Annals of the University of Craiova, Mathematics and Computer Science Series
Volume 43(1), 2016, Pages 33–52
ISSN: 1223-6934

Dedicated to Marius Iosifescu
on the occasion of his 80th anniversary

Estimation of noisy cubic spline using a natural basis

Azzouz Dermoune and Cristian Preda

Abstract. We define a new basis of cubic splines such that the coordinates of a natural cubic spline are sparse. We use it to analyse and to extend the classical Schoenberg and Reinsch result and to estimate a noisy cubic spline. We also discuss the choice of the smoothing parameter. All our results are illustrated graphically.

Key words and phrases. Cubic spline, BLUP, Smoothing parameter, Linear regression, Bayesian model, Stein's Unbiased Risk Estimate.

1. Introduction

We consider, for $n \geq 1$, the regression model

$$y_i = f(t_i) + w_i, \quad i = 1, \dots, n+1, \tag{1}$$

where $y_1, \dots, y_{n+1}, t_1 < \dots < t_{n+1}$ are real-valued observations, $w_1, \dots, w_{n+1}$ are measurement errors and $f : [t_1, t_{n+1}] \to \mathbb{R}$ is an unknown element of the infinite dimensional space H^2 of all functions with square integrable second derivative. The approximation of f by cubic splines considers the regression model

$$y_i = s(t_i) + w_i, \quad i = 1, \dots, n+1, \tag{2}$$

where s is an unknown element of the finite dimensional space of cubic splines. Schoenberg [19] introduced in 1946 the terminology spline for a certain type of piecewise polynomial interpolant. The ideas have their roots in the aircraft and shipbuilding industries. Since that time, splines have been shown to be applicable and effective for a large number of tasks in interpolation and approximation. Various aspect of splines and their applications can be found in [1], [2], [13], [17], [14] and [18]. See also the references therein.

Let us first define properly the cubic splines approximation and introduce our notations. A map s belongs to the set S_3 of cubic splines with the knots $t_1 < ... < t_{n+1}$ if there exist $(p_1, \dots, p_{n+1})$ in $\mathbb{R}^{n+1}$, $(q_1, \dots, q_n)$, $(u_1, \dots, u_n)$, $(v_1, \dots, v_n)$ in $\mathbb{R}^n$ such that, for $i = 1, \dots, n$ and $t \in [t_i, t_{i+1})$,

$$s(t) = p_i + q_i(t - t_i) + \frac{u_i}{2}(t - t_i)^2 + \frac{v_i}{6}(t - t_i)^3. \tag{3}$$

We are intereseted in the set $S_3 \cap C^2$ of C^2-cubic splines. A cubic spline s, having its second derivatives $s''(t_1+) = s''(t_{n+1}-) = 0$, is called natural. A well known result tells us that if $f \in H^2$ and $s \in S_3 \cap C^2$ are such that $f(t_i) = s(t_i)$ for all $i = 1, \dots, n+1$,

Received May 5, 2016.

then $\int_{t_1}^{t_{n+1}} |s(t) - f(t)|^2 dt = O(h^4)$ with $h = \max((t_{i+1} - t_i)^4 : i = 1, \ldots, n+1)$. See e.g. [1], [22]. Hence, by paying the cost $O(h^4)$ we can replace the model (1) by (2).

It is well known that any natural cubic spline of $S_3 \cap C^2$ can be expressed using the *all* the $n+3$ elements of the cubic B-spline basis, see e.g. [17]. In Section 2 we construct a new basis of $S_3 \cap C^2$ in which any natural cubic spline needs only $n+1$ elements. In Sections 3-6 we treat the problem of estimation a noisy cubic spline.

2. The natural basis for C^2-Cubic splines

Usually, the B-splines are used as a basis. The aim of this section is to construct a new basis which is more suitable for the natural cubic splines. Before going further, we need some notations. Let for $i = 1, \ldots, n$, $h_i = t_{i+1} - t_i$. The spline s, defined in (3), is of class C^2 if and only if

$$p_i + q_i h_i + \frac{u_i}{2} h_i^2 + \frac{v_i}{6} h_i^3 = p_{i+1}, \quad i = 1, \ldots, n, \tag{4}$$

$$q_i + u_i h_i + \frac{v_i}{2} h_i^2 = q_{i+1}, \quad i = 1, \ldots, n-1, \tag{5}$$

$$u_i + h_i v_i = u_{i+1}, \quad i = 1, \ldots, n. \tag{6}$$

We introduce the column vectors $\boldsymbol{q} = (q_1, \ldots, q_n)^T$, $\boldsymbol{p} = (p_1, \ldots, p_{n+1})^T$, and $\boldsymbol{u} = (u_1, \ldots, u_{n+1})^T$, where $\mathbf{M}^T$ is the transpose of the matrix $\mathbf{M}$. Using (4), (5), (6), we can show that there exist three matrices $\mathbf{Q}, \mathbf{U}, \mathbf{V}$ such that

$$\boldsymbol{q} = \mathbf{Q} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix}, \tag{7}$$

$$\boldsymbol{u} = \mathbf{U} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix}, \tag{8}$$

$$\boldsymbol{v} = \mathbf{V} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix}. \tag{9}$$

Let us define, for each $i = 1, \ldots, n$, the piecewise functions,

$$\chi_i(t) = 1_{[t_i, t_{i+1})}(t), \quad \chi_i^1(t) = (t - t_i) 1_{[t_i, t_{i+1})}(t), \quad \chi_i^2(t) = (t - t_i)^2 1_{[t_i, t_{i+1})}(t),$$
$$\chi_i^3(t) = (t - t_i)^3 1_{[t_i, t_{i+1})}(t), \quad \chi_0 = 0, \quad \chi_{n+1} = 1_{t_{n+1}}, \quad \chi_{n+2} = 0.$$

Here 1_A denotes the indicator function of the set A. Clearly, the set $[\chi_i, \chi_i^k : i = 1, \ldots, n+1,\ k = 1, 2, 3]$ forms a basis of the set of cubic splines S_3. The map s has the coordinates $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{u}, \boldsymbol{v}$ in this basis, i.e.

$$s = \left([\chi_1 \cdots \chi_n\, \chi_{n+1}] \quad [\chi_1^1 \cdots \chi_n^1] \quad \tfrac{1}{2}[\chi_1^2 \cdots \chi_n^2\, 0] \quad \tfrac{1}{6}[\chi_1^3 \cdots \chi_n^3] \right) \begin{pmatrix} \boldsymbol{p} \\ \boldsymbol{q} \\ \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}.$$

If s is C^2, then from (7), (8), (9) we have

$$s = ([\chi_0\, \chi_1 \cdots \chi_n\, \chi_{n+1}\, \chi_{n+2}] + [\chi_1^1 \cdots \chi_n^1]\mathbf{Q} + \frac{1}{2}[\chi_1^2 \cdots \chi_n^2\, 0]\mathbf{U} + \frac{1}{6}[\chi_1^3 \cdots \chi_n^3]\mathbf{V})\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix}$$

The C^2 cubic spline s can be rewritten in the following new basis:

$$s = [\varphi_0 \ldots \varphi_{n+2}] \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix},$$

where, for $j = 0, \ldots, n+2$,

$$\varphi_j = \chi_j + [\chi_1^1 \cdots \chi_n^1]\boldsymbol{q}_{\cdot j+1} + \frac{1}{2}[\chi_1^2 \cdots \chi_n^2\, 0]\boldsymbol{u}_{\cdot j+1} + \frac{1}{6}[\chi_1^3 \cdots \chi_n^3]\boldsymbol{v}_{\cdot j+1}. \tag{10}$$

Here $\boldsymbol{a}_{\cdot j}$ denotes the jth column of the matrix $\mathbf{A}$. Each element of the new basis is a C^2 cubic spline.

From (10), we derive that the set of natural cubic splines is spanned by the basis $(\varphi_j : j = 1, \ldots, n+1)$.

• The spline φ_0 is the unique C^2 cubic spline interpolating the points $(t_1, 0)$, …, $(t_{n+1}, 0)$ and such that $\varphi_0''(t_1+) = 1$, $\varphi_0''(t_{n+1}-) = 0$. Hence, φ_0 is not a natural cubic spline.

• The spline φ_j, for $j = 1, \ldots, n+1$, is the unique natural cubic spline interpolating the points $(t_j, 1)$, $((t_i, 0), i \neq j)$.

• The spline φ_{n+2} is the unique C^2 cubic spline interpolating the points $(t_1, 0)$, …, $(t_{n+1}, 0)$ and such that $\varphi_{n+2}''(t_1+) = 0$, $\varphi_{n+2}''(t_{n+1}-) = 1$. Hence, φ_{n+2} is not a natural cubic spline.

Observe that the natural cubic spline interpolating the points $(t_i, 0)$, $i = 1, \ldots, n+1$ is the null map

$$s_0 = [\varphi_0, \ldots, \varphi_{n+2}] \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

As an illustration, in Figure 1 we plot, for $n = 7$, $t_i = \frac{i-1}{n}, i = 1, \ldots, n+1$, the basis $\{\varphi_0, \ldots, \varphi_{n+2}\}$ and their derivatives in Figure 2 and Figure 3. We can show that our basis has the reverse time property (see Figure 1), i.e.

$$\varphi_j(t_{n+1} - t) = \varphi_{n+3-j}(t), \quad \forall\, j = 0, \ldots, n+2, \quad t \in [t_1, t_{n+1}].$$

Observe that our new basis is very different of the classical cubic B-spline basis.

