Existence and Uniqueness of Penalized Least Square Estimation for Smoothing Spline Nonlinear Nonparametric Regression Models

Chunlei Ke and Yuedong Wang *

March 10, 2004

1 The Model

A smoothing spline nonlinear nonparametric regression model (SSNNRM) assumes that

$$y_i = \mathcal{N}_i(g_1, \cdots, g_r) + \epsilon_i, \quad i = 1, \cdots, n, \tag{1}$$

where $\mathcal{N}_i$ are known nonlinear functionals, $\boldsymbol{g} = (g_1, \cdots, g_r)$ are unknown functions, and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ are random errors. Without loss of generality, we assume that $r = 2$. As in O'Sullivan (1990), we express design points x explicitly in the functional $\mathcal{N}_i$: $\mathcal{N}_i(g_1, g_2) = \eta(g_1, g_2; x_i)$, where η is a known nonlinear functional. In the following sections, $\eta(g_1, g_2; x)$ is sometimes also represented by $\eta(g_1, g_2)$ or η when the meaning is clear.

g_1 and g_2 are estimated as minimizers of the following penalized least squares (PLS)

$$l_{n\boldsymbol{\lambda}}(g_1, g_2) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \eta(g_1, g_2; x_i))^2 + \lambda_1 J_1(g_1) + \lambda_2 J_2(g_2), \tag{2}$$

where $g_j \in \mathcal{H}_j$, $j = 1, 2$, $\mathcal{H}_j$ are Hilbert spaces, λ_1, $\lambda_2 \geq 0$ are smoothing parameters and J_1, J_2 are penalty functionals.

In this supplement, adapting the frameworks in Cox (1988), Cox and O'Sullivan (1990) and O'Sullivan (1990), we establish the existence and uniqueness of the solution to (2). To save space, some details are omitted in the following sections.

*Chunlei Ke (email: cke@sjm.com) is Statistician, St. Jude Medical, Cardiac Rhythm Management Division, Sylmar, CA 91342. Yuedong Wang (email: yuedong@pstat.ucsb.edu) is Professor, Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106. Yuedong Wang's research was supported by NIH Grant R01 GM58533. Address for correspondence: Yuedong Wang, Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106.

2 Some Assumptions

We make the following assumptions. See O'Sullivan (1990) for discussion on these assumptions.

Assumption A.1 (Observational model)

(i) $x_i \in \Omega_0 \subset R^d$, where Ω_0 is a bounded, open and simply connected set with C^∞ boundary.

(ii) $\{x_i, i = 1, \cdots, n\}$ is a random sample from a probability density function f, which is strictly bounded away from zero and infinity on Ω_0. Denote F and F_n as the CDF and empirical CDF of sample x_i's.

(iii) The x_i's and ϵ_i's are independent.

Denote the true functions in model (1) as g_{10} and g_{20}. It is obvious that g_{10} and g_{20} are minimizers of

$$l_0(g_1, g_2) = \int_{\Omega_0} (\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x))^2 dF(x). \tag{3}$$

As Cox and O'Sullivan (1990) and O'Sullivan (1990), we define $l_{\boldsymbol{\lambda}}(g_1, g_2)$ as the limiting regularization functional of $l_{n\boldsymbol{\lambda}}$:

$$l_{\boldsymbol{\lambda}}(g_1, g_2) = \int_{\Omega_0} (\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x))^2 dF(x) + \lambda_1 J_1(g_1) + \lambda_2 J_2(g_2). \tag{4}$$

Assumption A.2 (Parameter space)

(i) $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces with norms $||\cdot||_1$, $||\cdot||_2$ and inner products $< \cdot >_1$ and $< \cdot >_2$. For simplicity of notation, the subscripts will be dropped when there is no confusion. Let $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ with the norm $||(g_1, g_2)|| = ||g_1|| + ||g_2||$.

(ii) The penalty functionals are of a quadratic form $J_1(g_1) =< g_1, W_1 g_1 >$ and $J_2(g_2) =< g_2, W_2 g_2 >$, where W_1 and W_2 are nonnegative definite linear operators on $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively.

(iii) There are bounded linear operators $L_1 : \mathcal{H}_1 \to \mathcal{L}_2(\Omega_0)$ and $L_2 : \mathcal{H}_2 \to \mathcal{L}_2(\Omega_0)$ with zero null-space. There are strictly positive constants M_1 and M_2 such that for all $g_i \in \mathcal{H}_i$,

$$M_1 ||g_i||^2 \leq ||L_i g_i||^2_{\mathcal{L}_2} + < g_i, W_i g_i > \leq M_2 ||g_i||^2, \quad i = 1, 2.$$

We then define bounded linear operators U_1 and U_2 by

$$(L_i g_i, L_i h_i)_{\mathcal{L}_2} =< g_i, U_i h_i >, \quad g_i, h_i \in \mathcal{H}_i, \quad i = 1, 2.$$

U_1 and U_2 are compact, and satisfy

$$M_1||g_i||^2 \leq < g_i, U_i g_i > + < g_i, W_i g_i > \leq M_2||g_i||^2, \quad g_i \in \mathcal{H}_i, \quad i = 1, 2.$$

We now define normed spaces based on U_i and W_i as O'Sullican (1990). For $i = 1, 2$, there is a sequence $\{\phi_{vi} : v = 1, 2, \cdots\}$ of eigenfunctions and $\{\gamma_{vi}, v = 1, 2, \cdots\}$ of eigenvalues that satisfy

$$< \phi_{\mu i}, U_i \phi_{vi} >= \delta_{v\mu}, \ \ < \phi_{\mu i}, W_i \phi_{vi} >= \gamma_{vi}\delta_{v\mu},$$

where v and μ are any positive integers and $\delta_{v\mu}$ is Kronecker's delta. For $b \geq 0$, define

$$||g_i||_{bi} = \{\sum_{v=1}^{\infty}(1 + \gamma_{vi}^b) < g_i, U_i\phi_{vi} >_i^2\}^{1/2},$$

and let $\mathcal{H}_{bi}$ be the normed linear space obtained by completing $\{g \in \mathcal{H}_i : ||g||_{bi} < \infty\}$. $\mathcal{H}_{bi}$ is a Hilbert space with inner product

$$< g, h >_{bi}= \sum_{v=1}^{\infty}(1 + \gamma_{vi}^b) < g, U_i\phi_{vi} >< h, U_i\phi_{vi} > .$$

It is easy to see that $\mathcal{H}_i = \mathcal{H}_{1i}$.

