ACTA UNIVERSITATIS APULENSIS No 11/2006

Proceedings of the *International Conference on Theory and Application of Mathematics and Informatics ICTAMI 2005* - Alba Iulia, Romania

# ON THE SPLINE FUNCTIONS IN DATA ANALYSIS FRAMEWORK

Nicoleta Breaz and Daniel Breaz

Abstract. The paper is focused on the understanding of the spline function as a modern tool in data analysis. We consider three numerical methods that offer framework for the spline function's use as the interpolation, fitting and smoothing of the data. For each of these three methods, we present the corresponding spline and the conditions required by data for using the appropriate type of spline.

*Keywords:* spline function, interpolation, fitting, smoothing, least squares

2000 *Mathematics Subject Classification:* 65D07, 65D05, 65D10, 62J99, 62G08

## 1. Introduction

The data analysis constitute one of the most important fields in numerical analysis, importance winned by their wide applicability in many practical problems. Various data analysis techniques were developed in time, all finally coming down to the same simple target, namely, the extraction from the data of some information regarding the studied phenomenon, quantified by one or more variables. Even if it insn't always motivated by the practical significance of the phenomenon, the most used model is the linear one. Without diminishing the charm of the linear (or more general, polynomials) simplicity, another class of functions, much more flexible but with the same good approximation properties, arose about sixty years ago. Thus, in 1946, I.J.Schoenberg ([13]), introduced for the first time, the mathematical version of a spline, namely the spline function, although the notion in itself was a wellknown one, at that time, in the drafting branch. The mechanical spline was consist in a thin, elastic bar on which some weights were placed in order to draw curves that were forced to go through given points.

The image of this instrument naturally leads to Schoenberg's definition of a spline function, as a piecewise polynomial function with the pieces joined at the breaks together with a number of its derivatives. From this elementary definition, up to what can be meant today by a spline function, the way of scientifical researches has beared a lot of ramifications, as a result of multiple posibilities for generalizations and extensions provided by this notion. For example, we can preserve the piecewise nature of the initial notion but we change the polynomials with other elementary functions like exponential or rational functions. Even the piecewise nature was finally proved to be removable feature for a spline function, numerous mathematiciens prefering to define a spline function as the solution of a variational problem.

Although a relative recent one, the theory of spline functions benefits by a real interest from the mathematiciens focused on applied mathematics, in this regard, a large number of papers (over 700) related to the subject, giving evidence. A paper having bibliographical feature, useful in the synthesis of various subjects discussed in this domain, is the paper [9], the result of an extensive research in spline functions, by the romanian mathematician, Gh. Micula. From the rich literature that exists in this domain, we also remind here the papers [1], [5], [12] and [14] of de Boor, Prenter, Greville and Wahba, respectively, as well as, the papers [6] and [11], of the romanian mathematicians, D.V. Ionescu and T. Popoviciu, respectively, The importance of the spline functions is revealed by the numerous fields of modern numerical analysis, in which these functions are applied, namely, in fitting, interpolation and approximation, in numerical integration and differentiation, in differential and partial differential equations, in statistics and implicitely, further, in the applicative researching fields as geology, meteorology, biomedical sciences, sociology, economy, etc.

Our paper is focused on the applications of spline functions in data processing. We will consider three approaches to data, namely the interpolation, fitting and smoothing of the data. According to these approaches we can speak about three type of spline functions, namely the interpolation spline, the fitting spline and the smoothing spline. The use of one type of spline functions in spite of the others two, will be motivated here by the nature of data and approximation problem.

Beside this introductory section the paper contains other four sections. In the section 2, we deal with the issue of data processing by the three numerical methods: interpolation, fitting and smoothing. Some elementary concepts related to the spline functions theory are presented in the section 3. The clear delimitation realised in the section 2, between the notions of data interpolation, data fitting and data smoothing will encourage the next approach, namely the delimitation of three spline functions types: the interpolation spline, the (least squares) fiting spline and the smoothing (penalized least squares) spline. In the section 4, we focused on the spline function as a modern statistics tool, hence we deal just with fitting and smoothing spline in the context of the regression models. Thus, we present the least squares spline (the fitting spline) as a regression function in the parametric regression and the penalized least squares spline (smoothing spline) as a regression function in the nonparametric regression. In the last section we make some remarks on how to choose between interpolating, fitting or smoothing of the data, based on the data nature and the data modeling problem type.