3. The new basis and Schoenberg-Reinsch optimization

In this section we use the new basis to review the well known results concerning the L^2 penalty and the optimal property of cubic splines.

Let p_i, $i = 1, \ldots, n+1$, be a set of points in $\mathbb{R}$. The famous result of Schoenberg 1964 [19] and Reinsch 1967 [18] tells us that the minimizer

$$I_2(\boldsymbol{p}) \overset{def}{=} \arg\min_{f \in H^2} \{\int_{t_1}^{t_{n+1}} |f''(t)|^2 dt : \quad f(t_i) = p_i \quad, i = 1, ..., n+1\} = \sum_{j=1}^{n+1} p_j \varphi_j$$

is the natural C^2 cubic spline which interpolates the points $(t_i, p_i), i = 1, \ldots, n+1$.

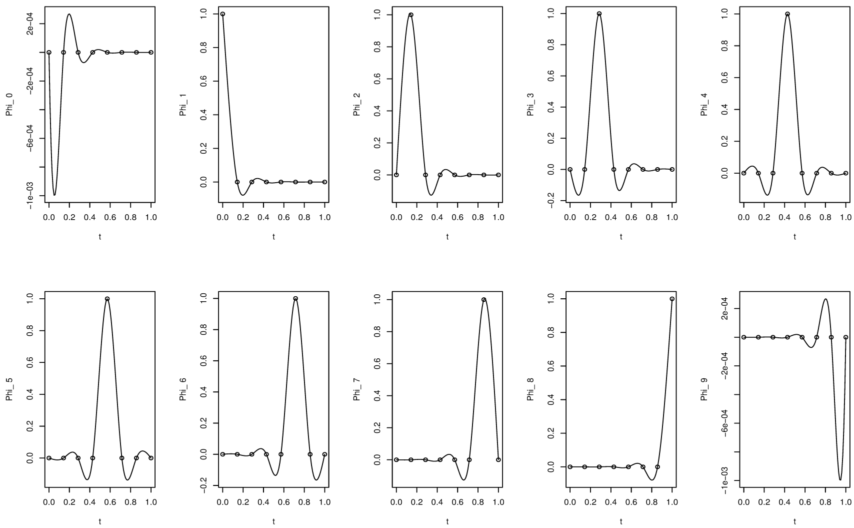

FIGURE 1. The graph of the 10 elements of the natural basis. Here $n = 7$ and $t_i = \frac{i-1}{n}$ for $i = 1, \ldots, 8$.

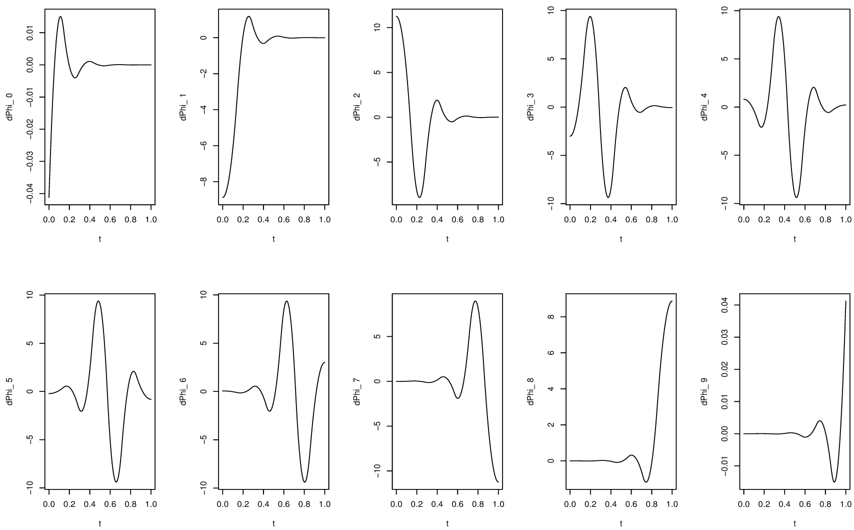

FIGURE 2. The graph of the first derivative of the 10 elements of the natural basis. Here $n = 7$ and $t_i = \frac{i-1}{n}$ for $i = 1, \ldots, 8$.

It follows, for $j = 1, \ldots, n+1$, that

$$I_2(\delta_j) \quad = \quad \arg\min_{f \in H^2} \{ \int_{t_1}^{t_{n+1}} |f''(t)|^2 dt : \quad f(t_i) = \delta_j^i \quad , i = 1, ..., n+1\} = \varphi_j,$$

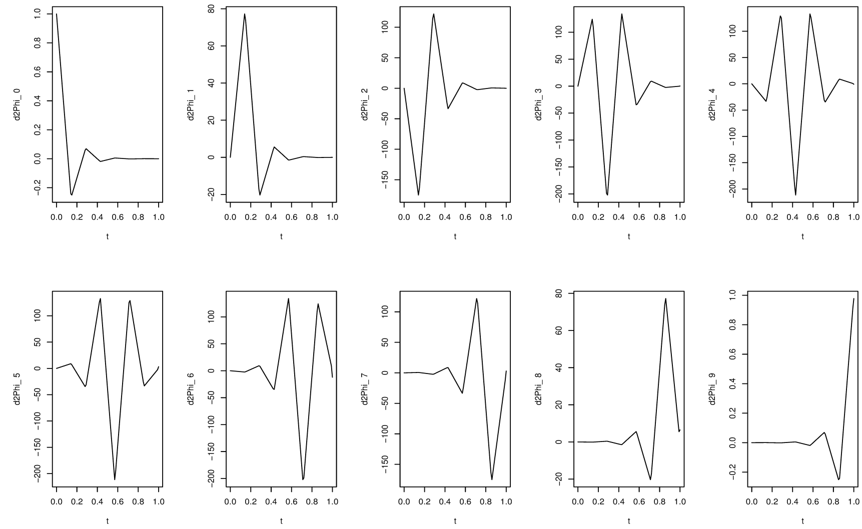

FIGURE 3. The graph of the second derivative of the 10 elements of the natural basis. Here $n = 7$ and $t_i = \frac{i-1}{n}$ for $i = 1, \ldots, 8$.

where $\delta_j \in \mathbb{R}^{n+1}$ and has the component $\delta_j^i = 1$ if $i = j$ and 0 otherwise.

The aim of this section is to interpret Schoenberg and Reinsch result using the natural basis. As a by-product, we will show that φ_0 and φ_{n+2} are respectively solution of the following optimization problems:

$$\min_{f \in S_3 \cap C^2} \{ \int_{t_1}^{t_{n+1}} |f''(t)|^2 dt : \quad f''(t_1) = 1, \; f(t_i) = 0 \quad , i = 1, ..., n+1 \}, \tag{11}$$

$$\min_{f \in S_3 \cap C^2} \{ \int_{t_1}^{t_{n+1}} |f''(t)|^2 dt : \quad f''(t_{n+1}) = 1, \; f(t_i) = 0 \quad , i = 1, ..., n+1 \}. \tag{12}$$

3.1. Revisiting Schoenberg and Reinsch result.

Proposition 3.1. Let us introduce, for $\boldsymbol{u} \in \mathbb{R}^{n+1}$, the quadratic form

$$J_2(\boldsymbol{u}) = \sum_{i=1}^{n} \frac{h_i}{3} (u_i^2 + u_i u_{i+1} + u_{i+1}^2).$$

The minimization

$$\min_{s \in C^2 \cap S_3} \{ \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt : \quad s(t_i) = p_i \quad , i = 1, ..., n+1 \}$$

is equivalent to

$$\min_{u_1, u_{n+1}} \{ J_2(\mathbf{U}(u_1, \boldsymbol{p}, u_{n+1})^T) \}.$$

Proof. Schoenberg and Reinsch result tells us that $I_2(\boldsymbol{p}) = \sum_{j=1}^{n+1} p_j\varphi_j$ is the minimizer of

$$\min_{s\in C^2\cap S_3}\{\int_{t_1}^{t_{n+1}} |s''(t)|^2 dt : \quad s(t_i) = p_i \quad , i = 1, ..., n+1\}. \tag{13}$$

If $s \in C^2 \cap S_3$, using (3) and (8), then

$$\begin{aligned} \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt &= \sum_{i=1}^{n} h_i \int_0^1 |tu_i + (1-t)u_{i+1}|^2 dt \\ &= J_2(\boldsymbol{u}). \end{aligned}$$

Now the equality (8) achieves the proof. □

3.2. Some consequences of Schoenberg-Reinsch result. First, let us rewrite

$$\begin{aligned} J_2(\boldsymbol{u}) &= \sum_{i=1}^{n} \frac{h_i}{3}(u_i^2 + u_iu_{i+1} + u_{i+1}^2) && (14) \\ &= (u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{C}\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1}\end{pmatrix}, && (15) \end{aligned}$$

where

$$\mathbf{C} = \frac{h_1}{3}\mathbf{U}_1^T\mathbf{U}_1 + \frac{2}{3}\sum_{i=2}^{n} h_i\mathbf{U}_i^T\mathbf{U}_i + \frac{h_n}{3}\mathbf{U}_{n+1}^T\mathbf{U}_{n+1} + \\ \frac{1}{6}\sum_{i=1}^{n} h_i[\mathbf{U}_i^T\mathbf{U}_{i+1} + \mathbf{U}_{i+1}^T\mathbf{U}_i]. \tag{16}$$

Now, we summarize the properties of the matrix $\mathbf{C}$.

