1. Introduction

A constant anxiety in the scientific research is to represent the real phenomena by the most precise method. The certainty of these phenomena is only known by tarnished measures of the uncertainty due to the limited exactness of used measuring apparatus. At that time, it is essential to be able to represent by the smoothest curve the phenomena of which one knows only some certain data. In order to obtain the ideal curves many authors have used the spline function theory. For the last 30 years, especially last 15 years, many mathematicians have been particularly interested in the spline functions. The word spline function has been used at first by Schoenberg (1946) [19] to designate the piecewise polynomial functions which are connected at the same time to certainty of their derivatives at the joining points. The second order (cubic degree) mathematical spline is a translation of the draftsman’s spline. This is a function of class $C^2$ of which the third derivative has first kind of discontinuity at the joining points. The methods of spline fitting for data points have been studied in different ways by Mehlum (1964) [15]; Carasso (1967) [5]. Recently one can see the studies of Paihua Montes (1978) [18] and Utreras (1979) [22]. The idea of spline fitting from data which are the vertical segments is due to Atteia (1967) [3]. Recently one can see also the studies of Laurent (1969) [10], (1975) [12], (1976) [13]. Presently the theory and the numerical methods for the construction of general spline fit from data which are horizontal segments, the rectangles of incertitude, and the combination between them, are not well known. In this paper an attempt is made to study theoretically the numerical methods for the construction of a general spline fit for the data which are erroneous in abscissa and ordinate.

2. Mathematical formulation of the problem

In all experimental measuring, no one knows the exact value of function
at a point but his value or his mean value on a small interval or on a point in the coordinate system. More precisely, for the value $t_i$, we have $z_i=x(t_i)$ in the cartesian system, but each $t_i$ is the result of a measurement with a precision $\eta_i$. That is, $t_i\in [a_i, b_i]$ with $b_i-a_i=2\eta_i$ $(i=1, \cdots, n)$. Similarly, $z_i\in [c_i, d_i]$ with $d_i-c_i=2\varepsilon_i$. One remarks that $\eta_i$ and $\varepsilon_i$ can be often $0$. Therefore the data are either the rectangles of incertitude $(\eta_i \neq 0, \varepsilon_i \neq 0)$, or the vertical segments $(\eta_i \neq 0, \varepsilon_i = 0)$, or the horizontal segments $(\eta_i = 0, \varepsilon_i \neq 0)$, or the points $(\eta_i = 0, \varepsilon_i = 0)$, or a combination of them. We want to find a smooth representation of the preceding data $(t_i, z_i)$ with errors $\eta_i$ and $\varepsilon_i$. It is interesting to note that the curve may not through the exact points $(t_i, z_i)$. Before we attempt to solve this problem, we need to make a mathematical formulation using the theory of spline function. We have to make a compromise between the smoothness of function and the approximation of data $z_i$.

Let $H^2[a, b]$ be the Hilbert space of real functions defined on $[a, b]$ which have absolutely continuous first derivatives, and square integrable second derivatives on $[a, b]$, with the scalar product.

$$\langle f | g \rangle = \sum_{i=1}^{n} \int_{a}^{b} f^{(i)}(t) g^{(i)}(t) dt$$

where $f^{(i)}$ is the $i$-th derivative of $f$, and the norm:

$$\|f\| = (\langle f | f \rangle)^{1/2}.$$  

Suppose that the smooth character of the function $x \in H^2[a, b]$ is measured by

(2.1)  

$$\int_{a}^{b} (x''(t))^2 dt.$$  


The value $z_i$ can be considered as a local mean value of the real function $x(t)$ on the interval $[a_i, b_i]$. That is,

$$z_i \approx \frac{1}{b_i-a_i} \int_{a_i}^{b_i} x(t) \ dt, \text{ if } \eta_i > 0,$$

