leq \sqrt{K}$,

$$\mathbb{V}(p_k(Z)p_\ell(Z)I_\Delta(Z)) \leq \mathbb{E}(p_k(Z)p_\ell(Z)I_\Delta(Z))^2 \lesssim \frac{1}{K}\cdot K^2 = K.$$

So for all $z \in [0,1]^d$,

$$\begin{aligned}
|p_k(z)p_\ell(z)I_\Delta(z)| &\le \frac{\sup_{z\in\Delta}|p_k(z)p_\ell(z)|}{(\mathbb{E}(p_k(Z)p_\ell(Z)I_\Delta(Z))^2)^{1/2}} \cdot (\mathbb{E}(p_k(Z)p_\ell(Z)I_\Delta(Z))^2)^{1/2} \\
&\le A_{K,\Delta}(\mathbb{E}(p_k(Z)p_\ell(Z)I_\Delta(Z))^2)^{1/2} \\
&\lesssim \sqrt{K}A_{K,\Delta}.
\end{aligned}$$

By Bernstein's inequality (Pollard 1984), for any $j,\ell = k_\Delta, ..., k_\Delta$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n p_k(Z_i)p_\ell(z_i)I_\Delta(Z_i) - \mathbb{E}p_k(z)p_\ell(z)\right| > t\right) \le 2\exp\left(-\frac{1}{2}\frac{(nt)^2}{ncK + \frac{1}{3}c\sqrt{K}A_{K,\Delta}nt}\right),$$

for a constant c. So letting $r_\Delta = K_\Delta - k_\Delta + 1$,

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n p_k(Z_i)p_\ell(Z_i)I_\Delta(Z_i) - \mathbb{E}p_k(z)p_\ell(z)I_\Delta(z)\right| > t, \text{ for all } k,\ell = k_\Delta, ..., K_\Delta\right) \le$$
$$\sum_{k=k_\Delta}^{K_\Delta} 2\exp\left(-\frac{1}{2}\frac{(nt)^2}{ncK + \frac{1}{3}c\sqrt{K}A_{K,\Delta}nt}\right) \le$$
$$2r_\Delta^2 \exp\left(-\frac{1}{2}\frac{(nt)^2}{ncK + \frac{1}{3}c\sqrt{K}A_{K,\Delta}nt}\right).$$

Given $f,g \in \mathcal{S}_{n,r}$, if $\left|\frac{1}{n}\sum_{i=1}^n p_k(Z_i)p_\ell(Z_i)I_\Delta(Z_i) - \mathbb{E}p_k(z)p_\ell(z)I_\Delta(z)\right| \le t/r_\Delta$, for all $k,\ell = k_\Delta, K_\Delta$,

$$\left|\frac{1}{n}\sum_{i=1}^n f(Z_i)g(Z_i)I_\Delta(Z_i) - \mathbb{E}f(z)g(z)I_\Delta(z))\right| =$$
$$\left|\frac{1}{n}\sum_{i=1}^n\sum_{k=k_\Delta}^{K_\Delta}\sum_{\ell=k_\Delta}^{K_\Delta} f_k g_\ell p_k(Z_i)p_\ell(Z_i) - \mathbb{E}f_k g_\ell p_k(z)p_\ell(z)\right| =$$
$$\left|\sum_{k=k_\Delta}^{K_\Delta}\sum_{\ell=k_\Delta}^{K_\Delta} f_k g_\ell\left(\frac{1}{n}\sum_{i=1}^n p_k(Z_i)p_\ell(Z_i) - \mathbb{E}p_k p_\ell\right)\right| \le$$
$$\sum_{k=k_\Delta}^{K_\Delta}\sum_{\ell=k_\Delta}^{K_\Delta}|f_k||g_\ell|\frac{t}{r_\Delta} \le$$
$$r_\Delta^{1/2}\left(\sum_{k=k_\Delta}^{K_\Delta} f_k^2\right)^{1/2} r_\Delta^{1/2}\left(\sum_{\ell=k_\Delta}^{K_\Delta} g_\ell^2\right)^{1/2}\frac{t}{r_\Delta} \lesssim$$
$$t\|f\|_\Delta\|g\|_\Delta,$$

where the penultimate line follows from Cauchy-Schwarz and the last line follow from (1). So since f and g were arbitrary, if $\left|\frac{1}{n}\sum_{i=1}^{n} p_k(Z_i)p_\ell(Z_i) - \mathbb{E}p_k(z)p_\ell(z)\right| \leq t/r_\Delta$ for all $k, \ell = k_\Delta, ..., K_\Delta$, then
$\sup_{f,g\in\mathcal{S}_{n,r}} \left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i) - \mathbb{E}f(z)g(z)I_\Delta(z)\right| \lesssim t\|f\|_\Delta\|g\|_\Delta$. Thus,

$$P\left(\sup_{f,g\in\mathcal{S}_{n,r}} \left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i) - \mathbb{E}f(z)g(z)I_\Delta(z)\right| > t\|f\|_\Delta\|g\|_\Delta\right) \leq$$
$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} p_k(Z_i)p_\ell(Z_i)I_\Delta(Z_i) - \mathbb{E}p_k(z)p_\ell(z)I_\Delta(z)\right| > t/r_\Delta \text{ for all } k, \ell = k_\Delta, ..., K_\Delta\right) \leq$$
$$2r_\Delta^2 \exp\left(-\frac{1}{2}\frac{(nt)^2}{ncK + \frac{1}{3}c\sqrt{K}A_{K,\Delta}nt}\right),$$

Also, for all $f, g \in \mathcal{S}_{n,r}$, if $\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i) - \mathbb{E}f(z)g(z)I_\Delta(z)\right| \leq t\|f\|_\Delta\|g\|_\Delta$ for all $\Delta \in \{\Delta_k\}_{k=1}^{K}$, then

