mathematics

Article

Principle of Duality in Cubic Smoothing Spline

Ruixue Du [1] and Hiroshi Yamada [2,*]

1 Graduate School of Social Sciences, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima 739-8525, Japan; tozuisetsu@yahoo.co.jp
2 School of Informatics and Data Science, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima 739-8525, Japan
* Correspondence: yamada@hiroshima-u.ac.jp; Tel.: +81-82-424-7214; Fax: +81-82-424-7212

Received: 20 September 2020; Accepted: 15 October 2020; Published: 19 October 2020

Abstract: Fitting a cubic smoothing spline is a typical smoothing method. This paper reveals a principle of duality in the penalized least squares regressions relating to the method. We also provide a number of results derived from them, some of which are illustrated by a real data example.

Keywords: cubic smoothing spline; principle of duality; penalized regression; right-inverse matrix; ridge regression

MSC: 62G05

1. Introduction

Fitting a cubic smoothing spline, which was developed by [1,2] and others, is a typical smoothing method. The cubic smoothing spline fitted to a scatter plot of ordered pairs (x_i, y_i) for $i = 1, \dots, n$ is a function such that

$$\widehat{f}(x) = \arg \min_{f \in \mathcal{W}} \sum_{i=1}^{n} \{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f''(x)\}^2 dx, \tag{1}$$

where $x_1, \dots, x_n$ are points satisfying $a < x_1 < \cdots < x_n < b$, $\mathcal{W}$ denotes a function space that contains all functions of which the second derivative is square integrable over $[a, b]$, and λ is a positive smoothing/tuning parameter, which controls the trade-off between goodness of fit and smoothness.

Let $\widehat{\boldsymbol{f}} = [\widehat{f}(x_1), \dots, \widehat{f}(x_n)]^\top$. Then, given $\widehat{f}(x)$ is a natural cubic spline of which the knots are $x_1, \dots, x_n$ (see, e.g., [3,4]), it follows that

$$\widehat{\boldsymbol{f}} = \arg \min_{\boldsymbol{f} \in \mathbb{R}^n} \|\boldsymbol{y} - \boldsymbol{f}\|^2 + \lambda \boldsymbol{f}^\top \boldsymbol{C}^\top \boldsymbol{R}^{-1} \boldsymbol{C} \boldsymbol{f} \tag{2}$$

$$= \left(\boldsymbol{I}_n + \lambda \boldsymbol{C}^\top \boldsymbol{R}^{-1} \boldsymbol{C}\right)^{-1} \boldsymbol{y}, \tag{3}$$

where $\boldsymbol{y} = [y_1, \dots, y_n]^\top$, $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix, and $\boldsymbol{C}$ and $\boldsymbol{R}$ are explicitly presented later. Then, as shown in [3], $\widehat{f}(x)$ in (1) is uniquely determined by $\widehat{\boldsymbol{f}} \in \mathbb{R}^n$ in (3). Thus, estimating $\widehat{f}(x)$ is equivalent to estimating $\widehat{\boldsymbol{f}}$.

Let $\boldsymbol{\Pi} = [\boldsymbol{\iota}_n, \boldsymbol{x}] \in \mathbb{R}^{n \times 2}$, where $\boldsymbol{\iota}_n = [1, \dots, 1]^\top \in \mathbb{R}^n$ and $\boldsymbol{x} = [x_1, \dots, x_n]^\top$. Note that since $x_1 < \cdots < x_n$, $\boldsymbol{\iota}_n$ and $\boldsymbol{x}$ are linearly independent and thus $\boldsymbol{\Pi}$ is of full column rank. Let

$$\widehat{\boldsymbol{\tau}} = \boldsymbol{\Pi}(\boldsymbol{\Pi}^\top \boldsymbol{\Pi})^{-1} \boldsymbol{\Pi}^\top \boldsymbol{y}. \tag{4}$$

Denote the difference between $\widehat{f}$ and $\widehat{\tau}$ (resp. y and $\widehat{f}$) by $\widehat{c}$ (resp. $\widehat{u}$):

$$\widehat{c} = \widehat{f} - \widehat{\tau}, \quad \widehat{u} = y - \widehat{f}. \tag{5}$$

Accordingly, we have

$$y = \widehat{\tau} + \widehat{c} + \widehat{u}. \tag{6}$$

In this paper, we present a comprehensive list of penalized least squares regressions relating to (6). One such example is the ridge regression [5] that leads to $\widehat{c}$. Then, we reveal a principle of duality in them. In addition, based on them, we provide a number of theoretical results, for example, $\iota_n^\top \widehat{c} = 0$.

This paper is organized as follows. Section 2 fixes some notations and gives key preliminary results used to derive the main results in the paper. Section 3 provides a comprehensive list of penalized least squares regressions relating to (6), and reveals a principle of duality in them. Section 4 shows some results that are obtainable from the regressions shown in Section 3. Section 5 illustrates some results provided in Sections 3 and 4 by a real data example. Section 6 deals with the cases such that the other right-inverse matrices are used. Section 7 concludes the paper.

2. Preliminaries

In this section, we give key preliminary results used to derive the main results of this paper. Before stating them, we fix some notations.

2.1. Notations

Let $\widehat{f}_i$ (resp. $\widehat{\tau}_i$) denote the ith entry of $\widehat{f}$ (resp. $\widehat{\tau}$) for $i = 1, \dots, n$, $\delta_i = x_{i+1} - x_i$, which is positive by definition, for $i = 1, \dots, n-1$, $\Delta = \mathsf{diag}(\delta_1, \dots, \delta_{n-1}) \in \mathbb{R}^{(n-1)\times(n-1)}$, and for a full-*row*-rank matrix $M \in \mathbb{R}^{m\times n}$, $M^\top(MM^\top)^{-1} \in \mathbb{R}^{n\times m}$, which is a right-inverse matrix of M, be denoted by M_{r}^{-1}. For a full-*column*-rank matrix $W \in \mathbb{R}^{n\times p}$, let $\mathcal{S}(W)$ (resp. $\mathcal{S}^\perp(W)$) denote the column space of W (resp. the orthogonal complement of $\mathcal{S}(W)$) and P_W (resp. Q_W) denote the orthogonal projection matrix to the space $\mathcal{S}(W)$ (resp. $\mathcal{S}^\perp(W)$). Explicitly, they are $P_W = W(W^\top W)^{-1}W^\top$ and $Q_W = I_n - P_W$. $D_{(i)} \in \mathbb{R}^{(n-i)\times(n-i+1)}$ is a Toeplitz matrix of which the first (resp. last) row is $[-1, 1, 0, \dots, 0]$ (resp. $[0, \dots, 0, -1, 1]$) and we define matrices $C \in \mathbb{R}^{(n-2)\times n}$ and $R \in \mathbb{R}^{(n-2)\times(n-2)}$ as follows:

$$C = \left[\begin{array}{c|c|c|c|c|c}
\delta_1^{-1} & -\delta_1^{-1} - \delta_2^{-1} & \delta_2^{-1} & 0 & \cdots & 0 \\ \hline
0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \hline
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \hline
0 & \cdots & 0 & \delta_{n-2}^{-1} & -\delta_{n-2}^{-1} - \delta_{n-1}^{-1} & \delta_{n-1}^{-1}
\end{array}\right] \tag{7}$$

and

$$R = \left[\begin{array}{c|c|c|c|c}
\frac{1}{3}(\delta_1+\delta_2) & \frac{1}{6}\delta_2 & 0 & \cdots & 0 \\ \hline
\frac{1}{6}\delta_2 & \frac{1}{3}(\delta_2+\delta_3) & \ddots & \ddots & \vdots \\ \hline
0 & \ddots & \ddots & \ddots & 0 \\ \hline
\vdots & \ddots & \ddots & \ddots & \frac{1}{6}\delta_{n-2} \\ \hline
0 & \cdots & 0 & \frac{1}{6}\delta_{n-2} & \frac{1}{3}(\delta_{n-2}+\delta_{n-1})
\end{array}\right]. \tag{8}$$

Finally, we denote the eigenvalues of R by $\omega_1, \dots, \omega_{n-2}$ in descending order.

2.2. Key Preliminary Results

Lemma 1.

(i) C *can be factorized as* $C = D_{(2)}\Delta^{-1}D_{(1)}$.

(ii) *We have the following inequalities:*

$$\omega_{n-2} \geq \min\left\{\frac{1}{3}\delta_1 + \frac{1}{6}\delta_2, \frac{1}{6}(\delta_2+\delta_3), \dots, \frac{1}{6}(\delta_{n-3}+\delta_{n-2}), \frac{1}{6}\delta_{n-2} + \frac{1}{3}\delta_{n-1}\right\} > 0.$$

Proof of Lemma 1.

(i) Let $w = [w_1, \dots, w_n]^\top$ be an n-dimensional column vector. Then, by definition of C, it follows that

$$\begin{aligned} Cw &= \begin{bmatrix} -\frac{w_2-w_1}{\delta_1} + \frac{w_3-w_2}{\delta_2} \\ \vdots \\ -\frac{w_{n-1}-w_{n-2}}{\delta_{n-2}} + \frac{w_n-w_{n-1}}{\delta_{n-1}} \end{bmatrix} = D_{(2)} \begin{bmatrix} \frac{w_2-w_1}{\delta_1} \\ \vdots \\ \frac{w_n-w_{n-1}}{\delta_{n-1}} \end{bmatrix} = D_{(2)}\Delta^{-1} \begin{bmatrix} -w_1+w_2 \\ \vdots \\ -w_{n-1}+w_n \end{bmatrix} \\ &= D_{(2)}\Delta^{-1}D_{(1)}w \in \mathbb{R}^{n-2}, \end{aligned}$$

which leads to $C = D_{(2)}\Delta^{-1}D_{(1)}$.

(ii) The first inequality follows by applying the Gershgorin circle theorem and the second inequality holds from $\delta_i > 0$ for $i = 1, \dots, n-1$.

□

Remark 1. *In Appendix A, we give some remarks on a special case such that* $x = [1, \dots, n]^\top$.

Lemma 2.

(i) $\mathcal{S}(C^\top)$ *equals* $\mathcal{S}^\perp(\Pi)$ *and*

(ii) $\mathcal{S}(C_{\mathrm{r}}^{-1})$ *equals* $\mathcal{S}^\perp(\Pi)$.

Proof of Lemma 2.

(i) Given that $\delta_i > 0$ for $i = 1, \dots, n-1$, both Π and $C^\top$ are of full column rank. In addition, $[\Pi, C^\top]$ is a square matrix. Thus, if $(C^\top)^\top\Pi = C\Pi = 0$, then it follows that $\mathcal{S}(C^\top) = \mathcal{S}^\perp(\Pi)$. From $D_{(1)}\iota_n = 0$, we have $C\iota_n = D_{(2)}\Delta^{-1}D_{(1)}\iota_n = 0$. Likewise, from $\Delta^{-1}D_{(1)}x = \Delta^{-1}\Delta\iota_{n-1} = \iota_{n-1}$ and $D_{(2)}\iota_{n-1} = 0$, we obtain $Cx = D_{(2)}\Delta^{-1}D_{(1)}x = 0$. Accordingly, we have $C\Pi = 0$, which completes the proof.

(ii) Recall that $C_{\mathrm{r}}^{-1} = C^\top(CC^\top)^{-1}$. It is clear that C_{r}^{-1} is a full-column-rank matrix such that $[\Pi, C_{\mathrm{r}}^{-1}]$ is a square matrix. In addition, $(C_{\mathrm{r}}^{-1})^\top\Pi = (CC^\top)^{-1}C\Pi = 0$. Thus, it follows that $\mathcal{S}(C_{\mathrm{r}}^{-1}) = \mathcal{S}^\perp(\Pi)$.

