Estimation of noisy cubic spline using a natural basis

Azzouz Dermoune, Cristian Preda
Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524.
UFR de Mathématiques, Bât. M2, 59655 Villeneuve d’Ascq Cédex, France.
azzouz.dermoune@univ-lille1.fr, cristian.preda@univ-lille1.fr

March 7, 2014

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.

keyword. Cubic spline. BLUP. Smoothing parameter. Natural cubic spline. Linear regression. Bayesian model. Stein’s Unbiased Risk Estimate. Prediction errors.

1 Introduction

We consider, for \( n \geq 1 \), the regression model

\[
y_i = f(t_i) + w_i, \quad i = 1, \ldots, n + 1,
\]

where \( y_1, \ldots, y_{n+1}, t_1 < \ldots < t_{n+1} \) are real-valued observations, \( w_1, \ldots, w_{n+1} \) 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

\[
y_i = s(t_i) + w_i, \quad i = 1, \ldots, n + 1,
\]

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 [15]. 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 \( t_1 < \ldots < t_{n+1} \) if there exist
\((p_1, \ldots, p_{n+1}) \in \mathbb{R}^{n+1}, \ (q_1, \ldots, q_n), \ (u_1, \ldots, u_n), \ (v_1, \ldots, v_n) \in \mathbb{R}^n\) such that, for \(i = 1, \ldots, n\) and \(t \in [t_i, t_{i+1}],\)

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

We are interested 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 \(i = 1, \ldots, n+1\), 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.

\section{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

\[
\begin{align*}
p_i + q_i h_i + \frac{u_i}{2} h_i^2 + \frac{v_i}{6} h_i^3 &= p_{i+1}, \quad i = 1, \ldots, n, \\
qu_i + u_i h_i + \frac{v_i}{2} h_i^2 &= q_{i+1}, \quad i = 1, \ldots, n-1, \\
u_i + h_i v_i &= u_{i+1}, \quad i = 1, \ldots, n.
\end{align*}
\]

We introduce the column vectors \(q = (q_1, \ldots, q_n)^T, \ p = (p_1, \ldots, p_{n+1})^T, \) and \(u = (u_1, \ldots, u_{n+1})^T\), where \(M^T\) is the transpose of the matrix \(M\). Using [11], [12], [13], we can show that there exist three matrices \(Q, U, V\) such that

\[
\begin{align*}
q &= Q \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix}, \\
u &= U \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix}, \\
v &= V \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix}.
\end{align*}
\]

See the appendices 1 and 2 for the details.
Let us define, for each \( i = 1, \ldots, n \), the piecewise functions,
\[
\begin{align*}
\phi_i(t) &= 1_{[t_i, t_{i+1})}(t), \\
\phi_1^2(t) &= (t-t_i)1_{[t_i, t_{i+1})}(t), \\
\phi_i^3(t) &= (t-t_i)^31_{[t_i, t_{i+1})}(t), \\
\chi_0 &= 0, \\
\chi_{n+1} &= 1_{t_{n+1}}, \\
\chi_{n+2} &= 0.
\end{align*}
\]
Here \( 1_A \) denotes the indicator function of the set \( A \). Clearly, the set \([\chi_i, \phi_i^k : i = 1, \ldots, n+1, k = 1, 2, 3] \) forms a basis of the set of cubic splines \( S_q \). The map \( s \) has the coordinates \( p, q, u, v \) in this basis, i.e.
\[
s = \left( \begin{array}{cccc}
[\chi_1 \ldots \chi_n \chi_{n+1}] & [\chi_1^1 \ldots \chi_n^1] & \frac{1}{2}[\chi_1^2 \ldots \chi_n^2] & 0 \\
[\chi_1^1 \ldots \chi_n^1] & \frac{1}{2}[\chi_1^2 \ldots \chi_n^2] & 0 & \frac{1}{6}[\chi_1^3 \ldots \chi_n^3]
\end{array} \right) \left( \begin{array}{c}
p \\
q \\
u \\
v
\end{array} \right).
\]
If \( s \) is \( C^2 \), then from (69), (69), (60) (see Appendix 1), we have
\[
s = ([\chi_0 \chi_1 \ldots \chi_n \chi_{n+1} \chi_{n+2}] + [\chi_1^1 \ldots \chi_n^1])Q + \left[ \frac{1}{2}[\chi_1^2 \ldots \chi_n^2] \right]U + \left[ \frac{1}{6}[\chi_1^3 \ldots \chi_n^3] \right]V \left( \begin{array}{c}
u_1 \\
\nu_{n+1}
\end{array} \right).
\]
The \( C^2 \) cubic spline \( s \) can be rewritten in the following new basis:
\[
s = [\varphi_0 \ldots \varphi_{n+2}] \left( \begin{array}{c}
u_1 \\
\nu_{n+1}
\end{array} \right),
\]
where, for \( j = 0, \ldots, n+2 \),
\[
\varphi_j = \chi_j + [\chi_1^1 \ldots \chi_n^1]q_{j+1} + \frac{1}{2}[\chi_1^2 \ldots \chi_n^2]u_{j+1} + \frac{1}{6}[\chi_1^3 \ldots \chi_n^3]v_{j+1}.
\]
Here \( a_j \) denotes the \( j \)th column of the matrix \( A \). Each element of the new basis is a \( C^2 \) cubic spline.

From (58), we derive that the set of natural cubic splines is spanned by the basis \( (\varphi_j : j = 1, \ldots, n+1) \).

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

- The spline \( \varphi_j \), 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 \( \varphi_{n+2} \) is the unique \( C^2 \) cubic spline interpolating the points \((t_1, 0), \ldots, (t_{n+1}, 0)\) and such that \( \varphi_{n+2}''(t_1+) = 0, \varphi_{n+2}''(t_{n+1}+) = 1 \). Hence, \( \varphi_{n+2} \) 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
\[
s_0 = [\varphi_0, \ldots, \varphi_{n+2}] \left( \begin{array}{c}
0 \\
0 \\
0
\end{array} \right).
\]
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.

**Figure 1:** The graph of the 10 elements of the natural basis. Here \( n = 7 \) and \( t_i = \frac{i-1}{n} \) for \( i = 1, \ldots, 8 \).

### 3 Our basis and Schoenberg-Reinsch optimization

In this section we use our 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

\[
I_2(p) \triangleq \arg \min_{f \in H^2} \left\{ \int_{t_1}^{t_{n+1}} |f''(t)|^2 dt : \quad f(t_i) = p_i, \quad i = 1, \ldots, n + 1 \right\}
\]

\[
= \sum_{j=1}^{n+1} p_j \varphi_j
\]

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

$I_2(\delta_j) = \arg \min_{f \in H^2} \left\{ \int_{t_1}^{t_{n+1}} |f''(t)|^2 dt : f(t_i) = \delta_j, i = 1, \ldots, n + 1 \right\} = \varphi_j$.
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 $\varphi_0$ and $\varphi_{n+2}$ are respectively solution of the following optimization problems:

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

(11)

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

(12)

### 3.1 Revisiting Schoenberg and Reinsch result

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

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

The minimization

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

is equivalent to

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

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

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

(13)

If $s \in C^2 \cap S_3$, using (3) and (53), then

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

Now the equality (56) achieves the proof.