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i) - \mathbb{E}fg\right| &\leq \sum_\Delta |\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z-i) - \mathbb{E}f(z)g(z)I_\Delta(z))| \\ &\leq \sum_\Delta t\|f\|_\Delta\|g\|_\Delta \leq t\|f\|\|g\|,\end{aligned}$$

by Cauchy-Schwarz and $\|f\|^2 = \sum_\Delta \|f\|_\Delta^2$. So

$$\sup_{f,g\in\mathcal{S}_{n,r}} \frac{\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i) - \mathbb{E}f(z)g(z)\right|}{\|f\|\|g\|} \leq$$
$$\sup_\Delta \sup_{f,g\in\mathcal{S}_{n,r}} \frac{\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i) - \mathbb{E}f(z)g(z)I_\delta(z)\right|}{\|f\|_\Delta\|g\|_\Delta},$$

and thus

$$
\begin{aligned}
&P\left(\sup_{f,g\in\mathcal{S}_{n,r}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)-\mathbb{E}f(z)g(z)\right|>t\|f\|\|g\|\right)\le\\
&P(\sup_{\Delta}\sup_{f,g\in\mathcal{S}_{n,r}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i)-\mathbb{E}f(z)g(z)I_\delta(z)\right|>t\|f\|_\Delta\|g\|_\Delta)\le\\
&\sum_{\Delta}P(\sup_{f,g\in\mathcal{S}_{n,r}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i)g(Z_i)I_\Delta(Z_i)-\mathbb{E}f(z)g(z)I_\delta(z)\right|>t\|f\|_\Delta\|g\|_\Delta)\le\\
&\sum_{\Delta}2r_\Delta^2\exp\left(-\frac{1}{2}\frac{(nt)^2}{ncK+\frac{1}{3}c\sqrt{K}A_{K,\Delta}nt}\right)\lesssim\\
&K\exp\left(-\frac{1}{C}\frac{(nt)^2}{nK+\frac{1}{3}\sqrt{K}A_K nt}\right),
\end{aligned}
$$

for some positive constant C, since r_Δ is bounded.

Lemma 2.1-2: For all $\Delta\in\{\Delta\}_{k=1}^K$, $A_{K,\Delta}\lesssim\sqrt{K}$.

Proof: Given $g=\sum_{k=1}^{K} g_k p_k\in\mathcal{S}_{n,r}$,

$$
\begin{aligned}
\frac{\sup_{z\in\Delta}|g(z)|}{(\mathbb{E}g^2(Z)I_\Delta(Z))^{1/2}} &= \frac{\sup_{z\in\Delta}\left|\sum_{k=k_\Delta}^{K_\Delta} g_k p_k(z)\right|}{\left(\mathbb{E}\left(\sum_{k=k_\Delta}^{K_\Delta} g_k p_k(Z)\right)^2\right)^{1/2}}\\
&\le \frac{\sup_{z\in\Delta}\left(\sum_{k=k_\Delta}^{K_\Delta} g_k^2\right)^{1/2}\left(\sum_{k=k_\Delta}^{K_\Delta} p_k(z)^2\right)^{1/2}}{\left(\sum_{k=k_\Delta}^{K_\Delta} g_k^2\right)^{1/2}}\\
&\lesssim \sqrt{K}.
\end{aligned}
$$

So $\sup_{g\in\mathcal{S}_{n,r}}\frac{\sup_{z\in\Delta}|g(z)|}{(Eg^2(Z)I_\Delta(Z))^{1/2}}\lesssim\sqrt{K}$, as desired.

Lemma 2.1-3: $\sup_{f,g\in\mathcal{S}_{n,r}}\frac{|(E_n-E)(fg)|}{\|f\|\|g\|}=O_p(\sqrt{K\log n/n})$, and thus if $K\log n/n\to 0$,

$$
c_1+o_p(1)\le\tilde{\lambda}_{\min}\le\tilde{\lambda}_{\max}\le c_2+o_p(1),
$$

for some constants $c_1,c_2>0$, where $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ are the minimum and maximum eigenvalues of $P'P/n$, respectively.

Proof: Choosing $t = \sqrt{\tilde{c}K \log n/n}$ for sufficiently large $\tilde{c}$ in the above expression, we have

$$
\begin{aligned}
& P\left(\sup_{f,g\in\mathcal{S}_{n,r}} \frac{\left|\frac{1}{n}\sum_{i=1}^n f(Z_i)g(Z_i) - \mathbb{E}f(z)g(z)\right|}{\|f\|\|g\|} > \sqrt{\frac{K\log n}{n}}\right) \leq \\
& K\exp\left(-\frac{1}{C}\frac{n^2\tilde{c}K\log n/n}{nK + \frac{1}{3}Kn\sqrt{\tilde{c}K\log n/n}}\right) = \\
& K\exp\left(-\frac{1}{C}\frac{\tilde{c}\log n}{1 + \frac{1}{3}\sqrt{\tilde{c}K\log n/n}}\right) = \\
& O(K\exp(-(\tilde{c}/C)\log n))) = \\
& O(K/n^{\tilde{c}/C}) = \\
& o(1).
\end{aligned}
$$

So $P\left(\sup_{f,g\in\mathcal{S}_{n,r}} \frac{\left|\frac{1}{n}\sum_{i=1}^n f(Z_i)g(Z_i) - \mathbb{E}f(z)g(z)\right|}{\|f\|\|g\|} \Big/ \sqrt{\frac{K\log n}{n}} > 1\right) = o(1)$, and thus

$$
\sup_{f,g\in\mathcal{S}_{n,r}} \frac{\left|\frac{1}{n}\sum_{i=1}^n f(Z_i)g(Z_i) - \mathbb{E}f(z)g(z)\right|}{\|f\|\|g\|} = O_p(\sqrt{K\log n/n}).
$$

