PENALIZED LIKELIHOOD FUNCTIONAL REGRESSION

Pang Du and Xiao Wang

Virginia Tech and Purdue University

Supplementary Material

The supplementary material collects the proofs for all the lemmas in Appendix A.2 of the paper.

S1 Proof of Lemmas

Proof of Lemma A.1. Denote

$$L(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\Big\{Y_i\eta(X_i;\beta) - b(\eta(X_i;\beta))\Big\} + \frac{\lambda}{2}\int_0^1 [\beta^{(m)}(t)]^2 dt.$$

For any $\delta > 0$ and $\beta_1 \in W_2^m$,

$$L(\beta+\delta\beta_1) - L(\beta_1) = -\frac{1}{n}\sum_{i=1}^{n}\Big\{\delta Y_i\eta(X_i;\beta_1) - [b(\eta(X_i;\beta+\delta\beta_1) - b(\eta(X_i;\beta))]\Big\} + \delta\lambda\int_0^1 \beta^{(m)}(t)\beta_1^{(m)}(t)dt + O(\delta^2)$$

So, by letting $\delta \to 0$, it is easy to see that the necessary and sufficient condition for β to minimize (2.2) is, for any $\beta_1 \in W_2^m$, $L_1(\beta, \beta_1) = 0$, where

$$L_1(\beta,\beta_1) = -\frac{1}{n}\sum_{i=1}^{n}\Big\{Y_i\eta(X_i;\beta_1) - b'(\eta(X_i;\beta))\eta(X_i;\beta_1)\Big\} + \lambda\int_0^1 \beta^{(m)}(t)\beta_1^{(m)}(t)dt. \tag{S1.1}$$

In (S1.1), letting $\beta_1(t) = t^k$, $k = 0, 1, \ldots, m$, we obtain m equalities in (A.12). For example, when $\beta_1(t) = 1$, we have

$$\frac{1}{n}\sum_{i=1}^{n} b'(\eta(X_i;\beta))X_i^{(-1)}(1) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i^{(-1)}(1).$$

Further, since

$$\eta(X_i; 1-t) = \int_0^1 X_i(t)\int_t^1 ds dt = \int_0^1 X_i^{(-1)}(t)dt = X_i^{(-2)}(1),$$

when $\beta_1(t) = t$,

$$\begin{aligned} -\frac{1}{n}\sum_{i=1}^{n}\Big\{Y_i\eta(X_i;t) - b'(\eta(X_i;\beta))\eta(X_i;t)\Big\} =& \frac{1}{n}\sum_{i=1}^{n}\Big\{Y_i\eta(X_i;1-t) - b'(\eta(X_i;\beta))\eta(X_i;1-t)\Big\} \\ =& \frac{1}{n}\sum_{i=1}^{n} Y_i X_i^{(-2)}(1) - \frac{1}{n}\sum_{i=1}^{n} b'(\eta(X_i;\beta))X_i^{(-2)}(1). \end{aligned}$$

Hence,

$$\frac{1}{n}\sum_{i=1}^{n} b'(\eta(X_i;\beta))X_i^{(-2)}(1) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i^{(-2)}(1).$$

Following the same procedure, it may be shown that (A.12) holds.

Next, using these equalities, we show that $L_1(\beta, \beta_1) = \int_0^1 L_2(\beta)\beta_1^{(m)}(t)dt$, where

$$L_2(\beta) = \lambda\beta^{(m)}(t) + (-1)^m\Big\{\frac{1}{n}\sum_{i=1}^{n} b'(\eta(X_i;\hat\beta))X_i^{(-m)}(t) - \frac{1}{n}\sum_{i=1}^{n} Y_i X_i^{(-m)}(t)\Big\}.$$

Note that

$$\begin{aligned} \int_0^1 X_i(s)\beta_1(s) &= \beta_1(1)X_i^{(-1)}(1) - \int_0^1 X_i^{(-1)}(s)\beta'(s)ds \\ &= \beta_1(1)X_i^{(-1)}(1) - \beta'(1)X_i^{(-2)}(1) + \int_0^1 X_i^{(-2)}(s)\beta''(s)ds \\ &= \cdots \\ &= \sum_{k=0}^{m-1}(-1)^k\beta_1^{(k)}(1)X_i^{(-k-1)} + (-1)^m\int_0^1 X_i^{(-m)}(s)\beta^{(m)}(s)ds. \end{aligned}$$