Proposition 3.2. The matrix $\mathbf{C}$ is symmetric, and non-negative definite. The quadratic form $(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{C}\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1}\end{pmatrix} = 0$ if and only if $u_1 = u_{n+1} = 0$ and $\boldsymbol{p}$ belongs to the range $R(\mathbf{L})$ of the matrix

$$\mathbf{L} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_{n+1}\end{pmatrix}. \tag{17}$$

It follows that, for all $j = 1, \ldots, n+1$, that $c_{1,j+1} = c_{n+3,j+1} = 0$, i.e. the matrix

$$\mathbf{C} = \begin{pmatrix} c_{1,1} & 0 & c_{1,n+3} \\ 0 & \mathbf{C}(2, n+2) & 0 \\ c_{n+3,1} & 0 & c_{n+3,n+3}\end{pmatrix},$$

where $\mathbf{C}(2, n+2) = [c_{ij} : i, j = 2, \ldots n+2]$. The sub-matrix

$$\begin{pmatrix} c_{1,1} & c_{1,n+3} \\ c_{n+3,1} & c_{n+3,n+3}\end{pmatrix} \tag{18}$$

is symmetric, positive definite. The null-space of the sub-matrix $\mathbf{C}(2, n+2)$ is equal to $R(\mathbf{L})$. Moreover, from the decomposition $s = u_1\varphi_0 + \sum_{j=1}^{n+1} p_j\varphi_j + u_{n+1}\varphi_{n+2}$, we have

$$\int_{t_1}^{t_{n+1}} |s''(t)|^2 dt = u_1^2 c_{1,1} + u_{n+1}^2 c_{n+3,n+3} + 2c_{1,n+3}u_1u_{n+1} + \boldsymbol{p}^T\mathbf{C}(2,n)\boldsymbol{p}. \tag{19}$$

From (19), we derive that the second derivatives $\{\varphi_j'' : j = 0, \ldots, n+2\}$ of the new basis satisfy

$$\int_{t_1}^{t_{n+1}} \varphi_i''(t)\varphi_j''(t)dt = c_{i+1,j+1}, \quad i, j = 0, \ldots, n+2.$$

We can show numerically that φ_0'' (respectively φ_{n+2}'') is orthogonal to φ_j'' for all $j = 1, \ldots, n+2$ (respectively to φ_j'' for all $j = 0, \ldots, n+1$). As an example the matrix $\mathbf{C}$ has the following form, for $n = 7$, $t_i = \frac{i-1}{n}, i = 1, \ldots, n+1$,

$$\mathbf{C} = \begin{pmatrix}
0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 551.44 & -1250.64 & 886.54 & -237.54 & 63.63 & -16.97 & 4.24 & -0.71 & 0.00 \\
0.00 & -1250.64 & 3387.82 & -3261.27 & 1425.26 & -381.77 & 101.80 & -25.45 & 4.24 & 0.00 \\
0.00 & 886.54 & -3261.27 & 4813.08 & -3643.03 & 1527.06 & -407.22 & 101.80 & -16.97 & 0.00 \\
0.00 & -237.54 & 1425.26 & -3643.03 & 4914.88 & -3668.49 & 1527.06 & -381.77 & 63.63 & 0.00 \\
0.00 & 63.63 & -381.77 & 1527.06 & -3668.49 & 4914.88 & -3643.03 & 1425.26 & -237.54 & 0.00 \\
0.00 & -16.97 & 101.80 & -407.22 & 1527.06 & -3643.03 & 4813.08 & -3261.27 & 886.54 & 0.00 \\
0.00 & 4.24 & -25.45 & 101.80 & -381.77 & 1425.26 & -3261.27 & 3387.82 & -1250.64 & 0.00 \\
0.00 & -0.71 & 4.24 & -16.97 & 63.63 & -237.54 & 886.54 & -1250.64 & 551.44 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.04
\end{pmatrix}.$$

Observe that the fact that sub-matrix (18) is diagonal is not expected.

Now we derive easily the following results.

Proposition 3.3. The splines φ_0 and φ_{n+2} are respectively solution of the optimizations

$$\min_{s \in S_3 \cap C^2} \{\int_{t_1}^{t_{n+1}} |s''(t)|^2 dt : \quad s''(t_1) = 1,\ s(t_i) = 0 \quad , i = 1, ..., n+1\} \tag{20}$$

$$\min_{s \in S_3 \cap C^2} \{\int_{t_1}^{t_{n+1}} |s''(t)|^2 dt : \quad s''(t_{n+1}) = 1,\ s(t_i) = 0 \quad , i = 1, ..., n+1\}. \tag{21}$$

Proof. The optimizations (20),(21) are equivalent to

$$\min_{u_1, \boldsymbol{p}, u_{n+1}} \{(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{C}\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1}\end{pmatrix} : \quad u_1 = 1, \quad \boldsymbol{p} = 0\},$$

$$\min_{u_1, \boldsymbol{p}, u_{n+1}} \{(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{C}\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1}\end{pmatrix} : \quad u_{n+1} = 1, \quad \boldsymbol{p} = 0\},$$

and then have respectively the solutions, $(u_1 = 1, \boldsymbol{p} = 0, u_{n+1} = 0)$ and $(u_1 = 0, \boldsymbol{p} = 0, u_{n+1} = 1)$. $\square$

More generally we have the following result.

Proposition 3.4. Let k be a positive integer, $\mathbf{M}$ a $n+3$ by $n+3$ matrix, $\mathbf{A}$ be a k by $n+3$ matrix and $\begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} := \boldsymbol{c} \in \mathbb{R}^k$ all are given. Suppose that the null spaces

$N(\mathbf{A})$, $N(\mathbf{M})$ do not overlap, i.e. $N(\mathbf{A}) \cap N(\mathbf{M}) = \{0\}$. Then the optimization

$$\min_{u_1, \boldsymbol{p}, u_{n+1}} \{(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{M} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} : \quad \mathbf{A} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} = \boldsymbol{c}\} \tag{22}$$

has a unique solution. More precisely, there exist a unique couple $(\boldsymbol{l}, \boldsymbol{v})$ such that

$$\mathbf{M}\boldsymbol{v} = \mathbf{A}^T \boldsymbol{l}, \quad \mathbf{A}\boldsymbol{v} = \boldsymbol{c}.$$

The vector $\boldsymbol{v}$ is the minimizer, and $\boldsymbol{l}$ is the Lagrange multiplier.

4. Natural cubic spline estimate

Let $s : \mathbb{R} \to \mathbb{R}$ be a natural cubic spline known with imprecision on the knots $t_1, \ldots, t_{n+1}$, i.e.

$$y_i = s(t_i) + w_i, \quad i = 1, \ldots, n+1, \tag{23}$$

where w_i is the noise added to the true value $s(t_i) = p_i$. In the sequel $\boldsymbol{y} = (y_1, \ldots, y_{n+1})^T$. Schoenberg and Reinsch optimization, for each $\lambda > 0$,

$$\arg \min_{f \in S_3 \cap C^2} \{\lambda \int_{t_1}^{t_{n+1}} |f''(t)|^2 dt + \sum_{i=1}^{n+1} |f(t_i) - y_i|^2\},$$

provides an estimator of s. The parameter $\lambda > 0$ is called the smoothing parameter. Using the same arguments and notations as in Proposition 3.1 and (14) the latter optimization problem is equivalent to

$$\min\{\lambda(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{C} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} + \|\boldsymbol{p} - \boldsymbol{y}\|^2 : \quad \boldsymbol{p} \in \mathbb{R}^{n+1}, u_1, u_{n+1} \in \mathbb{R}\}, \tag{24}$$

where $\|\cdot\|$ denotes the Euclidean norm.

We have easily the following result.

Proposition 4.1. The equality

$$(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{C} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} = \quad u_1^2 c_{11} + u_{n+1}^2 c_{n+3,n+3} + 2u_1 u_{n+1} c_{1,n+3} + \boldsymbol{p}^T \mathbf{C}(2, n+2)\boldsymbol{p}$$

implies that the minimizer of (24) is $u_1 = u_{n+1} = 0$ and $\boldsymbol{p}$ is solution of the following system:

$$(\mathbf{I} + \lambda \mathbf{C}(2, n+2))\boldsymbol{p} = \boldsymbol{y}.$$

The solution $\boldsymbol{p}$ is given by

$$\hat{\boldsymbol{p}} = \mathbf{H}(\lambda)\boldsymbol{y}, \tag{25}$$

where $\mathbf{H}(\lambda) := (\mathbf{I} + \lambda \mathbf{C}(2, n+2))^{-1}$ is called the hat matrix.

Now, we discuss the limits of (24) as $\lambda \to 0$ and $\lambda \to +\infty$.

Corollary 4.2. 1) The problem (24), when $\lambda \to 0$, becomes

$$\min\{(u_1, \boldsymbol{y}^T, u_{n+1})\mathbf{C}\begin{pmatrix} u_1 \\ \boldsymbol{y} \\ u_{n+1} \end{pmatrix} : \quad \boldsymbol{p} = \boldsymbol{y}, u_1, u_{n+1} \in \mathbb{R}\}. \tag{26}$$

Its minimizer is $u_1 = u_{n+1} = 0, \boldsymbol{p} = \boldsymbol{y}$ i.e.

$$\lim_{\lambda \to 0+} \mathbf{H}(\lambda) = \mathbf{I}_{n+1}.$$

where $\mathbf{I}_{n+1}$ is the $(n+1) \times (n+1)$ identity matrix.