Furthermore, given $s = \sum_{k=1}^K a_k p_k \in \mathcal{S}_{n,r}$ with $\sum_{k=1}^K a_k^2 = 1$, define a_{jk}, $j = 1, ..., d$, $k = 1, ..., K^{1/d}$ in the same way that f_{jk} was defined. Note that since

$$
1 = \sum_{k=1}^K a_k^2 = \prod_{j=1}^d \sum_{k_j=1}^{K^{1/d}} a_{jk_j}^2,
$$

we have $\sum_{k=1}^{K^{1/d}} a_{jk}^2 = 1$, $j = 1, ..., d$. As shown in de Boor (1978, p. 155), see also Zhou et. al. (1998, equation (13)), $\int_0^1 \left(\sum_{k_j=1}^{K^{1/d}} a_{jk_j} p_{jk_j}(z_j)\right)^2 dz_j \leq K^{1/d} \sum_{k_j=1}^{K^{1/d}} a_{jk_j}^2 (t_{j,k_j} - t_{j,k_j-r})$. So since the density of z is bounded above,

$$\begin{aligned}
\mathbb{E}s^2 &\asymp \prod_{j=1}^{d} \int_0^1 \left(\sum_{k_j=1}^{K^{1/d}} a_{jk_j} p_{jk_j}(z_j) \right)^2 dz_j \\
&\leq \prod_{j=1}^{d} K^{1/d} \sum_{k_j=1}^{K^{1/d}} a_{jk_j}^2 (t_{j,k_j} - t_{j,k_j-r}) \\
&\asymp \prod_{j=1}^{d} \sum_{k_j=1}^{K^{1/d}} a_{jk_j}^2 \\
&= \sum_{k=1}^{K} a_k^2 \\
&= 1,
\end{aligned}$$

since each $p_{jk_j}^2$ is nonzero on an interval of length $\asymp K^{1/d}$. So

$$\begin{aligned}
\sup_{\sum a_k^2=1} \left| \frac{1}{n} \sum_{i=1}^{n} s(Z_i)^2 - \mathbb{E}s(z)^2 \right| &\lesssim \sup_{\sum a_k^2=1} \frac{\left| \frac{1}{n} \sum_{i=1}^{n} s(Z_i)^2 - \mathbb{E}s(z)^2 \right|}{\|s\|^2} \\
&\leq \sup_{f,g \in \mathcal{S}_{n,r}} \frac{\left| \frac{1}{n} \sum_{i=1}^{n} f(Z_i)g(Z_i) - \mathbb{E}f(z)g(z) \right|}{\|f\|\|g\|} \\
&= o_p(1).
\end{aligned}$$

As already shown, $c_1 \leq \mathbb{E}s^2 \leq c_2$, for some positive constants c_1 and c_2. So since $\frac{1}{n} \sum_{i=1}^{n} s(Z_i)^2 = \mathbb{E}s(z)^2 + \left(\frac{1}{n} \sum_{i=1}^{n} s(Z_i)^2 - \mathbb{E}s(z)^2 \right) = \mathbb{E}s(z)^2 + o_p(1)$, we have

$$c_1 + o_p(1) \leq \frac{1}{n} \sum_{i=1}^{n} s(Z_i)^2 \leq c_2 + o_p(1).$$

Then since

$$\tilde{\lambda}_{\max} = \max_{\sum_{k=1}^{K} a_k^2=1} a'(P'P/n)a = \max_{\sum_{k=1}^{K} a_k^2=1} \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{k=1}^{K} a_k p_k(Z_i) \right)^2 = \max_{\sum_{k=1}^{K} a_k^2=1} E_n s^2$$

and similarly for $\tilde{\lambda}_{\min}$, we have

$$c_1 + o_p(1) \leq \tilde{\lambda}_{\min} \leq \tilde{\lambda}_{\max} \leq c_2 + o_p(1),$$

as desired.

APPENDIX B

Eigenvalues of the Design Matrix

Proof of Lemma II.3: The proof is given by Volker Elling, University of Michigan.

Beta and Gamma functions

Gamma function:

$$\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt \tag{B.1}$$

Well-known: $\Gamma(k+1) = k!$. Beta function: to generalize

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \tag{B.2}$$

to real numbers it is natural to write

$$B(a, b) := 1 / \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}. \tag{B.3}$$

(extra inverse and coefficients not quite same...) Convenient formula:

$$\Gamma(a)\Gamma(b) = \int_0^\infty \int_0^\infty t^{a-1} s^{b-1} e^{-t-s} ds \underset{t=xz, s=(1-x)z}{\overset{z=t+s, x=t/z}{=}} \int_0^\infty \int_0^1 x^{a-1}(1-x)^{b-1} z^{a+b-2} e^{-z} z \; dx \; dz \tag{B.4}$$

$$= \int_0^1 x^{a-1}(1-x)^{b-1} dx \int_0^\infty z^{a+b-1} e^{-z} dz = \int_0^1 x^{a-1}(1-x)^{b-1} dx \Gamma(a+b) \tag{B.5}$$

Hence

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx \tag{B.6}$$

More convenient for us:

$$\int_0^1 x^c (1-x)^d dx = B(c+1, d+1) \frac{\Gamma(1+c)\Gamma(1+d)}{\Gamma(2+c+d)} \tag{B.7}$$

ℓ^p norm unit ball volumes

Let $V(d,p;r)$ be the volume of a d-dimensional p-ball of radius r. Then

$$V(d,p;r) = r^d \underbrace{V(d,p;1)}_{=:V(d,p)} \tag{B.8}$$

Recursion:

$$V(1,p;r) = 2r \tag{B.9}$$

Idea: the dimension d ball is composed of slices of intervals $\times$ dimension $d-1$ balls that have at x_d radius $(1-x_d^p)^{1/p}$ (so that $x_1^p + ... + x_d^p < 1$). So

$$V(d,p;1) = \int_{-1}^1 V(d-1,p;(1-x_d^p)^{1/p}) dx_d = \int_{-1}^1 V(d-1,p)(1-x_d^p)^{(d-1)/p} dx_d \tag{B.10}$$

(could also use $V(0,p;r)=1$... but unclean because a 0-sphere is 2 points as defined, but 1 in reduced homology...) so we get

$$V(d,p;1)=\underbrace{V(1,p;1)}_{=2}\prod_{\ell=1}^{d-1}\int_{-1}^{1}(1-x^p)^{\ell/p}dx=\prod_{\ell=0}^{d-1}\int_{-1}^{1}(1-x^p)^{\ell/p}dx \tag{B.11}$$