Assumption A.3 (Property of η, g_{10} and g_{20}) For some $\alpha_1, \alpha_2 \in (0, 1]$, there are $g_{10} \in \mathcal{H}_{\alpha_1}$, $g_{20} \in \mathcal{H}_{\alpha_2}$ and their neighbourhoods $\mathcal{N}_{10} \subset \mathcal{H}_{\alpha_1}$ and $\mathcal{N}_{20} \subset \mathcal{H}_{\alpha_2}$, such that

(i) $\eta(g_1, g_2; x)$ is three times continuously Fréchet differentiable with respect to g_1 and g_2 in $\mathcal{N}_{10} \times \mathcal{N}_{20}$. (g_{10}, g_{20}) are the unique root of $D_1 l_0(g_1, g_2) = 0$ and $D_2 l_0(g_1, g_2) = 0$.

(ii) For some s such that $m\alpha > s > d/2$, there exists $M > 0$ such that for $g_1 \in \mathcal{N}_{10}$ and $g_2 \in \mathcal{N}_{20}$, $||\eta(g_1, g_2; x)||_{W_2^s} < M$, where W_2^s is a Sobolev space with noninteger order s.

For simplicity of notations, we assume that $\alpha_1 = \alpha_2 = \alpha$. Construction and proof are similar for the case of $\alpha_1 \neq \alpha_2$.

3 Linearizations

In this section, we approximate the systematic and stochastic errors in the PLS estimates using Taylor series expansions. We first derive a Taylor series expansion for a bivariate nonlinear operator.

Theorem 1 *Let $f : D(f) \subseteq X \times Y \to Z$, where X, Y and Z are Banach spaces. If f'' exists at (x, y), then the partial Fréchet derivatives f_{xx}, f_{xy}, f_{yx} and f_{yy} exist at (x, y) and for any $h, a \in X$, $k, b \in Y$,*

$$f''(x, y)(h, k)(a, b) = f_{xx}(x, y)ha + f_{xy}(x, y)ka + f_{yx}(x, y)hb + f_{yy}(x, y)kb. \tag{5}$$

[Proof] By definition, for any $(h,k),(a,b) \in X \times Y$,

$$f'(x+h,y+k)(a,b) = f'(x,y)(a,b) + f''(x,y)(h,k)(a,b) + o(||(h,k)||),$$

as $||(h,k)|| \to 0$. For the norm on the product space $X \times Y$, choose $||(h,k)|| = ||h|| + ||k||$. Setting $b = k = 0$, we have

$$f'(x+h,y)(a,0) = f'(x,y)(a,0) + f''(x,y)(h,0)(a,0) + o(||h||).$$

Applying equation (15) in Chapter 4 of Zeidler (1985), we have

$$f_x(x+h,y)a = f_x(x,y)a + f''(x,y)(h,0)(a,0) + o(||h||).$$

Therefore,

$$f_{xx}(x,y)ha = f''(x,y)(h,0)(a,0).$$

Similarly by setting $a = h = 0$, $(h,b) = (0,0)$ and $(a,k) = (0,0)$ respectively, we get

$$\begin{aligned} f_{yy}(x,y)kb &= f''(x,y)(0,k)(0,b), \\ f_{xy}(x,y)ka &= f''(x,y)(0,k)(a,0), \\ f_{yx}(x,y)hb &= f''(x,y)(h,0)(0,b). \end{aligned}$$

Addition gives (5).

Based on the generalized Taylor's theorem (Theorem 4.A (b), Zeidler, 1985), equation (15) in Chapter 4 of Zeidler (1985) and Theorem 1, we have the first order Taylor series expansion

$$f(x+h,y+k) = f(x,y) + \int_0^1 [f_x(x+\tau h, y+\tau k)h + f_y(x+\tau h, y+\tau k)k]d\tau, \tag{6}$$

and the second order Taylor series expansion

$$f(x+h,y+k) = f(x,y) + f_x(x,y)h + f_y(x,y)k + R, \tag{7}$$

where the remainder

$$\begin{aligned} R &= \int_0^1 (1-\tau) f''(x+\tau h, y+\tau k)(h,k)(h,k)d\tau \\ &= \int_0^1 (1-\tau)[f_{xx}(x+\tau h, y+\tau k)hh + f_{xy}(x+\tau h, y+\tau k)kh + \\ &\qquad f_{yx}(x+\tau h, y+\tau k)hk + f_{yy}(x+\tau h, y+\tau k)kk]d\tau. \end{aligned}$$

We now use Taylor expansions to approximate the systematic and stochastic errors of the estimates. Denote D_1 and D_2 as the partial Fréchet derivatives of η with respect to g_1 and g_2 respectively, D_{11} and D_{22} as the second partial Fréchet derivatives of η with respect to g_1 and

g_2 respectively, and D_{12} as the second partial Fréchet derivative of η with respect to g_1 and g_2 (Zeidler, 1985). Higher Fréchet partial derivatives are denoted similarly. Let

$$Z_i(g_1, g_2) = \frac{1}{2} D_i l_{\boldsymbol{\lambda}}(g_1, g_2), \quad i = 1, 2.$$