2) The problem (24), when $\lambda \to +\infty$, becomes

$$\min\{\|\boldsymbol{y} - \boldsymbol{p}\|^2 : \quad \mathbf{C}\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} = 0\}. \tag{27}$$

Its minimizer is $u_1 = u_{n+1} = 0$ and

$$\begin{aligned} \boldsymbol{p} &= \lim_{\lambda \to +\infty} \mathbf{H}(\lambda)\boldsymbol{y} \\ &= \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\boldsymbol{y} \\ &= \pi_{R(\mathbf{L})}\boldsymbol{y}, \end{aligned}$$

where $\mathbf{L}$ is the linear model matrix (17), and $\pi_{R(\mathbf{L})}$ is the orthogonal projection on the range of $\mathbf{L}$.

Remark 4.1. We have easily $\mathbf{C}(2, n+2)\mathbf{L} = 0$. From that we derive that $\mathbf{H}(\lambda)\mathbf{L} = \mathbf{L}$ and then $\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T[\mathbf{H}(\lambda) - \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T] = 0$. It follows that $\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\boldsymbol{y}$ and $[\mathbf{H}(\lambda) - \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T]\boldsymbol{y}$ are orthogonal. The component $Lreg\boldsymbol{y} := \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\boldsymbol{y}$ is the linear regression i.e. $Lreg\boldsymbol{y}$ is the orthogonal projection of the data $\boldsymbol{y}$ on the linear space $R(\mathbf{L})$ (the range of $\mathbf{L}$). By introducing the orthogonal projections $\pi_{R(\mathbf{L})}$ and $\pi_{R(\mathbf{L})^\perp}$ respectively on $R(\mathbf{L})$ and $R(\mathbf{L})^\perp$, the minimizator

$$\mathbf{H}(\lambda)\boldsymbol{y} = \arg\min\{\lambda \boldsymbol{p}^T\mathbf{C}(2, n+2)\boldsymbol{p} + \|\boldsymbol{p} - \boldsymbol{y}\|_2^2 : \quad \boldsymbol{p} \in \mathbb{R}^{n+1}\}$$

is the sum of

$$\arg\min\{\lambda \boldsymbol{p}_1^T\mathbf{C}(2, n+2)\boldsymbol{p}_1 + \|\boldsymbol{p}_1 - \pi_{R(\mathbf{L})^\perp}\boldsymbol{y}\|_2^2 : \quad \boldsymbol{p}_1 \in R(\mathbf{L})^\perp\},$$

and

$$\arg\min\{\|\boldsymbol{p}_1 - \pi_{R(\mathbf{L})}\boldsymbol{y}\|^2 : \quad \boldsymbol{p}_1 \in R(\mathbf{L})\} = \pi_{R(\mathbf{L})}\boldsymbol{y}.$$

Hence, the component

$$\begin{aligned} [\mathbf{H}(\lambda) - \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T]\boldsymbol{y} = \arg\min\{ \quad & \lambda \boldsymbol{p}_1^T\mathbf{C}(2, n+2)\boldsymbol{p}_1 + \|\boldsymbol{p}_1 - \pi_{R(L)^\perp}\boldsymbol{y}\|_2^2 : \\ & \boldsymbol{p}_1 \in R(L)^\perp\} \end{aligned}$$

is the penalized projection of $\boldsymbol{y}$ on $R(L)^\perp$, i.e. $[\mathbf{H}(\lambda) - \mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T]\boldsymbol{y}$ is the nearest vector $\boldsymbol{p} \in R(L)^\perp$ to $\pi_{R(\mathbf{L})^\perp}\boldsymbol{y}$ under the constraint $\boldsymbol{p}^T\mathbf{C}(2, n+2)\boldsymbol{p} \le \delta$. Thanks to Lagrange multiplier, the positive constant δ and the smoothing parameter λ are related by the equation $\boldsymbol{y}^T\mathbf{H}(\lambda)\mathbf{C}(2, n+2)\mathbf{H}(\lambda)\boldsymbol{y} = \delta$. The penalty $\boldsymbol{p}^T\mathbf{C}(2, n+2)\boldsymbol{p} = \|\mathbf{C}^{1/2}(2, n+2)\boldsymbol{p}\|^2$ measures the deviation of the vector $\boldsymbol{p}$ with respect to the linear space $R(\mathbf{L})$. The vector $\mathbf{C}^{1/2}(2, n+2)\boldsymbol{p}$ can be seen as an oblique projection on $R(\mathbf{L})^\perp$.

5. Choice of the smoothing parameter λ

5.1. Deterministic noise. We have, for any $\lambda > 0$, the estimated model of (23)

$$\boldsymbol{y} = \mathbf{H}(\lambda)\boldsymbol{y} + [\mathbf{I} - \mathbf{H}(\lambda)]\boldsymbol{y}.$$

We proposed $\mathbf{H}(\lambda)\boldsymbol{y}$ as an estimator of $\boldsymbol{p}$ and therefore, $[\mathbf{I} - \mathbf{H}(\lambda)]\boldsymbol{y}$ is an estimator of the noise $\boldsymbol{w} = (w_1, \ldots, w_{n+1})^T$.

The equality $[\mathbf{I} - \mathbf{H}(\lambda)]\boldsymbol{y} = \boldsymbol{w}$ holds only for $\lambda = 0$, $\boldsymbol{w} = 0$ and $\boldsymbol{p}$ is a straightline, i.e. $p_i = a + bt_i$ for all i. A natural way to link the smoothing parameter and the size of the noise is to solve the equation

$$\|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2 = \|\boldsymbol{w}\|^2. \tag{28}$$

The following result shows that the equation (28) has a solution only for "small noise".

Proposition 5.1. The map $\lambda \in (0, +\infty) \to \psi(\lambda) = \|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2$ is concave, varies between 0 and $\|\boldsymbol{y} - Lreg(\boldsymbol{y})\|^2$. The equation (28) has a solution if and only if

$$\|\boldsymbol{y} - Lreg(\boldsymbol{y})\|^2 > \|\boldsymbol{w}\|^2. \tag{29}$$

The proof is a consequence of the fact that $\mathbf{H}(\lambda) \to \mathbf{I}$ as $\lambda \to 0$ and $\mathbf{H}(\lambda) \to Lreg$ as $\lambda \to +\infty$.

Observe that (29) is equivalent to

$$\sum_{i=1}^{n+1} y_i^2 \|\boldsymbol{e}_i - Lreg(\boldsymbol{e}_i)\|^2 + 2\sum_{i<j} y_i y_j \langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle > \|\boldsymbol{w}\|^2, \tag{30}$$

and the smoothing parameter is solution of the equation

$$\sum_{i=1}^{n+1} y_i^2 \|\boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda)\|^2 + 2\sum_{i<j} y_i y_j \langle \boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda), \boldsymbol{e}_j - \mathbf{H}_{\cdot j}(\lambda)\rangle = \|\boldsymbol{w}\|^2. \tag{31}$$

It follows that the size of the weights $\|\boldsymbol{e}_i - Lreg(\boldsymbol{e}_i)\|^2$, $\langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle$, $\|\boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda)\|^2$, $\langle \boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda), \boldsymbol{e}_j - \mathbf{H}_{\cdot j}(\lambda)\rangle$, are crucial in the existence of the smoothing parameter (28). In Figure 4 we plot for $i = 1, \ldots, 8$ the graph of $\lambda \to \|\boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda)\|^2$ for $n = 7$ and $t_i = \frac{i-1}{n}$.

Remark that for the model $\boldsymbol{y} = y_i \boldsymbol{e}_i$ the ith column $\mathbf{H}_{\cdot i}(\lambda)$ is an estimator of the signal. Thanks to the equation (31), the quantity $\|\boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda)\|^2$ represents the noise-to-signal ratio (NSR), i.e. the smoothing parameter is solution of

$$\|\boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda)\|^2 = \frac{\|\boldsymbol{w}\|^2}{y_i^2}. \tag{32}$$

For large noise, there is no smoothing parameter solution of (32).

In Figure 5 we plot the "rainbow" $\mathbf{H}_{\cdot i}(\lambda)$ for $i = 1, \ldots, n+1$ and $\lambda \in \{0.1, 0.5, 1\}$.

Concerning the weights $\|\boldsymbol{e}_i - Lreg(\boldsymbol{e}_i)\|^2$, $\langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle$, we can calculate them explicitly as following. We recall that the linear regression of the data $\boldsymbol{e}_i$ is given by

$$Lreg(\boldsymbol{e}_i)(t) = \beta_1^i + t\beta_2^i,$$

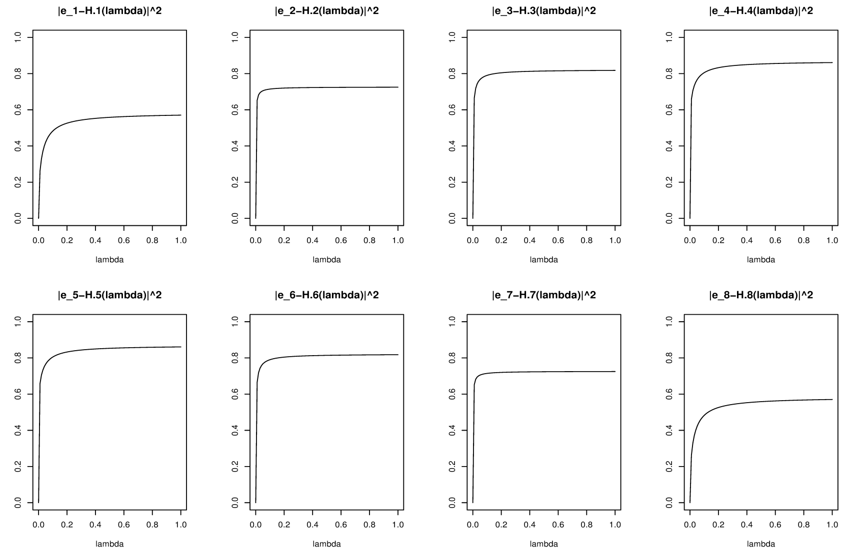

FIGURE 4. $\lambda \to \|\boldsymbol{e}_i - \mathbf{H}_{\cdot i}(\lambda)\|^2$.