and

$$z_i = x(t_i) \text{ if } \eta_i = 0.$$

Easily we can see that if $\varepsilon_i$ converges to $0$,

$$\frac{1}{b_i-a_i} \int_{a_i}^{b_i} x(t) \ dt$$

converges to $z_i$, and that if $\varepsilon_i=0$,

$$z_i = \frac{1}{b_i-a_i} \int_{a_i}^{b_i} x(t) \ dt.$$
Therefore we obtain a criterion:
\[
\left| \frac{1}{b_i - a_i} \int_{a_i}^{b_i} x(t) dt - z_i \right| \leq \varepsilon_i.
\]
Suppose then that the approximation of values \( z_i \) at points \( t_i \) is measured by
(2.2) \[
\sum_{i=1}^{n} \frac{1}{\varepsilon_i^2} \left( \frac{1}{b_i - a_i} \int_{a_i}^{b_i} x(t) dt - z_i \right)^2.
\]

The term of smoothness (2.1) and the term of approximation (2.2) lead us to the following mathematical formulation of the problem:

Find the spline function \( \sigma \in H^2[a, b] \) such that
(2.3) \[
\int_a^b (\sigma''(t))^2 dt + \rho \sum_{i=1}^{n} \frac{1}{\varepsilon_i^2} \left( \frac{1}{b_i - a_i} \int_{a_i}^{b_i} \sigma(t) dt - z_i \right)^2
\]
for all \( a \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_n \leq b_n = b \), \( \varepsilon_i > 0 \) and \( \rho > 0 \).

The given constant \( \rho > 0 \) determines the relative importance that one takes into account with the smoothness and the approximation. We know that there exists a unique solution of the problem (2.3). Such a solution \( \sigma \) is called a cubic-quartic (second order) spline function for fitting.

3. Characterization of the problem

We desire to modify (2.3) into a simple form in order to solve the problem. We take \( X = H^2[a, b] \), \( Y = L^2[a, b] = H^0[a, b] \). From the Riesz theorem, let \( k_i (i=1, \ldots, n) \) be the elements of \( X \) such that
(3.1) \[
\langle k_i | x \rangle_X = \frac{1}{b_i - a_i} \int_{a_i}^{b_i} x(t) dt, \text{ for all } x \in X.
\]
We provide the space \( \mathcal{Z} = \mathbb{R}^n \) with the scalar product:
(3.2) \[
\langle x | y \rangle_z = \sum_{i=1}^{n} \frac{1}{\varepsilon_i^2} x_i y_i,
\]
where \( x = (x_1, \ldots, x_n) \), \( y = (y_1, \ldots, y_n) \), and the norm:
\[
\|x\|_z^2 = \sum_{i=1}^{n} \frac{1}{\varepsilon_i^2} x_i^2.
\]
Let \( T \) be the continuous linear operator of \( X \) into \( Y \) defined by \( T x(t) = x''(t) \), and let \( A \) be the mapping of \( X \) into \( \mathcal{Z} \) defined by
(3.3) \[
A(x) = [\langle k_1 | x \rangle_X, \ldots, \langle k_n | x \rangle_X] \in \mathcal{Z}.
\]
Then the expression (2.3) can be written as
(3.4) \[
\|T(\sigma)\|_Y^2 + \rho \|A(\sigma) - z\|_z^2 = \min_{x \in X} \|T(x)\|_Y^2 + \rho \|A(x) - z\|_z^2.
\]
We form the product Hilbert space \( \mathcal{V} = Y \times \mathcal{Z} \) on which the scalar product is defined to be
where \( w_1 = [y_1, z_1] \) and \( w_2 = [y_2, z_2] \in \mathbb{Q}, \)
and on which the associated norm is
\[
\| w \| = \langle w | w \rangle^{1/2}
\]
\[
= (\langle y | y \rangle + \rho \langle z | z \rangle)^{1/2}
\]
where \( w = [y, z] \in \mathbb{Q}. \)

We have also the continuous linear operator \( L \) of \( X \) into \( \mathbb{Q} \) defined by
\[
L(x) = [T(x), A(x)] \in \mathbb{Q}.
\]

Let \( w_0 = [0, z] \in \mathbb{Q} \) (0 designates the origin of \( \Psi \)).