Note that Z_i also depend on $\boldsymbol{\lambda}$, which is not expressed explicitly. For $g_{10} + h_1 \in \mathcal{N}_{10}$, $g_{20} + h_2 \in \mathcal{N}_{20}$ and $i = 1, 2$,

$$\begin{aligned}
& Z_i(g_{10} + h_1, g_{20} + h_2) \\
= \; & Z_i(g_{10}, g_{20}) + D_1 Z_i(g_{10}, g_{20}) h_1 + D_2 Z_i(g_{10}, g_{20}) h_2 \\
& + \int_0^1 [D_{11} Z_i(g_{10} + \tau h_1, g_{20} + \tau h_2) h_1 h_1 + D_{12} Z_i(g_{10} + \tau h_1, g_{20} + \tau h_2) h_1 h_2 \\
& + D_{21} Z_i(g_{10} + \tau h_1, g_{20} + \tau h_2) h_2 h_1 + D_{22} Z_i(g_{10} + \tau h_1, g_{20} + \tau h_2) h_2 h_2](1 - \tau) d\tau.
\end{aligned}$$

Define $U_1(g_1, g_2)$, $U_2(g_1, g_2)$, $U_{12}(g_1, g_2)$ and $U_{21}(g_1, g_2)$ by

$$\begin{aligned}
&< u_1, U_1(g_1, g_2) v_1 >_1 = (D_1 \eta(g_1, g_2) u_1, D_1 \eta(g_1, g_2) v_1)_{\mathcal{L}_2}, \\
&< u_2, U_2(g_1, g_2) v_2 >_2 = (D_2 \eta(g_1, g_2) u_2, D_2 \eta(g_1, g_2) v_2)_{\mathcal{L}_2}, \\
&< u_1, U_{21}(g_1, g_2) v_2 >_1 = (D_1 \eta(g_1, g_2) u_1, D_2 \eta(g_1, g_2) v_2)_{\mathcal{L}_2},
\end{aligned}$$

for any $u_i, v_i \in \mathcal{H}_i$, $i = 1, 2$. U_{12} is the adjoint operator of U_{21}, i.e. $U_{12} = U_{21}^*$. Let

$$G_i(g_1, g_2) = U_i(g_1, g_2) + \lambda_i W_i, \quad i = 1, 2.$$

Then we have

$$\begin{aligned}
& Z_1(g_{10} + h_1, g_{20} + h_2) \\
= \; & Z_1(g_{10}, g_{20}) + G_1(g_{10}, g_{20}) h_1 + U_{12}(g_{10}, g_{20}) h_2 \\
& + \int_0^1 [e_{11}(g_{10} + \tau h_1, g_{20} + \tau h_2) h_1 h_1 + e_{12}(g_{10} + \tau h_1, g_{20} + \tau h_2) h_1 h_2 \\
& + e_{13}(g_{10} + \tau h_1, g_{20} + \tau h_2) h_2 h_1 + e_{14}(g_{10} + \tau h_1, g_{20} + \tau h_2) h_2 h_2](1 - \tau) d\tau,
\end{aligned}$$

where $e_{1i}(g_1, g_2) uvw = \int e_{1i}(g_1, g_2; x) uvw dF(x)$, $i = 1, 2, 3, 4$ and

$$\begin{aligned}
e_{11}(g_1, g_2; x) uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)] D_{111} \eta uvw \\
&\quad + D_1 \eta u D_{11} \eta vw + D_1 \eta v D_{11} \eta uw + D_1 \eta w D_{11} \eta uv, \\
e_{12}(g_1, g_2; x) uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)] D_{112} \eta uvw \\
&\quad + D_1 \eta u D_{12} \eta vw + D_2 \eta w D_{11} \eta uv + D_1 \eta v D_{12} \eta uw, \\
e_{13}(g_1, g_2; x) uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)] D_{121} \eta uvw \\
&\quad + D_1 \eta u D_{21} \eta vw + D_1 \eta w D_{12} \eta uv + D_2 \eta v D_{11} \eta uw, \\
e_{14}(g_1, g_2; x) uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)] D_{122} \eta vw D_1 \eta u \\
&\quad + D_1 \eta u D_{22} \eta vw + D_2 \eta v D_{12} \eta uw + D_2 \eta w D_{12} \eta uv.
\end{aligned}$$

Similarly, for Z_2 we have

$$\begin{aligned}
& Z_2(g_{10}+h_1, g_{20}+h_2) \\
= \ & Z_2(g_{10}, g_{20}) + G_2(g_{10}, g_{20})h_2 + U_{21}(g_{10}, g_{20})h_1 \\
& + \int_0^1 [e_{21}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_1h_1 + e_{22}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_1h_2 \\
& + e_{23}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_2h_1 + e_{24}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_2h_2](1-\tau)d\tau,
\end{aligned}$$

where $e_{2i}(g_1, g_2)uvw = \int e_{2i}(g_1, g_2; x)uvwdF(x)$, $i = 1, 2, 3, 4$ and

$$\begin{aligned}
e_{21}(g_1, g_2; x)uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)]D_{211}\eta uvw \\
&\quad + D_2\eta u D_{11}\eta vw + D_1\eta v D_{21}\eta uw + D_1\eta w D_{21}\eta uv, \\
e_{22}(g_1, g_2; x)uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)]D_{212}\eta uvw \\
&\quad + D_2\eta u D_{12}\eta vw + D_2\eta w D_{21}\eta uv + D_1\eta v D_{22}\eta uw, \\
e_{23}(g_1, g_2; x)uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)]D_{221}\eta uvw \\
&\quad + D_2\eta u D_{21}\eta vw + D_1\eta w D_{22}\eta uv + D_2\eta v D_{21}\eta uw, \\
e_{24}(g_1, g_2; x)uvw &= -[\eta(g_{10}, g_{20}; x) - \eta(g_1, g_2; x)]D_{222}\eta vw D_1\eta u \\
&\quad + D_2\eta u D_{22}\eta vw + D_2\eta v D_{22}\eta uw + D_2\eta w D_{22}\eta uv.
\end{aligned}$$