## 2. Interpolation, fitting and smoothing of data

With respect to what sort of the closeness between $f(x_i)$ and $y_i, i = \overline{1,n}$ , we have taken it in consideration, we will speak about the interpolation, fitting and smoothing of the data.

Definition 1. *It is called an interpolation problem related to the data* $(x_i, y_i), i = \overline{1,n}$, *the problem which consists of the determination of a function* $f : I \mapsto R, I \subset R, x_i \in I, i = \overline{1,n}$, *whose values at the data sites* $x_i$ , *"come close" to data* $y_i$, *in the following sense:*

$$f(x_i) = y_i, i = \overline{1,n}. \tag{1}$$

Definition 2. *i)It is called a fitting problem related to the data* $(x_i, y_i), i = \overline{1,n}$, *the problem which consists of the determination of a function* $f : I \mapsto R, I \subset R, x_i \in I, i = \overline{1,n}$, *whose values at the data sites* $x_i$ , *"come close" to data* $y_i$, *whithout leading necesarily to equality:*

$$f(x_i) \cong y_i, i = \overline{1,n}. \tag{2}$$

*ii)We say that a fitting problem is the better closeness to data fitting problem with respect to the criteria* $E$, *if it consists of the determination of a function for which* $E(f)$ *is minimum. The criteria* $E$ *is chosen such that its minimization coresponds to the closeness to data.*

Definition 3. *It is called the least squares problem related to the data* $(x_i, y_i), i = \overline{1,n}$*, the problem which consists of the determination of a function (from a settled functions space),* $f : I \mapsto R, I \subset R, x_i \in I, i = \overline{1,n}$*, that is the solution to the minimization problem:*

$$E(f) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [y_i - F(x_i)]^2 = \min . \tag{3}$$

Definition 4. *We define the smoothing problem related to the data* $(x_i, y_i)$*,* $i = \overline{1,n}$*, as a problem which consists of the determination of a function* $f : I \mapsto R, I \subset R, x_i \in I$*,* $i = \overline{1,n}$*, whose values at the data sites* $x_i$*, "come close" to data* $y_i$*, so much that the function remains smooth. In other words, we are searching for f as a solution to the following minimum criteria:*

$$E(f) + \lambda J(f) \to \min . \tag{4}$$

Here, $E(f)$is a functional that reflects, by minimizing, the closeness to data(fitting), and $J(f)$ is related to the smoothing condition, the minimization of this functional leading to a function with some smoothness properties. The parameter $\lambda$, called smoothing parameter, takes values in the interval $(0, \infty)$.

## 3. Elementary spline functions. Interpolating, fitting and smoothing spline functions

In the data analysis framework, arises the piecewise polynomial version of a spline function:

Definition 5.*Let be the following partition of the real line:*

$$\Delta : -\infty \leq a < t_1 < t_2 < ... < t_N < b \leq \infty. \tag{5}$$

*The function* $s : [a, b] \longmapsto R$ *is called spline function of m degree (order* $m + 1$*), with the breaks (knots of the function),* $t_1 < t_2 < ... < t_N$ *, if the following conditions are satisfied:*

*i)* $s \in P^m, t \in [t_i, t_{i+1}], i = \overline{0, N}, t_0 = a, t_{N+1} = b,$

*ii)* $s \in C^{m-1}$*,* $t \in [a, b]$*.*

Here, $P^m$ is the class of polinomyals of degree $m$ or less and $C^{m-1}$ is the class of functions with $m - 1$continuous derivatives.