### 3.2 Some consequences of Schoenberg-Reinsch result

First, let us rewrite

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

(14)

$$= (u_1,p^T,u_{n+1}) C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix},$$

(15)
where
\[
C = \frac{h_1}{3} U_1^T U_1 + \frac{2}{3} \sum_{i=2}^{n} h_i U_i^T U_i + \frac{h_n}{3} U_{n+1}^T U_{n+1} + \frac{1}{6} \sum_{i=1}^{n} h_i [ U_i^T U_{i+1} + U_{i+1}^T U_i ].
\]  
(16)

Now, we summarize the properties of the matrix \( C \).

**Proposition 3.2.** The matrix \( C \) is symmetric, and non-negative definite. The quadratic form
\[
(u_1, p^T, u_{n+1}) C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} = 0
\]
if and only if \( u_1 = u_{n+1} = 0 \) and \( p \) belongs to the range \( R(L) \) of the matrix
\[
L = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_{n+1} \end{pmatrix}.
\]  
(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
\[
C = \begin{pmatrix} c_{1,1} & 0 & c_{1,n+3} \\ 0 & C(2, n + 2) & 0 \\ c_{n+3,1} & 0 & c_{n+3,n+3} \end{pmatrix},
\]
where \( C(2, n + 2) = [c_{ij} : i, j = 2, \ldots, n + 2] \). The sub-matrix
\[
\begin{pmatrix} c_{1,1} & c_{1,n+3} \\ c_{n+3,1} & c_{n+3,n+3} \end{pmatrix}
\]  
(18)
is symmetric, positive definite. The null-space of the sub-matrix \( C(2, n + 2) \) is equal to \( R(L) \). 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
\[
\int_{t_1}^{t_{n+1}} |s''(t)|^2 dt = u_1^2 c_{1,1} + u_{n+1}^2 c_{n+3,n+3} + 2c_{1,n+3} u_1 u_{n+1} + p^T C(2, n) p.
\]  
(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 \( \varphi''_0 \) (respectively \( \varphi''_{n+2} \)) is orthogonal to \( \varphi''_j \) for all \( j = 1, \ldots, n + 2 \) (respectively to \( \varphi''_j \) for all \( j = 0, \ldots, n + 1 \)). As an example the matrix \( C \) has the following form,
for \( n = 7, \ t_i = \frac{t_i - 1}{n}, i = 1, \ldots, n + 1, \)

\[
C = \begin{pmatrix}
0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.04 & -1250.64 & -886.54 & -237.54 & 63.63 & -16.97 \\
0.00 & -1250.64 & 3387.82 & -3261.27 & 1425.26 & -3668.49 & 1527.06 \\
0.00 & 3387.82 & -3261.27 & 1425.26 & -3668.49 & 1527.06 & -381.77 \\
0.00 & -3261.27 & 1425.26 & -3668.49 & 1527.06 & -381.77 & 101.80 \\
0.00 & 1425.26 & -3668.49 & 1527.06 & -381.77 & 101.80 & -25.45 \\
0.00 & -3668.49 & 1527.06 & -381.77 & 101.80 & -25.45 & 4.24 \\
0.00 & 1527.06 & -381.77 & 101.80 & -25.45 & 4.24 & -6.74 \\
\end{pmatrix}_{(20)}
\]

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 \( \varphi_0 \) and \( \varphi_{n+2} \) are respectively solution of the optimizations

\[
\begin{align}
\min_{s \in S_0 \cap C^2} \{ \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt : \ s''(t_1) = 1, \ s(t_i) = 0, \ i = 1, \ldots, n + 1 \} \quad (21) \\
\min_{s \in S_1 \cap C^2} \{ \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt : \ s''(t_{n+1}) = 1, \ s(t_i) = 0, \ i = 1, \ldots, n + 1 \}. \quad (22)
\end{align}
\]

**Proof.** The optimizations (21), (22) are equivalent to

\[
\begin{align}
\min_{u_1, p, u_{n+1}} \{ (u_1, p^T, u_{n+1})C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} : u_1 = 1, \ p = 0 \}, \\
\min_{u_1, p, u_{n+1}} \{ (u_1, p^T, u_{n+1})C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} : u_{n+1} = 1, \ p = 0 \},
\end{align}
\]

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

More generally we have the following result.

**Proposition 3.4.** Let \( k \) be a positive integer, \( M \) a \( n + 3 \) by \( n + 3 \) matrix, \( A \) be a \( k \) by \( n + 3 \) matrix and \( \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} := c \in \mathbb{R}^k \) all are given. Suppose that the null spaces \( N(A), N(M) \) do not overlap, i.e. \( N(A) \cap N(M) = \{0\} \). Then the optimization

\[
\begin{align}
\min_{u_1, p, u_{n+1}} \{ (u_1, p^T, u_{n+1})M \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} : A \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} = c \}
\end{align}
\]

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

\[
Mv = A^Tl, \quad Av = c.
\]

The vector \( v \) is the minimizer, and \( l \) 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.
\[
y_i = s(t_i) + w_i, \quad i = 1, \ldots, n + 1,
\]
where \( w_i \) is the noise added to the true value \( s(t_i) = p_i \). In the sequel \( y = (y_1, \ldots, y_{n+1})^T \).

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
\[
\min\{ \lambda (u_1, p^T, u_{n+1})C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} + ||p - y||^2 : \quad p \in \mathbb{R}^{n+1}, u_1, u_{n+1} \in \mathbb{R} \}, \tag{25}
\]
where \( || \cdot || \) denotes the Euclidean norm.

We have easily the following result.

**Proposition 4.1.** The equality
\[
(u_1, p^T, u_{n+1})C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} = u_1^2 c_{11} + u_{n+1}^2 c_{n+3,n+3} + 2u_1u_{n+1}c_{1,n+3} + p^T C(2,n+2)p
\]
implies that the minimizer of (25) is \( u_1 = u_{n+1} = 0 \) and \( p \) is solution of the following system:
\[(I + \lambda C(2,n+2))p = y.\]
The solution \( p \) is given by
\[
\hat{p} = H(\lambda)y, \tag{26}
\]
where \( H(\lambda) := (I + \lambda C(2,n+2))^{-1} \) is called the hat matrix.

Now, we discuss the limits of (25) as \( \lambda \to 0 \) and \( \lambda \to +\infty \).