Plugging this into $L_1(\beta, \beta_1) = 0$, together with (A.12), we have $L_1(\beta, \beta_1) = \int_0^1 L_2(\beta)\beta_1^{(m)}(t)dt$. Finally, since $L_1(\beta, \beta_1) = 0$ for any $\beta_1 \in W_2^m$, we have $L_2(\beta) = 0$ a.e., which completes the proof of the lemma. □

Proof of Lemma A.2. Observe that

$$\int_0^1 Z_i(s)\tilde{\beta}(s)ds = \sum_{k=0}^{m-1}(-1)^k\hat{\beta}^{(k)}(1)Z_i^{(-k-1)}(1) + (-1)^m\int_0^1 Z_i^{(-m)}(s)\hat{\beta}^{(m)}(s)ds.$$

Hence, for $j = 1, \ldots, m$,

$$\int_0^1 \hat{G}^{(-j,0)}(1,s)\tilde{\beta}(s)ds = \sum_{k=0}^{m-1}(-1)^k\tilde{\beta}^{(k)}(1)\hat{G}^{(-j,-k)}(1,1)+(-1)^m\int_0^1 \hat{G}^{(-j,-m)}(1,s)\tilde{\beta}^{(m)}(s)ds.$$

From (A.16), we have

$$\hat{H}\tilde{\beta}_v(1) = \frac{1}{n}\sum_{i=1}^n \epsilon_i\widetilde{Z}(1) - (-1)^m\int_0^1 \hat{G}^{(-m\bullet)}(s)\tilde{\beta}^{(m)}(s)ds, \tag{S1.2}$$

where $\tilde{\beta}_v(1) = \left[\tilde{\beta}(1), -\tilde{\beta}'(1), \cdots, (-1)^{m-1}\tilde{\beta}^{(m-1)}(1)\right]^T$. Hence, (A.17) follows by plugging (S1.2) into (A.15). □

Proof of Lemma A.3. Direct calculation yields

$$\frac{1}{n}\sum_{i=1}^n \epsilon_i\tilde{Z}(1) = \hat{H}\beta_{0v}(1) + (-1)^m\int_0^1 \hat{G}^{(-m\bullet)}(s)\beta_0^{(m)}(s)ds + \frac{1}{n}\sum_{i=1}^n \epsilon_i\widetilde{Z}_i(1).$$

Combining this with (S1.2) leads

$$\tilde{\beta}_v(1) - \beta_{0v}(1) = (-1)^{m+1}\int_0^1 \hat{H}^{-1}\hat{G}^{(-m\bullet)}(s)\left(\tilde{\beta}^{(m)}(s) - \beta_0^{(m)}(s)\right)ds + \frac{1}{n}\sum_{i=1}^n \epsilon_i\hat{H}^{-1}\widetilde{Z}_i(1).$$

Therefore,

$$\begin{aligned} & \int_0^1 Z(s)\left(\hat{\beta}(s) - \beta_0(s)\right)ds \qquad \text{(S1.3)} \\ = \;& \widetilde{Z}(1)^T\left(\tilde{\beta}_v(1) - \beta_{0v}(1)\right) + (-1)^m\int_0^1 Z^{(-m)}(s)\left(\tilde{\beta}^{(m)}(s) - \beta_0^{(m)}(s)\right)ds \\ = \;& (-1)^m\int_0^1 \hat{U}(s;Z)\left(\tilde{\beta}^{(m)}(s) - \beta_0^{(m)}(s)\right)ds + \frac{1}{n}\sum_{i=1}^n \epsilon_i\widetilde{Z}(1)^T\hat{H}^{-1}\widetilde{Z}_i(1). \end{aligned}$$

From Lemma A.2,

$$\tilde{\beta}^{(m)} - \beta_0^{(m)} = -\lambda\hat{Q}^+\beta_0^{(m)} + (-1)^m\frac{1}{n}\sum_{i=1}^n \epsilon_i\hat{Q}^+\hat{U}(t;Z_i).$$

Plugging this into (S1.3) leads to (A.18). This completes the proof of the lemma. □