Let $e_1 = \int_0^1 [e_{11}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_1h_1 + e_{12}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_1h_2 + e_{13}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_2h_1 + e_{14}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_2h_2](1-\tau)d\tau$, and $e_2 = \int_0^1 [e_{21}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_1h_1 + e_{22}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_1h_2 + e_{23}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_2h_1 + e_{24}(g_{10}+\tau h_1, g_{20}+\tau h_2)h_2h_2](1-\tau)d\tau$. Then we have the following systems of equations

$$\begin{aligned}
Z_1(g_{10}+h_1, g_{20}+h_2) &= Z_1(g_{10}, g_{20}) + G_1(g_{10}, g_{20})h_1 + U_{12}(g_{10}, g_{20})h_2 + e_1, \\
Z_2(g_{10}+h_1, g_{20}+h_2) &= Z_2(g_{10}, g_{20}) + G_2(g_{10}, g_{20})h_2 + U_{21}(g_{10}, g_{20})h_1 + e_2.
\end{aligned} \tag{8}$$

In the following the dependence of Z_i, U_{ij} and G_i on g_1 and g_2 are sometimes dropped when there is no confusion. Let

$$\begin{aligned}
G_{11} &= G_1 - U_{12}G_2^{-1}U_{21}, \\
G_{22} &= G_2 - U_{21}G_2^{-1}U_{12}.
\end{aligned}$$

Define the systematic errors as $g_{\lambda_1} - g_{10}$ and $g_{\lambda_2} - g_{20}$, where g_{λ_i}'s are the local unique root of $Z_i(g_1, g_2) = 0$, $i = 1, 2$, in $\mathcal{N}_{10} \times \mathcal{N}_{20}$. Ignoring e_1 and e_2 in the system of equations (8) and assuming the existence of $G_{11}^{-1}(g_{10}, g_{20})$ and $G_{22}^{-1}(g_{10}, g_{20})$, we have linear approximations to the systematic errors

$$\begin{aligned}
\bar{g}_{\lambda_1} - g_{10} &= -G_{11}^{-1}(g_{10}, g_{20})(Z_1(g_{10}, g_{20}) - U_{12}(g_{10}, g_{20})G_2^{-1}(g_{10}, g_{20})Z_2(g_{10}, g_{20})), \\
\bar{g}_{\lambda_2} - g_{20} &= -G_{22}^{-1}(g_{10}, g_{20})(Z_2(g_{10}, g_{20}) - U_{21}(g_{10}, g_{20})G_1^{-1}(g_{10}, g_{20})Z_1(g_{10}, g_{20})).
\end{aligned} \tag{9}$$

The stochastic errors are defined as $g_{n\lambda_1} - g_{\lambda_1}$ and $g_{n\lambda_2} - g_{\lambda_2}$. Let

$$Z_{ni}(g_1, g_2) = \frac{1}{2} D_i l_{n\boldsymbol{\lambda}}(g_1, g_2), \quad i = 1, 2.$$

Approximations of the stochastic errors will be obtained based on expansions of Z_{n1} and Z_{n2}. For $g_{\lambda_1} + h_1 \in \mathcal{N}_{10}$ and $g_{\lambda_2} + h_2 \in \mathcal{N}_{20}$, we have

$$\begin{aligned} Z_{n1}(g_{\lambda_1} + h_1, g_{\lambda_2} + h_2) &= Z_{n1}(g_{\lambda_1}, g_{\lambda_2}) + D_1 Z_{n1}(g_{\lambda_1}, g_{\lambda_2})h_1 + D_2 Z_{n1}(g_{\lambda_1}, g_{\lambda_2})h_2 + e_{n1}, \\ Z_{n2}(g_{\lambda_1} + h_1, g_{\lambda_2} + h_2) &= Z_{n2}(g_{\lambda_1}, g_{\lambda_2}) + D_1 Z_{n2}(g_{\lambda_1}, g_{\lambda_2})h_1 + D_2 Z_{n2}(g_{\lambda_1}, g_{\lambda_2})h_2 + e_{n2}, \end{aligned}$$

where e_{n1} and e_{n2} are defined similarly as e_1 and e_2. Again, ignoring remainder terms and assuming the existence of $G_{11}^{-1}(g_{\lambda_1}, g_{\lambda_2})$ and $G_{22}^{-1}(g_{\lambda_1}, g_{\lambda_2})$, we approximate $g_{n\lambda_1}$ and $g_{n\lambda_2}$ by

$$\begin{aligned} \bar{g}_{n\lambda_1} - g_{\lambda_1} &= -G_{11}^{-1}(g_{\lambda_1}, g_{\lambda_2})(Z_{n1}(g_{\lambda_1}, g_{\lambda_2}) - U_{12}(g_{\lambda_1}, g_{\lambda_2})G_2^{-1}(g_{\lambda_1}, g_{\lambda_2})Z_{n2}(g_{\lambda_1}, g_{\lambda_2})), \\ \bar{g}_{n\lambda_2} - g_{\lambda_2} &= -G_{22}^{-1}(g_{\lambda_1}, g_{\lambda_2})(Z_{n2}(g_{\lambda_1}, g_{\lambda_2}) - U_{21}(g_{\lambda_1}, g_{\lambda_2})G_1^{-1}(g_{\lambda_1}, g_{\lambda_2})Z_{n1}(g_{\lambda_1}, g_{\lambda_2})). \end{aligned} \tag{10}$$