**Corollary 4.2.** 1) The problem (25), when \( \lambda \to 0 \), becomes
\[
\min\{ (u_1, y^T, u_{n+1})C \begin{pmatrix} u_1 \\ y \\ u_{n+1} \end{pmatrix} : \quad p = y, u_1, u_{n+1} \in \mathbb{R} \}. \tag{27}
\]
Its minimizer is $u_1 = u_{n+1} = 0, p = y$ i.e.

$$\lim_{\lambda \to 0^+} H(\lambda) = I_{n+1}.$$ 

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

2) The problem (25), when $\lambda \to +\infty$, becomes

$$\min\{\|y - p\|^2 : C \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} = 0\}. \quad (28)$$

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

$$p = \lim_{\lambda \to +\infty} H(\lambda)y = L(L^T L)^{-1}L^T y = \pi_{R(L)}y,$$

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

Remark 4.3. We have easily $C(2, n + 2)L = 0$. From that we derive that $H(\lambda)L = L$ and then $L(L^T L)^{-1}L^T[H(\lambda) - L(L^T L)^{-1}L^T] = 0$. It follows that $L(L^T L)^{-1}L^T y$ and $[H(\lambda) - L(L^T L)^{-1}L^T]y$ are orthogonal. The component $L_{\text{reg}}y := L(L^T L)^{-1}L^T y$ is the linear regression i.e. $L_{\text{reg}}y$ is the orthogonal projection of the data $y$ on the linear space $R(L)$ (the range of $L$).

By introducing the orthogonal projections $\pi_{R(L)}$ and $\pi_{R(L)^\perp}$ respectively on $R(L)$ and $R(L)^\perp$, the minimizer

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

is the sum of

$$\arg\min\{\lambda p_1^T C(2, n + 2)p_1 + \|p_1 - \pi_{R(L)^\perp}y\|^2_2 : \ p_1 \in R(L)^\perp\},$$

and

$$\arg\min\{\|p_1 - \pi_{R(L)}y\|^2 : \ p_1 \in R(L)\} = \pi_{R(L)}y.$$

Hence, the component

$$[H(\lambda) - L(L^T L)^{-1}L^T]y = \arg\min\{\lambda p_1^T C(2, n + 2)p_1 + \|p_1 - \pi_{R(L)^\perp}y\|^2_2 : \ p_1 \in R(L)^\perp\}$$

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

10
5 Choice of the smoothing parameter $\lambda$

5.1 Deterministic noise

We have, for any $\lambda > 0$, the estimated model of (24)

$$y = H(\lambda)y + [I - H(\lambda)]y.$$ 

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

The equality $[I - H(\lambda)]y = w$ holds only for $\lambda = 0$, $w = 0$ and $p$ 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

$$\|y - H(\lambda)y\|^2 = \|w\|^2. \tag{29}$$

The following result shows that the equation (29) has a solution only for "small noise".

**Proposition 5.1.** The map $\lambda \in (0, +\infty) \rightarrow \psi(\lambda) = \|y - H(\lambda)y\|^2$ is concave, varies between 0 and $\|y - L_{reg}(y)\|^2$. The equation (29) has a solution if and only if

$$\|y - L_{reg}(y)\|^2 > \|w\|^2. \tag{30}$$

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

Observe that (30) is equivalent to

$$\sum_{i=1}^{n+1} y_i^2 \|e_i - L_{reg}(e_i)\|^2 + 2 \sum_{i<j} y_i y_j \langle e_i - L_{reg}(e_i), e_j - L_{reg}(e_j) \rangle > \|w\|^2, \tag{31}$$

and the smoothing parameter is solution of the equation

$$\sum_{i=1}^{n+1} y_i^2 \|e_i - H_i(\lambda)\|^2 + 2 \sum_{i<j} y_i y_j \langle e_i - H_i(\lambda), e_j - H_j(\lambda) \rangle = \|w\|^2. \tag{32}$$

It follows that the size of the weights $\|e_i - L_{reg}(e_i)\|^2$, $\langle e_i - L_{reg}(e_i), e_j - L_{reg}(e_j) \rangle$, $\|e_i - H_i(\lambda)\|^2$, $\langle e_i - H_i(\lambda), e_j - H_j(\lambda) \rangle$, are crucial in the existence of the smoothing parameter (29).

In Figure 4 we plot for $i = 1, \ldots, 8$ the graph of $\lambda \rightarrow \|e_i - H_i(\lambda)\|^2$ for $n = 7$ and $t_i = i - 1$.

Remark that for the model $y = y_i e_i$ the $i$th column $H_i(\lambda)$ is an estimator of the signal. Thanks to the equation (32), the quantity $\|e_i - H_i(\lambda)\|^2$ represents the noise-to-signal ratio (NSR), i.e. the smoothing parameter is solution of

$$\|e_i - H_i(\lambda)\|^2 = \frac{\|w\|^2}{y_i^2}. \tag{33}$$
For large noise, there is no smoothing parameter solution of (33).

In Figure 5 we plot the “rainbow” $H_i(\lambda)$ for $i = 1, \ldots, n + 1$ and $\lambda \in \{0.1, 0.5, 1\}$.

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

$$Lreg(e_i)(t) = \beta_1^i + t\beta_2^i,$$

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 ($t \to Lreg(e_i)(t) : i = 1, \ldots, n+1$) have the common point $(\bar{t}, \frac{1}{n+1})$. Moreover we have, for $i \neq j$, that

$$\|e_i - Lreg(e_i)\|^2 = 1 - Lreg(e_i)(t_i),$$

$$\langle e_i - Lreg(e_i), e_j - Lreg(e_j) \rangle = -Lreg(e_i)(t_j),$$

$$= -Lreg(e_j)(t_i).$$

From all that we get the following result.

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

$$\|e_i - Lreg(e_i)\|^2 = \frac{n}{n+1} - \frac{(t_i - \bar{t})^2}{\sum_{j=1}^{n+1} (t_j - \bar{t})^2},$$

$$\langle e_i - Lreg(e_i), e_j - Lreg(e_j) \rangle = -\frac{1}{n+1} \frac{(t_i - \bar{t})(t_j - \bar{t})}{\sum_{j=1}^{n+1} (t_j - \bar{t})^2}.$$

It follows that the most important weights $\|e_i - Lreg(e_i)\|^2$ are when the $t_i$’s are close to $\bar{t}$. The most important negative correlation $\langle e_i - Lreg(e_i), e_j - Lreg(e_j) \rangle$ is given by the couple of end-points $(t_1, t_{n+1})$. The most important positive correlations $\langle e_i - Lreg(e_i), e_j - Lreg(e_j) \rangle$ 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 5 shows that the straightlines ($t \to Lreg(e_j)(t) : j = 1, \ldots, n+1$) turn in the trigonometric sense around their common point. Remark that the Figure 5 illustrates also the Proposition 5.2. convergence of $\|e_i - Lreg(e_i)\|^2$ as $\lambda \to +\infty$.