Change of variables:

$$\int_{-1}^{1}(1-x^p)^{\ell/p}dx=2\int_{0}^{1}(1-x^p)^{\ell/p}dx \quad \overset{x=y^{1/p}}{=} 2\int_{0}^{1}(1-y)^{\ell/p}\frac{1}{p}y^{\frac{1}{p}-1}dy \tag{B.12}$$

$$\overset{(B.7)}{=} \frac{2}{p}\frac{\Gamma(1/p)\Gamma(1+\ell/p)}{\Gamma(1+(\ell+1)/p)} \tag{B.13}$$

$$V(d,p;1)=(\frac{2\Gamma(1/p)}{p})^d\prod_{\ell=0}^{d-1}\frac{\Gamma(1+\ell/p)}{\Gamma(1+(\ell+1)/p)} \tag{B.14}$$

$$=(\frac{2\Gamma(1/p)}{p})^d\frac{\Gamma(1+0/p)}{\Gamma(1+1/p)}\frac{\Gamma(1+1/p)}{\Gamma(1+2/p)}...\frac{\Gamma(1+(d-1)/p)}{\Gamma(1+d/p)} \tag{B.15}$$

$$=(\frac{2\Gamma(1/p)}{p})^d\frac{1}{\Gamma(1+d/p)} \tag{B.16}$$

<u>Higher-order 1d elliptic</u>

Want to solve

$$(-\partial^2)^m w=\mu w \quad \text{on } [a,b] \tag{B.17}$$

for Dirichlet conditions

$$w=w'=...=w^{(m-1)}=0 \quad \text{in } a,b \tag{B.18}$$

or Neumann conditions

$$w^{(m)}=...=w^{(2m-1)}=0 \quad \text{in } a,b. \tag{B.19}$$

Characteristic:

$$(-z^2)^m = \mu = r^{2m} \tag{B.20}$$

with $r > 0$; we already know the eigenvalues must be > 0 except for maybe some at $r = 0$ which we discuss later. Solution

$$z_k = iru^k \quad , \quad k = 0, ..., 2m-1 \tag{B.21}$$

where

$$u = \exp(\frac{2\pi i}{2m}) = \exp(\frac{\pi i}{m}) \tag{B.22}$$

is the first $2m$th root of 1. Particular solutions

$$\exp(z_k x) = \exp(iru^k x) \tag{B.23}$$

General solution

$$w(x) = \sum_{k=0}^{2m-1} c_k \exp(z_k x) \tag{B.24}$$

$$w^{(n)}(x) = \sum_{k=0}^{2m-1} c_k z_k^n \exp(z_k x) \tag{B.25}$$

Boundary conditions: form system $Mc = 0$ where $c = (c_0, ..., c_{2m-1})$ and

$$M = \begin{bmatrix} z_0^0 \exp(z_0 a) & \dots & z_{2m-1}^0 \exp(z_{2m-1} a) \\ z_0^1 \exp(z_0 a) & \dots & z_{2m-1}^1 \exp(z_{2m-1} a) \\ \vdots & & \vdots \\ z_0^{m-1} \exp(z_0 a) & \dots & z_{2m-1}^{m-1} \exp(z_{2m-1} a) \\ z_0^0 \exp(z_0 b) & \dots & z_{2m-1}^0 \exp(z_{2m-1} b) \\ z_0^1 \exp(z_0 b) & \dots & z_{2m-1}^1 \exp(z_{2m-1} b) \\ \vdots & & \vdots \\ z_0^{m-1} \exp(z_0 b) & \dots & z_{2m-1}^{m-1} \exp(z_{2m-1} b) \end{bmatrix} \tag{B.26}$$

$$= \begin{bmatrix} (ri)^0 u^{0\cdot 0} \exp(z_0 a) & \dots & (ri)^0 u^{(2m-1)\cdot 0} \exp(z_{2m-1} a) \\ (ri)^1 u^{0\cdot 1} \exp(z_0 a) & \dots & (ri)^1 u^{(2m-1)\cdot 1} \exp(z_{2m-1} a) \\ \vdots & & \vdots \\ (ri)^{m-1} u^{0\cdot (m-1)} \exp(z_0 a) & \dots & (ri)^{m-1} u^{(2m-1)(m-1)} \exp(z_{2m-1} a) \\ (ri)^0 u^{0\cdot 0} \exp(z_0 b) & \dots & (ri)^0 u^{(2m-1)\cdot 0} \exp(z_{2m-1} b) \\ (ri)^1 u^{0\cdot 1} \exp(z_0 b) & \dots & (ri)^1 u^{(2m-1)\cdot 1} \exp(z_{2m-1} b) \\ \vdots & & \vdots \\ (ri)^{m-1} u^{0\cdot (m-1)} \exp(z_0 b) & \dots & (ri)^{m-1} u^{(2m-1)(m-1)} \exp(z_{2m-1} b) \end{bmatrix} \tag{B.27}$$

To determine M we need to find zeros r of $\det M$. Row j contains $(ri)^{j-1}$, so we may factor those out:

$$\det M = (\underbrace{\prod_{j=0}^{m-1} (ri)^j}_{=(ri)^{(m-1)m/2}})^2 \det \begin{bmatrix} u^{0\cdot 0} \exp(z_0 a) & \dots & u^{(2m-1)\cdot 0} \exp(z_{2m-1} a) \\ \vdots & & \vdots \\ u^{0\cdot (m-1)} \exp(z_0 a) & \dots & u^{(2m-1)(m-1)} \exp(z_{2m-1} a) \\ u^{0\cdot 0} \exp(z_0 b) & \dots & u^{(2m-1)\cdot 0} \exp(z_{2m-1} b) \\ \vdots & & \vdots \\ u^{0\cdot (m-1)} \exp(z_0 b) & \dots & u^{(2m-1)(m-1)} \exp(z_{2m-1} b) \end{bmatrix} \tag{B.28}$$

This leading factor indicates there is an eigenvalue 0 with multiplicity $(m-1)m$.

$$\frac{z_k}{r} = iu^k = \exp(\frac{i\pi}{2})\exp(\frac{2\pi i}{2m}k) = \exp(\frac{2i\pi m}{4m})\exp(\frac{2\pi i}{4m}2k) \tag{B.29}$$

$$= \exp(\frac{2\pi i}{2m}(\frac{1}{2}+k)) \quad (k = 0, ..., 2m-1) \tag{B.30}$$

Note: $\Re z_k = 0$ for $k = 0$ and $k = m$. For $k = 1, ..., m-1$ we have $\Re z_k > 0$, while $\Re z_k < 0$ for $k = m+1, ..., 2m-1$.