For $0 \le b \le \alpha$, $\lambda_1 > 0$, $\lambda_2 > 0$, $g_1, h_1 \in \mathcal{N}_{10}$, $g_2, h_2 \in \mathcal{N}_{20}$, unit elements $u_1, u_2 \in \mathcal{H}_{1\alpha}$ ($||u_i||_\alpha = 1$) and unit elements $v_1, v_2 \in \mathcal{H}_{2\alpha}$, let

$$\begin{aligned}
K_{11} &= \sup_{g_1,g_2,h_1,h_2} \sup_{u_1,u_2} ||G_{11}(h_1,h_2)^{-1}[e_{11}(g_1,g_2)u_1u_2 - U_{12}(h_1,h_2)G_2^{-1}(h_1,h_2)e_{21}(g_1,g_2)u_1u_2]||_b, \\
K_{12} &= \sup_{g_1,g_2,h_1,h_2} \sup_{u_1,v_1} ||G_{11}(h_1,h_2)^{-1}[e_{12}(g_1,g_2)u_1v_1 - U_{12}(h_1,h_2)G_2^{-1}(h_1,h_2)e_{22}(g_1,g_2)u_1v_1]||_b, \\
K_{13} &= \sup_{g_1,g_2,h_1,h_2} \sup_{u_1,v_1} ||G_{11}(h_1,h_2)^{-1}[e_{13}(g_1,g_2)v_1u_1 - U_{12}(h_1,h_2)G_2^{-1}(h_1,h_2)e_{23}(g_1,g_2)v_1u_1]||_b, \\
K_{14} &= \sup_{g_1,g_2,h_1,h_2} \sup_{v_1,v_2} ||G_{11}(h_1,h_2)^{-1}[e_{14}(g_1,g_2)v_1v_2 - U_{12}(h_1,h_2)G_2^{-1}(h_1,h_2)e_{24}(g_1,g_2)v_1v_2]||_b, \\
K_{21} &= \sup_{g_1,g_2,h_1,h_2} \sup_{u_1,u_2} ||G_{22}(h_1,h_2)^{-1}[e_{21}(g_1,g_2)u_1u_2 - U_{21}(h_1,h_2)G_1^{-1}(h_1,h_2)e_{11}(g_1,g_2)u_1u_2]||_b, \\
K_{22} &= \sup_{g_1,g_2,h_1,h_2} \sup_{u_1,v_1} ||G_{22}(h_1,h_2)^{-1}[e_{22}(g_1,g_2)u_1v_1 - U_{21}(h_1,h_2)G_1^{-1}(h_1,h_2)e_{12}(g_1,g_2)u_1v_1]||_b, \\
K_{23} &= \sup_{g_1,g_2,h_1,h_2} \sup_{u_1,v_1} ||G_{22}(h_1,h_2)^{-1}[e_{23}(g_1,g_2)v_1u_1 - U_{21}(h_1,h_2)G_1^{-1}(h_1,h_2)e_{13}(g_1,g_2)v_1u_1]||_b, \\
K_{24} &= \sup_{g_1,g_2,h_1,h_2} \sup_{v_1,v_2} ||G_{22}(h_1,h_2)^{-1}[e_{24}(g_1,g_2)v_1v_2 - U_{21}(h_1,h_2)G_1^{-1}(h_1,h_2)e_{14}(g_1,g_2)v_1v_2]||_b.
\end{aligned}$$

Similarly we can define $K_{ij,n}$, $i, j = 1, 2$. Denote

$$A_{ij}(g_1, g_2) = D_j Z_{ni}(g_1, g_2) - D_j Z_i(g_1, g_2), \quad i, j = 1, 2.$$

Let

$$\begin{aligned}
E_{21,n}^1 &= \sup_{g_1,g_2} \sup_u ||G_{11}^{-1}(g_1,g_2)U_{12}(g_1,g_2)G_2^{-1}(g_1,g_2)A_{21}(g_1,g_2)u||_b, \\
E_{21,n}^2 &= \sup_{g_1,g_2} \sup_u ||G_{11}^{-1}(g_1,g_2)A_{11}(g_1,g_2)u||_b, \\
E_{31,n}^1 &= \sup_{g_1,g_2} \sup_u ||G_{11}^{-1}(g_1,g_2)U_{12}(g_1,g_2)G_2^{-1}(g_1,g_2)A_{22}(g_1,g_2)u||_b, \\
E_{31,n}^2 &= \sup_{g_1,g_2} \sup_u ||G_{11}^{-1}(g_1,g_2)A_{12}(g_1,g_2)u||_b, \\
E_{22,n}^1 &= \sup_{g_1,g_2} \sup_u ||G_{22}^{-1}(g_1,g_2)U_{21}(g_1,g_2)G_1^{-1}(g_1,g_2)A_{11}(g_1,g_2)u||_b, \\
E_{22,n}^2 &= \sup_{g_1,g_2} \sup_u ||G_{22}^{-1}(g_1,g_2)A_{21}(g_1,g_2)u||_b, \\
E_{32,n}^1 &= \sup_{g_1,g_2} \sup_u ||G_{22}^{-1}(g_1,g_2)U_{11}(g_1,g_2)G_1^{-1}(g_1,g_2)A_{12}(g_1,g_2)u||_b, \\
E_{32,n}^2 &= \sup_{g_1,g_2} \sup_u ||G_{22}^{-1}(g_1,g_2)A_{22}(g_1,g_2)u||_b.
\end{aligned}$$

Let $K_1 = K_{11} + K_{12} + K_{13} + K_{14}$, $K_2 = K_{21} + K_{22} + K_{23} + K_{24}$, $E_{21,n} = E^1_{21,n} + E^2_{21,n}$ $E_{31,n} = E^1_{31,n} + E^2_{31,n}$, $E_{22,n} = E^1_{22,n} + E^2_{22,n}$ and $E_{32,n} = E^1_{32,n} + E^2_{32,n}$. Standard analysis based on Taylor series expansions leads to the estimate of the error terms

$$\begin{aligned}
&||\bar{g}_{\lambda_1} - g_{10}||_b \le K_1||h_1|| \cdot ||h_2||,\\
&||\bar{g}_{\lambda_2} - g_{20}||_b \le K_2||h_1|| \cdot ||h_2||,\\
&||\bar{g}_{n\lambda_1} - g_{\lambda_1}||_b \le E_{21,n}||h_1|| + E_{31,n}||h_2|| + K_{11,n}||h_1|| \cdot ||h_2||,\\
&||\bar{g}_{n\lambda_2} - g_{\lambda_2}||_b \le E_{22,n}||h_1|| + E_{32,n}||h_2|| + K_{12,n}||h_1|| \cdot ||h_2||.
\end{aligned}$$