Then the problem (3.4) is to characterize \( \sigma(t) \in X \) such that
(3.5) \[
\| L(\sigma) - w_0 \| = \min_{x \in \mathbb{X}} \| L(x) - w_0 \| .
\]

Note that \( \text{Ker}(A) \) is the kernel of \( A \) and \( \text{Im}(A) = A(X) \) is the image of \( A. \)

If one takes a subspace \( \mathbb{Q} \) of \( \mathbb{Q} \) such that
\[
\mathbb{Q} = L(\mathbb{Q}) = \text{Im}(L)
\]
\[
= \{ [y, z] \in \mathbb{Q} \times \mathbb{Z} | y = T(x), z = A(x) \text{ with } x \in \mathbb{X} \},
\]
the problem (3.5) is equivalent to find \( \bar{w} \in \mathbb{Q} \) such that
(3.6) \[
\| \bar{w} - w_0 \| = \min_{w \in \mathbb{Q}} \| w - w_0 \| .
\]

From the theorem of the characterization of the best approximation, one knows that if \( n \geq 2 \), there exists a unique solution \( \sigma(t) \) which satisfies the problem (3.6). For the convenience, one takes \( n \geq 3 \) in order to obtain the numerical solution of the problem.

We remark that \( \text{Ker}(T) \) is finite dimensional and the dimension of \( \mathbb{Z} = \mathbb{R}^* \) is determined by the number of the given data conditions.

4. Numerical construction to the projection method

According to the theorem of characterization of the best approximation (the property of the orthogonality) to the problem (3.6), we know that a necessary and sufficient condition that there exists a unique solution \( \sigma(t) \) for \( \forall z \in \mathbb{Z} \) (relatively to \( T, A, z \) and \( \rho > 0 \)) is that
(4.1) \[
\langle T(\sigma) | T(x) \rangle + \rho \langle A(\sigma) - z | A(x) \rangle = 0, \text{ for all } x \in \mathbb{X}.
\]

Let \( T^*, A^* \) and \( L^* \) be the adjoint operators of \( T, A \) and \( L \) respectively.

Then the equation (4.1) becomes
\[
T^* T(\sigma) = \rho A^*(z - A(\sigma)), \text{ or } L^* L(\sigma) = \rho A^*(z).
\]

With this condition, we define the space \( \mathcal{J} \) of spline functions:
(4.2) \[
\mathcal{J} = \{ \sigma \in \mathbb{X} | \langle T(\sigma) | T(x) \rangle + \rho \langle A(\sigma) - z | A(x) \rangle = 0, \forall x \in \mathbb{X} \}
\]
\[
= \{ \sigma \in \mathbb{X} | T^* T(\sigma) = A^*(\lambda), \lambda = \rho (z - A(\sigma)) \in \mathbb{Z} \}.
\]
From the definition of $d$, we have
\[ T(d) = T(Ker(A))^\perp \text{ and } T^* T(d) = \text{Im}(T^*) \cap \text{Im}(A^*). \]

We now consider the three subspaces:
\[ \mathcal{K} = \text{Im}(T^*) \cap \text{Im}(A^*) \text{ in } \mathcal{X}, \]
\[ \mathcal{F} = T(Ker(A))^\perp \text{ in } \mathcal{Y}, \]
\[ \mathcal{E} = A(Ker(T))^\perp \text{ in } \mathcal{Z}. \]

These three spaces are connected to bijective and bicontinuous way by the operators $T^*$ and $A^*$. We have then
\[ (4.3) \quad \mathcal{K} = T^*(\mathcal{F}) = A^*(\mathcal{E}), \quad T(d) = \mathcal{F} \text{ and } T^* T(d) = \mathcal{K}. \]

It can be shown that the spaces $\mathcal{K}$, $\mathcal{F}$ and $\mathcal{E}$ are of dimension $n-2$.