What can we do if the condition ([30]) does not hold ? In this case for all $\lambda \geq 0$, $\|y - H(\lambda) y\|^2$ represents only a part of the noise i.e.

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

**5.2 Gaussian white noise**

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

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

13
holds with the probability

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

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 $\lambda$ solution of the following constraint

$$(n + 1)(1 - \varepsilon)\sigma^2 \leq \|y - H(\lambda)y\|^2 \leq (n + 1)(1 + \varepsilon)\sigma^2.$$ \hspace{1cm} (34)

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

$$\|y - H(\lambda)y\|^2 = (n + 1)(1 - \varepsilon)\sigma^2,$$ \hspace{1cm} (35)

and

$$\|y - H(\lambda)y\|^2 = (n + 1)(1 + \varepsilon)\sigma^2.$$ \hspace{1cm} (36)

The solution of (35) exists under the hypothesis

$$(1 - \varepsilon)\sigma^2 < \frac{\|y - L\text{regy}\|^2}{n + 1}.$$ \hspace{1cm} (37)

The solution of (36) exists under the hypothesis

$$(1 + \varepsilon)\sigma^2 < \frac{\|y - L\text{regy}\|^2}{n + 1}.$$ \hspace{1cm} (38)
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 (37) and (38), we are going to study the quantity \( \| y - \text{Lreg} y \|_2 \) as a function of the signal \( p \) and the noise \( w \).

If the model is \( y_i = a + b t_i + w_i \) with \( i = 1, \ldots, n + 1 \), then \( \text{Lreg} y = \mathbf{H}(+\infty) y \) is the maximum likelihood estimator of the vector \( \mathbf{L}(a, b)^T \). Moreover, \( \frac{\| y - \text{Lreg} y \|^2}{n - 1} \) is an unbiased consistent estimator of the variance \( \sigma^2 \). More precisely, \( \frac{\| y - \text{Lreg} y \|^2}{n - 1} \) has the \( \chi^2_{n - 1} \)-distribution.

Hence, the constraint (38) holds with the probability \( P(\chi^2_{n - 1} > (n + 1)(1 + \varepsilon)) \to 0 \) as \( n \to +\infty \).

But for \( \varepsilon > \frac{2}{n + 1} \), the constraint (37) holds with the probability \( P(\chi^2_{n - 1} > (n - 1)(1 - \varepsilon)) \to 1 \) as \( n \to +\infty \).

In the general case we have the following result.

**Proposition 5.3.** Let \( y = p + w \), where \( w \) is the Gaussian white noise with the variance \( \sigma^2_w \). We have

\[
\mathbb{E}[\| y - \text{Lreg} y \|^2] = (n - 1)\sigma^2_w + \| p - \text{Lreg} p \|^2.
\]

If the noise is fixed, then \( p \to \mathbb{E}[\| y - \text{Lreg} y \|^2] \) is minimal at the straightlines, i.e. \( p_i = a + b t_i \) for all \( i = 1, \ldots, n + 1 \). The minimal value is equal to

\[
\mathbb{E}[\| y - \text{Lreg} y \|^2] = (n - 1)\sigma^2.
\]  \( \text{(39)} \)

**Proof.** From the equality \( y = p + w \), we have

\[
\| y - \text{Lreg} y \|^2 = \| p - \text{Lreg} p \|^2 + \| w - \text{Lreg} w \|^2 + 2\langle p - \text{Lreg} p, w - \text{Lreg} w \rangle.
\]

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

Roughly speaking, Proposition 5.3 combined with (37) and (38) tell us that the smoothing parameter

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

exists under the constraint

\[
\| p - \text{Lreg} p \|^2 \approx 2\sigma^2.
\]

### 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 \( p \) by \( \mathbf{H}(\lambda) y \) is equal to

\[
\| \mathbf{H}(\lambda) y - s \|^2 = \sum_{i=1}^{n+1} |\mathbf{H}(\lambda) y(i) - s(t_i)|^2,
\]
and the residual sum of squares is defined by

\[ \text{RSS}(\lambda) := \| y - H(\lambda)y \|^2. \]

The mean square risk is equal to

\[
R(H(\lambda)y, p) = \mathbb{E}[\| H(\lambda)y - p \|^2] \\
= \mathbb{E}[\| y - p \|^2] + \mathbb{E}[\text{RSS}(\lambda)] - 2\text{cov}(w - H(\lambda)w, w) \\
= \mathbb{E}[\| y - p \|^2] + \mathbb{E}[\text{RSS}(\lambda) + 2\sigma^2(\text{Trace}(H(\lambda)) - (n + 1))] \\
= \mathbb{E}[\text{RSS}(\lambda) + 2\sigma^2\text{Trace}(H(\lambda)) - (n + 1)\sigma^2].
\]

The quantity

\[ \text{RSS}(\lambda) + 2\sigma^2\text{Trace}(H(\lambda)) \]

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 \( \lambda_{\text{SURE}} \).

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

\[
\mathbb{E}[\| y^* - H(\lambda)y \|^2] = \mathbb{E}[\| p - H(\lambda)y \|^2] + (n + 1)\sigma^2 \\
= R(H(\lambda)y, p) + (n + 1)\sigma^2 \\
= \mathbb{E}[\text{RSS}(\lambda) + 2\sigma^2\text{Trace}(H(\lambda))].
\]

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

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

6 Cubic spline estimate: General case

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

\[
(\hat{u}_1, \hat{p}_1, \ldots, \hat{p}_{n+1}, \hat{u}_{n+1}) = \arg \min \{ \lambda(u_1, p^T, u_{n+1})P_{\text{pen}} \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} + \| y - p \|^2 : u_1, p_1, \ldots, p_{n+1}, u_{n+1} \}
\]

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 [10] to rediscover a non natural spline.
Figure 7: Plot of $\lambda \in (0, +\infty) \rightarrow \text{Trace}(H(\lambda))$.

**Proposition 6.1.** 1) The minimizer of (40) 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

$$
\begin{pmatrix}
    \hat{u}_1 \\
    \hat{p} \\
    \hat{u}_{n+1}
\end{pmatrix} = (\lambda P_{\text{pen}} + \Pi^T \Pi)^{-1} \begin{pmatrix}
    0 \\
    y \\
    0
\end{pmatrix}
$$

(41)

$$
: = H_{\text{pen}}(\lambda) \begin{pmatrix}
    0 \\
    y \\
    0
\end{pmatrix}
$$

(42)

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

17
2) The condition
\[(p_{1,j} : j = 2, \ldots, n + 2) \neq 0, \quad (43)\]
respectively
\[(p_{n+3,j} : j = 2, \ldots, n + 2) \neq 0, \quad (44)\]
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 (41) as \( \lambda \to 0 \) and \( \lambda \to +\infty \).