4 Existence and Uniqueness

In this section, we show that $l_{\boldsymbol{\lambda}}$ and $l_{n\lambda}$ have unique minimizers. Let

$$\begin{aligned}
S_{1,(g_1,g_2)}(r,b) &= \{h \in \mathcal{H}_{b1} : ||h - g_{10}||_b \le r\},\\
S_{2,(g_1,g_2)}(r,b) &= \{h \in \mathcal{H}_{b2} : ||h - g_{20}||_b \le r\},\\
S_1(r,b) &= S_{1,(0,0)}(r,b),\\
S_2(r,b) &= S_{2,(0,0)}(r,b),\\
d_1(\boldsymbol{\lambda},b) &= ||\bar{g}_{\lambda_1} - g_{10}||_b,\\
d_2(\boldsymbol{\lambda},b) &= ||\bar{g}_{\lambda_2} - g_{20}||_b,\\
r_1(\boldsymbol{\lambda},b) &= (K_{11} + K_{21})d_1(\boldsymbol{\lambda},\alpha) + (K_{12} + K_{22})d_2(\boldsymbol{\lambda},\alpha),\\
r_2(\boldsymbol{\lambda},b) &= (K_{13} + K_{23})d_1(\boldsymbol{\lambda},b) + (K_{14} + K_{24})d_2(\boldsymbol{\lambda},b).
\end{aligned}$$

The following theorem guarantees the existence and uniqueness of g_{λ_1} and g_{λ_2}.

Theorem 2 *If $d_i(\boldsymbol{\lambda},\alpha) \to 0$ and $r_i(\boldsymbol{\lambda},\alpha) \to 0$ as $\lambda_1 \to 0$ and $\lambda_2 \to 0$, then for λ_1 and λ_2 sufficiently small, we have*

(a) There are unique $g_{\lambda_1} \in S_{1,(g_{10},g_{20})}(2d_1(\boldsymbol{\lambda},\alpha),\alpha)$ and $g_{\lambda_2} \in S_{2,(g_{10},g_{20})}(2d_2(\boldsymbol{\lambda},\alpha),\alpha)$ satisfying $Z_1(g_{\lambda_1}, g_{\lambda_2}) = 0$ and $Z_2(g_{\lambda_1}, g_{\lambda_2}) = 0$.

(b) $||\bar{g}_{\lambda_1} - g_{\lambda_1}||_b + ||\bar{g}_{\lambda_2} - g_{\lambda_2}||_b \le 4r_1(\boldsymbol{\lambda},b)d_1(\boldsymbol{\lambda},\alpha) + 4r_2(\boldsymbol{\lambda},b)d_2(\boldsymbol{\lambda},\alpha)$ for $0 \le b \le \alpha$.

[Proof] For simplicity of notations, we let $S_i = S_i(r,b)$, $d_i = d_i(\boldsymbol{\lambda},\alpha)$ and $r_i = r_i(\boldsymbol{\lambda},b)$, $i = 1,2$. Let

$$\begin{aligned}
F_1(h,k) &= h - G_{11}^{-1}(g_{10},g_{20})(Z_1(g_{10}+h, g_{20}+k)\\
&\quad -U_{12}(g_{10},g_{20})G_2^{-1}(g_{10},g_{20})Z_2(g_{10}+h, g_{20}+k)),\\
F_2(h,k) &= k - G_{22}^{-1}(g_{10},g_{20})(Z_2(g_{10}+h, g_{20}+k)\\
&\quad -U_{21}(g_{10},g_{20})G_1^{-1}(g_{10},g_{20})Z_1(g_{10}+h, g_{20}+k)).
\end{aligned}$$

$\boldsymbol{F}(h,k) = (F_1(h,k), F_2(h,k))$ is a linear operator on the product Hilbert space $\mathcal{H}$. It is not difficult to show that $\boldsymbol{F}(S_1 \times S_2) \subset S_1 \times S_2$. Then we need to show that $\boldsymbol{F}$ is a contraction on $S_1 \times S_2$.

For $h_1, h_2 \in S_1$ and $k_1, k_2 \in S_2$,

$$\begin{aligned} & F_1(h_1,k_1) - F_1(h_2,k_2) \\ = \; & h_1 - h_2 - G_{11}^{-1}(g_{10}, g_{20})[Z_1(g_{10}+h_1, g_{20}+k_1) - Z_1(g_{10}+h_2, g_{20}+k_2) \\ & -U_{12}(g_{10}, g_{20})G_2^{-1}(g_{10}, g_{20})(Z_2(g_{10}+h_1, g_{20}+k_1) - Z_2(g_{10}+h_2, g_{20}+k_2))]. \end{aligned}$$

Applying Taylor series expansion (6), we have

$$\begin{aligned} & Z_1(g_{10}+h_2, g_{20}+k_2) \\ = \; & Z_1(g_{10}+h_1, g_{20}+k_1) + \int_0^1 [D_1 Z_1(g_{10}+h_1+\tau(h_2-h_1), g_{20}+k_1+\tau(k_2-k_1))(h_2-h_1) \\ & +D_2 Z_1(g_{10}+h_1+\tau(h_2-h_1), g_{20}+k_1+\tau(k_2-k_1))(k_2-k_1)]d\tau. \end{aligned}$$