Proof of Lemma A.4. Let $H(K)$ denote the reproducing kernel space associated withe the kernel K. For two covariance kernel K and L on $[0,1]^2$ we write $K \ll L$ if $cL - K$ is nonnegative definite for some positive constant c. Then, $K \ll L$ implies $H(K) \subset H(L)$. From (A.21),

$$Q \ll G \Longleftrightarrow H(Q) \subset H(G).$$

Let $\lambda_1(G) \geq \lambda_2(G) \geq \cdots > 0$ be the eigenvalues of G. Let ϕ_k be the eigenfunction of Q which corresponds to κ_k such that $Q\phi_k = \kappa_k \phi_k$. The minimax principle [see Weidmann(1980), Theorem 7.3] yields

$$Q \ll G \Longrightarrow \kappa_k \leq c\lambda_k(G) \text{ for some positive constant } c. \tag{S1.4}$$

We may write $H(G) = H(Q) \oplus H(G - Q)$, and $H(G - Q)$ is the orthogonal complement of $H(Q)$. Note that $H(G - Q)$ is a finite dimensional space with rank m. Let $f_1, \ldots, f_m$ be an orthonormal base of $H(G - Q)$, and let $\mathcal{F}^\perp$ denote the sets of normalized functions in L_2 that are orthogonal to $f_1, \ldots, f_m$. Let $\Xi_k^\perp = \text{span}\{\phi_k, \phi_{k+1}, \ldots, \}$. The minimax principle implies

$$\lambda_{k+m}(G) \leq \sup_{f \in \mathcal{F}^\perp \cap \Xi_k^\perp} \langle Gf, f\rangle = \sup_{f \in \mathcal{F}^\perp \cap \Xi_k^\perp} \langle Qf, f\rangle \leq \sup_{f \in \Xi_k^\perp} \langle Qf, f\rangle = \kappa_k. \tag{S1.5}$$

From A1, Ritter et al. (1995) showed that $\lambda_k(G) \asymp k^{-2(m+r)}$. Further, (S1.4) and (S1.5) yield that $\kappa_k \asymp k^{-2(m+r)}$. Since $H(G - Q)$ is not an empty set, the operator Q is not a strictly positive definite operator, and Q has m zero eigenvalues. □

Proof of Lemma A.5. In the lemma, we discuss the relationship between $\hat{\beta}$ and $\tilde{\beta}$. Denote

$$\Delta(\beta_1, \beta_2; X_i) = b'(\eta(X_i; \beta_1)) - b'(\eta(X_i; \beta_2)) - b''(\eta(X_i; \beta_2))\eta(X_i; \beta_1 - \beta_2).$$

Let δ_1 satisfy

$$(-1)^m \lambda \delta_1^{(m)}(t) + \frac{1}{n}\sum_{i=1}^n b''(\eta(X_i; \tilde{\beta}))\eta(X_i; \delta_1) X_i^{(-m)}(t) = -\frac{1}{n}\sum_{i=1}^n X_i^{(-m)}(t)\Delta(\tilde{\beta}, \beta_0; X_i).$$

For $k \geq 2$, let δ_k satisfy

$$(-1)^m \lambda \delta_k^{(m)}(t) + \frac{1}{n}\sum_{i=1}^n b''(\eta(X_i;\delta_{k-1}))\eta(X_i;\delta_k)X_i^{(-m)}(t) = -\frac{1}{n}\sum_{i=1}^n X_i^{(-m)}(t)\Delta(\tilde{\beta}+\sum_{j=1}^{k-1}\delta_j, \tilde{\beta}+\sum_{j=1}^{k-2}\delta_j; X_i).$$

By summing all these equations together, it is easy to verify that $\hat{\beta} = \tilde{\beta} + \sum_{k=1}^{\infty}\delta_k$. Following the same discussion in Lemma A.1, for any $\beta_1 \in W_2^m$,

$$\frac{1}{n}\sum_{i=1}^n \left\{ b''(\eta(X_i;\tilde{\beta}))\eta(X_i;\delta_1)\eta(X_i;\beta_1) - \sqrt{b''(\eta(X_i;\tilde{\beta}))}\,\eta(X_i;\beta_1)\Delta_i \right\} + \lambda \int_0^1 \delta_1^{(m)}(t)\beta_1^{(m)}(t)dt = 0,$$

where $\Delta_i = \Delta(\tilde{\beta}, \beta_0; X_i)$. By choosing $\beta_1 = \delta_1$,

$$\frac{1}{n}\sum_{i=1}^n b''(\eta(X_i;\tilde{\beta}))[\eta(X_i;\delta_1)]^2 - \frac{1}{n}\sum_{i=1}^n \sqrt{b''(\eta(X_i;\tilde{\beta}))}\,\eta(X_i;\beta_1)\Delta_i + \lambda\int_0^1 [\delta_1^{(m)}(t)]^2 dt = 0.$$