**Corollary 6.2.** 1) The limit
\[
H_{P_{pen}}(\lambda) \begin{pmatrix}
0 \\
y \\
0
\end{pmatrix} \to \begin{pmatrix}
u_1^0 \\
y \\
u_{n+1}^0
\end{pmatrix},
\]
as \( \lambda \to 0 \). Here \( u_1^0, u_{n+1}^0 \) is a minimizer of the objective function
\[(u_1, u_{n+1}) \to (u_1, y^T, u_{n+1})P_{pen} \begin{pmatrix}
u_1 \\
y \\
u_{n+1}
\end{pmatrix},\]
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 \( y \) is not orthogonal to the space spanned by the vectors \( (p_{1,j} : j = 2, \ldots, n + 2)^T, (p_{n+3,j} : j = 2, \ldots, n + 2)^T, \) 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
\[
H_{P_{pen}}(\lambda) \begin{pmatrix}
0 \\
y \\
0
\end{pmatrix} \to \arg\min\{\|y - p\|^2 : P_{pen} \begin{pmatrix}
u_1 \\
p \\
u_{n+1}
\end{pmatrix} = 0\}, \quad (46)
\]
as \( \lambda \to +\infty \).

**Examples.** The matrix \( P_{pen} = C \) given in (16) corresponds to the penalty
\[(u_1, p^T, u_{n+1})^T C(u_1, p^T, u_{n+1})^T = \int_{t_1}^{t_{n+1}} |s''(t)|^2 dt\]
where the cubic spline \( s = u_1 \varphi_0 + \sum_{i=1}^{n+1} p_i \varphi_i + u_{n+1} \varphi_{n+2} \). A natural way to construct new matrices \( \mathbf{P}_{\text{pen}} \) is to consider the more general penalization

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

where \( a_0, a_1, a_2 \) are given real numbers. Hence

\[
\mathbf{P}_{\text{pen}} = a_0^2 \mathbf{C}_{00} + a_1^2 \mathbf{C}_{11} + a_2^2 \mathbf{C}_{22} + a_0 a_1 [\mathbf{C}_{01} + \mathbf{C}_{10}^T] + a_0 a_2 [\mathbf{C}_{02} + \mathbf{C}_{20}^T] + a_1 a_2 [\mathbf{C}_{12} + \mathbf{C}_{21}^T],
\]

where \( \mathbf{C}_{00}, \mathbf{C}_{01}, \mathbf{C}_{02}, \mathbf{C}_{11}, \mathbf{C}_{12}, \mathbf{C}_{22} \) are respectively defined by

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

and the matrix \( \mathbf{C}_{22} = \mathbf{C} \) defined in (43).

The matrix \( \mathbf{P}_{\text{pen}} \), 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.

\[
\mathbf{P}_{\text{pen}} = \begin{pmatrix}
2.6 & -239.22 & 614.68 & -559.93 & 255.91 & -96.85 & 33.253 & -9.951 & 1.841 & -0.01 \\
-239.22 & 398.834 & -895.24 & 9892.35 & -57644.4 & 23760.7 & -8518.95 & 2611.28 & -488.75 & 1.84 \\
614.68 & -895.24 & 27842.5 & -345243.9 & 388016.7 & -113691.9 & 43414.47 & -13742.88 & 2611.71 & -9.94 \\
-559.93 & 9892.35 & -345243.9 & 516426.7 & -459007.5 & 281456.3 & -127439.9 & 43417.03 & -8521.09 & 33.24 \\
255.91 & -57644.4 & 238016.7 & -459007.5 & 559860.4 & -472755.5 & 281452.9 & -113702.2 & 23768.8 & -96.81 \\
-96.85 & 23760.7 & -113691.9 & 281456.3 & -472755.5 & 559860.4 & -459017.7 & 228055.1 & -345387.2 & 23768.8 \\
-9.951 & 2611.28 & -13742.88 & 43417.03 & -113702.2 & 228055.1 & -345387.2 & 27677.6 & -8946.04 & 612.72 \\
1.841 & -488.75 & 2611.71 & -8521.09 & 23768.8 & -57664.7 & 99036.12 & -8946.04 & 3129.29 & -236.99 \\
-0.01 & 1.84 & -9.95 & 33.24 & -96.81 & 255.76 & -559.40 & 612.72 & -236.99 & 3.85
\end{pmatrix}
\]

Observe that the conditions (43) and (44) are satisfied.

### 7 Bayesian Model and statistical analysis

The aim of this section is to give the Bayesian interpretation of the matrix penalization \( \mathbf{P}_{\text{pen}} \).

Let us first set the noisy cubic spline estimate in the context of the general linear model:

\[
y = \mathbf{F} \beta + \mathbf{R} \eta + \mathbf{w},
\]

where \( \beta \) is an unknown parameters, \( \mathbf{F} \) and \( \mathbf{R} \) are known matrices. The random effects \( \eta \) and the noise \( \mathbf{w} \) are unknown, centred and independent random vectors. Their covariance matrices
\( \text{cov}(\eta) := \Sigma_{\eta}, \text{cov}(w) := \Sigma_{w} \) are known. The term \( F \beta \) is called the fixed effects and \( R \eta \) 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], [10], [11], [12], [15], [16], [20] and references herein.

**BLUE of \( \beta \).** The BLUE of \( \beta \) is the estimator \( \hat{\beta} = \hat{M}_{\beta} y \), with \( \hat{M}_{\beta} \) (called the hat matrix of \( \hat{\beta} \)) being the matrix such that \( \hat{M}_{\beta} F = I \) (the identity matrix) and \( \text{cov}(My) - \text{cov}(\hat{M}_{\beta} y) \) is positive semi-definite for all matrix \( M \) subject to \( MF = 0 \).

**BLUP of \( \eta \).** The BLUP of \( \eta \) is the estimator \( \hat{\eta} = \hat{M}_{\eta} y \), with \( \hat{M}_{\eta} \) (called the hat matrix of \( \hat{\eta} \)) being the matrix such that \( \hat{M}_{\eta} F = 0 \) and \( \text{cov}(My) - \text{cov}(\hat{M}_{\eta} y) \) is positive semi-definite for all matrix \( M \) subject to \( MF = 0 \).