In what follows, we denote by $\delta_m(\Delta)$, the space of spline functions of $m$ degree with the breaks given by the partition $\Delta$. Also, for the so-called truncated power function we use the notation:

$$t_+^m = \begin{cases} 0, t \leq 0 \\ t^m, t > 0. \end{cases} \quad (6)$$

The following formula uniquely gives the reprezentation of an element from $\delta_m(\Delta)$ , by means of the truncated power functions basis:

$$s(t) = p(t) + \sum_{i=1}^{N} c_i(t - t_i)_+^m, t \in [a, b], p \in P^m, c_i \in R. \quad (7)$$

Definition 6. *Let be $\Delta$ the partition from (5) and the space of odd degree spline functions, $\delta_{2m-1}(\Delta), 2m-1, m \geq 1$. It is called the natural spline function of $2m-1$ degree (order $2m$), an element $s$, from the space $\delta_{2m-1}(\Delta)$, which satisfies the condition $s \in P^{m-1}, t \in [a, t_1] \cup [t_N, b]$.*

We denote by $S_{2m-1}(\Delta)$, the space of natural spline functions of $2m-1$ degree, related to the partition $\Delta$.

If we link the spline function with the three ways of data approaches from the section 2, we will have three types of spline functions, namely, interpolating, fitting and smoothing spline function.

Definition 7. *An element from $\delta_m(\Delta)$ is called the interpolating spline function (Lagrange), if it is the solution to the data interpolation problem, presented in the Definition 1.*

We will use the following notation:

$$H^{m,2}[a,b] = \left\{ f : [a,b] \to R \mid f, f', \cdots, f^{(m)} \text{absolutely continuous}, f^{(m)} \in L_2[a,b] \right\}. \quad (8)$$

The following variational property of the interpolating natural spline functions is a previous step for the introduction of the smoothing spline function, used in statistics:

Theorem 8. *Let $a \leq x_1 < x_2 < ... < x_n \leq b$, be a partition of $[a, b]$, $y_i, i = \overline{1,n}, n \geq m$, real numbers and the set $J(y) = \left\{ f \in H^{m,2}[a,b] : f(x_i) = y_i, i = \overline{1,n} \right\}$. Then there exists a unique $s \in J(y)$ , such that,*

$$\int_a^b \left[s^{(m)}(x)\right]^2 dx = \min\left\{\int_a^b \left[f^{(m)}(x)\right]^2 dx, f \in J(y)\right\}. \quad (9)$$

*Moreover, the following statements hold:*
i) $s \in C^{2m-2}([a,b])$,
ii) $s\mid_{[x_i,x_{i+1}]} \in P^{2m-1}, i = \overline{1, n-1}$,
iii) $s\mid_{[a,x_1]} \in P^{m-1}$ and $s\mid_{[x_n,b]} \in P^{m-1}$.

Definition 9. *An element from the $\delta_m(\Delta)$ is called fitting spline function if it is a solution to the fitting problem presented in the Definition 2.i.*

Definition 10. *We define the (general) smoothing spline function as a function from an appropriate smooth function space, that is a solution to the data smoothing spline problem, presented in the Definition 4.*

It is interesting that the searching for interpolating and fitting spline functions starting from the spline functions space, at the same time, that the smoothing spline function is obtained as a result of a searching for the solution to the smoothing spline problem, in more general smooth functions spaces.

## 4. Spline in statistics

Here we will consider just the last two approaches of the data, fitting and smoothing, these approaches being of interest in the statistics field, especially in the regression models.

### 4.1. Least squares spline regression as a parametric regression

In this section, we approach an application of the spline function in statistics, namely the least squares regression with a spline function, regression that involves a linearizable model. On wide data range, the polynomial regression works only as a local approximation method, leading to the necessity of finding a more flexible model. A solution to this problem could be the statement of a polynomial regression problem on each subinterval from a partition of the data range, dealing with a switching regression model that bears structural changes in some data points (see [10] and [4]).

