Strong noise estimation in cubic splines

A. Dermoune\textsuperscript{a}, A. El Kaabouchi\textsuperscript{b}

\textsuperscript{a}Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524. UFR de Mathématiques, Bât. M2, 59655 Villeneuve d’Ascq Cedex, France
\textsuperscript{b}Institut Supérieur des Matériaux et Mécaniques Avancés, 44, Avenue Bartholdi, 72000, Le Mans, France

Abstract

The data \((y_i, x_i) \in \mathbb{R} \times [a, b], i = 1, \ldots, n\) satisfy \(y_i = s(x_i) + e_i\) where \(s\) belongs to the set of cubic splines. The unknown noises \((e_i)\) are such that \(\text{var}(e_I) = 1\) for some \(I \in \{1, \ldots, n\}\) and \(\text{var}(e_i) = \sigma^2\) for \(i \neq I\). We suppose that the most important noise is \(e_I\), i.e. the ratio \(r_I = \frac{1}{\sigma^2}\) is larger than one. If the ratio \(r_I\) is large, then we show, for all smoothing parameter, that the penalized least squares estimator of the \(B\)-spline basis recovers exactly the position \(I\) and the sign of the most important noise \(e_I\).

Key words: B-spline functions, Cubic splines, hat matrix, Moore-Penrose pseudoinverse.

1991 MSC: 2000, 15A09, 41A15, 62J05.

1 Linear inverse problem: General setting

The data \((y_i, x_i) \in \mathbb{R} \times [a, b], i = 1, \ldots, n\) satisfy

\[ y_i = s(x_i) + e_i. \tag{1.1} \]

The map \(s : [a, b] \to \mathbb{R}\) is unknown, and \((e_i)\) are the error of measurements, also called the noise and is unknown. We suppose that \(s\) belongs to a set \(\mathcal{C}\) of functions, and we are interested in the estimation of \(s \in \mathcal{C}\) using the data \((y_i, x_i), i = 1, \ldots, n\). Suppose that \(\mathcal{C}\) has a basis \((b_j)_{j=1,\ldots,d}\). In this case, each map \(s \in \mathcal{C}\) is determined by its coordinates \(\beta = (\beta_1 \ldots \beta_d)^T\) in the latter basis,

\textit{Email addresses: Azzouz.Dermoune@math.univ-lille1.fr} (A. Dermoune), \texttt{aek@ismans.fr} (A. El Kaabouchi).
i.e., \( \forall x \in [a, b], \ s(x) = \sum_{j=1}^{d} \beta_j b_j(x) \). Hence, for each \( i \), \( s(x_i) = b(x_i)\beta \) with \( b(x_i) = (b_1(x_i) \ldots b_d(x_i)) \) is the \((1, d)\) matrix. If we introduce the \((n, d)\) matrix

\[
B = \begin{pmatrix} b(x_1) \\ \vdots \\ b(x_n) \end{pmatrix},
\]

(1.2)

then, the data \( y = (y_1 \ldots y_n)^T \) and the noise \( e = (e_1 \ldots e_n)^T \) satisfy the linear system

\[
y = B\beta + e.
\]

(1.3)

The latter is known as the linear regression in Statistic community and the linear inverse problem in the Inverse problem community. This problem is ill-posed, because the transformation \( \beta \mapsto B\beta \) is not invertible. Moreover, the noise \( e \) is not known.

One way to estimate the parameter \( \beta \) and the noise \( e \) is to use the generalized penalized least square estimators. It works as following. We fix a matrix \( M \) having \( n \) columns, and we consider the ellipsoide quasi-norm \( \| \cdot \|_M \) defined by \( \| x \|_M^2 = x^T M^T M x \). We propose, for each \( \lambda > 0 \) and for each matrix \( L \) having \( d \) columns, the minimizers

\[
\hat{\beta}(\lambda, M, L) \in \arg \min_\beta \{ \| y - B\beta \|_M^2 + \lambda \| L\beta \|_2^2 \}
\]

(1.4)

as an estimator of the vector \( \beta \). The quantity \( \| y - B\beta \|_M^2 \) is the square of the residual error with respect to the metric defined by the quasi-norm \( \| \cdot \|_M \), and \( \| L\beta \|_2^2 \) is called the penalty. The parameter \( \lambda \) is called the smoothing parameter. We have easily the following results.