Hence, if we take $a = -b$, the upper half of columns $2, ..., m$ have exponential growth in r while the lower half has decay, and the opposite for columns $m+2, ..., 2m-1$. In the $r \to +\infty$ limit the remaining determinant is asymptotic to

$$\det \begin{bmatrix} \vec{a}_1 & A & \vec{a}_2 & 0 \\ \vec{b}_1 & 0 & \vec{b}_2 & B \end{bmatrix} \tag{B.31}$$

where

$$\vec{a}_1 = \begin{bmatrix} u^{0\cdot 0}\exp(z_0 a) \\ \vdots \\ u^{0\cdot(m-1)}\exp(z_0 a) \end{bmatrix}, A = \begin{bmatrix} u^{1\cdot 0}\exp(z_1 a) & \dots & u^{(m-1)\cdot 0}\exp(z_m a) \\ \vdots & & \vdots \\ u^{1\cdot(m-1)}\exp(z_1 a) & \dots & u^{(m-1)\cdot(m-1)}\exp(z_m a) \end{bmatrix}, \tag{B.32}$$

$$\vec{a}_2 = \begin{bmatrix} u^{m\cdot 0}\exp(z_m a) \\ \vdots \\ u^{m\cdot(m-1)}\exp(z_m a) \end{bmatrix}, \vec{b}_1 = \begin{bmatrix} u^{0\cdot 0}\exp(z_0 b) \\ \vdots \\ u^{0\cdot(m-1)}\exp(z_0 b) \end{bmatrix}, \vec{b}_2 = \begin{bmatrix} u^{m\cdot 0}\exp(z_m b) \\ \vdots \\ u^{m\cdot(m-1)}\exp(z_m b) \end{bmatrix} \tag{B.33}$$

$$B = \begin{bmatrix} u^{(m+1)\cdot 0}\exp(z_{m+1} b) & \dots & u^{(2m-1)\cdot 0}\exp(z_{2m} b) \\ \vdots & & \vdots \\ u^{(m+1)\cdot(m-1)}\exp(z_{m+1} b) & \dots & u^{(2m-1)\cdot(2-1)}\exp(z_{2m} b) \end{bmatrix} \tag{B.34}$$

Laplace expansion across column 1 yields that the remaining determinant is ($\vec{b}_1$ is swapped across

$$\det\begin{bmatrix}\vec{a}_1 & A\end{bmatrix}\cdot\det\begin{bmatrix}\vec{b}_2 & B\end{bmatrix}-\det\begin{bmatrix}A & \vec{a}_2\end{bmatrix}\det\begin{bmatrix}B & \vec{b}_1\end{bmatrix} \tag{B.35}$$

$$=\det\begin{bmatrix}u^{0\cdot 0}\exp(z_0a) & u^{1\cdot 0}\exp(z_1a) & \dots & u^{(m-1)\cdot 0}\exp(z_{m-1}a)\\ \vdots & & & \\ u^{0\cdot(m-1)}\exp(z_0a) & u^{1\cdot(m-1)}\exp(z_1a) & \dots & u^{(m-1)\cdot(m-1)}\exp(z_{m-1}a)\end{bmatrix} \tag{B.36}$$

$$\cdot\det\begin{bmatrix}u^{m\cdot 0}\exp(z_mb) & u^{(m+1)\cdot 0}\exp(z_{m+1}b) & \dots & u^{(2m-1)\cdot 0}\exp(z_{2m-1}b)\\ \vdots & \vdots & & \vdots\\ u^{m\cdot(m-1)}\exp(z_mb) & u^{(m+1)\cdot(m-1)}\exp(z_{m+1}b) & \dots & u^{(2m-1)\cdot(2-1)}\exp(z_{2m-1}b)\end{bmatrix} \tag{B.37}$$

$$-\det\begin{bmatrix}u^{1\cdot 0}\exp(z_1a) & \dots & u^{(m-1)\cdot 0}\exp(z_{m-1}a) & u^{m\cdot 0}\exp(z_ma)\\ \vdots & & \vdots & \vdots\\ u^{1\cdot(m-1)}\exp(z_1a) & \dots & u^{(m-1)\cdot(m-1)}\exp(z_{m-1}a) & u^{m\cdot(m-1)}\exp(z_ma)\end{bmatrix} \tag{B.38}$$

$$\cdot\det\begin{bmatrix}u^{(m+1)\cdot 0}\exp(z_{m+1}b) & \dots & u^{(2m-1)\cdot 0}\exp(z_{2m-1}b) & u^{0\cdot 0}\exp(z_0b)\\ \vdots & & \vdots & \\ u^{(m+1)\cdot(m-1)}\exp(z_{m+1}b) & \dots & u^{(2m-1)\cdot(2-1)}\exp(z_{2m-1}b) & u^{0\cdot(m-1)}\exp(z_0b)\end{bmatrix} \tag{B.39}$$

If we take $a = -b$, we may use $z_{m+k} = -z_k$ to get

$$\det \begin{bmatrix} u^{0\cdot 0}\exp(z_0 a) & \dots & u^{(m-1)\cdot 0}\exp(z_{m-1}a) \\ \vdots & & \vdots \\ u^{0\cdot(m-1)}\exp(z_0 a) & \dots & u^{(m-1)\cdot(m-1)}\exp(z_{m-1}a) \end{bmatrix} \times \tag{B.40}$$

$$\det \begin{bmatrix} u^{m\cdot 0}\exp(z_0 a) & \dots & u^{(2m-1)\cdot 0}\exp(z_{m-1}a) \\ \vdots & & \vdots \\ u^{m\cdot(m-1)}\exp(z_0 a) & \dots & u^{(2m-1)\cdot(2-1)}\exp(z_{m-1}a) \end{bmatrix} \tag{B.41}$$

$$-\det \begin{bmatrix} u^{1\cdot 0}\exp(z_1 a) & \dots & u^{m\cdot 0}\exp(z_m a) \\ \vdots & & \vdots \\ u^{1\cdot(m-1)}\exp(z_1 a) & \dots & u^{m\cdot(m-1)}\exp(z_m a) \end{bmatrix} \times \tag{B.42}$$

$$\det \begin{bmatrix} u^{(m+1)\cdot 0}\exp(z_1 a) & \dots & u^{0\cdot 0}\exp(z_m a) \\ \vdots & & \vdots \\ u^{(m+1)\cdot(m-1)}\exp(z_1 a) & \dots & u^{0\cdot(m-1)}\exp(z_m a) \end{bmatrix} \tag{B.43}$$