The reason for the least squares spline regression is given, beside the obvious flexibility of the spline functions, by the Taylor theorem. If we suposse that the regression function is from $H^{m,2}[a,b]$, then, for a sample with $n$ observations, the observational model, $y_i = f(x_i) + \varepsilon_i$, becomes

$$y_i = \sum_{k=0}^{m} \alpha_k x_i^k + R_m(x_i) + \varepsilon_i, i = \overline{1,n}, R_m(x) = (m!)^{-1} \int_0^1 (x-t)_+^m f^{(m+1)}(t)\, dt.$$

If $R_m(x_i), i = \overline{1,n}$ , are small, then the $m$ degree polynomial regression is reasonable, but if these remainders are great, it must be found a model that compensate for the possible inadequacy of a polynomial model. Thus, aproximating the integrand from $R_m(x)$, with the quadrature formula $\sum_{k=1}^{N} \beta_k (x-t_k)_+^m$, for the coefficients $\beta_k, k = \overline{1,n}$, depending on the values $f^{(m+1)}(t_k), k = \overline{1,n}$ and for the points $0 < t_1 < t_2 < ... < t_N < 1$, one can obtain the following approximation of the regression function (truncated power functions based form of a spline of $m$ degree, with the breaks, $t_1, t_2, ..., t_N$):

$$f(x) = \sum_{k=0}^{m} \alpha_k x^k + \sum_{k=1}^{N} \beta_k (x-t_k)_+^m.$$

Definition 11. *It is called the fitting spline (simple) regression model the model*

$$y = f(x) + \varepsilon,$$

*where $f$ is the spline function (Definition 5), of $m$ degree, $m \geq 1$, with the breaks $t_1 < t_2 < ... < t_N$ . If the model is based on the least squares criteria then we use the term of least squares spline regression.*

In a polynomial regression, the degree must be selected. Similarly, in the fitting spline regression, the following aspects must be selected: the function degree, $m$, the number of breaks (knots of the function), $N$, the breaks, $t_k, k = \overline{1,N}$ and the location of these breaks in respect with data sites, $x_i, i = \overline{1,n}$. For the beginning, we have considered all these aspects, fixed.

The spline model can be reduced to a linear one, by the substitution $X^k = Z_k, k = \overline{1,m}$ and $(X-t_k)_+^m = U_k, k = \overline{1,N}$, thus obtaining the linear model with a constant term,

$$Y = \alpha_0 + \alpha_1 Z_1 + ... + \alpha_m Z_m + \beta_1 U_1 + ... + \beta_N U_N. \tag{10}$$

Further, for a sample with $n$ observations, we have the sample matrix before the linearization, $x$ and the sample matrix after the linearization, $z$, respectively:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix}, \; z = \begin{pmatrix} z_{11} & z_{12} & ... & z_{1m} & u_{11} & u_{12} & ... & u_{1N} & 1 \\ z_{21} & z_{22} & ... & z_{2m} & u_{21} & u_{22} & ... & u_{2N} & 1 \\ ... & ... & ... & ... & ... & ... & ... & ... & ... \\ z_{n1} & z_{n2} & ... & z_{nm} & u_{n1} & u_{n2} & ... & u_{nN} & 1 \end{pmatrix}.$$

Here, likewise in the polynomial regression, we are interested in obtaining the uniquenes and existence conditions for the least squares estimators, in terms of the initial sample matrix, $x$. The following result was published in the paper [2].

Theorem 12. *If among the $n$ data sites related to the variable $X$, there is at least one value situated in each of the $N+1$ open intervals delimited by the breaks and there are another $m$ distinct values situated in $(-\infty, t_1)$, then the solution of the least squares fitting corresponding to the model (1), is uniquely given by the formulae*

$$a_* = (a_1, a_2, ..., a_m, b_1, b_2, ..., b_N)' = \left( \widetilde{V_*}' (x) \widetilde{V_*} (x) \right)^{-1} \widetilde{V_*}' (x) \widetilde{y}, \tag{11}$$

$$a_0 = \overline{y} - \sum_{k=1}^{m} a_k \overline{x^k} - \sum_{k=1}^{N} b_k \overline{(x-t_k)_+^m}. \tag{12}$$