□

Denote the spectral decomposition of R by $V\Omega V^\top$ and let $R^{-1/2} = V\Omega^{-1/2}V^\top$, where $\Omega^{-1/2} = \mathrm{diag}\left(1/\sqrt{\omega_1}, \dots, 1/\sqrt{\omega_{n-2}}\right)$. Then, $R^{-1/2}$ is a positive definite matrix such that $R^{-1/2}R^{-1/2} = R^{-1}$. Define

$$D = R^{-1/2}C. \tag{9}$$

Then, given that $C^\top$ is of full column rank and $R^{-1/2}$ is nonsingular, $D \in \mathbb{R}^{(n-2)\times n}$ is also of full row rank. In addition, we have

$$D^\top D = C^\top R^{-1}C. \tag{10}$$

(We provide Matlab/GNU Octave and R functions for calculating $\boldsymbol{C}$, $\boldsymbol{R}$, and $\boldsymbol{D}$ in Appendix A).

Lemma 3.

(i) $\mathcal{S}(\boldsymbol{D}^\top)$ *equals* $\mathcal{S}^\perp(\boldsymbol{\Pi})$ *and*
(ii) $\mathcal{S}(\boldsymbol{D}_\mathrm{r}^{-1})$ *equals* $\mathcal{S}^\perp(\boldsymbol{\Pi})$.

Proof of Lemma 3. Both (i) and (ii) may be proved similarly to Lemma 2 (ii). For example, given $\boldsymbol{C\Pi} = \boldsymbol{0}$, we have $(\boldsymbol{D}^\top)^\top\boldsymbol{\Pi} = \boldsymbol{D\Pi} = \boldsymbol{R}^{-1/2}\boldsymbol{C\Pi} = \boldsymbol{0}$. □

Denote the eigenvalues of $\boldsymbol{C}^\top\boldsymbol{R}^{-1}\boldsymbol{C}$ by $g_1, \ldots, g_n$ in ascending order and the spectral decomposition of $\boldsymbol{C}^\top\boldsymbol{R}^{-1}\boldsymbol{C}$ by $\boldsymbol{UGU}^\top$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$ and $\boldsymbol{G} = \mathsf{diag}(g_1, \ldots, g_n)$. Let $\boldsymbol{T} = [\boldsymbol{u}_1, \boldsymbol{u}_2] \in \mathbb{R}^{n\times 2}$, $\boldsymbol{E}^\top = [\boldsymbol{u}_3, \ldots, \boldsymbol{u}_n] \in \mathbb{R}^{n\times(n-2)}$, and $\boldsymbol{S} = \mathsf{diag}(g_3, \ldots, g_n) \in \mathbb{R}^{(n-2)\times(n-2)}$.

Lemma 4.

(i) $\mathcal{S}(\boldsymbol{T})$ *equals* $\mathcal{S}(\boldsymbol{\Pi})$,
(ii) $\mathcal{S}(\boldsymbol{E}^\top)$ *equals* $\mathcal{S}^\perp(\boldsymbol{\Pi})$, *and*
(iii) $\mathcal{S}(\boldsymbol{E}_\mathrm{r}^{-1})$ *equals* $\mathcal{S}^\perp(\boldsymbol{\Pi})$.

Proof of Lemma 4. (i) Since $\boldsymbol{C}^\top\boldsymbol{R}^{-1}\boldsymbol{C} \in \mathbb{R}^{n\times n}$ is a nonnegative definite matrix of which the rank is $n-2$, we have $0 = g_1 = g_2 < g_3 < \cdots < g_n$. In addition, given $\boldsymbol{C\Pi} = \boldsymbol{0}$, it follows that $\boldsymbol{C}^\top\boldsymbol{R}^{-1}\boldsymbol{C\Pi} = 0 \cdot \boldsymbol{\Pi}$, which completes the proof. (ii) and (iii) may be proved similarly to Lemma 2 (ii). □

Given $g_1 = g_2 = 0$, we have

$$\boldsymbol{E}^\top\boldsymbol{SE} = \boldsymbol{C}^\top\boldsymbol{R}^{-1}\boldsymbol{C}. \tag{11}$$

Define

$$\boldsymbol{F} = \boldsymbol{S}^{1/2}\boldsymbol{E}, \tag{12}$$

where $\boldsymbol{S}^{1/2} = \mathsf{diag}(\sqrt{g_3}, \ldots, \sqrt{g_n}) \in \mathbb{R}^{(n-2)\times(n-2)}$. Then, we have

$$\boldsymbol{F}^\top\boldsymbol{F} = \boldsymbol{C}^\top\boldsymbol{R}^{-1}\boldsymbol{C}. \tag{13}$$

Lemma 5.

(i) $\mathcal{S}(\boldsymbol{F}^\top)$ *equals* $\mathcal{S}^\perp(\boldsymbol{\Pi})$ *and*
(ii) $\mathcal{S}(\boldsymbol{F}_\mathrm{r}^{-1})$ *equals* $\mathcal{S}^\perp(\boldsymbol{\Pi})$.

Proof of Lemma 5. Both (i) and (ii) may be proved similarly to Lemma 2 (ii). For example, given $\boldsymbol{E\Pi} = \boldsymbol{0}$, we have $(\boldsymbol{F}^\top)^\top\boldsymbol{\Pi} = \boldsymbol{F\Pi} = \boldsymbol{S}^{1/2}\boldsymbol{E\Pi} = \boldsymbol{0}$. □

Lemma 6. *There exists an orthogonal matrix* $\mathbf{Y} \in \mathbb{R}^{(n-2)\times(n-2)}$ *such that* $\boldsymbol{F}^\top = \boldsymbol{D}^\top\mathbf{Y}$.

Proof of Lemma 6. Recall that both $\boldsymbol{D}^\top \in \mathbb{R}^{n\times(n-2)}$ and $\boldsymbol{F}^\top \in \mathbb{R}^{n\times(n-2)}$ are of full column rank and $\mathcal{S}(\boldsymbol{D}^\top) = \mathcal{S}(\boldsymbol{F}^\top)$. Accordingly, there exists a nonsingular matrix $\mathbf{Y} \in \mathbb{R}^{(n-2)\times(n-2)}$ such that $\boldsymbol{F}^\top = \boldsymbol{D}^\top\mathbf{Y}$. Given that $\boldsymbol{D}^\top\boldsymbol{D} = \boldsymbol{F}^\top\boldsymbol{F}$, we have $\boldsymbol{D}^\top(\boldsymbol{I}_{n-2} - \mathbf{YY}^\top)\boldsymbol{D} = \boldsymbol{0}$. Then, from $\boldsymbol{D}_\mathrm{r}^{-1\top}\boldsymbol{D}^\top(\boldsymbol{I}_{n-2} - \mathbf{YY}^\top)\boldsymbol{D}\boldsymbol{D}_\mathrm{r}^{-1} = \boldsymbol{I}_{n-2} - \mathbf{YY}^\top = \boldsymbol{0}$, we have $\mathbf{Y}^\top = \mathbf{Y}^{-1}$. □

Let (i) $\mathcal{A} = \boldsymbol{D}, \boldsymbol{F}$, (ii) $(\mathcal{B}, \mathcal{Q}) = (\boldsymbol{C}, \boldsymbol{R}), (\boldsymbol{E}, \boldsymbol{S}^{-1})$, (iii) $\mathcal{D} = \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, \boldsymbol{F}$, and (iv) $\mathcal{P} = \boldsymbol{\Pi}, \boldsymbol{T}$. From the results above, we immediately obtain the following results:

Proposition 1.

(i) $C^\top R^{-1} C = \mathcal{A}^\top \mathcal{A} = \mathcal{B}^\top \mathcal{Q}^{-1} \mathcal{B}$,

(ii) $\mathcal{D}\mathcal{P} = \mathcal{D}_{\mathrm{r}}^{-1\top}\mathcal{P} = \mathbf{0}$,

(iii) *both* $[\mathcal{P}, \mathcal{D}^\top]$ *and* $[\mathcal{P}, \mathcal{D}_{\mathrm{r}}^{-1}]$ *are nonsingular, and*

(iv) $P_{\mathcal{D}^\top} = P_{\mathcal{D}_{\mathrm{r}}^{-1}} = Q_{\mathcal{P}}$.

3. Several Regressions Relating to (6) and Principle of Duality in Them

In this section, we provide a comprehensive list of penalized least squares regressions relating to (6), and reveal a principle of duality in them. The penalized regressions are, more precisely, those to compute $\widehat{c}$, $\widehat{u}$, $\widehat{\tau}$, $\widehat{\tau} + \widehat{c}$, $\widehat{c} + \widehat{u}$, and $\widehat{\tau} + \widehat{u}$.

3.1. Penalized Regressions to Compute $\widehat{\tau} + \widehat{c}$

Concerning $\widehat{\tau} + \widehat{c}$, we have the following results:

Lemma 7. *It follows that*

$$\widehat{\tau} + \widehat{c} = \arg\min_{f \in \mathbb{R}^n} \|y - f\|^2 + \lambda \|\mathcal{A} f\|^2 = \left(I_n + \lambda \mathcal{A}^\top \mathcal{A}\right)^{-1} y \tag{14}$$

$$= \arg\min_{f \in \mathbb{R}^n} \|y - f\|^2 + \lambda f^\top \mathcal{B}^\top \mathcal{Q}^{-1} \mathcal{B} f = \left(I_n + \lambda \mathcal{B}^\top \mathcal{Q}^{-1} \mathcal{B}\right)^{-1} y. \tag{15}$$

Proof of Lemma 7. From Proposition 1, we have $C^\top R^{-1} C = \mathcal{A}^\top \mathcal{A} = \mathcal{B}^\top \mathcal{Q}^{-1} \mathcal{B}$. Then, (2) and (3) can be represented as follows:

$$\widehat{f} = \arg\min_{f \in \mathbb{R}^n} \|y - f\|^2 + \lambda \|\mathcal{A} f\|^2 = \left(I_n + \lambda \mathcal{A}^\top \mathcal{A}\right)^{-1} y$$

$$= \arg\min_{f \in \mathbb{R}^n} \|y - f\|^2 + \lambda f^\top \mathcal{B}^\top \mathcal{Q}^{-1} \mathcal{B} f = \left(I_n + \lambda \mathcal{B}^\top \mathcal{Q}^{-1} \mathcal{B}\right)^{-1} y.$$

In addition, by definition of $\widehat{c}$, we have $\widehat{f} = \widehat{\tau} + \widehat{c}$. Hence, we obtain (14) and (15). □

3.2. Penalized Regressions to Compute $\widehat{c}$

Concerning $\widehat{c}$, we have the following results:

Lemma 8. *Consider the following penalized regressions:*

$$\widehat{\gamma} = \arg\min_{\gamma \in \mathbb{R}^{n-2}} \|y - \mathcal{A}_{\mathrm{r}}^{-1} \gamma\|^2 + \lambda \|\gamma\|^2 = (\mathcal{A}_{\mathrm{r}}^{-1\top} \mathcal{A}_{\mathrm{r}}^{-1} + \lambda I_{n-2})^{-1} \mathcal{A}_{\mathrm{r}}^{-1\top} y, \tag{16}$$

$$\widehat{\kappa} = \arg\min_{\kappa \in \mathbb{R}^{n-2}} \|y - \mathcal{B}_{\mathrm{r}}^{-1} \kappa\|^2 + \lambda \kappa^\top \mathcal{Q}^{-1} \kappa = (\mathcal{B}_{\mathrm{r}}^{-1\top} \mathcal{B}_{\mathrm{r}}^{-1} + \lambda \mathcal{Q}^{-1})^{-1} \mathcal{B}_{\mathrm{r}}^{-1\top} y. \tag{17}$$

Then, we have

$$\widehat{c} = \mathcal{A}_{\mathrm{r}}^{-1} \widehat{\gamma} = \mathcal{B}_{\mathrm{r}}^{-1} \widehat{\kappa}. \tag{18}$$