Now let $h_1, \cdots, h_{n-2}$ be a base of $\mathcal{K}$ and $f^1, \cdots, f^{n-2}$ a base of $\mathcal{F}$ such that
\[ T^*(f^j) = h^j, \quad j=1, \cdots, n-2, \]
and let $b_1, \cdots, b^{n-2}$ be a base of $\mathcal{E}$ such that
\[ A^*(b^j) = h^j, \quad j=1, \cdots, n-2. \]

Then we have
\[ (4.4) \quad h^j = T^*(f^j) = A^*(b^j), \quad j=1, \cdots, n-2. \]

We set
\[ t_i = \frac{a_i + b_i}{2}, \quad i=1, \cdots, n. \]

From (3.2) and (3.3), the operator $A^*$ is given by
\[ A^*(z) = \sum_{i=1}^{n} \frac{1}{\varepsilon_i^2} z_i k_i \in \mathcal{X}, \quad \text{with } z = (z_1, \cdots, z_n) \in \mathcal{Z}. \]

We have then
\[ (4.5) \quad h^j = A^*(b^j) = \sum_{i=1}^{n} \frac{1}{\varepsilon_i^2} b^j_i k_i. \]

Let $b^1, \cdots, b^{n-2}$ be the vectors of $\mathcal{Z}$ defined by
\[ b^j_i = \begin{cases} 0 & \text{for } i \in \{1, \cdots, n\} \setminus \{j, j+1, j+2\}, \\ \frac{\varepsilon_j^2}{d_j^j} & \text{for } i=j, \\ \frac{\varepsilon_{j+1}^2}{d_{j+1}^j} & \text{for } i=j+1, \\ \frac{\varepsilon_{j+2}^2}{d_{j+2}^j} & \text{for } i=j+2, \end{cases} \]

with $j=1, \cdots, n-2$, and $d_i^j = \prod_{l=i}^{j+2} (t_i - t_l)$, and $b^j_i$ is the $i$-th compasont of vector $b^j$.

We want to verify that the $b^j$ form a base of $\mathcal{E} = A(Ker(T))^\perp \subset \mathcal{Z}$.

In view of (4.5), the following lemma is an easy consequence of the fact that $Ker(T)$ is the set of polynomials of degree $\leq 1$, and that $\dim(\mathcal{E}) = n-2$. 

LEMMA 1. The vectors $b_1, b_2, \ldots, b^{n-2}$ form a base of $\mathcal{E}$.

REMARK 2. From (4.5) and (4.6), the expression of the element $h^j \in \mathcal{X}$ is

$$h^j = b_j^i k_j + b_{j+1}^i k_{j+1} + b_{j+2}^i k_{j+2} = \sum_{i=j}^{i+2} b_i^i k_i.$$ 

We recall the notation:

$$(f)_+ = \begin{cases} f & \text{if } f \geq 0, \\ 0 & \text{if } f < 0. \end{cases}$$

We now want to find the explicit expression of $f^j \in \mathcal{Y}$ such that

$$h^j = T^*(f^j), \quad j = 1, \ldots, n-2.$$ 

LEMMA 2. The functions $f^j(t) = \sum_{i=j}^{i+2} f_i^j(a_i, b_i, t)$, where

$$f_i^j(a_i, b_i, t) = \begin{cases} \frac{(b_i - t)^2 - (a_i - t)^2}{2 d_i^i (b_i - a_i)} & \text{if } a_i \neq b_i \\ \frac{(a_i - t)^2}{d_i^i} & \text{if } a_i = b_i, \end{cases}$$

form a base of $\mathcal{E} = T(Ker(A)) \subset \mathcal{Y}$.

Proof. Let $x \in \mathcal{X}$. We have

$$x(t) = p(t) + \int_a^b (t - \tau)_+ x''(\tau) d\tau$$

where $p(t)$ is a polynomial of degree 1. Then we have

$$\langle A^*(b^j) | x \rangle_x = \langle A^*(b^j) | p \rangle_x + \int_a^b \langle A^*(b^j) | (\tau - t)_+ \rangle_x x''(t) dt.$$ 

From Lemma 1, we have

$$\langle A^*(b^j) | x \rangle_x = \int_a^b \langle A^*(b^j) | (\tau - t)_+ \rangle_x x''(t) dt.$$ 

In fact we have, from (4.5), that

$$\langle A^*(b^j) | (\tau - t)_+ \rangle_x = \sum_{i=j}^{i+2} \frac{k_i | (\tau - t)_+ \rangle_x}{d_i^i}.$$ 