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

\[
\begin{align*}
\hat{\beta} & = (F^T(\Sigma_{\eta}R + \Sigma_{w})^{-1}F)^{-1}F^T(\Sigma_{\eta}R + \Sigma_{w})^{-1}y, \\
\hat{\eta} & = (R^T\Sigma_{\eta}^{-1}R + \Sigma_{w}^{-1})^{-1}R^T\Sigma_{\eta}^{-1}F(F^T(\Sigma_{w} + R\Sigma_{\eta}R^T)^{-1}F)^{-1}F^T(\Sigma_{w} + R\Sigma_{\eta}R^T)^{-1}y.
\end{align*}
\]

Now we are able to give a Bayesian interpretation of the hat matrix \( H_{P_{\text{pen}}}(\lambda) \) (41). Let \( P_1 \) be an \( n+3 \) by \( n+3 - \dim(N(P_{\text{pen}})) \) matrix such that

\[
P_1^TP_{\text{pen}}P_1 = I_{(n+3-\dim(N(P_{\text{pen}})))\times(n+3-\dim(N(P_{\text{pen}})))},
\]

respectively \( P_0 \) an \( n+3 \) by \( \dim(N(P_{\text{pen}})) \) matrix such that

\[
P_{\text{pen}}P_0 = 0, \quad \text{and its columns form a basis of } N(P_{\text{pen}}).
\]

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

\[
\begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix} = P_0\beta + P_1\eta.
\]

Hence the model \( y = p + w \) becomes

\[
y = \Pi P_0\beta + \Pi P_1\eta + w := F\beta + R\eta + w.
\]

We suppose that \( \beta \) is the fixed effect and \( \eta \) is independent of the noise \( w \) and drawn from a centred distribution having the covariance matrix \( \sigma_{\eta}^2 I_{n+3-\dim(N(P))} \).

Now, we are able to give our Bayesian interpretation.
Proposition 7.1. The components $(\hat{\beta}, \hat{\eta})$ of
\[
\arg\min\left\{ \frac{\sigma_w^2}{\sigma^2} \|\eta\|^2 + \|y - (F\beta + R\eta)\|^2 : \beta \in \mathbb{R}^{\dim(N(P_{\text{pen}}))}, \eta \in \mathbb{R}^{n+3-\dim(N(P_{\text{pen}}))} \right\}
\]
are respectively the BLUE of $\beta$ and the BLUP of $\eta$. Moreover, we have
\[
F\hat{\beta} + R\hat{\eta} = \Pi H_{P_{\text{pen}}} \left( \frac{\sigma_w^2}{\sigma^2} \right) \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix},
\]
\[
P_0\hat{\beta} + P_1\hat{\eta} = H_{P_{\text{pen}}} \left( \frac{\sigma_w^2}{\sigma^2} \right) \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix}.
\]

The proof is a consequence of the change of variable formula (52) and (49). See [3] Proposition 2.2 for a similar proof.

Corollary 7.2. Let $P = C$ be the matrix (10) and $P_0, P_1$ be the corresponding matrices defined by (51), (50). We have
\[
P_0\hat{\beta} = \begin{pmatrix} 0 \\ Lregy \\ 0 \end{pmatrix},
\]
\[
P_1\hat{\eta} = \begin{pmatrix} 0 \\ (H(\frac{\sigma_\nu^2}{\sigma^2}) - Lreg)y \\ 0 \end{pmatrix}.
\]

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.

8 References

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), pp. 1-26.
[3] A. Dermoune, N. Rahmania, T. Wei: General Linear mixed model and signal extraction problem with constraint, Journal of Multivariate Analysis, 2012, Elsevier, Vol. 105 (1), pp. 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, vol. 13, No 3, (2009).

[5] A. Dermoune, B. Djehiche, N. Rahmania: A consistent estimator of the smoothing parameter in the Hodrick-Prescott filter J. Japan Statist. Soc., Vol. 38 (No. 2), pp. 225-241, (2008).

[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 16, (1985) 71-84.

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

[9] C.R. Henderson: Estimation of genetic parameters, Ann. Math. Stat 21, (1950) 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, Biometrics 16, (1959) 192-218.

[12] C.R. Henderson: Best linear unbiased estimation and prediction under a selection model, Biometrics 31, (1975) 423-447.

[13] N. Krämer, A. L. Boulesteix, G. Tutz: Penalized Partial Least Squares Based on B-Splines Transformations, Chemometrics and Intelligent Laboratory Systems, Vol. 94, Issue 1, 2008, pp. 60-69.

[14] R. Paige, A. A. Trindade: The Hodrick-Prescott Filter: A special case of penalized spline smoothing, Electronic Journal of Statistics, Vol. 4 (2010) pp. 856-874.

[15] C.R. Rao: Representation of the best linear unbiased estimators in the Gauss-Markoff model with a singular dispersion matrix, Journal of Multivariate analysis 3, (1973) 276-292.

[16] G.K. Robinson: That BLUP is a good thing: The estimation of random effects, Statistical Science 6(1), (1991) 15-51.

[17] Micula G, Micula S (1999), Handbook of splines, Kluwer Academic Publishers, Dordrecht, the Netherlands.
Appendix 1: Construction of the matrices $Q$, $U$, $V$

From (6) we can solve for $v_1, \ldots, v_n$ in terms of $u_1, \ldots, u_{n+1}$, i.e.

$$v_i = \frac{u_{i+1} - u_i}{h_i}, \quad i = 1, \ldots, n.$$  \hfill (53)

Now, using (5) and (4) we can solve $q$ in terms of $p, u$. We get

$$q = Q_1 p + Q_2 u,$$

where the $n$ by $n+1$ matrices

$$
Q_1 = [-\frac{1}{h_i} e_i^T + \frac{1}{h_i} e_{i+1}^T, \quad \quad i = 1, \ldots, n],
$$

$$
Q_2 = [-\frac{h_i}{3} e_i^T - \frac{h_i}{6} e_{i+1}^T, \quad \quad i = 1, \ldots, n].
$$

Here the column vectors $(e_i : i = 1, \ldots, n + 1)$ denote the canonical basis of $\mathbb{R}^{n+1}$. If we plug $q, v = (v_1, \ldots, v_n)^T$ in the continuity equations (4), then we get

$$Su = \Delta p,$$  \hfill (54)

where the $n-1$ by $n+1$ matrix

$$
\Delta = \left[ e_i^T \left( \frac{1}{h_i} + \frac{1}{h_{i+1}} \right) e_{i+1}^T + \frac{1}{h_{i+1}} e_{i+2}^T, \quad i = 1, \ldots, n - 1 \right].
$$