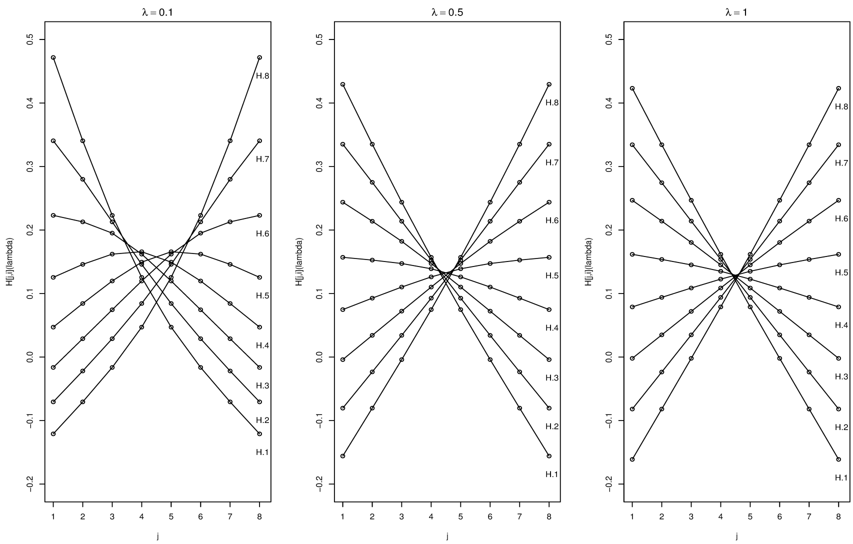

FIGURE 5. Plot of $j \to \mathbf{H}_{ji}(\lambda)$. Here $n = 7$ and $t_i = \frac{i-1}{n}$ for $i, j = 1, \ldots, 8$.

with $\beta_2^i = \frac{t_i - \bar{t}}{\sum_{j=1}^{n+1}(t_j - \bar{t})^2}$, $\beta_1^i = \frac{1}{n+1} - \beta_2^i \bar{t}$, and $\bar{t}$ denotes the empirical mean of the knots. The straightlines $(t \to Lreg(\boldsymbol{e}_i)(t) : i = 1, \ldots, n+1)$ have the common point

$(\bar{t}, \frac{1}{n+1})$. Moreover we have, for $i \neq j$, that

$$\begin{aligned} \|\boldsymbol{e}_i - Lreg(\boldsymbol{e}_i)\|^2 &= 1 - Lreg(\boldsymbol{e}_i)(t_i), \\ \langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle &= -Lreg(\boldsymbol{e}_i)(t_j), \\ &= -Lreg(\boldsymbol{e}_j)(t_i). \end{aligned}$$

From all that we get the following result.

Proposition 5.2. We have, for $i \neq j$,

$$\begin{aligned} \|\boldsymbol{e}_i - Lreg(\boldsymbol{e}_i)\|^2 &= \frac{n}{n+1} - \frac{(t_i - \bar{t})^2}{\sum_{j=1}^{n+1}(t_j - \bar{t})^2}, \\ \langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle &= -\frac{1}{n+1} - \frac{(t_i - \bar{t})(t_j - \bar{t})}{\sum_{j=1}^{n+1}(t_j - \bar{t})^2}. \end{aligned}$$

It follows that the most important weights $\|\boldsymbol{e}_i - Lreg(\boldsymbol{e}_i)\|^2$ are when the t_i's are close to $\bar{t}$. The most important negative correlation $\langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle$ is given by the couple of end-points (t_1, t_{n+1}). The most important positive correlations $\langle \boldsymbol{e}_i - Lreg(\boldsymbol{e}_i), \boldsymbol{e}_j - Lreg(\boldsymbol{e}_j)\rangle$ are given by the begining (t_1, t_2) and the ending (t_n, t_{n+1}) of the knots. The message of these remarks is that the allowed size of the noise depends on the values of data at the end-points (t_1, t_{n+1}) and at center i.e. near $\bar{t}$.

Figure 6 shows that the straightlines $(t \to Lreg(\boldsymbol{e}_j)(t) : j = 1, \ldots, n+1)$ turn in the trigonometric sense around their common point. Remark that the Figure 5 illustrates also the convergence of $\mathbf{H}_{\cdot i}(\lambda) \to [Lreg(\boldsymbol{e}_i)(t_j) : j = 1, \ldots, n+1]^T$ as $\lambda \to +\infty$.

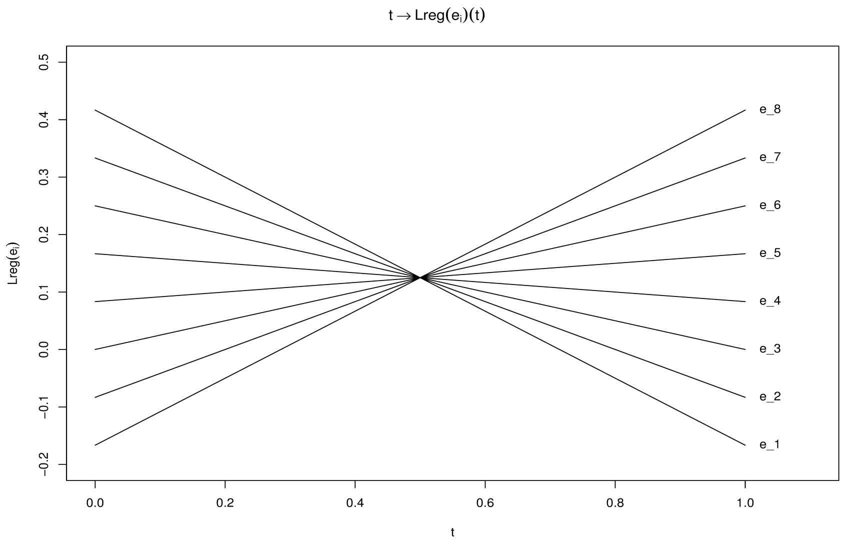

FIGURE 6. Plot of $t \to Lreg(\boldsymbol{e}_i)(t)$. Here $n = 7$ and $t_j = \frac{j-1}{n}$ for $i, j = 1, \ldots, 8$.

What can we do if the condition (29) does not hold ? In this case for all $\lambda \geq 0$, $\|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2$ represents only a part of the noise i.e.

$$\frac{\|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2}{\|\boldsymbol{w}\|^2} < \frac{\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2}{\|\boldsymbol{w}\|^2} \in [0, 1).$$

5.2. Gaussian white noise. We suppose that the noise $\boldsymbol{w}$ is Gaussian and white with the variance σ_w^2. In this case $\sum_{i=1}^{n+1} \frac{w_i^2}{\sigma_w^2}$ has the χ^2_{n+1} probability distribution $((\frac{w_i}{\sigma} : \quad i = 1, \ldots, n+1)$ are i.i.d. with the common distribution $\mathcal{N}(0,1)$, the standard Gaussian distribution). For all $\varepsilon > 0$, the event

$$(n+1)(1-\varepsilon) \leq \frac{\|\boldsymbol{w}\|^2}{\sigma^2} \leq (n+1)(1+\varepsilon)$$

holds with the probability

$$\mathbb{P}((1-\varepsilon) \leq \frac{\chi^2_{n+1}}{n+1} \leq (1+\varepsilon)).$$

The latter probability is close to 1 as n becomes large.

A first way to link the smoothing parameter to the noise is to choose λ solution of the following constraint

$$(n+1)(1-\varepsilon)\sigma^2 \leq \|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2 \leq (n+1)(1+\varepsilon)\sigma^2. \tag{33}$$

We denote respectively $\lambda^-(\sigma^2, \varepsilon, n+1)$, $\lambda^+(\sigma^2, \varepsilon, n+1)$ the solution of the equations

$$\|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2 = (n+1)(1-\varepsilon)\sigma^2, \tag{34}$$

and

$$\|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2 = (n+1)(1+\varepsilon)\sigma^2. \tag{35}$$

The solution of (34) exists under the hypothesis

$$(1-\varepsilon)\sigma^2 < \frac{\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2}{n+1}. \tag{36}$$

The solution of (35) exists under the hypothesis

$$(1+\varepsilon)\sigma^2 < \frac{\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2}{n+1}. \tag{37}$$

Remark that if $\lambda^+(\sigma^2, \varepsilon, n+1)$ exists then $\lambda^-(\sigma^2, \varepsilon, n+1)$ also exists. But in general the opposite is false. To understand the constraints (36) and (37), we are going to study the quantity $\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2$ as a function of the signal $\boldsymbol{p}$ and the noise $\boldsymbol{w}$.