**Proposition 1** The set of the minimizers of the latter optimization is given by the following normal equation

\[
(B^T M^T M B + \lambda L^T L)\hat{\beta} = B^T M^T M y.
\]

(1.5)

If \( N(MB) \cap N(L) = \{0\} \), then \( \hat{\beta}(\lambda, M, L) \) is unique and is given by

\[
\hat{\beta}(\lambda, M, L) = (B^T M^T M B + \lambda L^T L)^{-1} B^T M^T M y := H(\lambda, M, L)y.
\]

(1.6)

Here, \( N(A) \) denotes the null-space of the matrix \( A \).

The minimizer \( H(\lambda, M, L)y \) is proposed as an estimator of the parameter \( \beta \). Hence, \( BH(\lambda, M, L)y \) is an estimator of \( B\beta \) and \( y - BH(\lambda, M, L)y \) is an estimator of the noise \( e \). The map \( x \in [a, b] \mapsto \sum_{j=1}^{d} H(\lambda, M, L)y(j)b_j(x) \) is an
estimator of the map \(s\). The performance of these estimators depends clearly on the matrix \(H(\lambda, M, L)\) known as the hat matrix in Statistics community.

**Proposition 2** The limit \(\lim_{\lambda \to 0^+} H(\lambda, M, L) := H(0, M, L)\) exists and is equal to the ML-weighted pseudoinverse of \(B\) defined by

\[
B_{ML}^+ = (I - (LP_{MB})^+L)(MB)^+M. \tag{1.7}
\]

Here \(A^+\) denotes the Moore-Penrose inverse of \(A\) and

\[
P_A = I - A^+A \tag{1.8}
\]

denotes the orthogonal projection on \(N(A)\).

It follows that the estimator \(\hat{\beta}(\lambda, M, L)\) converges to \(B_{ML}^+y := \hat{\beta}(0, M, L)\) as \(\lambda \to 0\). Moreover, we can show that

\[
\hat{\beta}(0, M, L) = \arg\min_{\beta} \{\|L\beta\|^2 : \beta \in \arg\min \|y - B\beta\|^2 \}. \tag{1.9}
\]

In particular, if \(B\) has maximal rank, i.e., \(R(B) = R^n\), then

\[
\hat{\beta}(0, M, L) = \arg\min_{\beta} \{\|L\beta\|^2 : y = B\beta\}, \tag{1.10}
\]

or equivalently

\[
BB_{ML}^+B = B. \tag{1.11}
\]

Observe that, if \(M\) is invertible, then

\[
B_{ML}^+B = B_{L}^+B. \tag{1.12}
\]

On the other hand, the limit

\[
\lim_{\lambda \to +\infty} (B^TM^TMB + \lambda L^TL)^{-1}B^TM^TM \tag{1.13}
\]

exists and is equal to

\[
C_{MBL} := (MBP_L)^+M. \tag{1.14}
\]

In the classical case \(M = I\) and \(L = I\), the matrix

\[
H(\lambda, I, I) := (B^TB + \lambda I)^{-1}B^T \tag{1.15}
\]

is known as the Tikhonov regularization of the Moore-Penrose inverse of the matrix \(B\).

**PROOF.** It is a consequence of known results\(^2\).

In the sequel, we suppose that the noise \(e\) is Gaussian with the covariance matrix \(C = diag(\sigma_i^2)\). In this case, the natural choice of \(M\) is the weight
matrix $C^{-1/2}$. We suppose that the variance $\sigma_I^2 = 1$ for some $I$ and $\sigma_i^2 = \sigma^2$ for all $i \neq I$. The set of functions $C$ is the set of cubic splines. The true signal is an element of the $B$-spline basis. We consider, for each $\lambda > 0$, the noise estimator $\hat{e}(\lambda) = y - BH(\lambda, M, L)y$. The aim of our work is to show that if $\sigma^2$ is small, then $\hat{e}(\lambda)$ recovers exactly the position and the sign of the most important noise. Section 2 recalls some cubic splines results. In Section 3 we present our numerical results.

2 Cubic splines

Schoenberg introduced in [6] 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 tasks in interpolation and approximation. Various aspect of splines and their applications can be found in [3].