The $n-1$ by $n+1$ matrix

$$
S = \left[ \frac{h_i}{6} e_i^T + \frac{h_i + h_{i+1}}{3} e_{i+1}^T + \frac{h_{i+1}}{6} e_{i+2}^T, \quad \quad i = 1, \ldots, n - 1 \right].$$  \hfill (55)
Observe that the \( n + 3 \) by \( n + 3 \) matrix
\[
S_{1,n+1} := \begin{pmatrix}
e^T_1 \\
S \\
e^T_{n+1}
\end{pmatrix}
\]
is invertible. Hence, we can solve for \( u \) in terms of \((u_1, p, u_{n+1})\) as follows:
\[
u = U \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix}
\]
where the \( n + 1 \) by \( n + 3 \) matrix
\[
U = S_{1,n+1}^{-1} \begin{pmatrix}
e^T_1 \\ 0 & \Delta & 0 \\ e^T_{n+1}
\end{pmatrix}
\]
The matrix \( U \) tells us, for all \( i = 1, \ldots, n + 1 \), that the second derivative \( u_i \) is a linear combination of \( u_1, u_{n+1} \) and \( p \) with the weight \( U = [u_{i,j}] \), i.e.
\[
u_i = u_1 u_{i,1} + \sum_{j=1}^{n+1} p_j u_{i,j+1} + u_{n+1} u_{i,n+3}.
\]
The coefficient \( u(i, 1) \) is the weight of \( u_1 \), \((u_{i,j+1}, j = 1, \ldots, n + 1)\) are the weight of the observations \((p_j, j = 1, \ldots, n + 1)\) respectively. The equality \((58)\) tells us that the initial and the terminal second derivatives do not depend on the observations \( p \). We can show numerically and we will prove it rigorously (Proposition \( \text{(58)} \)) that the mean weight of the observalls on each second derivative is equal to zero, i.e.,
\[
\sum_{j=1}^{n+1} u_{i,j+1} = 0, \quad \forall i = 1, \ldots, n + 1.
\]

Now, we come back to the first and the third derivatives \( q, v \). We can solve for \( q \) and \( v \) in terms of \((u_1, p, u_{n+1})\), i.e.
\[
q = Q \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix},
\]
with
\[
Q = \begin{pmatrix} 0 & Q_1 & 0 \end{pmatrix} + Q_2 U.
\]
Similarly,

\[ v = V \begin{pmatrix} u_1 \\ p \\ u_{n+1} \end{pmatrix}, \]  

(60)

with the \( n \) by \( n + 3 \) matrix

\[ V = \tilde{V}U, \]

where the \( n \) by \( n + 1 \) matrix

\[ \tilde{V} = [-\frac{1}{h_i}e_i^T + \frac{1}{h_i}e_{i+1}^T, \quad i = 1, \ldots, n]. \]

The matrices \( Q, U, V \) satisfy the equality

\[ Q + \text{diag}(\frac{h}{2})U + \text{diag}(\frac{h^2}{6})V = [0 \quad D \quad 0], \]  

(61)

where the \( n \) by \( n + 1 \) matrix

\[ D = [\frac{1}{h_i}(e_{i+1}^T - e_i^T), \quad i = 1, \ldots, n]. \]

**Appendix 2: Analysis and interpretation of the matrices \( Q, U, V \)**

How to interpret the columns?

- The derivatives up to order 3, for \( j = 0, \ldots, n+2 \), of \( \varphi_j \) at the knots \( t_1, \ldots, t_{n+1} \) are respectively the \((j+1)\)th columns of \( P, Q, U, V \). Here \( P \) is the \( n+1 \) by \( n+3 \) matrix

\[ P = [\varphi_j(t_i) : i = 1, \ldots, n+1, j = 0, \ldots, n+2]. \]

How to interpret the rows?

- The \( i \)th row of the matrices \( P, Q, U, V \) represents respectively the row \((\varphi_j(t_i) : j = 0, \ldots, n+2), (\varphi'_j(t_i) : j = 0, \ldots, n+2), (\varphi''_j(t_i) : j = 0, \ldots, n+2)\) and \((\varphi'''_j(t_i+) : j = 0, \ldots, n+2)\).

How to interpret the basis elements?

- The natural cubic spline interpolating \((t_i, 0), i = 1, \ldots, n+1\) is the null map denoted by \( s_0 \). The natural cubic spline interpolating \((t_j, 1), (t_i, 0), i \neq j\) is equal to \( \varphi_j \). It is the unique \( C^2 \)-cubic spline such that, for \( i = 1, \ldots, n+1 \),

\[ s(t_i) = \delta_i^j, \quad s'(t_i) = a_{i,j+1}, \quad s''(t_i) = u_{i,j+1}, \quad s'''(t_i+) = v_{i,j+1}. \]

It can be seen as the perturbation of the null map having the most important value at the knot \( t_j \) (see Figure 1). A perturbation of 1 at the knot \( t_j \), i.e. \( p_j \rightarrow p_j + 1 \), produces for each
We recall, for \( i \in \{ t_1, \ldots, t_{n+1} \} \) a perturbation \( \varphi_j(t) \in (-1, 1) \).

- The \( C^2 \)-cubic spline \( s \) interpolating \((t_i, 0), i = 1, \ldots, n+1 \) and such that \( s''(t_1) = 1 := u_{1,1} \) is equal to \( \varphi_0 \). It is the unique \( C^2 \)-cubic spline such that, for \( i = 1, \ldots, n+1 \),

\[
\begin{align*}
  s(t_i) &= 0, \quad s'(t_i) = q_{i,1}, \\
  s''(t_i) &= u_{i,1}, \quad s'''(t_i) = v_{i,1}.
\end{align*}
\]

It can be seen as the perturbation of the null map having the most important second derivative at the knot \( t_1 \) (see Figure 3). A perturbation of 1 on the second derivative at the knot \( t_1 \) produces for each \( t \in (t_1, t_{n+1}] \) a perturbation \( \varphi_0''(t) \in (-1, 1) \).

- The \( C^2 \)-cubic spline \( s \) interpolating \((t_i, 0), i = 1, \ldots, n+1 \) and such that \( s''(t_{n+1}) = 1 := u_{1,n+1} \) is equal to \( \varphi_{n+2} \). It is the unique \( C^2 \)-cubic spline such that, for \( i = 1, \ldots, n+1 \),

\[
\begin{align*}
  s(t_i) &= 0, \quad s'(t_i) = q_{i,n+2}, \\
  s''(t_i) &= u_{i,n+2}, \quad s'''(t_i) = v_{i,n+2}.
\end{align*}
\]

It can be seen as the perturbation of the null map having the most important second derivative at the knot \( t_{n+1} \) (see Figure 3). A perturbation of 1 on the second derivative at the knot \( t_1 \) produces for each \( t \in [t_1, t_{n+1}] \) a perturbation \( \varphi_{n+2}''(t) \in (-1, 1) \).