The exp factors are the same for each column, so we get

$$\prod_{k=0}^{m-1}\exp(2z_k a)\det \begin{bmatrix} u^{0\cdot 0} & \dots & u^{(m-1)\cdot 0} \\ \vdots & & \vdots \\ u^{0\cdot(m-1)} & \dots & u^{(m-1)\cdot(m-1)} \end{bmatrix} \det \begin{bmatrix} u^{m\cdot 0} & \dots & u^{(2m-1)\cdot 0} \\ \vdots & & \vdots \\ u^{m\cdot(m-1)} & \dots & u^{(2m-1)\cdot(2-1)} \end{bmatrix} \tag{B.44}$$

$$-\prod_{k=1}^{m}\exp(2z_k a)\det \begin{bmatrix} u^{1\cdot 0} & \dots & u^{m\cdot 0} \\ \vdots & & \vdots \\ u^{1\cdot(m-1)} & \dots & u^{m\cdot(m-1)} \end{bmatrix} \det \begin{bmatrix} u^{(m+1)\cdot 0} & \dots & u^{0\cdot 0} \\ \vdots & & \vdots \\ u^{(m+1)\cdot(m-1)} & \dots & u^{0\cdot(m-1)} \end{bmatrix} \tag{B.45}$$

We can regard each determinant as a Vandermonde

$$\det \begin{bmatrix} a_1^0 & \dots & a_m^0 \\ \vdots & & \vdots \\ a_1^{m-1} & \dots & a_m^{m-1} \end{bmatrix} = \prod_{1\le j<k\le m}(a_j - a_k) \tag{B.46}$$

in this case the first determinant is

$$U := \prod_{0\leq j<k\leq m-1} (u^j - u^k) \tag{B.47}$$

which is obviously $\neq 0$, the second is

$$\prod_{0\leq j<k\leq m-1} (u^{m+j} - u^{m+k}) = u^{m\frac{m(m-1)}{2}} U \tag{B.48}$$

the third is

$$\prod_{0\leq j<k\leq m-1} (u^{m+j} - u^{m+k}) = u^{\frac{m(m-1)}{2}} U \tag{B.49}$$

and the fourth is

$$\prod_{0\leq j<k\leq m-1} (u^{m+j} - u^{m+k}) = u^{(m+1)\frac{m(m-1)}{2}} U \tag{B.50}$$

so that the remaining determinant becomes

$$U^2\Big([\prod_{k=0}^{m-1} \exp(2z_k a)]1u^{m\frac{m(m-1)}{2}} - [\prod_{k=1}^{m} \exp(2z_k a)]u^{\frac{m(m-1)}{2}}u^{(m+1)\frac{m(m-1)}{2}}\Big) \tag{B.51}$$

$$= U^2\Big([\prod_{k=0}^{m-1} \exp(2z_k a)]u^{m\frac{m(m-1)}{2}} - [\prod_{k=1}^{m} \exp(2z_k a)]u^{(m+2)\frac{m(m-1)}{2}}\Big) \tag{B.52}$$

$$= U^2[\prod_{k=0}^{m-1} \exp(2z_k a)]u^{m\frac{m(m-1)}{2}}\Big(1 - \exp(2(z_m - z_0)a)u^{2\frac{m(m-1)}{2}}\Big) \tag{B.53}$$

All factors except the last one are obviously 0. We have to find solutions r of

$$1 = \exp(2(z_m - z_0)a)u^{2\frac{m(m-1)}{2}} = \exp[2ir(\underbrace{u^m}_{=-1} - u^0)a]u^{2\frac{m(m-1)}{2}} = \exp[-4ira]u^{2\frac{m(m-1)}{2}} \tag{B.54}$$

so

$$\exp[-4ira] = u^{-2\frac{m(m-1)}{2}} = \exp(\frac{2\pi i}{2m})^{-2\frac{m(m-1)}{2}} = \exp(-2\frac{m(m-1)}{2}\frac{2\pi i}{2m}) \tag{B.55}$$

$$= \exp((1-m)i\pi) = (-1)^{m-1} \tag{B.56}$$

$$-4ira = i\pi(1-m+2k) \quad , \quad (k \in \mathbb{Z}) \tag{B.57}$$

$$-4ra = \pi(1-m+2k) \quad , \quad (k \in \mathbb{Z}) \tag{B.58}$$

$$4ra = \pi(m-1+2k) \quad , \quad (k \in \mathbb{Z}) \tag{B.59}$$

$$r = \frac{\pi(m-1+2k)}{4a} \quad , \quad (k \in \mathbb{Z}) \tag{B.60}$$

If we pick $a = -L/2$, for a length L interval, then

$$r = \pi(\frac{m+1}{2} + k) \cdot \frac{1}{L} \quad , \quad (k \in \mathbb{Z}) \tag{B.61}$$

Maple (`maple/selegue/eigenvals`) suggests this is correct.

So the solutions r are at $\frac{\pi}{L}, \frac{2\pi}{L}, \frac{3\pi}{L}, ...$ for odd m, but $\frac{\pi}{2L}, \frac{3\pi}{2L}, \frac{5\pi}{2L}, ...$ for even m. The corresponding eigenvalues are

$$\mu = \begin{cases} (\frac{\pi}{L})^{2m}, (\frac{2\pi}{L})^{2m}, (\frac{3\pi}{L})^{2m}, ..., & m \text{ odd.} \\ (\frac{\pi}{2L})^{2m}, (\frac{3\pi}{2L})^{2m}, (\frac{5\pi}{2L})^{2m}, ..., & m \text{ even} \end{cases} \tag{B.62}$$

If the operator comes with a coefficient A, all eigenvalues are multiplied by $A^{\frac{1}{2m}}$.