In this theorem, $\widetilde{v}$ denotes the matrix or the vector resulted after a rescaling (centering) with the mean value, and $V_*(x)$ is a matrix obtained as a result of joining the Vandermonde like matrix, with $n$ rows and $q$ coloumns (without the coloumns of ones), with the matrix:

$$\begin{pmatrix} (x_1-t_1)_+^q & (x_1-t_2)_+^q & ... & (x_1-t_N)_+^q \\ (x_2-t_1)_+^q & (x_2-t_2)_+^q & ... & (x_2-t_N)_+^q \\ ... & ... & ... & ... \\ (x_n-t_1)_+^q & (x_n-t_2)_+^q & ... & (x_n-t_N)_+^q \end{pmatrix}.$$

Degree and breaks selection in the least squares spline regression

Selection of the degree, $m$, is usually made as a result of a graphical analysis of data and selection of the breaks number, $N$, depends on the desired amount of flexibility. Also, there are several elementary rules for selecting the breaks. Thus, for a linear spline, the breaks are placed at points where the data exhibits a change in slope, and for a quadratic or cubic spline, the breaks are placed near local maxima, minima or inflection points in the data. Besides these settings, there are also data driven methods for selecting these parameters. In this regard, we have proposed a data driven method for selection of the breaks, based on a function called the cross validation function.

Definition 13. *We call the cross validation function related to the selection of the number and location of the breaks for a spline regression, the following function*

$$CV\left(\lambda_N\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f_{\lambda_N}^{(-i)}\left(x_i\right)\right)^2, \tag{13}$$

*where* $f_{\lambda_N}^{(-i)}$ *is the least squares spline estimator, obtained for the breaks given by* $\lambda_N = \{t_1, t_2, ..., t_N\}$, *based on the subsample* $\left(x_j, y_j\right), j = \overline{1,n}, j \neq i$ .

The optimal set of breaks $\widehat{\lambda}_{\widehat{N}}$, will be derived by minimizing, for fixed $N$, the expression $CV\left(\lambda_N\right)$, thus resulting the estimator $\widehat{\lambda}_N$ and then by obtaining $\widehat{\lambda}_{\widehat{N}}$, with $\widehat{N}$, the minimizer of $CV\left(\widehat{\lambda}_N\right)$.

## 4.2. Smoothing spline regression as a nonparametric regression

This part is concerned with another application of the spline functions in statistics, namely, the smoothing of data by spline functions. The setting is the one of the nonparametric regression in which, one begins with a more general assumption about the regression function, without limiting it at some analitical form, but just assuming some smoothness properties for it. The tool used here is the smoothing spline function. The smoothing spline regression model, presented in this section, is based on an estimator, both flexible and smooth, appropriate for the cases when a parametric regression model is not sufficently motivated.

Definition 14. *Let $H^{m,2}[a,b]$ , be the functions space, defined in formula (8). We call the penalized least squares estimator of the regression function from the model $y_i = f(x_i) + \varepsilon_i$, $i = \overline{1,n}$, with $\varepsilon' = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) \sim N(0, \sigma^2 I)$, an element from $H^{m,2}[a,b]$, that minimizes the expression*

$$n^{-1}\sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda\int_a^b \left[f^{(m)}(x)\right]^2 dx, \lambda \geq 0. \tag{14}$$

*The related regression model will be called the smoothing spline (simple) regression model.*

When the data are noisy, it will be suitable that the estimator does not facilitate an exagerate closeness to data, these being suspected to be not real data. In such situations, the estimator will have to smooth the data, by coming so close to it as some imposed amount of smoothness allows.