Proof of Lemma 8. Let $\boldsymbol{K} = [\mathcal{P}, \mathcal{A}_{\mathrm{r}}^{-1}]$. From Proposition 1, it follows that $\mathcal{A}\mathcal{P} = \mathbf{0}$, $\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{P} = \mathbf{0}$, and $\boldsymbol{K}$ is nonsingular. Accordingly, given that $\boldsymbol{K}^\top\boldsymbol{K} = \mathrm{diag}(\mathcal{P}^\top\mathcal{P}, \mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}_{\mathrm{r}}^{-1})$ and $\mathcal{A}\boldsymbol{K} = [\mathcal{A}\mathcal{P}, \mathcal{A}\mathcal{A}_{\mathrm{r}}^{-1}] = [\mathbf{0}, \boldsymbol{I}_{n-2}]$, it follows that

$$
\begin{aligned}
\widehat{\boldsymbol{f}} &= \boldsymbol{K}\left(\boldsymbol{K}^\top\boldsymbol{K} + \lambda\boldsymbol{K}^\top\mathcal{A}^\top\mathcal{A}\boldsymbol{K}\right)^{-1}\boldsymbol{K}^\top\boldsymbol{y} \\
&= [\mathcal{P}, \mathcal{A}_{\mathrm{r}}^{-1}]\begin{bmatrix}(\mathcal{P}^\top\mathcal{P})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}_{\mathrm{r}}^{-1} + \lambda\boldsymbol{I}_{n-2})^{-1}\end{bmatrix}\begin{bmatrix}\mathcal{P}^\top \\ \mathcal{A}_{\mathrm{r}}^{-1\top}\end{bmatrix}\boldsymbol{y} \\
&= \mathcal{P}(\mathcal{P}^\top\mathcal{P})^{-1}\mathcal{P}^\top\boldsymbol{y} + \mathcal{A}_{\mathrm{r}}^{-1}(\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}_{\mathrm{r}}^{-1} + \lambda\boldsymbol{I}_{n-2})^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{y} = \widehat{\boldsymbol{\tau}} + \mathcal{A}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\gamma}},
\end{aligned}
$$

from which we have $\widehat{\boldsymbol{f}} - \widehat{\boldsymbol{\tau}} = \mathcal{A}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\gamma}}$. Given $\widehat{\boldsymbol{f}} - \widehat{\boldsymbol{\tau}} = \widehat{\boldsymbol{c}}$, we thus obtain $\widehat{\boldsymbol{c}} = \mathcal{A}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\gamma}}$. Similarly, we can obtain $\widehat{\boldsymbol{c}} = \mathcal{B}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\kappa}}$. □

Lemma 9. *$\widehat{\boldsymbol{c}}$ can be calculated by the following penalized regressions:*

$$
\widehat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}\in\mathbb{R}^n} \|(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}) - \boldsymbol{c}\|^2 + \lambda\|\mathcal{A}\boldsymbol{c}\|^2 = \left(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A}\right)^{-1}(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}) \tag{19}
$$

$$
= \arg\min_{\boldsymbol{c}\in\mathbb{R}^n} \|(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}) - \boldsymbol{c}\|^2 + \lambda\boldsymbol{c}^\top\mathcal{B}^\top\mathcal{Q}^{-1}\mathcal{B}\boldsymbol{c} = \left(\boldsymbol{I}_n + \lambda\mathcal{B}^\top\mathcal{Q}^{-1}\mathcal{B}\right)^{-1}(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}). \tag{20}
$$

Proof of Lemma 9. Given (14), $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{c}}$, and $\mathcal{A}\mathcal{P} = \mathbf{0}$, we have

$$
\boldsymbol{y} = (\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A})\widehat{\boldsymbol{f}} = (\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A})(\widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{c}}) = \widehat{\boldsymbol{\tau}} + (\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A})\widehat{\boldsymbol{c}},
$$

which leads to (19). Similarly, we can obtain (20). □

Remark 2. *We add some more exposition about (16). Let $\boldsymbol{K} = [\mathcal{P}, \mathcal{A}_{\mathrm{r}}^{-1}]$ as before. In addition, let $\boldsymbol{\theta} = [\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top]^\top \in \mathbb{R}^n$ be a vector such that $\boldsymbol{f} = \boldsymbol{K}\boldsymbol{\theta} = \mathcal{P}\boldsymbol{\beta} + \mathcal{A}_{\mathrm{r}}^{-1}\boldsymbol{\gamma}$. Then, it follows that $\mathcal{A}\boldsymbol{f} = \mathcal{A}(\mathcal{P}\boldsymbol{\beta} + \mathcal{A}_{\mathrm{r}}^{-1}\boldsymbol{\gamma}) = \boldsymbol{\gamma}$. Given that $\boldsymbol{f} = \mathcal{P}\boldsymbol{\beta} + \mathcal{A}_{\mathrm{r}}^{-1}\boldsymbol{\gamma}$ and $\mathcal{A}\boldsymbol{f} = \boldsymbol{\gamma}$, the minimization problem in (14) can be represented as follows:*

$$
\min_{\boldsymbol{\beta}\in\mathbb{R}^2, \boldsymbol{\gamma}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \mathcal{P}\boldsymbol{\beta} - \mathcal{A}_{\mathrm{r}}^{-1}\boldsymbol{\gamma}\|^2 + \lambda\|\boldsymbol{\gamma}\|^2. \tag{21}
$$

It is noteworthy that $\boldsymbol{\beta}$ is not penalized in (21) and $(\mathcal{A}_{\mathrm{r}}^{-1})^\top\mathcal{P} = \mathbf{0}$. Thus, the minimization problem (21) can be decomposed into (16) and (40). Moreover, (21) gives the best linear unbiased predictors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of the following linear mixed model:

$$
\boldsymbol{y} = \mathcal{P}\boldsymbol{\beta} + \mathcal{A}_{\mathrm{r}}^{-1}\boldsymbol{\gamma} + \boldsymbol{u}, \quad [\boldsymbol{u}^\top, \boldsymbol{\gamma}^\top]^\top \sim N\left(\mathbf{0}, \mathrm{diag}(\sigma_u^2\boldsymbol{I}_n, \sigma_v^2\boldsymbol{I}_{n-2})\right), \tag{22}
$$

where $\lambda = \sigma_u^2/\sigma_v^2$.

Remark 3. *By using $\boldsymbol{C}_{\mathrm{r}}^{-1}$, ref. [6] derived the following expressions in our notations:*

$$
\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\tau}} + \boldsymbol{C}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\kappa}}, \quad \widehat{\boldsymbol{\kappa}} = (\boldsymbol{C}_{\mathrm{r}}^{-1\top}\boldsymbol{C}_{\mathrm{r}}^{-1} + \lambda\boldsymbol{R}^{-1})^{-1}\boldsymbol{C}_{\mathrm{r}}^{-1\top}\boldsymbol{y}. \tag{23}
$$

Here, we make the following remarks on (23).

(i) *First, $\widehat{\boldsymbol{\kappa}}$ is the solution of the following penalized regression:*

$$
\min_{\boldsymbol{\kappa}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \boldsymbol{C}_{\mathrm{r}}^{-1}\boldsymbol{\kappa}\|^2 + \lambda\boldsymbol{\kappa}^\top\boldsymbol{R}^{-1}\boldsymbol{\kappa}. \tag{24}
$$

(ii) *Moreover, (23) is a special case of $\widehat{\boldsymbol{c}} = \mathcal{B}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\kappa}}$ in (18).*

3.3. Penalized Regressions to Compute $\widehat{\boldsymbol{u}}$

Concerning $\widehat{\boldsymbol{u}}$, we have the following results:

Lemma 10. *Consider the following penalized regressions:*

$$\widehat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \mathcal{A}^\top\boldsymbol{\eta}\|^2 + \lambda^{-1}\|\boldsymbol{\eta}\|^2 = (\mathcal{A}\mathcal{A}^\top + \lambda^{-1}\boldsymbol{I}_{n-2})^{-1}\mathcal{A}\boldsymbol{y}, \tag{25}$$

$$\widehat{\boldsymbol{v}} = \arg\min_{\boldsymbol{v}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \mathcal{B}^\top\boldsymbol{v}\|^2 + \lambda^{-1}\boldsymbol{v}^\top\mathcal{Q}\boldsymbol{v} = (\mathcal{B}\mathcal{B}^\top + \lambda^{-1}\mathcal{Q})^{-1}\mathcal{B}\boldsymbol{y}. \tag{26}$$

Then, it follows that

$$\widehat{\boldsymbol{u}} = \mathcal{A}^\top\widehat{\boldsymbol{\eta}} = \mathcal{B}^\top\widehat{\boldsymbol{v}}. \tag{27}$$

Proof of Lemma 10. Applying the matrix inversion lemma to $\left(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A}\right)^{-1}$, we have

$$\left(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A}\right)^{-1} = \boldsymbol{I}_n - \mathcal{A}^\top(\mathcal{A}\mathcal{A}^\top + \lambda^{-1}\boldsymbol{I}_{n-2})^{-1}\mathcal{A}. \tag{28}$$

Postmultiplying (28) by $\boldsymbol{y}$ yields $\boldsymbol{y} - \widehat{\boldsymbol{f}} = \mathcal{A}^\top\widehat{\boldsymbol{\eta}}$. Given $\boldsymbol{y} - \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{u}}$, we thus have $\widehat{\boldsymbol{u}} = \mathcal{A}^\top\widehat{\boldsymbol{\eta}}$. Similarly, we can obtain $\widehat{\boldsymbol{u}} = \mathcal{B}^\top\widehat{\boldsymbol{v}}$. □

Lemma 11. *$\widehat{\boldsymbol{u}}$ can be calculated by the following penalized regressions:*

$$\begin{aligned}\widehat{\boldsymbol{u}} &= \arg\min_{\boldsymbol{u}\in\mathbb{R}^n} \|(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}) - \boldsymbol{u}\|^2 + \lambda^{-1}\|\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{u}\|^2 \\ &= \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\right)^{-1}(\boldsymbol{y} - \widehat{\boldsymbol{\tau}})\end{aligned} \tag{29}$$

and

$$\begin{aligned}\widehat{\boldsymbol{u}} &= \arg\min_{\boldsymbol{u}\in\mathbb{R}^n} \|(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}) - \boldsymbol{u}\|^2 + \lambda^{-1}\boldsymbol{u}^\top\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\boldsymbol{u} \\ &= \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\right)^{-1}(\boldsymbol{y} - \widehat{\boldsymbol{\tau}}).\end{aligned} \tag{30}$$

Proof of Lemma 11. Given (34), $\widehat{\boldsymbol{g}} = \widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{u}}$, and $\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{P} = \boldsymbol{0}$, we have

$$\begin{aligned}\boldsymbol{y} &= (\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top})\widehat{\boldsymbol{g}} = (\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top})(\widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{u}}) \\ &= \widehat{\boldsymbol{\tau}} + (\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top})\widehat{\boldsymbol{u}},\end{aligned}$$

which leads to (29). Similarly, we can obtain (30). □

Remark 4. *In [2] and ([3], p. 20), there are equations expressed in our notation as follows:*

$$(\boldsymbol{R} + \lambda\boldsymbol{C}\boldsymbol{C}^\top)\boldsymbol{\phi} = \boldsymbol{C}\boldsymbol{y}, \quad \widehat{\boldsymbol{f}} = \boldsymbol{y} - \lambda\boldsymbol{C}^\top\boldsymbol{\phi}. \tag{31}$$

Here, we make the following remarks on (31).