For a d-dimensional problem, the eigenvalues would be obtained by separation: asymptotically, for any m,

$$\mu_k \approx |\frac{k\pi}{L}|_{2m}^{2m} = (\frac{\pi}{L})^{2m}|k|_{2m}^{2m} \tag{B.63}$$

For an operator $p^{-1}(-\Delta)^m$ on $[0, L]^d$ with constant p,

$$\mu_k \approx p^{-1}|\frac{k\pi}{L}|_{2m}^{2m} = p^{-1}(\frac{\pi}{L})^{2m}|k|_{2m}^{2m} \tag{B.64}$$

We want $C_D(s; m, d, L, p)$, the number of eigenvalues $\leq s$ for $p^{-1}(-\Delta)^m$ on $[0, L]^d$ with Dirichlet conditions.

$$|\mu_k| \leq s \quad \Leftrightarrow \quad |k|_{2m} \leq (ps(\pi/L)^{-2m})^{1/2m} = (ps)^{1/2m}(\pi/L)^{-1} \tag{B.65}$$

So

$$C_D(s; m, d, L, p) \approx 2^{-d}V(d, 2m; (ps)^{1/2m}\frac{L}{\pi}) \tag{B.66}$$

(2^{-d} because we only consider the $k_1, ..., k_d > 0$ quadrant of the d-dimensional $2m$-ball)

$$\stackrel{(B.16)}{=} (ps)^{d/2m}2^{-d}L^d\pi^{-d}V(d, 2m; 1) = (\frac{\Gamma(1/2m)}{2\pi m})^d\frac{1}{\Gamma(1+d/2m)}(ps)^{d/2m}L^d \tag{B.67}$$

Some values for $C_D(1; m, d, 1, 1)$:

m	d	$C_D(1; m, d, 1, 1)$
1	2	$\frac{1}{4\pi}$
2	2	$\frac{\pi^{-1/2}}{4\Gamma(3/4)^2}$
3	2	$\frac{\Gamma(2/3)}{2\pi\sqrt{3}\Gamma(5/6)^2}$
1	3	$\frac{1}{6\pi^2}$
2	3	$\frac{\sqrt{2}}{24\Gamma(3/4)^4}$

Agree with Courant/Hilbert (VI Theorem 14 and Theorem 15).

Box counting

Consider the operator $A = p(x)^{-1}(-\Delta)^m$ on $[0, L]^d$. Let $C_D(s; m, d, L, p)$ be the number of eigenvalues $\leq s$.

Partition $[0, L]^d$ into M^d cubes. We can take an eigenfunction for some μ on one cube and extend it by 0 to the other ones. The functions obtained in this way for different cubes are obviously orthogonal.

On each small cube, $p(x)^{-1}$ is almost constant, and

$$C_D(s; m, d, \frac{L}{M}, p(x)) \approx C_D(1; m, d, 1, 1)s^{d/2m}(\frac{L}{M})^d p(x).$$

So we obtain

$$\begin{aligned} \sum_{\text{cube at } x} C_D(s; m, d, \frac{L}{M}, p(x)) &= C_D(1; m, d, 1, 1)(p(x)s)^{d/2m} \sum_x (\frac{L}{M})^d && \text{(B.68)} \\ &\approx C_D(1; m, d, 1, 1)s^{d/2m} \int\limits_{[0,L]^d} p(x)^{d/2m} dx && \text{(B.69)} \end{aligned}$$

(This agrees with e.g. Courant-hilbert (p. 436, VI.4.3 eqn (32)) which gives $p^{3/2}$ in $d = 3, m = 1$ and p for $d = 2, m = 1$.)

The extended functions are orthogonal (on different cubes by extension by 0, on the same cube by construction) and satisfy the constraints for the $[0, L]^d$ variational problem which are *weaker*: derivatives up to $m - 1$ zero on $[0, L]^d$ boundary, but not necessarily individual cube boundaries. Therefore

$$\hat{C}_D(s; m, d, L, p) \geq C_D(1; m, d, 1, 1)s^{d/2m} \int\limits_{[0,L]^d} p(x)^{d/2m} dx.$$

The eigenvalue distribution for the Neumann problem has the same leading-order asymptotic term but the constraints for the $[0, 1]^d$ variational problem are *stronger* now (H^m not only on each side of a subcube, but *across* cube boundaries, so e.g. jumps across boundaries no longer allowed), so

$$\begin{aligned}\hat{C}_N(s;m,d,L,p) &\leq C_D(1;m,d,1,1)s^{d/2m}\int\limits_{[0,L]^d} p(x)^{d/2m}dx \\ &\approx C_N(1;m,d,1,1)s^{d/2m}\int\limits_{[0,L]^d} p(x)^{d/2m}dx.\end{aligned}$$

Finally,

$$\hat{C}_N(s;m,d,L,p) \leq \hat{C}_D(s;m,d,L,p) \tag{B.70}$$

since the Dirichlet problem minimizes over the smaller space ($H_0^m[0,L]^d$ as opposed to $H^m[0,L]^d$). But since the leading-order terms for the subcubes are the same we get

$$\hat{C}_N(s;m,d,L,p) \approx \hat{C}_D(s;m,d,L,p) \tag{B.71}$$

$$\overset{(B.67)}{\approx} (\frac{\Gamma(1/2m)}{2\pi m})^d \frac{1}{\Gamma(1+d/2m)} s^{d/2m}\int\limits_{[0,L]^d} p(x)^{d/2m}dx. \tag{B.72}$$

Eigenvalue distribution

Another formulation of the same result: number eigenvalues in non-decreasing fashion by $k \in \mathbb{N}$. The distribution formula says:

$$C(s) = \#\{k : \mu_k \leq s\} = (c + o(1))s^{d/2m} \tag{B.73}$$

Hence for any $\delta > 0$ and for s sufficiently large ($= k$ sufficiently large),

$$(c-\delta)s^{d/2m} \leq C(s) \leq (c+\delta)s^{d/2m} \tag{B.74}$$

hence

$$(\frac{C(s)}{c+\delta})^{2m/d} \leq s \leq (\frac{C(s)}{c-\delta})^{2m/d} \tag{B.75}$$

so since $C(\mu_k) = k$,

$$(\frac{k}{c+\delta})^{2m/d} \leq \mu_k \leq (\frac{k}{c-\delta})^{2m/d} \tag{B.76}$$

which means

$$\mu_k = (c^{-2m/d} + o_k(1))k^{2m/d}, \tag{B.77}$$

as desired.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adams, R. & Fournier, J. 2003. *Sobolev spaces*, Vol. 140, Academic Pr.