If the model is $y_i = a + bt_i + w_i$ with $i = 1, \ldots, n+1$, then $Lreg\boldsymbol{y} = \mathbf{H}(+\infty)\boldsymbol{y}$ is the maximum likelihood estimator of the vector $\mathbf{L}(a,b)^T$. Moreover, $\frac{\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2}{n-1}$ is an unbiased consistent estimator of the variance σ^2. More precisely, $\frac{\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2}{\sigma^2}$ has the χ^2_{n-1}-distribution. Hence, the constraint (37) holds with the probability $\mathbb{P}(\chi^2_{n-1} > (n+1)(1+\varepsilon)) \to 0$ as $n \to +\infty$. But for $\varepsilon > \frac{2}{n+1}$, the constraint (36) holds with the probability $\mathbb{P}(\chi^2_{n-1} > (n-1)(1-\varepsilon)) \to 1$ as $n \to +\infty$.

In the general case we have the following result.

Proposition 5.3. Let $\boldsymbol{y} = \boldsymbol{p} + \boldsymbol{w}$, where $\boldsymbol{w}$ is the Gaussian white noise with the variance σ_w^2. We have

$$\mathbb{E}[\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2] = (n-1)\sigma_w^2 + \|\boldsymbol{p} - Lreg\boldsymbol{p}\|^2.$$

If the noise is fixed, then $\boldsymbol{p} \to \mathbb{E}[\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2]$ is minimal at the straightlines, i.e. $p_i = a + bt_i$ for all $i = 1, \dots, n+1$. The minimal value is equal to

$$\mathbb{E}[\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2] = (n-1)\sigma^2. \tag{38}$$

Proof. From the equality $\boldsymbol{y} = \boldsymbol{p} + \boldsymbol{w}$, we have

$$\|\boldsymbol{y} - Lreg\boldsymbol{y}\|^2 = \|\boldsymbol{p} - Lreg\boldsymbol{p}\|^2 + \|\boldsymbol{w} - Lreg\boldsymbol{w}\|^2 + 2\langle \boldsymbol{p} - Lreg\boldsymbol{p}, \boldsymbol{w} - Lreg\boldsymbol{w}\rangle.$$

The rest of the proof is consequence of $\mathbb{E}(\boldsymbol{w}) = 0$ and $\mathbb{E}[\|\boldsymbol{w} - Lreg\boldsymbol{w}\|^2] = (n-1)\sigma^2$. □

Roughly speaking, Proposition 5.3 combined with (36) and (37) tell us that the smoothing parameter

$$\lambda \in [\lambda^-(\sigma^2, \varepsilon, n+1), \lambda^+(\sigma^2, \varepsilon, n+1)]$$

exists under the constraint

$$\|\boldsymbol{p} - Lreg\boldsymbol{p}\|^2 \approx 2\sigma^2.$$

5.3. Smoothing parameter, SURE and PE. A second way to choose the smoothing parameter is to consider Stein's unbiased risk estimate (SURE) and the predictive risk error (PE).

a) Stein's Unbiased Risk estimate (SURE) [8], [21]: The quadratic loss of the estimation of the vector $\boldsymbol{p}$ by $\mathbf{H}(\lambda)\boldsymbol{y}$ is equal to

$$\|\mathbf{H}(\lambda)\boldsymbol{y} - \boldsymbol{s}\|^2 = \sum_{i=1}^{n+1} |\mathbf{H}(\lambda)\boldsymbol{y}(i) - s(t_i)|^2,$$

and the residual sum of squares is defined by

$$RSS(\lambda) := \|\boldsymbol{y} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2.$$

The mean square risk is equal to

$$\begin{aligned}
&R(\mathbf{H}(\lambda)\boldsymbol{y}, \boldsymbol{p}) = \mathbb{E}[\|\mathbf{H}(\lambda)\boldsymbol{y} - \boldsymbol{p}\|^2] \\
&= \mathbb{E}[\|\boldsymbol{y} - \boldsymbol{p}\|^2] + \mathbb{E}[RSS(\lambda)] - 2cov(\boldsymbol{w} - \mathbf{H}(\lambda)\boldsymbol{w}, \boldsymbol{w}) \\
&= \mathbb{E}[\|\boldsymbol{y} - \boldsymbol{p}\|^2] + \mathbb{E}[RSS(\lambda) + 2\sigma^2(Trace(\mathbf{H}(\lambda)) - (n+1))] \\
&= \mathbb{E}[RSS(\lambda) + 2\sigma^2 Trace(\mathbf{H}(\lambda)) - (n+1)\sigma^2].
\end{aligned}$$

The quantity

$$RSS(\lambda) + 2\sigma^2 Trace(\mathbf{H}(\lambda)) - (n+1)\sigma^2$$

is an unbiased risk estimate (called Stein's Unbiased Risk estimate, SURE for short). By minimizing SURE with respect to $\lambda \in (0, +\infty)$ we provide a criterion for choosing the smoothing parameter λ_{SURE}.

b) Prediction and Training errors (PE). The prediction error is our error on a new observations $y_i^* = s(t_i) + w^*(t_i), i = 1, \ldots, n+1$ independent of $\boldsymbol{y}$. If we predict the vector $\boldsymbol{p}$ by $\mathbf{M}(\lambda)\boldsymbol{y}$, then the predictive risk PE is equal to

$$\begin{aligned}
&\mathbb{E}[\|\boldsymbol{y}^* - \mathbf{H}(\lambda)\boldsymbol{y}\|^2] = \mathbb{E}[\|\boldsymbol{p} - \mathbf{H}(\lambda)\boldsymbol{y}\|^2] + (n+1)\sigma^2 \\
&= R(\mathbf{H}(\lambda)\boldsymbol{y}, \boldsymbol{p}) + (n+1)\sigma^2 \\
&= \mathbb{E}[RSS(\lambda) + 2\sigma^2 Trace(\mathbf{H}(\lambda))].
\end{aligned}$$

Hence, $RSS(\lambda) + 2\sigma^2 Trace(\mathbf{H}(\lambda))$ is an unbiased estimate of the prediction error. It follows that minimizing SURE is equivalent to minimize PE and then $\lambda_{SURE} = \lambda_{PE}$.

In Figure 7 we plot, for $n = 7$, $i = 1, \ldots, n+1$, $t_i = \frac{i-1}{n}$, the map $\lambda \in (0, +\infty) \to Trace(\mathbf{H}(\lambda))$.

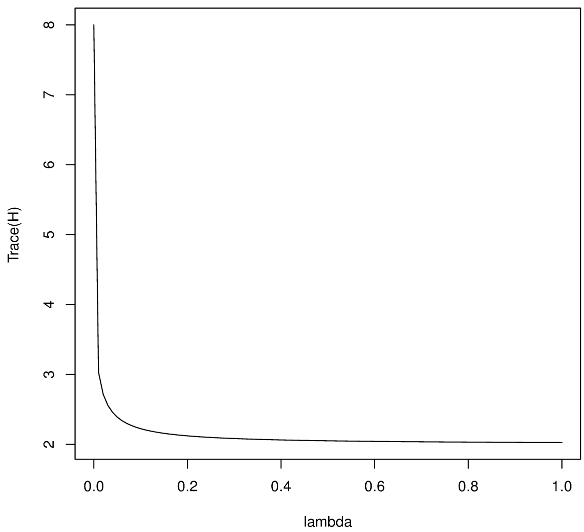

FIGURE 7. Plot of $\lambda \in (0, +\infty) \to Trace(\mathbf{H}(\lambda))$.

6. Cubic spline estimate: General case

In this section we propose to find suitable symmetric and non-negative definite matrices $\mathbf{P}_{pen} = [p_{ij} : i, j = 1, \ldots, n+3]$ such that the minimizer

$$(\hat{u}_1, \hat{p}_1, \ldots, \hat{p}_{n+1}, \hat{u}_{n+1}) = \arg\min\{\lambda(u_1, \boldsymbol{p}^T, u_{n+1})\mathbf{P}_{pen} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} + \|\boldsymbol{y} - \boldsymbol{p}\|^2 : \quad u_1, p_1, \ldots, p_{n+1}, u_{n+1}\} \qquad (39)$$

is a non natural cubic spline, i.e. $(\hat{u}_1, \hat{u}_{n+1}) \neq (0, 0)$. The following proposition addresses the uniqueness and the capacity of the estimator (39) to rediscover a non natural spline.

Proposition 6.1. 1) The minimizer of (39) is unique if and only if the sub-matrix $\begin{pmatrix} p_{1,1} & p_{1,n+3} \\ p_{n+3,1} & p_{n+3,n+3} \end{pmatrix}$ is invertible. In this case the minimizer is given by

$$\begin{pmatrix} \hat{u}_1 \\ \hat{\boldsymbol{p}} \\ \hat{u}_{n+1} \end{pmatrix} = (\lambda \mathbf{P}_{pen} + \Pi^T \Pi)^{-1} \begin{pmatrix} 0 \\ \boldsymbol{y} \\ 0 \end{pmatrix} \tag{40}$$

$$: = \mathbf{H}_{\mathbf{P}_{pen}}(\lambda) \begin{pmatrix} 0 \\ \boldsymbol{y} \\ 0 \end{pmatrix} \tag{41}$$

where $\Pi \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} = \boldsymbol{p}$ for all $u_1, \boldsymbol{p}, u_{n+1}$.

2) The condition

$$(p_{1,j} : \quad j = 2, \ldots, n+2) \neq 0, \tag{42}$$

respectively

$$(p_{n+3,j} : \quad j = 2, \ldots, n+2) \neq 0, \tag{43}$$

is the necessary condition which guaranties that $\hat{u}_1 \neq 0$ respectively $\hat{u}_{n+1} \neq 0$.