Applying Taylor series expansion again to the terms inside the integral, we have

$$\begin{aligned} & Z_1(g_{10}+h_2, g_{20}+k_2) - Z_1(g_{10}+h_1, g_{20}+k_1) \\ = \; & D_1 Z_1(g_{10}, g_{20})(h_2-h_1) + D_2 Z_1(g_{10}, g_{20})(k_2-k_1) \\ & + \int_0^1 \int_0^1 [D_{11} Z_1(g_{10}+\tau' h_*, g_{20}+k_*)\tau' h_* + D_{12} Z_1(g_{10}+\tau' h_*, g_{20}+\tau' k_*)k_*](h_2-h_1)d\tau d\tau' \\ & + \int_0^1 \int_0^1 [D_{21} Z_1(g_{10}+\tau' h_*, g_{20}+\tau' k_*)h_* + D_{22} Z_1(g_{10}+\tau' h_*, g_{20}+\tau' k_*)k_*](k_2-k_1)d\tau d\tau', \end{aligned}$$

where $h_* = h_1 + \tau(h_2 - h_1)$ and $k_* = k_1 + \tau(k_2 - k_1)$. Similar approximation to $Z_2(g_{10}+h_2, g_{20}+k_2) - Z_2(g_{10}+h_1, g_{20}+k_1)$ can be obtained. Plugging in and after some algebraic steps, we have

$$\begin{aligned} ||F_1(h_1,k_1) - F_1(h_2,k_2)||_b &\le 2(K_{11}d_1 + K_{12}d_2)||h_2-h_1||_b + 2(K_{13}d_1 + K_{14}d_2)||k_2-k_1||_b, \\ ||F_2(h_1,k_1) - F_2(h_2,k_2)||_b &\le 2(K_{21}d_1 + K_{22}d_2)||h_2-h_1||_b + 2(K_{23}d_1 + K_{24}d_2)||k_2-k_1||_b. \end{aligned}$$

Therefore,

$$\begin{aligned} & ||\boldsymbol{F}(h_1,h_2) - \boldsymbol{F}(k_1,k_2)||_b \\ \le \; & 2(K_{11}d_1 + K_{12}d_1 + K_{12}d_2 + K_{22}d_2)||h_2-h_1||_b \\ & +2(K_{13}d_1 + K_{23}d_1 + K_{14}d_2 + K_{24}d_2)||k_2-k_1||_b \\ \le \; & 2r_1||h_2-h_1||_b + 2r_2||k_2-k_1||_b. \end{aligned}$$

Choosing λ_{10} and λ_{20} such that for $\lambda_1 \in (0, \lambda_{10}]$ and $\lambda_2 \in (0, \lambda_{20}]$, we have $r_1, r_2 < C$ for some constant $C < 1$. For this choice of λ_{10} and λ_{20}, $\boldsymbol{F}$ is a contraction. The contraction theorem leads to unique h_{λ_1} and k_{λ_2} for which $\boldsymbol{F}(h_{\lambda_1}, k_{\lambda_2}) = 0$. Let $g_{\lambda_1} = g_{10} + h_{\lambda_1}$ and $g_{\lambda_2} = g_{20} + k_{\lambda_2}$, and $g_{\lambda_1} \in S_{1,(g_{10},g_{20})}(2d_1,\alpha)$ and $g_{\lambda_2} \in S_{2,(g_{10},g_{20})}(2d_2,\alpha)$ are unique solutions to $Z_1(g_{\lambda_1},g_{\lambda_2}) = 0$ and $Z_2(g_{\lambda_1},g_{\lambda_2}) = 0$.

To obtain the bound, note that

$$(\bar{g}_{\lambda_1} - g_{\lambda_1}, \bar{g}_{\lambda_2} - g_{\lambda_2}) = \boldsymbol{F}(h_{\lambda_1}, k_{\lambda_2}) - \boldsymbol{F}(0,0).$$

Thus

$$\begin{aligned}
& ||\bar{g}_{\lambda_1} - g_{\lambda_1}||_b + ||\bar{g}_{\lambda_2} - g_{\lambda_2}||_b \\
\leq\ & ||\boldsymbol{F}(h_{\lambda_1}, k_{\lambda_2}) - \boldsymbol{F}(0,0)||_b \\
\leq\ & 2r_1||h_{\lambda_1}||_b + 2r_2||k_{\lambda_2}||_b \\
\leq\ & 4(r_1 d_1 + r_2 d_2).
\end{aligned}$$

This completes the proof.

We next consider the solution to $l_{n\boldsymbol{\lambda}}$. For $g_{n\lambda_1} \in \mathcal{N}_{10}$ and $g_{n\lambda_2} \in \mathcal{N}_{20}$, define

$$\begin{aligned}
d_{n1}(\boldsymbol{\lambda}, b) &= ||\bar{g}_{n\lambda_1} - g_{\lambda_1}||_b, \\
d_{n2}(\boldsymbol{\lambda}, b) &= ||\bar{g}_{n\lambda_2} - g_{\lambda_2}||_b, \\
r_{n1}(\boldsymbol{\lambda}, b) &= (K_{11,n} + K_{21,n}) d_{n1}(\boldsymbol{\lambda}, \alpha) + (K_{12,n} + K_{22,n}) d_{n2}(\boldsymbol{\lambda}, \alpha) + E_{21,n} + E_{22,n}, \\
r_{n2}(\boldsymbol{\lambda}, b) &= (K_{13,n} + K_{23,n}) d_{n1}(\boldsymbol{\lambda}, b) + (K_{14,n} + K_{24,n}) d_{n2} + E_{31,n} + E_{32,n}.
\end{aligned}$$

Theorem 3 *If λ_{n1} and λ_{n2} tend to zero such that $g_{\lambda_{n1}} \in \mathcal{N}_{10}$ and $g_{\lambda_{n2}} \in \mathcal{N}_{20}$ for all n and $d_{n1}(\boldsymbol{\lambda}, b) \xrightarrow{p} 0$, $d_{n2}(\boldsymbol{\lambda}, b) \xrightarrow{p} 0$, $r_{n1}(\boldsymbol{\lambda}, b) \xrightarrow{p} 0$ and $r_{n2}(\boldsymbol{\lambda}, b) \xrightarrow{p} 0$, then with probability tending to unity as $n \to \infty$, we have*

(a) There is a unique root $(g_{n\lambda_{n1}}, g_{n\lambda_{n2}})$ satisfying $Z_{n1}(g_{n\lambda_{n1}}, g_{n\lambda_{n2}}) = 0$ and $Z_{n2}(g_{n\lambda_{n1}}, g_{n\lambda_{n2}}) = 0$.