Aerts, M., Claeskens, G. & Wand, M. 2002. Some theory for penalized spline generalized additive models, *Journal of statistical planning and inference* **103**(1-2): 455–470.

Agmon, S. 2010. *Lectures on elliptic boundary value problems*, American Mathematical Society (RI).

Aziz, A. & Babuska, I. 1972. *Part I, Survey lectures on the mathematical foundations of the finite element method, The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations (AK Aziz, ed.)*, Academic Press, New York, NY.

Cattaneo, M., Jansson, M. & Newey, W. 2010. Alternative asymptotics and the partially linear model with many regressors, *Technical report*, Working paper, MIT.

Claeskens, G., Krivobokova, T. & Opsomer, J. 2009. Asymptotic properties of penalized spline estimators, *Biometrika* **96**(3): 529–544.

Cox, D. 1984. Multivariate smoothing spline functions, *SIAM Journal on Numerical Analysis* pp. 789–813.

Craven, P. & Wahba, G. 1978. Smoothing noisy data with spline functions, *Numerische Mathematik* **31**(4): 377–403.

De Boor, C. 1976. Splines as linear combinations of b-splines. a survey.

De Boor, C. 2001. *A practical guide to splines*, Vol. 27, Springer Verlag.

Donald, S. & Newey, W. 1994. Series estimation of semilinear models, *Journal of Multivariate Analysis* **50**(1): 30–40.

Eilers, P. & Marx, B. 1996. Flexible smoothing with b-splines and penalties, *Statistical science* pp. 89–102.

Eubank, R. 1999. *Nonparametric regression and spline smoothing*, Vol. 157, CRC Press.

Fix, G. 1972. Effects of quadrature errors in finite element approximation of steady state, eigenvalue and parabolic problems, *Mathematical Foundations of the Finite Element Method* pp. 525–556.

Green, P. & Silverman, B. 1994. *Nonparametric regression and generalized linear models: a roughness penalty approach*, Vol. 58, Chapman & Hall/CRC.

Hall, P. & Opsomer, J. 2005. Theory for penalised spline regression, *Biometrika* **92**(1): 105–118.

Härdle, W., Müller, M., Sperlich, S. & Werwatz, A. 2004. *Nonparametric and semiparametric models*, Springer, Heidelberg, Germany.

Huang, J. 1998. Functional anova models for generalized regression, *Journal of multivariate analysis* **67**(1): 49–71.

Huang, J. 2003a. Asymptotics for polynomial spline regression under weak conditions, *Statistics & probability letters* **65**(3): 207–216.

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

Imbens, G. & Wooldridge, J. 2008. Recent developments in the econometrics of program evaluation, *Technical report*, National Bureau of Economic Research.

Kauermann, G., Krivobokova, T. & Fahrmeir, L. 2009. Some asymptotic results on generalized penalized spline smoothing, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(2): 487–503.

Kohler, M. & Krzyzak, A. 2001. Nonparametric regression estimation using penalized least squares, *Information Theory, IEEE Transactions on* **47**(7): 3054–3058.

Krivobokova, T., Kneib, T. & Claeskens, G. 2010. Simultaneous confidence bands for penalized spline estimators, *Journal of the American Statistical Association* **105**(490): 852–863.

Li, Y. & Ruppert, D. 2008. On the asymptotics of penalized splines, *Biometrika* **95**(2): 415–436.

Newey, W. 1997. Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics* **79**(1): 147–168.

Nychka, D. 1995. Splines as local smoothers, *The Annals of Statistics* pp. 1175–1197.

O'Sullivan, F. 1986. A statistical perspective on ill-posed inverse problems, *Statistical science* pp. 502–518.

Póo, J. 1999. Constrained smoothing splines, *Econometric theory* **15**(1): 114–138.

Robinson, P. 1988. Root-n-consistent semiparametric regression, *Econometrica: Journal of the Econometric Society* pp. 931–954.

Ruppert, D., Wand, M. & Carroll, R. 2003. *Semiparametric regression*, Vol. 12, Cambridge Univ Pr.

Schumaker, L. 1981. *Spline functions: basic theory.*

Schwetlick, H. & Kunert, V. 1993. Spline smoothing under constraints on derivatives, *BIT Numerical Mathematics* **33**(3): 512–528.

Speckman, P. 1981. The asymptotic integrated mean square error for smoothing noisy data by splines, *Unpublished manuscript* .

Speckman, P. 1985. Spline smoothing and optimal rates of convergence in nonparametric regression models, *The Annals of Statistics* pp. 970–983.

Stock, J. 1989. Nonparametric policy analysis, *Journal of the American Statistical Association* pp. 567–575.

Stone, C. 1994. The use of polynomial splines and their tensor products in multivariate function estimation, *The Annals of Statistics* pp. 118–171.

Utreras, F. 1979. Cross-validation techniques for smoothing spline functions in one or two dimensions, *Smoothing techniques for curve estimation* pp. 196–232.

Utreras, F. 1988. Convergence rates for multivariate smoothing spline functions, *Journal of approximation theory* **52**(1): 1–27.

Wahba, G. 1975. Smoothing noisy data with spline functions, *Numerische Mathematik* **24**(5): 383–393.

Wahba, G. 1990. *Spline models for observational data*, Vol. 59, Society for Industrial Mathematics.

Wand, M. 1999. On the optimal amount of smoothing in penalised spline regression, *Biometrika* **86**(4): 936–940.

Wand, M. & Ormerod, J. 2008. On semiparametric regression with o'sullivan penalized splines, *Australian & New Zealand Journal of Statistics* **50**(2): 179–198.

Yu, Y. & Ruppert, D. 2002. Penalized spline estimation for partially linear single-index models, *Journal of the American Statistical Association* **97**(460): 1042–1054.

Zhou, S., Shen, X. & Wolfe, D. 1998. Local asymptotics for regression splines and confidence regions, *The Annals of Statistics* **26**(5): 1760–1782.

Zhou, S. & Wolfe, D. 2000. On derivative estimation in spline regression, *Statistica Sinica* **10**(1): 93–108.