(i) First, these lead to a penalized least squares problem. Given that $\widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{f}}$, removing $\boldsymbol{\phi}$ from the above equations leads to

$$\begin{aligned}\widehat{\boldsymbol{u}} &= \boldsymbol{y} - \widehat{\boldsymbol{f}} = \lambda\boldsymbol{C}^\top(\boldsymbol{R} + \lambda\boldsymbol{C}\boldsymbol{C}^\top)^{-1}\boldsymbol{C}\boldsymbol{y} \\ &= \boldsymbol{C}^\top(\boldsymbol{C}\boldsymbol{C}^\top + \lambda^{-1}\boldsymbol{R})^{-1}\boldsymbol{C}\boldsymbol{y} = \boldsymbol{C}^\top\widehat{\boldsymbol{v}},\end{aligned} \tag{32}$$

where

$$\widehat{\boldsymbol{v}} = \arg\min_{\boldsymbol{v}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \boldsymbol{C}^\top \boldsymbol{v}\|^2 + \lambda^{-1}\boldsymbol{v}^\top \boldsymbol{R}\boldsymbol{v}. \tag{33}$$

(ii) *Moreover,* (32) *is a special case of* $\widehat{\boldsymbol{u}} = \mathcal{B}^\top \widehat{\boldsymbol{v}}$ *in* (27).

3.4. Penalized Regression to Compute $\widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{u}}$

Concerning $\widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{u}}$, we have the following results:

Lemma 12. *Let* $\widehat{\boldsymbol{g}} = \widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{u}}$*. Then, it follows that*

$$\widehat{\boldsymbol{g}} = \arg\min_{\boldsymbol{g}\in\mathbb{R}^n} \|\boldsymbol{y} - \boldsymbol{g}\|^2 + \lambda^{-1}\|\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{g}\|^2 = \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\right)^{-1}\boldsymbol{y} \tag{34}$$

$$= \arg\min_{\boldsymbol{g}\in\mathbb{R}^n} \|\boldsymbol{y} - \boldsymbol{g}\|^2 + \lambda^{-1}\boldsymbol{g}^\top\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\boldsymbol{g} = \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\right)^{-1}\boldsymbol{y}. \tag{35}$$

Proof of Lemma 12. Let $\boldsymbol{J} = [\mathcal{P}, \mathcal{A}^\top]$. From Proposition 1, it follows that $\mathcal{A}\mathcal{P} = \boldsymbol{0}$, $\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{P} = \boldsymbol{0}$, and $\boldsymbol{J}$ is nonsingular. Accordingly, given that $\boldsymbol{J}^\top\boldsymbol{J} = \mathrm{diag}(\mathcal{P}^\top\mathcal{P}, \mathcal{A}\mathcal{A}^\top)$ and $\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{J} = [\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{P}, \mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}^\top] = [\boldsymbol{0}, \boldsymbol{I}_{n-2}]$, it follows that

$$\begin{aligned}
&\left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\right)^{-1}\boldsymbol{y} \\
&\quad = \boldsymbol{J}\left(\boldsymbol{J}^\top\boldsymbol{J} + \lambda^{-1}\boldsymbol{J}^\top\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{J}\right)^{-1}\boldsymbol{J}^\top\boldsymbol{y} \\
&\quad = [\mathcal{P}, \mathcal{A}^\top]\begin{bmatrix}(\mathcal{P}^\top\mathcal{P})^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & (\mathcal{A}\mathcal{A}^\top + \lambda^{-1}\boldsymbol{I}_{n-2})^{-1}\end{bmatrix}\begin{bmatrix}\mathcal{P}^\top \\ \mathcal{A}\end{bmatrix}\boldsymbol{y} \\
&\quad = \mathcal{P}(\mathcal{P}^\top\mathcal{P})^{-1}\mathcal{P}^\top\boldsymbol{y} + \mathcal{A}^\top(\mathcal{A}\mathcal{A}^\top + \lambda^{-1}\boldsymbol{I}_{n-2})^{-1}\mathcal{A}\boldsymbol{y} = \widehat{\boldsymbol{\tau}} + \mathcal{A}^\top\widehat{\boldsymbol{\eta}}.
\end{aligned}$$

Given $\widehat{\boldsymbol{u}} = \mathcal{A}^\top\widehat{\boldsymbol{\eta}}$, we obtain (34). Similarly, we can obtain (35). □

Remark 5. *Similarly to Remark 2, we add some more exposition about (25). Let* $\boldsymbol{\xi} = [\boldsymbol{\beta}^\top, \boldsymbol{\eta}^\top]^\top \in \mathbb{R}^n$ *be such that* $\boldsymbol{g} = \boldsymbol{J}\boldsymbol{\xi} = \mathcal{P}\boldsymbol{\beta} + \mathcal{A}^\top\boldsymbol{\eta}$*. As stated,* $\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{J} = [\boldsymbol{0}, \boldsymbol{I}_{n-2}]$*. Then, it follows that*

$$\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{g} = \mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{J}\boldsymbol{\xi} = \boldsymbol{\eta}. \tag{36}$$

Given $\boldsymbol{g} = \mathcal{P}\boldsymbol{\beta} + \mathcal{A}^\top\boldsymbol{\eta}$ *and* $\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{g} = \boldsymbol{\eta}$*, the minimization problem (34) can be represented as follows:*

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^2, \boldsymbol{\eta}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \mathcal{P}\boldsymbol{\beta} - \mathcal{A}^\top\boldsymbol{\eta}\|^2 + \lambda^{-1}\|\boldsymbol{\eta}\|^2. \tag{37}$$

Again, it is noteworthy that $\boldsymbol{\beta}$ *is not penalized in (37). Moreover, it follows that* $(\mathcal{A}^\top)^\top\mathcal{P} = \mathcal{A}\mathcal{P} = \boldsymbol{0}$*. Thus, the minimization problem (37) can be decomposed into (25) and (40).*

3.5. Ordinary Regressions to Compute $\widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}}$ *and* $\widehat{\boldsymbol{\tau}}$

Concerning $\widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{\tau}}$, we have the following results:

Lemma 13.

(i) *Let* $\widehat{\boldsymbol{h}} = \mathcal{D}^\top\widehat{\boldsymbol{\alpha}}$*, where*

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^{n-2}} \|\boldsymbol{y} - \mathcal{D}^\top\boldsymbol{\alpha}\|^2 = (\mathcal{D}\mathcal{D}^\top)^{-1}\mathcal{D}\boldsymbol{y}. \tag{38}$$

Then, it follows that

$$\widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}} = \widehat{\boldsymbol{h}}. \tag{39}$$

(ii) *It follows that* $\widehat{\boldsymbol{\tau}} = \mathcal{P}\widehat{\boldsymbol{\beta}}$, *where*

$$\widehat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^2} \|\boldsymbol{y} - \mathcal{P}\boldsymbol{\beta}\|^2 = (\mathcal{P}^\top \mathcal{P})^{-1}\mathcal{P}^\top \boldsymbol{y}. \tag{40}$$

Proof of Lemma 13. Given Proposition 1, both results are easily obtainable. For example, the former result can be proved as follows:

$$\widehat{\boldsymbol{h}} = \mathcal{D}^\top \widehat{\boldsymbol{\alpha}} = \boldsymbol{P}_{\mathcal{D}^\top}\boldsymbol{y} = \boldsymbol{Q}_{\mathcal{P}}\boldsymbol{y} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}} = \widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}}.$$

□

Remark 6. *From Proposition 1, we also have* $\widehat{\boldsymbol{h}}(= \widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}}) = \mathcal{D}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\rho}}$, *where*

$$\widehat{\boldsymbol{\rho}} = \arg \min_{\boldsymbol{\rho} \in \mathbb{R}^{n-2}} \|\boldsymbol{y} - \mathcal{D}_{\mathrm{r}}^{-1}\boldsymbol{\rho}\|^2 = (\mathcal{D}_{\mathrm{r}}^{-1\top}\mathcal{D}_{\mathrm{r}}^{-1})^{-1}\mathcal{D}_{\mathrm{r}}^{-1\top}\boldsymbol{y}. \tag{41}$$

3.6. Principle of Duality in the Penalized Regressions

See the second columns of Tables 1 and 2. In the columns, the penalized regressions shown above are arranged in pairs that mirror one another. We reveal a principle of duality in the penalized regressions. As stated in Section 1, (D1) is obtainable by replacing $\mathcal{A}^\top, \lambda$ in (P1) by $\mathcal{A}_{\mathrm{r}}^{-1}, \lambda^{-1}$, respectively. Likewise, for example, (D6) in Table 2 is obtainable by replacing $\mathcal{B}^\top, \mathcal{Q}, \lambda^{-1}$ in (P6) by $\mathcal{B}_{\mathrm{r}}^{-1}, \mathcal{Q}^{-1}, \lambda$, respectively. In Tables 1 and 2, we may observe five other pairs of regressions that are duals of each other. From the seven dual pairs shown in Tables 1 and 2, we observe that the following principle exists:

Proposition 2 (**Principle of duality**)**.** *The regressions labeled with the letter D in Tables 1 and 2, for example, (D1), are obtainable by replacing each occurrence of* $\mathcal{A}^\top, \mathcal{B}^\top, \mathcal{D}^\top, \mathcal{Q}, \mathcal{Q}^{-1}, \lambda, \lambda^{-1}$ *in the regressions labeled with the letter P, for example, (P1), by* $\mathcal{A}_{\mathrm{r}}^{-1}, \mathcal{B}_{\mathrm{r}}^{-1}, \mathcal{D}_{\mathrm{r}}^{-1}, \mathcal{Q}^{-1}, \mathcal{Q}, \lambda^{-1}, \lambda$, *respectively.*

Table 1. Most of the main results (I).

	Regressions Relating to the Cubic Smoothing Spline	Average	$\lambda \to \infty$	$\lambda \to 0$	$\boldsymbol{y} \in \mathcal{S}(\mathcal{P})$	Sum	$\perp$
(P1)	$\widehat{\boldsymbol{f}}(=\widehat{\boldsymbol{\tau}}+\widehat{\boldsymbol{c}}) = \arg\min_{\boldsymbol{f}\in\mathbb{R}^n} \lVert\boldsymbol{y}-\boldsymbol{f}\rVert^2 + \lambda\lVert\mathcal{A}\boldsymbol{f}\rVert^2 = \left(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A}\right)^{-1}\boldsymbol{y}$	$\bar{y}$	$\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}$	$\boldsymbol{y}$	1	
(D1)	$\widehat{\boldsymbol{g}}(=\widehat{\boldsymbol{\tau}}+\widehat{\boldsymbol{u}}) = \arg\min_{\boldsymbol{g}\in\mathbb{R}^n} \lVert\boldsymbol{y}-\boldsymbol{g}\rVert^2 + \lambda^{-1}\lVert\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{g}\rVert^2 = \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\right)^{-1}\boldsymbol{y}$	$\bar{y}$	$\boldsymbol{y}$	$\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}$	1	
(P2)	$\widehat{\boldsymbol{u}} = \mathcal{A}^\top\widehat{\boldsymbol{\eta}}$, where $\widehat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}\in\mathbb{R}^{n-2}} \lVert\boldsymbol{y}-\mathcal{A}^\top\boldsymbol{\eta}\rVert^2 + \lambda^{-1}\lVert\boldsymbol{\eta}\rVert^2 = (\mathcal{A}\mathcal{A}^\top + \lambda^{-1}\boldsymbol{I}_{n-2})^{-1}\mathcal{A}\boldsymbol{y}$	0	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	$\boldsymbol{0}$	0	$\circ$
(D2)	$\widehat{\boldsymbol{c}} = \mathcal{A}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\gamma}}$, where $\widehat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{n-2}} \lVert\boldsymbol{y}-\mathcal{A}_{\mathrm{r}}^{-1}\boldsymbol{\gamma}\rVert^2 + \lambda\lVert\boldsymbol{\gamma}\rVert^2 = (\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}_{\mathrm{r}}^{-1} + \lambda\boldsymbol{I}_{n-2})^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{y}$	0	$\boldsymbol{0}$	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	0	$\circ$
(P3)	$\widehat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}\in\mathbb{R}^n} \lVert(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})-\boldsymbol{c}\rVert^2 + \lambda\lVert\mathcal{A}\boldsymbol{c}\rVert^2 = \left(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A}\right)^{-1}(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})$	0	$\boldsymbol{0}$	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	0	$\circ$
(D3)	$\widehat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}\in\mathbb{R}^n} \lVert(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})-\boldsymbol{u}\rVert^2 + \lambda^{-1}\lVert\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{u}\rVert^2 = \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\right)^{-1}(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})$	0	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	$\boldsymbol{0}$	0	$\circ$
(P4)	$\widehat{\boldsymbol{h}}(=\widehat{\boldsymbol{c}}+\widehat{\boldsymbol{u}}) = \mathcal{D}^\top\widehat{\boldsymbol{\alpha}}$, where $\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^{n-2}} \lVert\boldsymbol{y}-\mathcal{D}^\top\boldsymbol{\alpha}\rVert^2 = (\mathcal{D}\mathcal{D}^\top)^{-1}\mathcal{D}\boldsymbol{y}$	0	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	0	$\circ$
(D4)	$\widehat{\boldsymbol{h}}(=\widehat{\boldsymbol{c}}+\widehat{\boldsymbol{u}}) = \mathcal{D}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\rho}}$, where $\widehat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}\in\mathbb{R}^{n-2}} \lVert\boldsymbol{y}-\mathcal{D}_{\mathrm{r}}^{-1}\boldsymbol{\rho}\rVert^2 = (\mathcal{D}_{\mathrm{r}}^{-1\top}\mathcal{D}_{\mathrm{r}}^{-1})^{-1}\mathcal{D}_{\mathrm{r}}^{-1\top}\boldsymbol{y}$	0	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	0	$\circ$
	$\widehat{\boldsymbol{\tau}} = \mathcal{P}\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^2} \lVert\boldsymbol{y}-\mathcal{P}\boldsymbol{\beta}\rVert^2 = (\mathcal{P}^\top\mathcal{P})^{-1}\mathcal{P}^\top\boldsymbol{y}$	$\bar{y}$	$\widehat{\boldsymbol{\tau}}$	$\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}$	1	