### Appendix 3: Calculus of the matrices \( C_{ij}, i, j = 0, \ldots, 2 \)

We recall, for \( i = 1, \ldots, n, \) \( l = 0, \ldots, n+2 \) and \( t \in (t_i, t_{i+1}) \), that

\[
\varphi_l(t) = \delta_l(i) + q_{i,l+1}(t - t_i) + \frac{u_{i,l+1}}{2}(t - t_i)^2 + \frac{v_{i,l+1}}{6}(t - t_i)^3,
\]

\[
\varphi'_l(t) = q_{i,l+1} + u_{i,l+1}(t - t_i) + \frac{v_{i,l+1}}{2}(t - t_i)^2,
\]

\[
\varphi''_l(t) = u_{i,l+1} + v_{i,l+1}(t - t_i).
\]

#### Calculus of \( C_{00} \). We have, for \( l, k = 0, \ldots, n + 2 \), that

\[
\int_{t_1}^{t_{n+1}} \varphi_l(t) \varphi_k(t) dt = \sum_{i=1}^{n} \int_{0}^{1} \varphi_l(t,i) \varphi_k(t,i) dt,
\]

where for \( i = 1, \ldots, n \),

\[
\varphi_l(t,i) = \delta_l(i) + q_{i,l+1}t + \frac{u_{i,l+1}}{2}t^2 + \frac{v_{i,l+1}}{6}t^3
\]

\[
:= \sum_{p=0}^{3} a_p(l,i)t^p.
\]

Hence

\[
\varphi_l(t,i) \varphi_k(t,i) = \sum_{p=0}^{6} a_p(l,k,i)t^p,
\]

26
where
\[ a_p(l, k, i) = \sum_{p_1+p_2=p} a_{p_1}(l, i) a_{p_2}(k, i). \]

It follows that
\[ \int_0^1 \varphi_l(t, i) \varphi_k(t, i) dt = \sum_{p=0}^{6} \frac{a_p(l, k, i)}{p+1}, \]
and the matrix
\[ C_{00} = \left[ \sum_{i=1}^n h_i \frac{a_p(l-1, k-1, i)}{p+1} \right] : \quad l, k = 1, \ldots, n + 3. \]

**Calculus of C_{01}**. We have, for \( l, k = 0, \ldots, n + 2 \), that
\[ \int_{t_1}^{t_{n+1}} \varphi_l(t) \varphi'_k(t) dt = \sum_{i=1}^n h_i \int_0^1 \varphi_l(t, i) \varphi'_k(t, i) dt, \]
where, for \( i = 1, \ldots, n \),
\[ \varphi'_k(t, i) = \sum_{p=1}^3 p a_p(k, i) t^{p-1} \]
\[ := \sum_{p=1}^3 a_p^{(1)}(k, i) t^{p-1}. \]
Hence
\[ \varphi_l(t, i) \varphi'_k(t, i) = \sum_{p=1}^6 a_p^{(01)}(l, k, i) t^{p-1}, \]
where
\[ a_p^{(01)}(l, k, i) = \sum_{p_1+p_2=p} a_{p_1}(l, i) a_{p_2}^{(1)}(k, i). \]

It follows that
\[ \int_0^1 \varphi_l(t, i) \varphi'_k(t, i) dt = \sum_{p=1}^6 \frac{a_p^{(01)}(l, k, i)}{p} \]
and the matrix
\[ C_{01} = \left[ \sum_{i=1}^n \sum_{p=1}^6 h_i \frac{a_p^{(01)}(l-1, k-1, i)}{p} \right] : \quad l, k = 1, \ldots, n + 3. \]
Calculus of $C_{02}$. We have, for $l, k = 0, \ldots, n + 2$, that
\[
\int_{t_1}^{t_{n+1}} \varphi_l(t) \varphi_k'(t) dt = \sum_{i=1}^{n} h_i \int_{0}^{1} \varphi_l(t,i) \varphi_k'(t,i) dt,
\]
where, for $i = 1, \ldots, n$,
\[
\varphi_k''(t,i) = \sum_{p=2}^{3} p(p-1)a_p(k,i)t^{p-2} = \sum_{p=2}^{3} a_p^{(2)}(k,i)t^{p-2}.
\]
Hence
\[
\varphi_l(t,i)\varphi_k''(t,i) := \sum_{p=2}^{3} a_p^{(2)}(l,k,i)t^{p-2},
\]
where
\[
a_p^{(02)}(l,k,i) = \sum_{p_1+p_2=p} a_{p_1}(l,i)a_{p_2}(k,i).
\]
It follows that
\[
\int_{0}^{1} \varphi_l(t,i)\varphi_k''(t,i) dt = \sum_{p=2}^{3} \frac{a_p^{(02)}(l,k,i)}{p-1},
\]
and
\[
C_{02} = \left[ \sum_{i=1}^{n} \sum_{p=2}^{6} h_i a_p^{(02)}(l-1,k-1,i) \right]: l, k = 1, \ldots, n + 3.
\]

Calculus of $C_{11}$. We have, for $l, k = 0, \ldots, n + 2$, that
\[
\varphi_l'(t,i)\varphi_k'(t,i) = \sum_{p_1+p_2=1}^{3} a_{p_1}^{(1)}(l,i)a_{p_2}^{(1)}(k,i)t^{p_1+p_2-2} = \sum_{p=2}^{6} a_p^{(11)}(l,k,i)t^{p-2},
\]
where
\[
a_p^{(11)}(l,k,i) = \sum_{p_1+p_2=p} a_{p_1}^{(1)}(l,i)a_{p_2}^{(1)}(k,i).
\]
Hence,

\[
\int_{t_1}^{t_{n+1}} \varphi'_l(t) \varphi'_k(t) dt = \sum_{i=1}^{n} h_i \int_{0}^{1} \varphi'_l(t,i) \varphi'_k(t,i) dt
\]

\[
= \sum_{i=1}^{n} h_i \frac{6 \sum_{p=2}^{6} a_p^{(11)} (l, k, i)}{p - 1},
\]

and

\[
C_{11} = \left[ \sum_{i=1}^{n} h_i \frac{6 \sum_{p=2}^{6} a_p^{(11)} (l - 1, k - 1, i)}{p - 1} \right] : l, k = 1, \ldots, n+3.
\]

**Calculus of C_{12}**. We have, for \( l, k = 0, \ldots, n + 2 \), that

\[
\varphi'_l(t,i) \varphi''_k(t,i) = \sum_{p_1=1}^{3} \sum_{p_2=2}^{3} a_{p_1}^{(1)} (l, i) t^{p_1-1} a_{p_2}^{(2)} (k, i) t^{p_2-2}
\]

\[
= \sum_{p=3}^{6} a_p^{(12)} (l, k, i) t^{p-3},
\]

where

\[
a_p^{(12)} (l, k, i) = \sum_{p_1+p_2=p, p_1=1,2,3, p_2=2,3} a_{p_1}^{(1)} (l, i) a_{p_2}^{(2)} (k, i).
\]

Hence,

\[
\int_{t_1}^{t_{n+1}} \varphi'_l(t) \varphi''_k(t) dt = \sum_{i=1}^{n} h_i \int_{0}^{1} \varphi'_l(t,i) \varphi''_k(t,i) dt
\]

\[
= \sum_{i=1}^{n} \frac{6 \sum_{p=3}^{6} h_i a_p^{(12)} (l, k, i)}{p - 2},
\]

and

\[
C_{12} = \left[ \sum_{i=1}^{n} h_i \frac{6 \sum_{p=3}^{6} a_p^{(12)} (l - 1, k - 1, i)}{p - 2} \right] : l, k = 1, \ldots, n+3.
\]