On the other hand,

$$\langle k_i | (\tau - t)_+ \rangle_x = \frac{1}{b_i - a_i} \int_{a_i}^{b_i} (\tau - t)_+ d\tau = d_i^i | (a_i, b_i, t),$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, n-2$. Then we have

$$\langle A^*(b^j) | (\tau - t)_+ \rangle_x = \sum_{i=j}^{i+2} f_i^j(a_i, b_i, t) = f^j(t).$$ 

Thus we have

$$\langle A^*(b^j) | x \rangle_x = \int_a^b f^j(t) x''(t) dt = \langle f^j(t) | T(x) \rangle_y.$$ 

That is, $\langle b^j | A(x) \rangle_y = \langle f^j(t) | T(x) \rangle_y$. According to Lemma 1, we have
Numerical construction of cubic-quartic (second order) spline fits

\[ \langle f^j(t) | T(x) \rangle_q = 0. \quad \forall x \text{ such that } A(x) = 0. \]

Consequently, \( f^j(t) \in T(\text{Ker}(A)) \). Therefore, the functions \( f^j(t) \) are linearly independent in \( \mathcal{F} \) which is \( n-2 \) dimensional.

This completes the proof of the lemma.

**REMARK 2.** If \( a_i \neq b_i \) \((i=j, j+1, j+2)\), the functions \( f^j(t) \) on \( [a, b] \) are piecewise polynomial of degree 2 in each \( [a_i, b_i] \), of degree 1 between these intervals, and vanish outside \( (a_j, b_j+2) \). These polynomials are continuous at \( a_i \) and \( b_i \) up to the first derivative. (The second derivative has discontinuities at \( a_i \) and \( b_i \)). If \( a_i = b_i \), the polynomial of degree 2 in \( [a_i, b_i] \) disappears, and the other polynomials remain identically with the precedent. The curve \( f^j(t) \) is always continuous on \( [a, b] \).

According to Lemma 1 and Lemma 2, we easily obtain the following proposition.

**PROPOSITION 1.** For all \( i, j \) with \(|i-j| \geq 3\), we have

\[
\langle b^i | b^j \rangle_q = \sum_{k=1}^{n} \frac{1}{d_k^i} b_k^i b_k^j = 0
\]

\[
\langle f^i(t) | f^j(t) \rangle_q = \int_{a}^{b} f^i(t) f^j(t) dt = 0
\]

for \( i = 1, \cdots, n \) and \( j = 1, \cdots, n-2 \).

We now shall deduce the theorem of the characterization for the solution of our problem from the definition of the space \( \mathcal{J} \).

Let \( D \) be the continuous linear mapping of \( \mathcal{J} \) into \( \mathbb{C} \) which associates \( \lambda \) of (4.2) to \( s \in \mathcal{J} \). We have then

\[ D(s) = A^{-1} T^* T(s), \quad \text{for all } s \in \mathcal{J}. \]

From (4.3), \( \text{Im}(D) = D(\mathcal{J}) = \mathbb{C} \). We can deduce that for any \( z \in \mathbb{C} \), there exists a unique \( \sigma \in \mathcal{J} \) such that

\[ A(\sigma) + \frac{1}{\rho} D(\sigma) = z. \]

Let \( M \) be the mapping of \( \mathcal{J} \) into \( \mathcal{Q} \) defined by

\[ M(s) = [T(s), -\frac{1}{\rho} D(s)], \quad s \in \mathcal{J}. \]

Let \( \mathcal{Q} \) be the subspace of \( \mathcal{Q} \) such that \( \mathcal{Q} = \text{Ker}(L^*) \). Then \( \mathcal{Q} \) is of dimension \( n-2 \). We know also that

\[ \mathcal{Q} = M(\mathcal{J}) = \{[y, z] \in \mathcal{Q} | y \in \mathcal{J}, \, z = -\frac{1}{\rho} A^{-1} T^*(y)\}. \]