$\boldsymbol{y} = \widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}}$. $\mathcal{A} = \boldsymbol{D}, \boldsymbol{F}$. $\mathcal{D} = \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, \boldsymbol{F}$. $\boldsymbol{M}_{\mathrm{r}}^{-1} = \boldsymbol{M}^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-1}$ for $\boldsymbol{M} = \mathcal{A}, \mathcal{D}$. $\mathcal{P} = \boldsymbol{\Pi}, \boldsymbol{T}$. $\lambda > 0$ is a smoothing/tuning parameter. $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. $\mathcal{S}(\mathcal{P})$ denotes the column space of $\mathcal{P}$. 'Sum' denotes the sum of the entries in each row of the hat matrices. $\circ$ indicates that the corresponding component belongs to the orthogonal complement of $\mathcal{S}(\mathcal{P})$.

Table 2. Most of the main results (II).

	Regressions Relating to the Cubic Smoothing Spline	Average	$\lambda \to \infty$	$\lambda \to 0$	$\boldsymbol{y} \in \mathcal{S}(\mathcal{P})$	Sum	$\perp$
(P5)	$\widehat{\boldsymbol{f}}(=\widehat{\boldsymbol{\tau}}+\widehat{\boldsymbol{c}}) = \arg\min_{\boldsymbol{f}\in\mathbb{R}^n} \lVert\boldsymbol{y}-\boldsymbol{f}\rVert^2 + \lambda\boldsymbol{f}^\top\mathcal{B}^\top\mathcal{Q}^{-1}\mathcal{B}\boldsymbol{f} = \left(\boldsymbol{I}_n + \lambda\mathcal{B}^\top\mathcal{Q}^{-1}\mathcal{B}\right)^{-1}\boldsymbol{y}$	$\bar{y}$	$\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}$	$\boldsymbol{y}$	1	
(D5)	$\widehat{\boldsymbol{g}}(=\widehat{\boldsymbol{\tau}}+\widehat{\boldsymbol{u}}) = \arg\min_{\boldsymbol{g}\in\mathbb{R}^n} \lVert\boldsymbol{y}-\boldsymbol{g}\rVert^2 + \lambda^{-1}\boldsymbol{g}^\top\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\boldsymbol{g} = \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\right)^{-1}\boldsymbol{y}$	$\bar{y}$	$\boldsymbol{y}$	$\widehat{\boldsymbol{\tau}}$	$\boldsymbol{y}$	1	
(P6)	$\widehat{\boldsymbol{u}} = \mathcal{B}^\top\widehat{\boldsymbol{v}}$, where $\widehat{\boldsymbol{v}} = \arg\min_{\boldsymbol{v}\in\mathbb{R}^{n-2}} \lVert\boldsymbol{y}-\mathcal{B}^\top\boldsymbol{v}\rVert^2 + \lambda^{-1}\boldsymbol{v}^\top\mathcal{Q}\boldsymbol{v} = (\mathcal{B}\mathcal{B}^\top + \lambda^{-1}\mathcal{Q})^{-1}\mathcal{B}\boldsymbol{y}$	0	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	$\boldsymbol{0}$	0	$\circ$
(D6)	$\widehat{\boldsymbol{c}} = \mathcal{B}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\kappa}}$, where $\widehat{\boldsymbol{\kappa}} = \arg\min_{\boldsymbol{\kappa}\in\mathbb{R}^{n-2}} \lVert\boldsymbol{y}-\mathcal{B}_{\mathrm{r}}^{-1}\boldsymbol{\kappa}\rVert^2 + \lambda\boldsymbol{\kappa}^\top\mathcal{Q}^{-1}\boldsymbol{\kappa} = (\mathcal{B}_{\mathrm{r}}^{-1\top}\mathcal{B}_{\mathrm{r}}^{-1} + \lambda\mathcal{Q}^{-1})^{-1}\mathcal{B}_{\mathrm{r}}^{-1\top}\boldsymbol{y}$	0	$\boldsymbol{0}$	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	0	$\circ$
(P7)	$\widehat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}\in\mathbb{R}^n} \lVert(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})-\boldsymbol{c}\rVert^2 + \lambda\boldsymbol{c}^\top\mathcal{B}^\top\mathcal{Q}^{-1}\mathcal{B}\boldsymbol{c} = \left(\boldsymbol{I}_n + \lambda\mathcal{B}^\top\mathcal{Q}^{-1}\mathcal{B}\right)^{-1}(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})$	0	$\boldsymbol{0}$	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	0	$\circ$
(D7)	$\widehat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}\in\mathbb{R}^n} \lVert(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})-\boldsymbol{u}\rVert^2 + \lambda^{-1}\boldsymbol{u}^\top\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\boldsymbol{u} = \left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{B}_{\mathrm{r}}^{-1}\mathcal{Q}\mathcal{B}_{\mathrm{r}}^{-1\top}\right)^{-1}(\boldsymbol{y}-\widehat{\boldsymbol{\tau}})$	0	$\boldsymbol{y}-\widehat{\boldsymbol{\tau}}$	$\boldsymbol{0}$	$\boldsymbol{0}$	0	$\circ$

$\boldsymbol{y} = \widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}}$. $(\mathcal{B}, \mathcal{Q}) = (\boldsymbol{C}, \boldsymbol{R}), (\boldsymbol{E}, \boldsymbol{S}^{-1})$. $\mathcal{B}_{\mathrm{r}}^{-1} = \mathcal{B}^\top(\mathcal{B}\mathcal{B}^\top)^{-1}$. $\lambda > 0$ is a smoothing/tuning parameter. $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. $\mathcal{S}(\mathcal{P})$ denotes the column space of $\mathcal{P}$, where $\mathcal{P} = \boldsymbol{\Pi}, \boldsymbol{T}$. 'Sum' denotes the sum of the entries in each row of the hat matrices. $\circ$ indicates that the corresponding component belongs to the orthogonal complement of $\mathcal{S}(\mathcal{P})$.

4. Results That Are Obtainable from the Regressions

In this section, we show how the regressions listed in the previous section are of use to obtain a deeper understanding of the fitting a cubic smoothing spline. Before proceeding, recall $\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{c}}$ and so on.

First, given that (16) is a ridge regression, it immediately follows that $\lim_{\lambda\to\infty} \widehat{\boldsymbol{\gamma}} = \mathbf{0}$, which leads to $\lim_{\lambda\to\infty} \widehat{\boldsymbol{c}} = \mathcal{A}_{\mathrm{r}}^{-1} \lim_{\lambda\to\infty} \widehat{\boldsymbol{\gamma}} = \mathbf{0}$ and at the same time we have

$$\lim_{\lambda\to\infty} \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\tau}} + \lim_{\lambda\to\infty} \widehat{\boldsymbol{c}} = \widehat{\boldsymbol{\tau}}, \tag{42}$$

$$\lim_{\lambda\to\infty} \widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}} - \lim_{\lambda\to\infty} \widehat{\boldsymbol{c}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}}, \tag{43}$$

$$\lim_{\lambda\to\infty} \widehat{\boldsymbol{g}} = \widehat{\boldsymbol{\tau}} + \lim_{\lambda\to\infty} \widehat{\boldsymbol{u}} = \widehat{\boldsymbol{\tau}} + (\boldsymbol{y} - \widehat{\boldsymbol{\tau}}) = \boldsymbol{y}. \tag{44}$$

Second, (25) is again a ridge regression, we have $\lim_{\lambda\to 0} \widehat{\boldsymbol{\eta}} = \mathbf{0}$, which yields $\lim_{\lambda\to 0} \widehat{\boldsymbol{u}} = \mathcal{A}^{\top} \lim_{\lambda\to 0} \widehat{\boldsymbol{\eta}} = \mathbf{0}$ and accordingly we obtain

$$\lim_{\lambda\to 0} \widehat{\boldsymbol{f}} = \boldsymbol{y} - \lim_{\lambda\to 0} \widehat{\boldsymbol{u}} = \boldsymbol{y}, \tag{45}$$

$$\lim_{\lambda\to 0} \widehat{\boldsymbol{c}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}} - \lim_{\lambda\to 0} \widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}}, \tag{46}$$

$$\lim_{\lambda\to 0} \widehat{\boldsymbol{g}} = \widehat{\boldsymbol{\tau}} + \lim_{\lambda\to 0} \widehat{\boldsymbol{u}} = \widehat{\boldsymbol{\tau}}. \tag{47}$$

Third, from (19) and $\widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}} - \widehat{\boldsymbol{c}}$, we have

$$\widehat{\boldsymbol{c}} = \left(\boldsymbol{I}_n + \lambda\mathcal{A}^{\top}\mathcal{A}\right)^{-1} (\boldsymbol{y} - \widehat{\boldsymbol{\tau}}), \tag{48}$$

$$\widehat{\boldsymbol{u}} = \left\{\boldsymbol{I}_n - (\boldsymbol{I}_n + \lambda\mathcal{A}^{\top}\mathcal{A})^{-1}\right\} (\boldsymbol{y} - \widehat{\boldsymbol{\tau}}). \tag{49}$$

Thus, $\widehat{\boldsymbol{f}}$ can be represented as

$$\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\tau}} + \left(\boldsymbol{I}_n + \lambda\mathcal{A}^{\top}\mathcal{A}\right)^{-1} (\boldsymbol{y} - \widehat{\boldsymbol{\tau}}). \tag{50}$$

Here, we remark that, given that $\left(\boldsymbol{I}_n + \lambda\mathcal{A}^{\top}\mathcal{A}\right)^{-1}$ is a smoother matrix, the second term on the right-hand side of (50) represents a low-frequency part of $\boldsymbol{y} - \widehat{\boldsymbol{\tau}}$. In addition, from (49), $\widehat{\boldsymbol{u}}$ represents a high-frequency part of $\boldsymbol{y} - \widehat{\boldsymbol{\tau}}$. Thus, $\widehat{\boldsymbol{c}}$ is generally smoother than $\widehat{\boldsymbol{u}}$.