The beauty of this criteria, and implicitly of the related estimator, consists in its flexibility, provided by the parameter $\lambda$, called the smoothing parameter. This parameter can be chosen by the analyst, as he wants to give more importance to the fidelity to the data, $n^{-1}\sum\limits_{i=1}^{n}[y_i - f(x_i)]^2$ or to the smoothness of the solution, $\int\limits_a^b \left[f^{(m)}(x)\right]^2 dx$ . Also, in the related literature, several automatically data driven selection methods for $\lambda$ are developed. Therefore, the smoothing parameter controlls the tradeoff between the fidelity to data of the estimated curve and the smoothness of this curve. One can observe that, for $\lambda = 0$, the second term from (14) doesn't matter, consequently, the minimization of whole expression is reduced to the minimization "in force", of the sum of squares. This case leads to an estimator that interpolates the data. On the other side, the case $\lambda \to \infty$ makes the second term from (14) to grow, therefore, in compensation, in the minimizing process of the whole expression, the accent must be lay on $\int\limits_a^b \left[f^{(m)}(x)\right]^2 dx$. This case gives as estimator, the $m-1$ degree least squares fitting polynom, obtained from data (that is, the most possible smooth curve). Between these two extremes, a large value for $\lambda$ indicates the smoothness of the solution in spite of the closeness to the data, while a small value leads to a curve very close to data but which loses smoothness.

In what follows, we will present a motivation for the smoothing spline regression, based on the expansions in Taylor series of a function. Thus, as a result of the expansion of $f$ , the observational regressional model, $y_i = f(x_i) + \varepsilon_i$, becomes

$$y_i = \sum_{k=0}^{m-1} \alpha_k x_i^k + R_{m-1}(x_i) + \varepsilon_i, i = \overline{1,n},$$

with the remainder, $R_{m-1}(x) = [(m-1)!]^{-1} \int_0^1 (x-t)_+^{m-1} f^{(m)}(t)\, dt$ .

It can be shown that the following inequality holds:

$$\max_{1 \le i \le n} R_{m-1}(x_i)^2 \le \frac{J_m}{(2m-1)[(m-1)!]^2}$$

where $J_m(f) = \int_0^1 \left[f^{(m)}(x)\right]^2 dx$. If we know for example that $\rho \ge 0$ , then we will have information regarding the departure from a polynomial model and we can put this information in the estimation process. Therefore, the penalized least squares criteria is a natural criteria.

Now we want to point out the difference between the least squares spline regression model and the smoothing (penalized least squares) spline regression model, difference often ignored in the related romanian literature.

Remark 15.The fitting spline model is related to the parametric regression, claiming a certain form (piecewise polynomial, degree, breaks), while the smoothing spline model, in the nonparametric regression setting, claims just the affiliation to the space $H^{m,2}[a,b]$ . Also, the ways of approaching the data are different, the fitting model being suitable when we want to be as close as possible to some data, inaccesible to a polynomial regression, for example, and the smoothing model is appropriate when we want to get closer to the data, with caution, controlled by the smoothness of the estimator.

Anyway, in both cases, the estimator is a spline function. However, in the smoothing model we deal with only those polynomial spline function that are solutions to a variational problem (the natural spline, for example), while, in the fitting model, we can propose as estimators, other types of polynomial spline functions, too.

Also, while in the fitting model, the smoothness of the estimator is given by the number of breaks, in the smoothing model, this issue is controlled by the smoothing parameter, $\lambda$.

The cross validation function related to the spline smoothing problem, that by minimizing in respect with $\lambda$, gives an estimator of the smoothing parameter, is

$$CV(\lambda) = \frac{1}{n} \sum_{k=1}^{n} \left( y_k - f_{\lambda}^{(-k)}(x_k) \right)^2 , \tag{15}$$

where $f_{\lambda}^{(-k)}$ is the unique solution of the spline smoothing problem, stated for the data sample from which it was leaving out the $k$-th data.

## 5. Remarks and conclusions

The fitting and the smoothing are alternatives for the interpolating, much more attractive for the statistics field where an exact imitation of the data behaviour is not necessarily desired, giving the inherent errors contained in the data. We can state that from a statistical point of view, the spline functions constitute a link between the parametric regression and the nonparametric one, being estimating tools in both situations.