Since \( f^j(j=1, \cdots, n-2) \) form a base of \( \mathcal{F} \), we see that

\[ g^j = [f^j, -\frac{1}{\rho} A^{-1} T^*(f^j)] = [f^j, -\frac{1}{\rho} b^j], \quad j = 1, \cdots, n-2 \]

also form a base of \( \mathcal{Q} \).
THEOREM 1. A necessary and sufficient condition that \( \sigma \in \mathbb{X} \) is the cubic-quartic spline function for fitting of the problem (relatively to \( T, A, z \) and \( \rho > 0 \)) is that there exist the \( \mu_j(j=1, \ldots, n-2) \) such that
\[
T(\sigma) = \sigma''(t) = \sum_{j=1}^{n-2} \mu_j f^j(t), \quad A(\sigma) = z - \frac{1}{\rho} \sum_{j=1}^{n-2} \mu_j b^j.
\]
The coefficients \( \mu_j \) can be obtained by solving the following linear algebraic system of dimension \( n-2 \)
\[
\sum_{j=1}^{n-2} \omega_{ij} \mu_j = \langle z | b^j \rangle_z, \quad i = 1, \ldots, n-2,
\]
where \( \omega_{ij} = \langle f^i(t) | f^j(t) \rangle_y + \frac{1}{\rho} \langle b^i | b^j \rangle_z \).

Proof. From the definition of \( \mathcal{Q} \), the necessary and sufficient condition that \( \sigma \) is the cubic-quartic spline function for fitting of the problem is that \( \mathbb{Q} = [T(\sigma), A(\sigma) - z] \in \mathcal{Q} \). According to (4.7), \( \mathbb{Q} = [T(\sigma), A(\sigma) - z] \in \mathcal{Q} \) if and only if there exist the coefficients \( \mu_j(j=1, \ldots, n-2) \) such that
\[
T(\sigma) = \sigma''(t) = \sum_{j=1}^{n-2} \mu_j f^j(t),
\]
\[
A(\sigma) - z = - \frac{1}{\rho} \sum_{j=1}^{n-2} \mu_j b^j.
\]
To obtain the coefficients \( \mu_j \) numerically, we use the fact that \( \mathbb{Q} = [T(\sigma), A(\sigma)] \) is orthogonal to \( g^j \ (j=1, \ldots, n-2) \) in \( \mathcal{Q} \). Explicitly we obtain
\[
\langle \sum_{j=1}^{n-2} \mu_j f^j(t) | f^j(t) \rangle_y + \rho \langle z - \frac{1}{\rho} \sum_{j=1}^{n-2} \mu_j b^j | - \frac{1}{\rho} b^j \rangle_z = 0,
\]
for every \( i = 1, \ldots, n-2 \). Thus we have
\[
\sum_{j=1}^{n-2} \langle \langle f^i(t) | f^j(t) \rangle_y + \frac{1}{\rho} \langle b^i | b^j \rangle_z \rangle \mu_j = \langle z | b^j \rangle_z
\]
for every \( i = 1, \ldots, n-2 \).

This completes the proof of the theorem.

REMARK 3. Let the matrix \( \Omega = (\omega_{ij}) \). Then \( \Omega \) is a symmetric matrix of order \( n-2 \). According to Proposition 1, the matrix \( \Omega \) is pentadiagonal.

REMARK 4. The cubic-quartic spline functions \( \sigma \) for fitting of the problem is formed by the piecewise polynomial of degree 4 in each \( [a_i, b_i] \), and of degree 3 between these intervals.

REMARK 5. The projection method is very useful as we have seen in the proof of Theorem 1. This method has been proposed by Laurent and Ansleone (1968) \[14\]. The stability has been verified by Carasso (1966) \[4\] with diverse types of functions. It is also linked with the method proposed by Greville (1964) \[7\] which uses the divided differences. Munteanu and
Schumaker (1973) have discussed several similar methods which can be applied to more general interpolating spline.

Acknowledgement

The author wishes to express his appreciation to professor Claude Carasso for suggesting the problem and directing the research. This is a theoretical part of the research which was done as a part of the thesis of Doctor of third cycle submitted to University of Saint-Etienne.

References