Fourth, given $\mathcal{A}\mathcal{P} = \mathbf{0}$, $\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{P} = \mathbf{0}$, $\widehat{\boldsymbol{c}} = \mathcal{A}_{\mathrm{r}}^{-1}\widehat{\boldsymbol{\gamma}}$, and $\widehat{\boldsymbol{u}} = \mathcal{A}^{\top}\widehat{\boldsymbol{\eta}}$, we have

$$\widehat{\boldsymbol{\zeta}}^{\top}\widehat{\boldsymbol{\tau}} = 0, \quad \widehat{\boldsymbol{\zeta}} = \widehat{\boldsymbol{c}}, \widehat{\boldsymbol{u}}, \widehat{\boldsymbol{h}}. \tag{51}$$

Fifth, given $\mathcal{A}\mathcal{P} = \mathbf{0}$, $\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{P} = \mathbf{0}$, (28), and

$$\left(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\right)^{-1} = \boldsymbol{I}_n - \mathcal{A}_{\mathrm{r}}^{-1}(\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}_{\mathrm{r}}^{-1} + \lambda\boldsymbol{I}_{n-2})^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}, \tag{52}$$

if $\boldsymbol{y} \in \mathcal{S}(\mathcal{P})$, or in other words, if $\boldsymbol{y} = \mathcal{P}\boldsymbol{\psi}$, then we have

$$\widehat{\boldsymbol{\tau}} = \boldsymbol{y}, \quad \widehat{\boldsymbol{f}} = \boldsymbol{y}, \quad \widehat{\boldsymbol{g}} = \boldsymbol{y}, \quad \widehat{\boldsymbol{c}} = \mathbf{0}, \quad \widehat{\boldsymbol{u}} = \mathbf{0}, \quad \widehat{\boldsymbol{h}} = \mathbf{0}. \tag{53}$$

Sixth, given $\boldsymbol{\iota}_n \in \mathcal{S}(\mathcal{P})$, we have

$$\boldsymbol{P}_{\mathcal{P}}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n, \tag{54}$$
$$(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A})^{-1}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n, \tag{55}$$
$$(\boldsymbol{I}_n + \lambda^{-1}\mathcal{A}_{\mathrm{r}}^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top})^{-1}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n, \tag{56}$$
$$\mathcal{A}_{\mathrm{r}}^{-1}(\mathcal{A}_{\mathrm{r}}^{-1\top}\mathcal{A}_{\mathrm{r}}^{-1} + \lambda\boldsymbol{I}_{n-2})^{-1}\mathcal{A}_{\mathrm{r}}^{-1\top}\boldsymbol{\iota}_n = \boldsymbol{0}, \tag{57}$$
$$\mathcal{A}^\top(\mathcal{A}\mathcal{A}^\top + \lambda^{-1}\boldsymbol{I}_{n-2})^{-1}\mathcal{A}\boldsymbol{\iota}_n = \boldsymbol{0}, \tag{58}$$
$$\boldsymbol{P}_{\mathcal{D}^\top}\boldsymbol{\iota}_n = \boldsymbol{0}. \tag{59}$$

Note that $(\boldsymbol{I}_n + \lambda\mathcal{A}^\top\mathcal{A})^{-1}\boldsymbol{\iota}_n = \boldsymbol{\iota}_n$, for example, indicates that the sum of the entries in each row of the hat matrix of $\widehat{\boldsymbol{f}}$ equals unity.

Seventh, given (54)–(59), we have

$$\frac{1}{n}\boldsymbol{\iota}_n^\top\widehat{\boldsymbol{\zeta}} = \bar{y}, \quad \widehat{\boldsymbol{\zeta}} = \widehat{\boldsymbol{\tau}}, \widehat{\boldsymbol{f}}, \widehat{\boldsymbol{g}}, \tag{60}$$
$$\frac{1}{n}\boldsymbol{\iota}_n^\top\widehat{\boldsymbol{\zeta}} = 0, \quad \widehat{\boldsymbol{\zeta}} = \widehat{\boldsymbol{c}}, \widehat{\boldsymbol{u}}, \widehat{\boldsymbol{h}}, \tag{61}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. $\frac{1}{n}\boldsymbol{\iota}_n^\top\widehat{\boldsymbol{f}} = \bar{y}$, for example, shows that $\frac{1}{n}\sum_{i=1}^n \widehat{f}_i = \bar{y}$.

5. Illustrations of Some Results

In this section, we illustrate some of the results in the previous sections by a real data example.

Panel A of Figure 1 shows a scatter plot of North Pacific sea surface temperature (SST) anomalies (1891–2018). SST is an essential climate variable and has been used for the detection of climate change. See, for example, Høyer and Karagali [7] and the references therein. We obtained the data from the website of the Japan Meteorological Agency. The solid line in the panel plots $(x_i, \widehat{\tau}_i)$ for $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\tau}} = [\widehat{\tau}_1, \ldots, \widehat{\tau}_n]^\top$ in (4) and $n = 128$. Panel B of Figure 1 depicts a scatter plot of $(x_i, y_i - \widehat{\tau}_i)$ for $i = 1, \ldots, n$. The solid line in the panel plots $(x_i, \widehat{c}_i)$ for $i = 1, \ldots, n$, where $\widehat{\boldsymbol{c}} = [\widehat{c}_1, \ldots, \widehat{c}_n]^\top$ is calculated by (18) with $\lambda = 10^3$. The solid line in Panel C denotes $(x_i, \widehat{f}_i)$, where $\widehat{\boldsymbol{f}} = [\widehat{f}_1, \ldots, \widehat{f}_n]^\top$ is calculated by (14) with $\lambda = 10^3$. Panel D illustrates a scatter plot of $(x_i, y_i - \widehat{\tau}_i)$ for $i = 1, \ldots, n$. The solid line in the panel plots $(x_i, \widehat{u}_i)$ for $i = 1, \ldots, n$, where $\widehat{\boldsymbol{u}} = [\widehat{u}_1, \ldots, \widehat{u}_n]^\top$ is calculated by (27) with $\lambda = 10^3$. Figures 2–4 correspond to the cases such that $\lambda = 10^5, 10^{10}, 10^{-10}$, respectively.

Recall that concerning $\boldsymbol{y}$, $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{c}}$, $\widehat{\boldsymbol{f}}$, and $\widehat{\boldsymbol{u}}$, the following equations hold:

$$\widehat{\boldsymbol{\tau}} + \widehat{\boldsymbol{c}} = \widehat{\boldsymbol{f}}, \quad \widehat{\boldsymbol{c}} + \widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}}, \quad \lim_{\lambda\to\infty}\widehat{\boldsymbol{c}} = \boldsymbol{0}, \quad \lim_{\lambda\to\infty}\widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\tau}},$$
$$\lim_{\lambda\to\infty}\widehat{\boldsymbol{u}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}}, \quad \lim_{\lambda\to 0}\widehat{\boldsymbol{c}} = \boldsymbol{y} - \widehat{\boldsymbol{\tau}}, \quad \lim_{\lambda\to 0}\widehat{\boldsymbol{f}} = \boldsymbol{y}, \quad \lim_{\lambda\to 0}\widehat{\boldsymbol{u}} = \boldsymbol{0}.$$

From Figures 1–4, we can observe that these theoretical results are well illustrated in these figures. For example, from Panel D in Figure 4, we can observe that $\widehat{\boldsymbol{u}}$ almost equals $\boldsymbol{0}$ when $\lambda = 10^{-10}$.

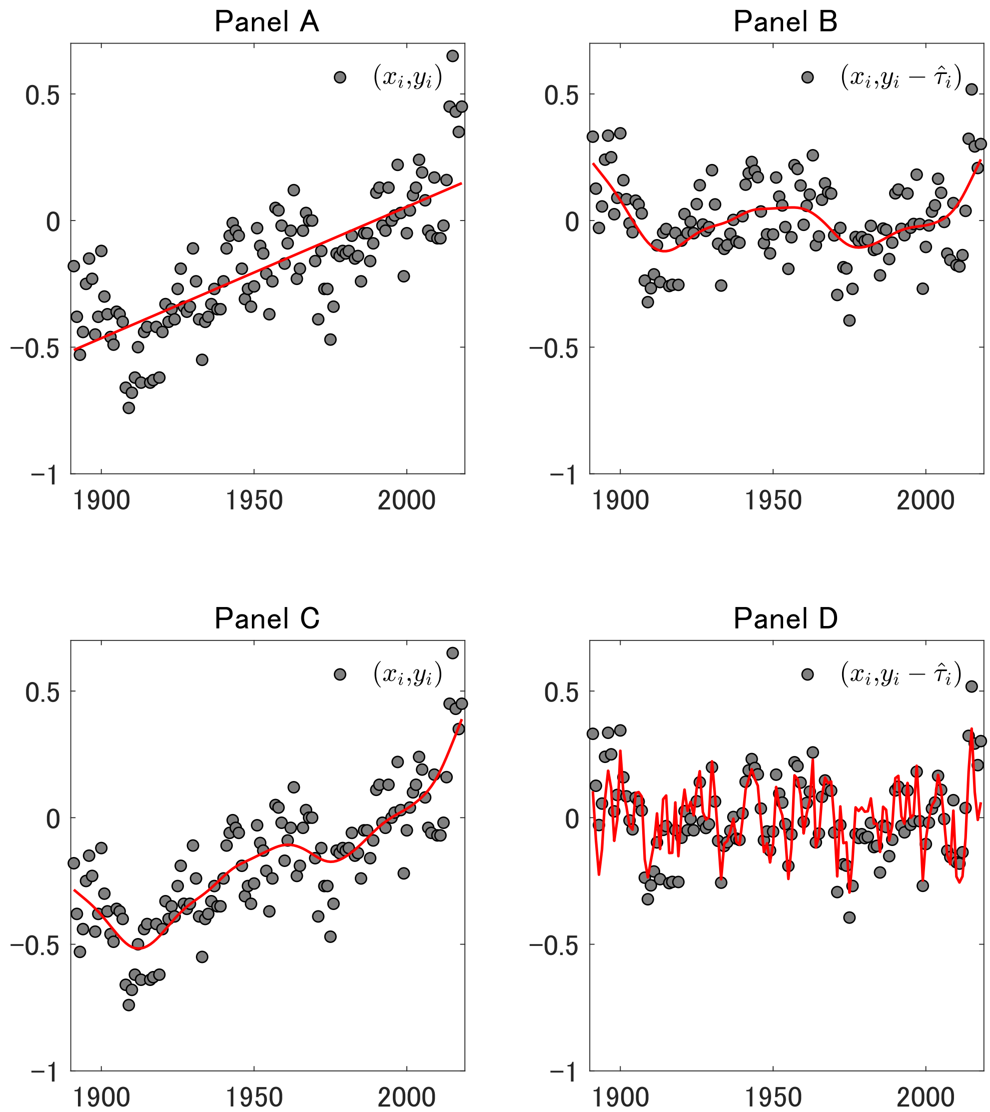

Figure 1. Panel A shows a scatter plot of North Pacific sea surface temperature anomalies (1891–2018). The solid line in the panel plots $(x_i, \widehat{\tau}_i)$ for $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\tau}} = [\widehat{\tau}_1, \ldots, \widehat{\tau}_n]^\top$ in (4) and $n = 128$. Panel B depicts a scatter plot of $(x_i, y_i - \widehat{\tau}_i)$ for $i = 1, \ldots, n$. The solid line in the panel plots $(x_i, \widehat{c}_i)$ for $i = 1, \ldots, n$, where $\widehat{\boldsymbol{c}} = [\widehat{c}_1, \ldots, \widehat{c}_n]^\top$ is calculated by (18) with $\lambda = 10^3$. The solid line in Panel C denotes $(x_i, \widehat{f}_i)$, where $\widehat{\boldsymbol{f}} = [\widehat{f}_1, \ldots, \widehat{f}_n]^\top$ is calculated by (14) with $\lambda = 10^3$. Panel D illustrates a scatter plot of $(x_i, y_i - \widehat{\tau}_i)$ for $i = 1, \ldots, n$. The solid line in the panel plots $(x_i, \widehat{u}_i)$ for $i = 1, \ldots, n$, where $\widehat{\boldsymbol{u}} = [\widehat{u}_1, \ldots, \widehat{u}_n]^\top$ is calculated by (27) with $\lambda = 10^3$.