Therefore,

$$\begin{aligned}\|\delta_1\|_{\Gamma_n}^2 &\leq \frac{c_2}{nc_1}\sum_{i=1}^n b''(\eta(X_i;\tilde{\beta}))[\eta(X_i;\delta_1)]^2 + \lambda\int_0^1 [\delta_1^{(m)}(t)]^2 dt \\ &\leq \frac{c_2}{nc_1}\sum_{i=1}^n \sqrt{b''(\eta(X_i;\tilde{\beta}))}\,\eta(X_i;\delta_1)\Delta_i \leq C_1\|\delta_1\|_{\Gamma_n}\left\{\frac{1}{n}\sum_{i=1}^n[\eta(X_i;\tilde{\beta}-\beta)]^4\right\}^{1/2}.\end{aligned}$$

Recall that the (ν_j, ψ_j) are the eigenvalue-eigenfunction pairs for the covariance kernel K. By

A4,

$$
\begin{aligned}
&E\Big\{\frac{1}{n}\sum_{i=1}^{n}[\eta(X_i;\tilde\beta-\beta)]^4\Big|\tilde\beta\Big\}\\
=&E\Big\{\sum_{j_1}^{\infty}\sum_{j_2}^{\infty}\sum_{j_3}^{\infty}\sum_{j_4}^{\infty}\zeta_{j_1}\zeta_{j_2}\zeta_{j_3}\zeta_{j_4}\eta(\psi_{j_1},\tilde\beta-\beta)\eta(\psi_{j_2},\tilde\beta-\beta)\eta(\psi_{j_3},\tilde\beta-\beta)\eta(\psi_{j_4},\tilde\beta-\beta)\Big|\tilde\beta\Big\}\\
=&\sum_{j=1}^{\infty}E(\zeta_j^4)[\eta(\psi_j,\tilde\beta-\beta)]^4+\sum_{j_1}\sum_{j_2\neq j_1}\nu_{j_1}\nu_{j_2}[\eta(\psi_{j_1},\tilde\beta-\beta)]^2[\eta(\psi_{j_2},\tilde\beta-\beta)]^2\\
\le&C_2\sum_{j=1}^{\infty}\nu_j^2[\eta(\psi_j,\tilde\beta-\beta)]^4+\sum_{j_1}\sum_{j_2\neq j_1}\nu_{j_1}\nu_{j_2}[\eta(\psi_{j_1},\tilde\beta-\beta)]^2[\eta(\psi_{j_2},\tilde\beta-\beta)]^2\\
\le&(1+C_2)\sum_{j_1=1}^{\infty}\nu_{j_1}[\eta(\psi_{j_1},\tilde\beta-\beta)]^2\sum_{j_2=1}^{\infty}\nu_{j_2}[\eta(\psi_{j_2},\tilde\beta-\beta)]^2\\
=&(1+C_2)\|\tilde\beta-\beta\|_K^4.
\end{aligned}
$$

So,

$$E\|\delta_1\|_\Gamma^2=E\|\delta_1\|_{\Gamma_n}^2\le C_2\Big(\lambda+n^{-1}\lambda^{-\frac{1}{2(m+r)}}+n^{-1}\Big)^2.$$

Similarly, we may establish, for $k\ge 1$,

$$E\|\delta_k\|_\Gamma^2=E\|\delta_1\|_{\Gamma_n}^2\le C_2\Big(\lambda+n^{-1}\lambda^{-\frac{1}{2(m+r)}}+n^{-1}\Big)^{2k}.$$

Therefore, $\sum_{k=1}^{\infty}\delta_k$ is of order $O_p(\lambda+n^{-1}\lambda^{-\frac{1}{2(m+r)}}+n^{-1})$. This shows the lemma. □