Now, we discuss the limits of (40) as $\lambda \to 0$ and $\lambda \to +\infty$.

Corollary 6.2. 1) The limit

$$\mathbf{H}_{\mathbf{P}_{pen}}(\lambda) \begin{pmatrix} 0 \\ \boldsymbol{y} \\ 0 \end{pmatrix} \to \begin{pmatrix} u_1^0 \\ \boldsymbol{y} \\ u_{n+1}^0 \end{pmatrix} \tag{44}$$

as $\lambda \to 0$. Here u_1^0, u_{n+1}^0 is a minimizer of the objective function

$$(u_1, u_{n+1}) \to (u_1, \boldsymbol{y}^T, u_{n+1}) \mathbf{P}_{pen} \begin{pmatrix} u_1 \\ \boldsymbol{y} \\ u_{n+1} \end{pmatrix},$$

i.e. u_1, u_{n+1} is solution of the linear system

$$u_1 p_{1,1} + u_{n+3} p_{1,n+3} = -\sum_{i=1}^{n+1} y_i p_{1,i+1}$$

$$u_1 p_{n+3,1} + u_{n+3} p_{n+3,n+3} = -\sum_{i=1}^{n+1} y_i p_{n+3,i+1}.$$

In particular, if the data $\boldsymbol{y}$ is not orthogonal to the space spanned by the vectors $(p_{1,j} : \quad j = 2, \ldots, n+2)^T$, $(p_{n+3,j} : \quad j = 2, \ldots, n+2)^T$, then $\hat{u}_1$, $\hat{u}_{n+1}$ can't be both equal to zero. Namely, the estimator (44) is not a natural spline.

2) The limit

$$\mathbf{H}_{\mathbf{P}_{pen}}(\lambda) \begin{pmatrix} 0 \\ \boldsymbol{y} \\ 0 \end{pmatrix} \to \arg\min\{\|\boldsymbol{y} - \boldsymbol{p}\|^2 : \quad \mathbf{P}_{pen} \begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} = 0\}, \tag{45}$$

as $\lambda \to +\infty$.

Examples. The matrix $\mathbf{P}_{pen} = \mathbf{C}$ given in (16) corresponds to the penalty

$$(u_1, \boldsymbol{p}^T, u_{n+1})^T \mathbf{C}(u_1, \boldsymbol{p}^T, u_{n+1})^T = \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt$$

where the cubic spline $s = u_1\varphi_0 + \sum_{i=1}^{n+1} p_j\varphi_i + u_{n+1}\varphi_{n+2}$. A natural way to construct new matrices $\mathbf{P}_{pen}$ is to consider the more general penalization

$$\int_{t_1}^{t_{n+1}} |a_2 s''(t) + a_1 s'(t) + a_0 s(t)|^2 dt = (u_1, \boldsymbol{p}^T, u_{n+1})^T \mathbf{P}_{pen}(u_1, \boldsymbol{p}^T, u_{n+1})^T,$$

where a_0, a_1, a_2 are given real numbers. Hence

$$\begin{aligned}\mathbf{P}_{pen} = a_0^2\mathbf{C}_{00} + a_1^2\mathbf{C}_{11} + a_2^2\mathbf{C}_{22} + a_0a_1[\mathbf{C}_{01} + \mathbf{C}_{01}^T] + a_0a_2[\mathbf{C}_{02} + \mathbf{C}_{02}^T] \\ +a_1a_2[\mathbf{C}_{12} + \mathbf{C}_{12}^T],\end{aligned} \tag{46}$$

where $\mathbf{C}_{00}, \mathbf{C}_{01}, \mathbf{C}_{02}, \mathbf{C}_{11}, \mathbf{C}_{12}$ are respectively defined by

$$\begin{aligned}
\mathbf{C}_{00} &= [\int_{t_1}^{t_{n+1}} \varphi_{i-1}(t)\varphi_{j-1}(t)dt: \quad i,j = 1,\ldots,n+3], \\
\mathbf{C}_{01} &= [\int_{t_1}^{t_{n+1}} \varphi_{i-1}(t)\varphi'_{j-1}(t)dt: \quad i,j = 1,\ldots,n+3], \\
\mathbf{C}_{02} &= [\int_{t_1}^{t_{n+1}} \varphi_{i-1}(t)\varphi''_{j-1}(t)dt: \quad i,j = 1,\ldots,n+3], \\
\mathbf{C}_{11} &= [\int_{t_1}^{t_{n+1}} \varphi'_{i-1}(t)\varphi'_{j-1}(t)dt: \quad i,j = 1,\ldots,n+3], \\
\mathbf{C}_{12} &= [\int_{t_1}^{t_{n+1}} \varphi'_{i-1}(t)\varphi''_{j-1}(t)dt: \quad i,j = 1,\ldots,n+3],
\end{aligned}$$

and the matrix $\mathbf{C}_{22} = \mathbf{C}$ defined in (16).

The matrix $\mathbf{P}_{pen}$, for $a_0 = a_1 = a_2 = 1$, $n = 7$, $t_i = \frac{i-1}{n}, i = 1, \ldots, n+1$, has the following form:

$$\begin{pmatrix}
2.6 & -239.22 & 614.68 & -559.93 & 255.91 & -96.85 & 33.253 & -9.951 & 1.841 & -0.01 \\
-239.22 & 30883.64 & -89524.08 & 98923.05 & -57634.4 & 23760.7 & -8518.95 & 2611.28 & -488.75 & 1.84 \\
614.68 & -89524.08 & 278142.5 & -345243.9 & 238016.7 & -113691.9 & 43414.47 & -13742.88 & 2611.71 & -9.94 \\
-559.93 & 98923.05 & -345243.9 & 516426.7 & -459007.5 & 281450.3 & -127439.9 & 43417.03 & -8521.09 & 33.24 \\
255.91 & -57634.4 & 238016.7 & -459007.5 & 559860.4 & -472755.5 & 281452.9 & -113702.2 & 23768.8 & -96.81 \\
-96.85 & 23760.7 & -113691.9 & 281450.3 & -472755.5 & 559863 & -459017.7 & 238055.1 & -57664.7 & 255.76 \\
33.253 & -8518.95 & 43414.47 & -127439.9 & 281452.9 & -459017.7 & 516465.1 & -345387.2 & 99036.12 & -559.40 \\
-9.951 & 2611.28 & -13742.88 & 43417.03 & -113702.2 & 238055.1 & -345387.2 & 278677.6 & -89946.04 & 612.72 \\
1.841 & -488.75 & 2611.71 & -8521.09 & 23768.8 & -57664.7 & 99036.12 & -89946.04 & 31219.29 & -236.99 \\
-0.01 & 1.84 & -9.95 & 33.24 & -96.819 & 255.76 & -559.40 & 612.722 & -236.99 & 3.85
\end{pmatrix}$$

Observe that the conditions (42) and (43) are satisfied.

7. Bayesian Model and statistical analysis

The aim of this section is to give the Bayesian interpretation of the matrix penalization $\mathbf{P}_{pen}$. Let us first set the noisy cubic spline estimate in the context of the general linear model:

$$\boldsymbol{y} = \mathbf{F}\beta + \mathbf{R}\eta + \boldsymbol{w},$$

where β is an unknown parameters, $\mathbf{F}$ and $\mathbf{R}$ are known matrices. The random effects η and the noise $\boldsymbol{w}$ are unknown, centred and independent random vectors.

Their covariance matrices $cov(\eta) := \mathbf{\Sigma}_\eta$, $cov(\boldsymbol{w}) := \mathbf{\Sigma}_w$ are known. The term $\mathbf{F}\beta$ is called the fixed effects and $\mathbf{R}\eta$ is the random effects.

Let us revisit the best linear unbiased predictors (BLUP) and the best linear unbiased estimators (BLUE). There is a long history and huge literature on this subject, see for instance [3], [4], [5], [9], [10], [11], [12], [15], [16], [20] and references herein.