Let $a = \kappa_0 < \kappa_1 < \ldots < \kappa_{K+1} = b$ be a sequence of increasing real numbers. Spline interpolation can be described as following. A map $s$ belongs to the set $S_3(\kappa_0, \ldots, \kappa_{K+1})$ if

$$s(x) = p_i + q_i(x - \kappa_i) + \frac{u_i}{2}(x - \kappa_i)^2 + \frac{v_i}{6}(x - \kappa_i)^3$$

(2.1)

for every $x \in [\kappa_i, \kappa_{i+1})$. Let $(b_j := S_{j,4} : j = -3, \ldots, K)$ be the $B$-spline basis functions of the set $S_3(\kappa_0, \ldots, \kappa_{K+1})$.

Before going further, let us recall the famous result of [6] and [5]. If the data $(y_i, \kappa_i) : i = 0, \ldots, K + 1$, then the minimizer of

$$\min_{s \in S_3(\kappa_0, \ldots, \kappa_{K+1})} \{\lambda \int_a^b |s^{(2)}(x)|^2 dx + \sum_{i=0}^{K+1} |s(\kappa_i) - y_i|^2\}$$

(2.2)

is the natural cubic spline, i.e. such that $s^{(2)}(\kappa_0) = s^{(2)}(\kappa_{K+1}) = 0$, where $s^{(2)}$ is the second derivative of $s$. Observe that the penalized matrix $L$ is defined by

$$\|L\beta\|^2 = \int_a^b |s^{(2)}(x)|^2 dx.$$

(2.3)

Let us calculate the matrix $L$. The unknown vector $\beta$ belongs to $\mathbb{R}^{K+4}$. From the derivative formula for $B$-spline functions [1], ch. X. we have

$$\sum_{j=-3}^{K} \beta_j S_{j,4}^{(2)}(x) = \sum_{j=-1}^{K} \beta_j^{(2)} S_{j,2}(x).$$

(2.4)
where the vector $\beta^{(2)} = \Delta_2 \beta$ and $\Delta_2$ denotes the matrix corresponding to the weighted difference operator. If we denote by $R$ the matrix with entries

$$ R_{ij} = \int_a^b S_{j,2}(x)S_{i,2}(x)dx, \quad i, j = -1, \ldots, K, $$

then

$$ \int_a^b |S^{(2)}(x)\beta|^2 dx = \beta^T \Delta_2^2 R \Delta_2 \beta = \|L\beta\|^2, $$

with $L = R^{1/2} \Delta_2$.

In the sequel we suppose that the data $(y_i, x_i), \quad i = 1, \ldots, n$ with $(x_i: \quad i = 1, \ldots, n)$ do not necessarily coincide with the knots $(\kappa_i: \quad i = 0, \ldots, K + 1)$.

We want to study the estimator $y - BH(\lambda, M, L)y$ with respect to the smoothing parameter $\lambda > 0$. More precisely we want to recover the position and the sign of the most important noise.

3 Numerical computation

We consider $a = 0$, $b = 1$, $K, n \in \mathbb{N}^*$ and $(\kappa_i)_{i=0, \ldots, K+1}$ with $\kappa_{i+1} - \kappa_i = \frac{b-a}{K+1}$ for all $i \in \{0, \ldots, K\}$. The data $(y_i, x_i)_{i=1, \ldots, n}$ are such that $x_{i+1} - x_i = \frac{b-a}{n-1}$ for all $i \in \{1, \ldots, n-1\}$. The model is $y = B \delta_j + e$.

The following show that for all $j$ and for each smoothing parameter $\lambda$ the noise estimator $[I - BH(\lambda, M)]y$ recovers exactly the position $I$ and the sign of the most important noise. We fix the variance $\sigma^2 \in (0, 1)$, and we consider, for each realization of the noise $e$, the maps $I(e)$ and $sgn(e)$ defined respectively by:

$$ \lambda \in (0, 10) \rightarrow \arg \max_i |[I - BH(\lambda, M)]y(i)| = I(e, \lambda) \in \{1, \ldots, n\}, \quad (3.1) $$

$$ \lambda \in (0, 10) \rightarrow sgn(e, \lambda) = sign([I - BH(\lambda, M)]y(I)). \quad (3.2) $$

where $sign(x) = -1$ if $x < 0$, $sign(x) = 1$ if $x > 0$.