Thus, in parametric regression, one must deal with a restrictive approximation class which provides estimators with known parametric form except for the coefficients involved. The corresponding spline model for this setting is the least squares fitting spline model. Like it was already suggested by the titulature of the estimator for the regression function used in this case, the data processing technique is the least squares fitting. Thus, the estimator comes closer to data as much as the assumed parametric forme (in this case, a spline one) allows. It is well known that the polynomial spline function is completely defined if the degree and the breaks are known. In case that we don't have any information about these, the function degree, the number of breaks and the location of them become aditional parameters that need to be estimated, in order to estimate the regression function, completely.

In the nonparametric regression, the class of function in which one is looking for the estimator, is extended to more general functions spaces, that do not assume a certain parametric form but just some smoothness properties (continuity, derivability, integrability) of the function. Also, one may be interested in changing the data approaching modality, based on the assumption that the data are suspected of containing errors. Thus, one will use an estimator which even if it has great flexibility will however smooth too perturbated data, assuming some smoothness. Consequently, the penalized least squares smoothing spline can be considered as an estimator. The penalty which is added to the least squares expression is based on a functional whose minimization will penalize the less smooth estimators. The form of this functional depends on the search space. In case of the $H^{m,2}$ space, containing functions that together with their first $m-1$ derivatives are continuous and have the $m$-th derivative square integrable, it is obtained as estimator the natural polynomial spline function. It is interesting to remark that although in the nonparametric regression, the search space is not directly related with the notion of spline (it just contains the natural spline functions space), the resulted estimator is a spline function. In any type of nonparametric regression, the complete determination of an estimator depends eventually on a small number of parameters, which however are not related to the parametric form but to some functionals optimization. Thus, in the nonparametric regression with splines we have only one smoothing parameter, $\lambda, \lambda > 0$, parameter that controlls the tradeoff between the closeness to data and the smoothness of the estimator.

## References

[1].de Boor C., *A Practical Guide to Splines*, Springer-Verlag, New York, 1978

[2].Breaz N., Breaz D., *Fitting of some linearisable regression model*, Studia Univ. Babeş-Bolyai, Mathematica, vol XLVIII, no. 2, 21-27, 2003

[3].Eubank R.L., *Nonparametrc Regression and Spline Smoothing-Second Edition*, Marcel Dekker, Inc., New York, Basel, 1999

[4].Ferreira P.E., *A Bayesian Analysis of a Switching Regression Model: Known Number of Regimes*, J. Amer. Statist. Assoc., Theory and Methods Sections, vol 70, no.350, 370-374, 1975

[5].Greville T., *Theory and Applications of Spline Functions*, University of Wisconsin Press, Madison, WI, 1968

[6].Ionescu D.V., *Introduction á la théorie des "fonctions spline"*, Acta Mathematica Academiae Scientiarum Hungaricae, no.21, 21-26, 1970

[7].Lange K., *Numerical Analysis for Statisticians*, Springer-Verlag, New York, 1999

[8].Micula Gh., *Funcţii spline şi aplicaţii*, Editura Tehnică, Bucureşti, 1978

[9].Micula Gh., Micula S., *Handbook of Splines*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1999

[10].Poirer D.J., *The Econometrics of Structural Change with Special Emphasis on Spline Functions*, North-Holland Publ. Comp., Amsterdam, 1976

[11].Popoviciu T., *Notes sur les fonctions convexes d'ordre supérieur*, Bull. Math. Soc. Sci.Math.,IX,43, 85-141,1941

[12].Prenter P., *Splines and Variational Methods*, John Wiley, New York, 1975

[13].Schoenberg I.J., *Contribution to the problem of approximation of equidistant data by analytic functions*, Parts A and B, Quart. Appl. Math. no. 4, 45-88, 112-441, 1946

[14].Wahba G., *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia, 1990

Nicoleta Breaz, Daniel Breaz
"1 Decembrie 1918" University of Alba Iuia
Alba Iulia, str. N. Iorga, No 11-13, 510009,
Alba, Romania
e-mail: nbreaz@uab.ro, dbreaz@uab.ro