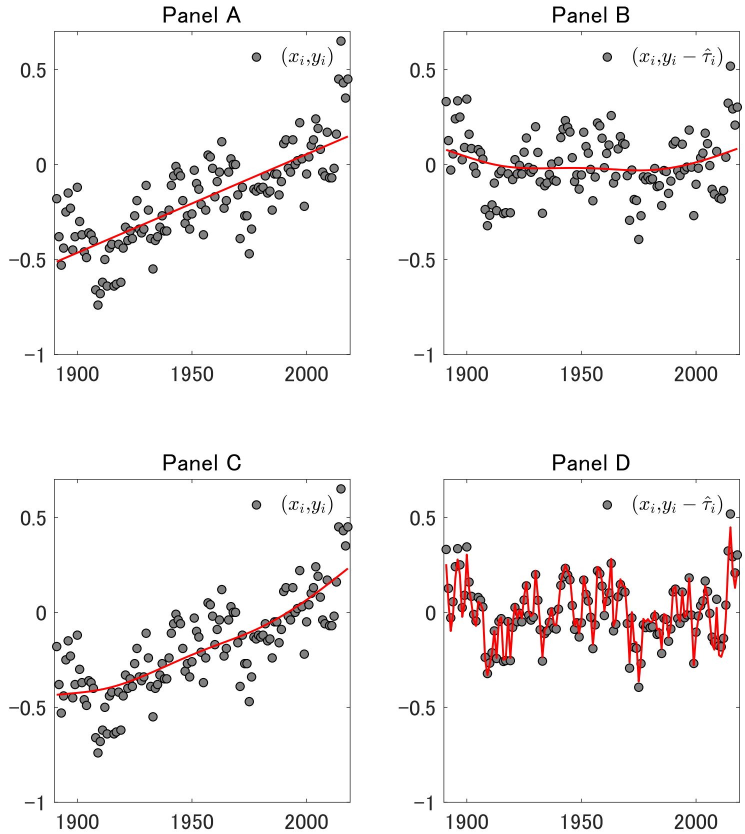

Figure 2. This figure corresponds to the case where $\lambda = 10^5$. For the other explanations, see Figure 1.

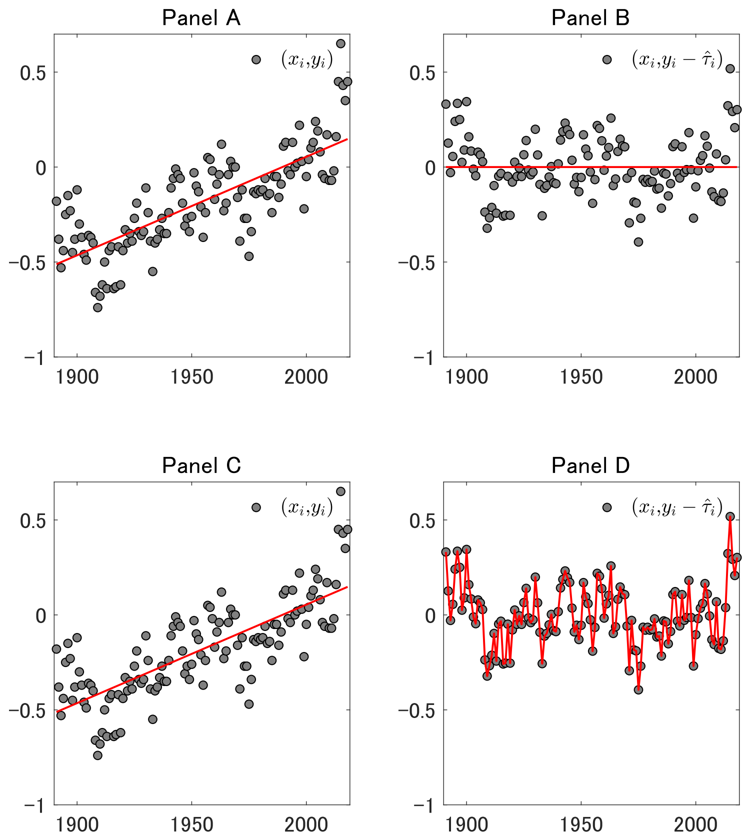

Figure 3. This figure corresponds to the case where $\lambda = 10^{10}$. For the other explanations, see Figure 1.

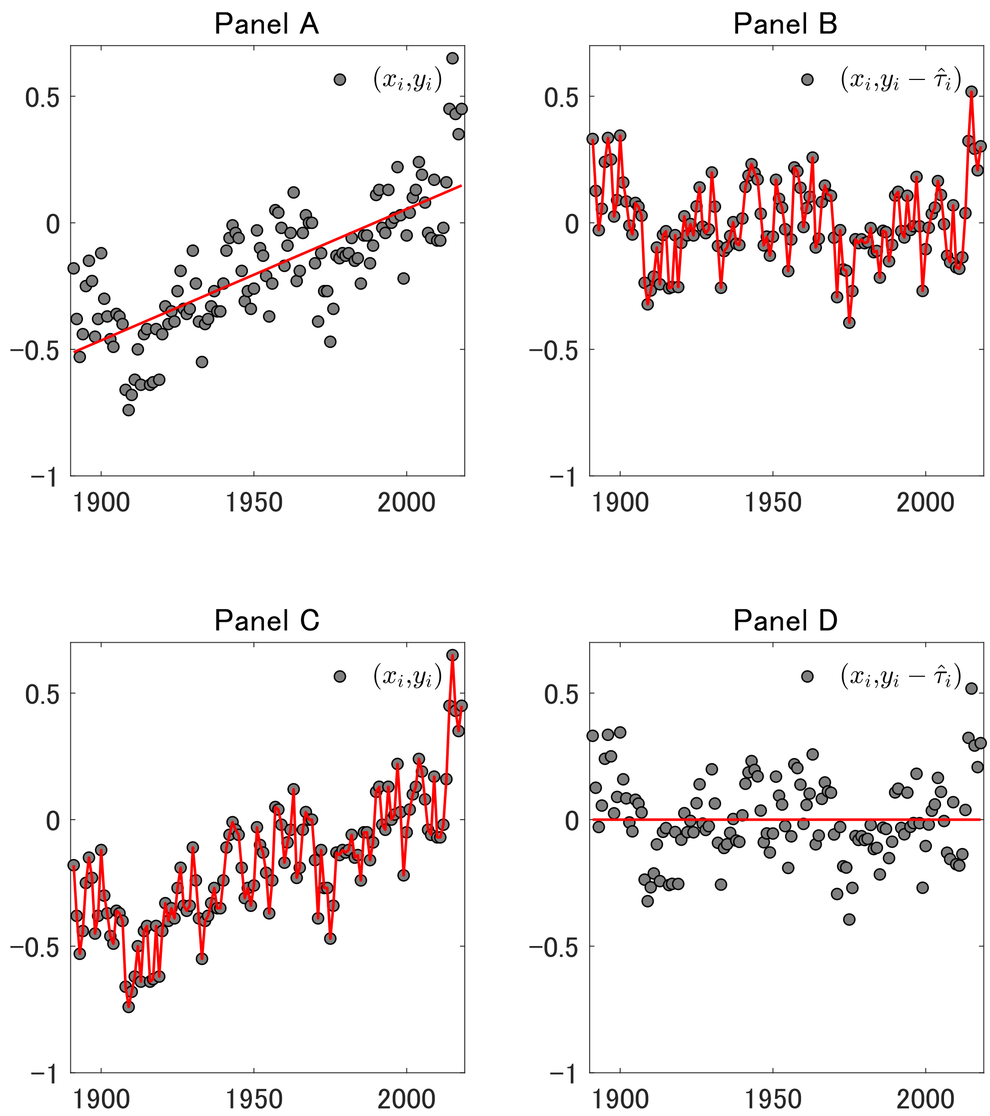

Figure 4. This figure corresponds to the case where $\lambda = 10^{-10}$. For the other explanations, see Figure 1.

6. The Cases Such That the Other Right-Inverse Matrices Are Used

In this section, we illustrate what happens if the other right-inverse matrices are used.

Let $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ be of full row rank. Recall that in this paper $\boldsymbol{M}_{\mathrm{r}}^{-1}$ denotes $\boldsymbol{M}^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-1}$, which is a right-inverse matrix of a full-row-rank matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$. Define a set of matrices

$$\Gamma_M = \{\boldsymbol{\Xi} \in \mathbb{R}^{n\times m} : \boldsymbol{M}\boldsymbol{\Xi} = \boldsymbol{I}_m\}.$$

Γ_M denotes the set of right-inverse matrices of $\boldsymbol{M}$ and accordingly $\boldsymbol{M}_{\mathrm{r}}^{-1}$ belongs to Γ_M.

Lemma 14. *$\boldsymbol{N} = \boldsymbol{M}_{\mathrm{r}}^{-1}$ if and only if $\boldsymbol{N} \in \Gamma_M$ and $\mathcal{S}(\boldsymbol{N}) = \mathcal{S}(\boldsymbol{M}^\top)$.*

Proof of Lemma 14. It is clear that if $\boldsymbol{N} = \boldsymbol{M}_{\mathrm{r}}^{-1}$, then $\boldsymbol{N} \in \Gamma_M$ and $\mathcal{S}(\boldsymbol{N}) = \mathcal{S}(\boldsymbol{M}^\top)$. Conversely, suppose that $\boldsymbol{N} \in \Gamma_M$ and $\mathcal{S}(\boldsymbol{N}) = \mathcal{S}(\boldsymbol{M}^\top)$. Then, $\boldsymbol{M}\boldsymbol{N} = \boldsymbol{I}_m$ and there exists a nonsingular matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ such that $\boldsymbol{N} = \boldsymbol{M}^\top\boldsymbol{\Sigma}$. By removing $\boldsymbol{N}$ from these equations, we have $\boldsymbol{\Sigma} = (\boldsymbol{M}\boldsymbol{M}^\top)^{-1}$, which leads to $\boldsymbol{N} = \boldsymbol{M}^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-1} = \boldsymbol{M}_{\mathrm{r}}^{-1}$. $\square$

From Lemma 14, if $\boldsymbol{N} \neq \boldsymbol{M}_{\mathrm{r}}^{-1}$, then $\boldsymbol{N} \notin \Gamma_M$ or $\mathcal{S}(\boldsymbol{N}) \neq \mathcal{S}(\boldsymbol{M}^\top)$. Accordingly, we have the following result:

Proposition 3. *If* $\boldsymbol{N} \in \Gamma_M \backslash \{\boldsymbol{M}_{\mathrm{r}}^{-1}\}$, *then* $\mathcal{S}(\boldsymbol{N}) \neq \mathcal{S}(\boldsymbol{M}^\top)$.