We repeat 100 realizations $(e^{(k)}: \quad k = 1, \ldots, 100)$. We calculate the proportion

$$ p_1(\sigma, \lambda, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[I(e^{(k)}, \lambda) \neq I]}, \quad (3.3) $$

i.e. the probability that the estimator $[I - BH(\lambda, M)]y$ does not recover the position $I$ of the strong noise $e_I$. We also calculate the probability that the estimator $[I - BH(\lambda, M)]y$ does not recover the $sgn(e_I)$ of the strong noise.
\[ p_2(\sigma, \lambda, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[\text{sgn}(e^{(k)}), \lambda) \neq \text{sign}(e(I))]} , \tag{3.4} \]

The probability that the path \( \lambda \in (0, 10) \rightarrow I(e^{(k)}, \lambda) \) does not coincide with the position \( I \) of the most important noise \( e_I \) is equal to

\[ p_3(\sigma, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[\text{sign}(e^{(k)}) \neq I]} . \tag{3.5} \]

The probability that the path \( \lambda \in (0, 10) \rightarrow \text{sign}(e^{(k)}, \lambda) \) does not coincide with the sign of the most important noise \( e_I \) is equal to

\[ p_4(\sigma, n) = \frac{1}{100} \sum_{k=1}^{100} 1_{[\text{sign}(e^{(k)}) \neq \text{sign}(e_I)]} . \tag{3.6} \]

Below we plot \( \lambda \in (0, 10) \rightarrow p_l(\sigma, \lambda, n) \) for \( l = 1, 2 \) and for fixed \( \sigma, \sigma \in (0, 1.5) \rightarrow p_l(\sigma, \lambda, n) \) for \( l = 1, 2 \) and for fixed \( \lambda \), and \( \sigma \rightarrow p_l(\sigma, n) \) for \( l = 3, 4 \).

The following example illustrates our results when \( K = 4, j = 3, I = 1, n = 6, n = 8, n = 10 \) and \( n = 20 \).

\begin{itemize}
  \item Plots of \( p_1(\sigma, \lambda, n) \) and \( p_2(\sigma, \lambda, n) \) for fixed \( \sigma \).
\end{itemize}

If the noise is white, then there is no dominating noise among \( e_1, \ldots, e_n \). Numerical result are coherent. It tells us that the probability that the estimator \( [I - \mathbf{BH}(\lambda, \mathbf{M})]y \) recovers the position of the most important noise is very small.
Numerical result shows that even the noise is white the probability that the estimator $[I - BH(\lambda, M)]y$ recovers the sign of the noise is nearly equal to 0.8.

If the noise has a dominating component, then the probability that the estimator $[I - BH(\lambda, M)]y$ recovers the position of the most important noise belongs to $(0.7, 0.9)$ for all $\lambda \in (0, 10)$.

If the noise has a dominating component, then the probability that the estimator $[I - BH(\lambda, M)]y$ recovers the sign of the most important noise belongs to $(0.8, 1)$ for all $\lambda \in (0, 10)$.

- Plots of $\sigma \rightarrow p_1(\sigma, \lambda, n)$ and $\sigma \rightarrow p_2(\sigma, \lambda, n)$ for fixed $\lambda$. Numerical results show that both are increasing, but $\sigma \rightarrow p_1(\sigma, \lambda, n)$ increases quickly than
\[ \sigma \rightarrow p_2(\sigma, \lambda, n). \]

- Plot of \( p_3(\sigma, n) \) and \( p_4(\sigma, n) \).
References

[1] C. de Boor. A Pratical Guide to Splines, New-York, 1978.

[2] L. Elden. A weighted pseudoinverse, generalized singular values, and constrained least squares problems, BIT, 22 (1982), pp. 487-501.

[3] Micula G, Micula S (1999): Handbook of splines, Kluwer Academic Publishers, Dordrecht, the Netherlands.

[4] Hochstenbach, Ninch, Reichel. Discrete ill-posed least-squares problems with a solution norm constraint. Linear Algebra and its Applications (2011). (2007) 935-958.

[5] C. H. Reinsch. Smoothing by Spline Functions, Numerische Mathematik 10, (1967) 177-183.

[6] I. J. Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions, Quart. Appl. Math., vol. 4, (1946) 4599 and 112141.