BLUE of β. The BLUE of β is the estimator $\hat{\beta} = \hat{\mathbf{M}}_\beta \boldsymbol{y}$, with $\hat{\mathbf{M}}_\beta$ (called the hat matrix of $\hat{\beta}$) being the matrix such that $\hat{\mathbf{M}}_\beta \mathbf{F} = \mathbf{I}$ (the identity matrix) and $cov(\mathbf{M}\boldsymbol{y}) - cov(\hat{\mathbf{M}}_\beta \boldsymbol{y})$ is positive semi-definite for all matrix $\mathbf{M}$ subject to $\mathbf{MF} = \mathbf{I}$.
BLUP of η. The BLUP of η is the estimator $\hat{\eta} = \hat{\mathbf{M}}_\eta \boldsymbol{y}$, with $\hat{\mathbf{M}}_\eta$ (called the hat matrix of $\hat{\eta}$) being the matrix such that $\hat{\mathbf{M}}_\eta \mathbf{F} = 0$ and $cov(\mathbf{M}\boldsymbol{y}) - cov(\hat{\mathbf{M}}_\eta \boldsymbol{y})$ is positive semi-definite for all matrix $\mathbf{M}$ subject to $\mathbf{MF} = 0$. We call, by convention, *predictors* of a random variable to distinguish them from *estimators* of a deterministic parameter. Henderson et al.(1959)[11] showed that the BLUE and the BLUP are respectively

$$\hat{\beta} = (\mathbf{F}^T(\mathbf{R}\mathbf{\Sigma}_\eta\mathbf{R}^T + \mathbf{\Sigma}_w)^{-1}\mathbf{F})^{-1}\mathbf{F}^T(\mathbf{R}\mathbf{\Sigma}_\eta\mathbf{R}^T + \mathbf{\Sigma}_w)^{-1}\boldsymbol{y}, \tag{47}$$

$$\begin{aligned}\hat{\eta} &= (\mathbf{R}^T\mathbf{\Sigma}_N^{-1}\mathbf{R} + \mathbf{\Sigma}_\eta^{-1})^{-1}[\mathbf{R}^T\mathbf{\Sigma}_w^{-1} - \mathbf{R}^T\mathbf{\Sigma}_w^{-1}\mathbf{F}(\mathbf{F}^T(\mathbf{\Sigma}_w + \mathbf{R}\mathbf{\Sigma}_\eta\mathbf{R}^T)^{-1}\mathbf{F})^{-1} \\ &\quad \mathbf{F}^T(\mathbf{\Sigma}_w + \mathbf{R}\mathbf{\Sigma}_\eta\mathbf{R}^T)^{-1}]\boldsymbol{y}.\end{aligned} \tag{48}$$

Now we are able to give a Baysian interpretation of the hat matrix $\mathbf{H}_{\mathbf{P}_{pen}}(\lambda)$ (40). Let $\mathbf{P}_1$ be an $n+3$ by $n+3-dim(N(\mathbf{P}_{pen}))$ matrix such that

$$\mathbf{P}_1^T\mathbf{P}_{pen}\mathbf{P}_1 = \mathbf{I}_{(n+3-dim(N(\mathbf{P}_{pen})))\times(n+3-dim(N(\mathbf{P}_{pen})))}, \tag{49}$$

respectively $\mathbf{P}_0$ an $n+3$ by $dim(N(\mathbf{P}_{pen}))$ matrix such that

$$\mathbf{P}_{pen}P_0 = 0, \quad \text{and its columns form a basis of} \quad N(\mathbf{P}_{pen}). \tag{50}$$

It follows that, for all vector $\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix}$, there exist a unique $\beta \in \mathbb{R}^{dim(N(\mathbf{P}_{pen}))}$ and $\eta \in \mathbb{R}^{n+3-dim(N(\mathbf{P}_{pen}))}$ such that

$$\begin{pmatrix} u_1 \\ \boldsymbol{p} \\ u_{n+1} \end{pmatrix} = \mathbf{P}_0\beta + \mathbf{P}_1\eta. \tag{51}$$

Hence the model $\boldsymbol{y} = \boldsymbol{p} + \boldsymbol{w}$ becomes

$$\begin{aligned}\boldsymbol{y} &= \Pi P_0\beta + \Pi P_1\eta + \boldsymbol{w} \\ &:= \mathbf{F}\beta + \mathbf{R}\eta + \boldsymbol{w}.\end{aligned}$$

We suppose that β is the fixed effect and η is independent of the noise $\boldsymbol{w}$ and drawn from a centred distribution having the covariance matrix $\sigma_s^2\mathbf{I}_{n+3-dim(N(P))}$.

Now, we are able to give our Bayesian interpretation.

Proposition 7.1. The components $(\hat{\beta}, \hat{\eta})$ of

$$\arg\min\{\frac{\sigma_w^2}{\sigma_s^2}\|\eta\|^2 + \|\boldsymbol{y} - (\mathbf{F}\beta + R\eta)\|^2 : \quad \beta \in \mathbb{R}^{dim(N(\mathbf{P}_{pen}))}, \eta \in \mathbb{R}^{n+3-dim(N(\mathbf{P}_{pen}))}\}$$

are respectively the BLUE of β and the BLUP of η. Moreover, we have

$$\begin{aligned}\mathbf{F}\hat{\beta}+\mathbf{R}\hat{\eta} &= \Pi\mathbf{H}_{\mathbf{P}_{pen}}(\frac{\sigma_w^2}{\sigma_s^2})\begin{pmatrix}0\\ \boldsymbol{y}\\ 0\end{pmatrix},\\ \mathbf{P}_0\hat{\beta}+\mathbf{P}_1\hat{\eta} &= \mathbf{H}_{P_{pen}}(\frac{\sigma_w^2}{\sigma_s^2})\begin{pmatrix}0\\ \boldsymbol{y}\\ 0\end{pmatrix}.\end{aligned}$$

The proof is a consequence of the change of variable formula (51) and (48). See [3] Proposition 2.2 for a similar proof.

Corollary 7.2. Let $P = \mathbf{C}$ be the matrix (16) and $\mathbf{P}_0, \mathbf{P}_1$ be the corresponding matrices defined by (50), (49). We have

$$\begin{aligned}\mathbf{P}_0\hat{\beta} &= \begin{pmatrix}0\\ Lreg\boldsymbol{y}\\ 0\end{pmatrix},\\ \mathbf{P}_1\hat{\eta} &= \begin{pmatrix}0\\ (\mathbf{H}(\frac{\sigma_w^2}{\sigma_s^2})-Lreg)\boldsymbol{y}\\ 0\end{pmatrix}.\end{aligned}$$

Conclusion

In this work we defined a new basis of the set of C^2-cubic splines. We revisited the estimation of a natural cubic spline using Schoenberg-Reinsch result and we extended their result to the estimation of any C^2-cubic spline. We studied the choice of the smoothing parameter when the noise is deterministic or white throughout several criteria. We also gave a Bayesian interpretation of our estimators.

References

[1] C. de Boor, *A practical guide to splines*, Springer.

[2] G. Claeskens, T. Krivobokova, J.D. Opsomer, Asymptotic properties of penalized splines estimators, *Biometrika*(2009) **96**(3), 529–544

[3] A. Dermoune, N. Rahmania, T. Wei, General Linear mixed model and signal extraction problem with constraint, *Journal of Multivariate Analysis* (2012) **105**(1), 311–321.

[4] A. Dermoune, B. Djehiche, N. Rahmania, Multivariate Extension of the Hodrick-Prescott Filter-Optimality and Characterization, *Studies in NonLinear dynamics and Econometrics* (2009) **13**(3), .

[5] A. Dermoune, B. Djehiche, N. Rahmania, A consistent estimator of the smoothing parameter in the Hodrick-Prescott filter, *J. Japan Statist. Soc.* (2008) **38**(2), 225–241.

[6] H. Drygas, *The coordinate-free approach to Gauss-Markov estimation*, Springer, 1970.

[7] H. Drygas, Linear sufficiency and some applications in multilinear estimation, *Journal of Multivariate Analysis* (1985) **16**, 71–84.

[8] B. Efron, How biased is the apparent error rate of prediction rule? *Journal of the American Statistical Association* (1986) **81** (394), 461–470.

[9] C.R. Henderson, Estimation of genetic parameters, *Ann. Math. Stat.* (1950) **21**, 309–310.

[10] C.R. Henderson, Selection index and expected genetic advance, *Statistical Genetics and Plant Breeding* (1963), 141–163.

[11] C.R. Henderson, O. Kempthorne, S.R. Searle, C. von Krosigk, The estimation of environmental and genetic trends from records subject to culling, *Biomctrics* (1959) **16**, 192–218.

[12] C.R. Henderson, Best linear unbiased estimation and prediction under a selection model, *Biometrics* (1975) **31**, 423–447.
[13] N. Krämer, A.L. Boulesteix, G. Tutz, Penalized Partial Least Squares Based on B-Splines Transformations, *Chemometrics and Intelligent Laboratory Systems* (2008) **94**(1), 60–69.
[14] R. Paige, A.A. Trindade, The Hodrick-Prescott Filter: A special case of penalized spline smoothing, *Electronic Journal of Statistics* (2010) **4**, 856–874.
[15] C.R. Rao, Representation of the best linear unbiased estimators in the Gauss-Markoff model with a singular disperssion matrix, *Journal of Multivariate analysis* (1973) **3**, 276–292.
[16] G.K. Robinson, That BLUP is a good thing: The estimation of random effects, *Statistical Science* (1991) **6**(1), 15–51.
[17] G. Micula, S. Micula, *Handbook of splines*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1999.
[18] C.H. Reinsch, Smoothing by Spline Functions, *Numerische Mathematik* (1967) **10**, 177–183.
[19] I.J. Schoenberg, Contributions to the problem of approximation of equidistant data by analytic functions, *Quart. Appl. Math.* (1946) **4**, 44–99 and 112–141.
[20] X.Q. Liu, J. Rong, X.Y. Liu, Best linear unbiased prediction for linear combinations in general mixed linear models, *Journal of Multivariate Analysis* (2008) **99**(8), 1503–1517.
[21] C.M. Stein, Estimation of the mean of a multivariate normal distribution, *The Annals of Statistics* (19810 **9**(6), 1135–1151.
[22] G. Sewell, *Computational methods of linear algebra*, Second edition, Wiley, 2005.

(Azzouz Dermoune) Laboratoire Paul Painlevé, UFR de Mathématiques,
Université de Lille 1, 59655 Villeneuve d'Ascq, France
E-mail address: azzouz.dermoune@univ-lille1.fr

(Cristian Preda) Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy: ISMMA Gheorghe Mihoc-Caius Iacob, Bucharest, 050711, Romania
Université de Lille 1, 59655 Villeneuve d'Ascq, France
E-mail address: cristian.preda@polytech-lille.fr