Based on the result, we illustrate what happens if the other right-inverse matrices are used. We give an example. Let $\boldsymbol{Z} \in \Gamma_D \backslash \{\boldsymbol{D}_{\mathrm{r}}^{-1}\}$. Then, from Proposition 3 and Lemma 3, it follows that $\mathcal{S}(\boldsymbol{Z}) \neq \mathcal{S}(\boldsymbol{D}_{\mathrm{r}}^{-1}) = \mathcal{S}^\perp(\boldsymbol{\Pi})$. Accordingly, letting $\boldsymbol{L} = [\boldsymbol{\Pi}, \boldsymbol{Z}]$, it follows that $\boldsymbol{Z}^\top \boldsymbol{\Pi} \neq \boldsymbol{0}$ and $\boldsymbol{D}\boldsymbol{L} = [\boldsymbol{D}\boldsymbol{\Pi}, \boldsymbol{D}\boldsymbol{Z}] = [\boldsymbol{0}, \boldsymbol{I}_{n-2}]$. In addition, given that $\boldsymbol{D}\boldsymbol{\Pi} = \boldsymbol{0}$, $\boldsymbol{D}\boldsymbol{Z} = \boldsymbol{I}_{n-2}$, and $\boldsymbol{\Pi}$ is of full column rank, $\boldsymbol{L}$ is nonsingular. Thus, from [8], for example, we have

$$\widehat{\boldsymbol{f}} = \boldsymbol{L}(\boldsymbol{L}^\top \boldsymbol{L} + \lambda \boldsymbol{L}^\top \boldsymbol{D}^\top \boldsymbol{D}\boldsymbol{L})^{-1} \boldsymbol{L}^\top \boldsymbol{y} = \boldsymbol{\Pi}\widehat{\boldsymbol{\pi}} + \boldsymbol{Z}\widehat{\boldsymbol{\varepsilon}}, \tag{62}$$

where

$$\widehat{\boldsymbol{\pi}} = \arg \min_{\boldsymbol{\pi} \in \mathbb{R}^2} \|(\boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\varepsilon}}) - \boldsymbol{\Pi}\boldsymbol{\pi}\|^2 = (\boldsymbol{\Pi}^\top \boldsymbol{\Pi})^{-1} \boldsymbol{\Pi}^\top (\boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\varepsilon}}) \tag{63}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\varepsilon}} &= \arg \min_{\boldsymbol{\varepsilon} \in \mathbb{R}^{n-2}} \|\boldsymbol{Q}_\Pi \boldsymbol{y} - \boldsymbol{Q}_\Pi \boldsymbol{Z}\boldsymbol{\varepsilon}\|^2 + \lambda \|\boldsymbol{\varepsilon}\|^2 \\ &= (\boldsymbol{Z}^\top \boldsymbol{Q}_\Pi \boldsymbol{Z} + \lambda \boldsymbol{I}_{n-2})^{-1} \boldsymbol{Z}^\top \boldsymbol{Q}_\Pi \boldsymbol{y}, \end{aligned} \tag{64}$$

which shows that we may obtain (penalized) regressions relating to the cubic smoothing spline even if we use the other right-inverse matrices of $\boldsymbol{D}$ such that $\boldsymbol{Z} \in \Gamma_D \backslash \{\boldsymbol{D}_{\mathrm{r}}^{-1}\}$. Nevertheless, as illustrated here, they are more complex than those shown in Tables 1 and 2.

7. Concluding Remarks

In this paper, we provided a comprehensive list of penalized least squares regressions relating to the cubic smoothing spline, and then revealed a principle of duality in them. This is the main contribution of this study. Such penalized regressions are tabulated in Tables 1 and 2 and the principle of duality revealed is stated in Proposition 2. In addition, we also provided a number of results derived from them, most of which are also tabulated in Tables 1 and 2 and some of which are illustrated in Figures 1–4.

Author Contributions: R.D. wrote an initial draft under the supervision of H.Y. and then H.Y. edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The Japan Society for the Promotion of Science supported this work through KAKENHI Grant Number 16H03606.

Acknowledgments: We thank Kazuhiko Hayakawa and three anonymous referees for their valuable comments on an earlier version of this paper. The usual caveat applies.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Some Remarks on a Special Case Such That $\boldsymbol{x} = [1, \dots, n]^\top$

(i) If $\boldsymbol{x} = [1, \dots, n]^\top$, then $\boldsymbol{C} = \boldsymbol{D}_{(2)}\boldsymbol{D}_{(1)} \in \mathbb{R}^{(n-2)\times n}$, which is a Toeplitz matrix of which the first (resp. last) row is $[1, -2, 1, 0, \dots, 0]$ (resp. $[0, \dots, 0, 1, -2, 1]$).

(ii) If $\boldsymbol{x} = [1, \dots, n]^\top$, then $(\boldsymbol{I}_n + \lambda \boldsymbol{C}^\top \boldsymbol{R}^{-1} \boldsymbol{C})^{-1}$ is bisymmetric (i.e., symmetric centrosymmetric), which may be proved as in Yamada (2020a).

(iii) If $\boldsymbol{x} = [1, \dots, n]^\top$, then $\boldsymbol{R}$ in (8) is not only a symmetric tridiagonal matrix but also a Toeplitz matrix. In the case, we have

$$\omega_k = \frac{2}{3} + \frac{1}{3}\cos\left(\frac{k\pi}{n-1}\right), \quad k = 1, \dots, n-2, \tag{A1}$$

and thus ω_{n-2}, which is the smallest eigenvalue of $\boldsymbol{R}$, satisfies the following inequality (see, e.g., [9]):

$$\omega_{n-2} = \frac{2}{3} + \frac{1}{3}\cos\left(\frac{n-2}{n-1}\pi\right) > \frac{1}{3}. \tag{A2}$$

(iv) If $\boldsymbol{x} = [1, \dots, n]^\top$ and $\boldsymbol{R} = \boldsymbol{I}_{n-2}$ in (2) and (3), then (2) and (3) reduce to

$$\begin{aligned}\widehat{\boldsymbol{f}} &= \arg\min_{\boldsymbol{f}\in\mathbb{R}^n} \|\boldsymbol{y}-\boldsymbol{f}\|^2 + \|\boldsymbol{D}_{(2)}\boldsymbol{D}_{(1)}\boldsymbol{f}\|^2 \\ &= \left\{\boldsymbol{I}_n + \lambda(\boldsymbol{D}_{(2)}\boldsymbol{D}_{(1)})^\top(\boldsymbol{D}_{(2)}\boldsymbol{D}_{(1)})\right\}^{-1}\boldsymbol{y}.\end{aligned} \tag{A3}$$

It is a type of the Whittaker–Henderson (WH) method of graduation, which was developed by Bohlmann [10], Whittaker [11] and others. See Weinert [12] for a historical review of the WH method of graduation. (A3) is also referred to as the Hodrick–Prescott (HP) [13] filtering in econometrics. For more details about the HP filtering, see, for example, Schlicht [14], Kim et al. [15], Paige and Trindade [16], and Yamada [17–21].

Appendix A.2. User-Defined Functions

Appendix A.2.1. A Matlab/GNU Octave Function to Make $\boldsymbol{C}$ in (7)

```
function C=makeCmat(x)
    n=length(x); D1=diff(eye(n)); D2=diff(eye(n-1));
    delta=diff(x); invDelta=diag(1./delta);
    C=D2*invDelta*D1;
end
```

Appendix A.2.2. A Matlab/GNU Octave Function to Make $\boldsymbol{R}$ in (8)

```
function R=makeRmat(x)
    n=length(x); delta=diff(x);
    R0=diag(delta(1:n-2)+delta(2:n-1))/3;
    R1=diag(delta(2:n-2),1)/6;
    R=R1'+R0+R1;
end
```

Appendix A.2.3. A Matlab/GNU Octave Function to Make $\boldsymbol{D}$ in (9)

```
function D=makeDmat(x)
    C=makeCmat(x); R=makeRmat(x); [P,L]=eig(R);
    invsqrtR=P*diag(1./sqrt(diag(L)))*P';
    D=invsqrtR*C;
end
```

Appendix A.2.4. A R Function to Make $\boldsymbol{C}$ in (7)

```
makeCmat <- function(x) {
# Note: x is an n x 1 matrix (not a vector).
    n <- length(x)
```

```
    D1 <- diff(diag(n)); D2 <- diff(diag(n-1))
    delta <- diff(x); invDelta <- diag(1/delta[1:(n-1),1])
    C <- D2%*%invDelta%*%D1
    return(C)
}
```

Appendix A.2.5. A R Function to Make ***R*** in (8)

```
makeRmat <- function(x) {
# Note: x is an n x 1 matrix (not a vector).
    n <- length(x); delta <- diff(x)
    R0 <- diag((delta[1:(n-2),1]+delta[2:(n-1),1])/3)
    R1 <- diag(0,n-2)
    R1[row(R1)==col(R1)-1] <- delta[2:(n-2),1]/6
    R <- t(R1)+R0+R1
    return(R)
}
```

Appendix A.2.6. A R Function to Make ***D*** in (9)

```
makeDmat <- function(x) {
# Note: x is an n x 1 matrix (not a vector).
    n <- length(x); C <- makeCmat(x); R <- makeRmat(x)
    z <- eigen(R); P <- z$vectors
    invsqrtR <- P%*%diag(1/sqrt(z$values))%*%t(P)
    D <- invsqrtR%*%C
    return(D)
}
```

References

1. Schoenberg, I.J. Spline functions and the problem of graduation. *Proc. Natl. Acad. Sci. USA* **1964**, *52*, 947–950. [CrossRef] [PubMed]
2. Reinsch, C. Smoothing by spline functions. *Numer. Math.* **1967**, *10*, 177–183. [CrossRef]
3. Green, P.J.; Silverman, B.W. *Nonparametric Regression and Generalized Linear Models: A roughness Penalty Approach*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1994.
4. Wood, S.N. *Generalized Additive Models: An Introduction with R*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2017.
5. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]
6. Verbyla, A.P.; Cullis, B.R.; Kenward, M.G.; Welham, S.J. The analysis of designed experiments and longitudinal data by using smoothing splines. *J. R. Stat. Soc.* **1999**, *48*, 269–311. [CrossRef]
7. Høyer, J.L.; Karagali, I. Sea surface temperature climate data record for the North Sea and Baltic Sea. *J. Clim.* **2016**, *29*, 2529–2541. [CrossRef]
8. Yamada, H. The Frisch–Waugh–Lovell theorem for the lasso and the ridge regression. *Commun. Stat. Theory Methods* **2017**, *46*, 10897–10902. [CrossRef]
9. Pesaran, M.H. Exact maximum likelihood estimation of a regression equation with a first-order moving-average error. *Rev. Econ. Stud.* **1973**, *40*, 529–535. [CrossRef]
10. Bohlmann, G. Ein ausgleichungsproblem. *Nachrichten von der Gesellschaft der Wissenschaften zu Gottingen, Mathematisch-Physikalische Klasse* **1899**, 1899, 260–271.
11. Whittaker, E.T. On a new method of graduation. *Proc. Edinb. Math. Soc.* **1923**, *41*, 63–75. [CrossRef]

12. Weinert, H.L. Efficient computation for Whittaker–Henderson smoothing. *Comput. Stat. Data Anal.* **2007**, *52*, 959–974. [CrossRef]
13. Hodrick, R.J.; Prescott, E.C. Postwar U.S. business cycles: An empirical investigation. *J. Money Credit. Bank.* **1997**, *29*, 1–16. [CrossRef]
14. Schlicht, E. Estimating the smoothing parameter in the so-called Hodrick–Prescott filter. *J. Jpn. Stat. Soc.* **2005**, *35*, 99–119. [CrossRef]
15. Kim, S.; Koh, K.; Boyd, S.; Gorinevsky, D. ℓ_1 trend filtering. *SIAM Rev.* **2009**, *51*, 339–360. [CrossRef]
16. Paige, R.L.; Trindade, A.A. The Hodrick–Prescott filter: A special case of penalized spline smoothing. *Electron. J. Stat.* **2010**, *4*, 856–874. [CrossRef]
17. Yamada, H. Ridge regression representations of the generalized Hodrick–Prescott filter. *J. Jpn. Stat. Soc.* **2015**, *45*, 121–128. [CrossRef]
18. Yamada, H. Why does the trend extracted by the Hodrick–Prescott filtering seem to be more plausible than the linear trend? *Appl. Econ. Lett.* **2018**, *25*, 102–105. [CrossRef]
19. Yamada, H. Several least squares problems related to the Hodrick–Prescott filtering. *Commun. Stat. Theory Methods* **2018**, *47*, 1022–1027. [CrossRef]
20. Yamada, H. A note on Whittaker–Henderson graduation: Bisymmetry of the smoother matrix. *Commun. Stat. Theory Methods* **2020**, *49*, 1629–1634. [CrossRef]
21. Yamada, H. A smoothing method that looks like the Hodrick–Prescott filter. *Econom. Theory* **2020**. [CrossRef]