(b) for $b \in [0,\alpha]$, $||\bar{g}_{n\lambda_1} - g_{n\lambda_1}||_b + ||\bar{g}_{n\lambda_2} - g_{n\lambda_2}||_b \leq 4r_{n1}(\boldsymbol{\lambda}, b) d_{n1}(\boldsymbol{\lambda}, \alpha) + 4r_{n2}(\boldsymbol{\lambda}, b) d_{n2}(\boldsymbol{\lambda}, \alpha)$.

[Proof] For simplicity of notations, denote $d_{ni} = d_{ni}(\boldsymbol{\lambda}, \alpha)$ and $r_{ni} = r_{ni}(\boldsymbol{\lambda}, b)$, $i = 1, 2$. Let

$$\begin{aligned}
F_{n1}(h,k) &= h - G_{11}^{-1}(g_{\lambda_1}, g_{\lambda_2})(Z_{n1}(g_{\lambda_1} + h, g_{\lambda_2} + k) \\
&\quad - U_{12}(g_{\lambda_1}, g_{\lambda_2}) G_2^{-1}(g_{\lambda_1}, g_{\lambda_2}) Z_{n2}(g_{10} + h, g_{20} + k)), \\
F_{n2}(h,k) &= k - G_{22}^{-1}(g_{\lambda_1}, g_{\lambda_2})(Z_{n2}(g_{\lambda_1} + h, g_{\lambda_2} + k) \\
&\quad - U_{21}(g_{\lambda_1}, g_{\lambda_2}) G_1^{-1}(g_{\lambda_1}, g_{\lambda_2}) Z_{n1}(g_{\lambda_1} + h, g_{\lambda_2} + k)).
\end{aligned}$$

The proof proceeds similarly to the proof of Theorem 2, with additional terms to approximate Z_{n1} and Z_{n2} by Z_1 and Z_2. Take n large enough so that with probability arbitrarily close to unity, $S_{1,(g_{\lambda_1},g_{\lambda_2})}(2d_{n1},\alpha) \subset \mathcal{N}_{10}$, $S_2(2d_{n2},\alpha) \subset \mathcal{N}_{20}$, $r_{n1} < \frac{1}{2}$ and $r_{n2} < \frac{1}{2}$. For the rest of the proof, we restrict to this event. It is not difficult to show that $\boldsymbol{F}_n(S_1(2d_{n1},\alpha)\times S_2(2d_{n2},\alpha))\subseteq S_1(2d_{n1},\alpha)\times S_2(2d_{n2},\alpha)$. We then need to show that $\boldsymbol{F}_n$ is a contraction on $S_1(2d_{n1},\alpha)\times S_2(2d_{n2},\alpha)$. Expand Z_{n1} and Z_{n2} as in Theorem 2 and Section 3. After some algebraic steps, for $h_1,h_2\in S_1(2d_{n1},\alpha)$ and $k_2,k_2\in S_2(2d_{n2},\alpha)$,

$$\begin{aligned}||F_{n1}(h_1,k_1)-F_{n1}(h_2,k_2)||_b &\le 2(K_{11,n}d_{n1}+K_{12,n}d_{n2}+E_{21,n})||h_2-h_1||_b\\ &\quad +2(K_{13,n}d_{n1}+K_{14,n}d_{n2}+E_{22,n})||k_2-k_1||_b,\\ ||F_{n2}(h_1,k_1)-F_{n2}(h_2,k_2)||_b &\le 2(K_{21,n}d_{n1}+K_{22,n}d_{n2}+E_{31,n})||h_2-h_1||_b\\ &\quad +2(K_{23,n}d_{n1}+K_{24,n}d_{n2}+E_{32,n})||k_2-k_1||_b.\end{aligned}$$

Therefore

$$||\boldsymbol{F}_n(h_1,h_2)-\boldsymbol{F}_n(k_1,k_2)||\le 2r_{n1}||h_1-h_2||_b+2r_{n2}||k_1-k_2||_b,$$

which indicates that $\boldsymbol{F}_n$ is a contraction.

To get the upper bound, notice that

$$||\bar{g}_{n\lambda_1}-g_{n\lambda_1}||_b+||\bar{g}_{n\lambda_2}-g_{n\lambda_2}||_b=||\boldsymbol{F}_n(h_{n\lambda_1},k_{n\lambda_2})||_b\le 4r_{n1}d_{n1}+4r_{n2}d_{n2}.$$

This completes proof of Theorem 3.

References

[1] Cox, D. D. (1988). Approximation of Method of Regularization estimators. *Ann. Statist.* **16**, 694-712.

[2] Cox, D. D. and O'Sullivan (1990). Asymptotic Analysis of Penalized Likelihood and Related Estimators. *Ann. Statist.* **18**, 1676-1695.

[3] O'Sullivan, F. (1990). Convergence Characteristics of Methods of Regularization Estimators for Nonlinear Operator Equations. *SIAM Journal on Numerical Analysis* **27**, 1635-1649.

[4] Zeidler, E. (1985). *Nonlinear functional analysis and its applications*. Springer: New York.