International Journal of Advances in Applied Sciences (IJAAS)
Vol.x, No.x, Month 201x, pp. xx~xx
ISSN: 2252-8814

A New, Robust and Applied Model for Interpolation of Huge Data

Mehdi Zamani*,
* Civil Engineering Department, Faculty of Technology and Engineering, Yasouj University, Yasouj, IRAN,

Article Info

Article history:

Received Jun 12th, 201x
Revised Aug 20th, 201x
Accepted Aug 26th, 201x

Keyword:

Approximation
Estimation
Cubic spline
Optimization
B-spline

ABSTRACT

A powerful and accurate model for interpolation of data is piecewise cubic spline method. In this model the obtained final curve consists of a series of local cubic spline curves which combine together with suitable continuity along their boundaries. The continuity at the boundaries of each local curve is c^2. The attraction of this research is to apply cubic spline method for approximation and estimation of data. Therefore; each local spline curve satisfies the minimization of sum of square errors along its length in addition of obtaining c^2 continuity at its edges or boundaries. Because in the local scale the approximation of any section of data by cubic spline is accurate; therefore, the presented model is applicable to any kind of data with highly nonlinear distributions.

Corresponding Author:

Author,
Civil Engineering Department, Faculty of Technology and Engineering,
Yasouj University, Moallem Square, Daneshjoo Avenue, Yasouj, IRAN,
E-mail: mahdi@mail.yu.ac.ir

1. INTRODUCTION

The methods of approximation are used for the estimation and prediction of engineering data especially experimental data. These methods are mostly linear and are solved by the optimization approach which minimizes the sum of square error. There are many cases in engineering problem which cannot be approximated by linear model. For some of those cases the logarithmic, exponential and polynomial models can be applied which can be transferred to linear ones. The above non-linear models can be applied only for a limited group of data while, in most cases of problems especially ones having huge data and their non-linear variation behavior, it cannot be implemented. In engineering activities there are many cases in which we encounter the data that could be experimental, statistical and field data. The simplest cases of data are two columns of n pair's data x_i and y_i where they are independent and dependant variables respectively. The aim is to determine a function $f(x)$ which could predict suitable and reasonable y_i values with respect to the variables x_i at the data points as well as the points between the data. The interpolation approach cannot be applied for data when systematic errors exist corresponding to measuring and values, also when there are several values of y_i at each x_i value. For such cases the approximated analysis is more suitable and accurate to define and explain the variability distribution of data. Therefore the best curves that indicate the variability behavior of data, does not cross the data points but passes near them.

2. LITERATURE REVIEW

Most of the literatures about this subject are mainly linear or simple non-linear models [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], and [13]. These methods are used most for computer graphic. Because of the waviness and the sinusoidal forms of their curves, they are not applicable to engineering problems. [14] has presented a simple piece wise cubic spline model for approximation of highly non-linear data. In his model the domain of problem is divided into n elements. The interval for the element i is $[l_i, l_{i+1}]$. The cubic spline for this element is

$$s(x) = a_i (x - l_i)^3 + b_i (x - l_i)^2 + c_i (x - l_i) + d_i \tag{1}$$

At the edges of each element c^1 continuity exists.

$$\begin{cases} s_i(l_i) = s_{i-1}(l_i) \\ s_i^{'}(l_i) = s_{i-1}^{'}(l_i) \end{cases} \tag{2}$$

If the continuity conditions of Eq. (2) are inserted into Eq. (1) the following equation will result.

$$s_i(x) = a_i(x - l_i)^3 + b_i(x - l_i)^2 + s_{i-1}^{'}(x - l_i) + s_{i-1}(l_i) \tag{3}$$

The values of $s_{i-1}(l_i)$ and $s_{i-1}^{'}(l_i)$ in the above equation can be calculated numerically. Therefore the four parameters of Eq. (1) are reduced to two. The parameters a_i and b_i on each element can be determined by the minimization of the sum of square errors on the element data as the following equations.

$$\begin{Bmatrix} b_i \\ a_i \end{Bmatrix} = \frac{1}{a_{11}a_{22} - a_{12}^2} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{12} & a_{11} \end{bmatrix} \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix} \tag{4}$$

in which

$$\begin{cases} a_{11} = \sum_{j=1}^{m_j} (x_j - l_i)^4 \\ a_{12} = \sum_{j=1}^{m_j} (x_j - l_i)^5 \\ a_{22} = \sum_{j=1}^{m_j} (x_j - l_i)^6 \end{cases} \tag{5}$$

and

$$\begin{cases} f_1 = s_{i-1}^{'}(l_i)\sum_{j=1}^{m_j}(x_j - l_i)^3 - s_{i-1}(l_i)\sum_{j=1}^{m_j}(x_j - l_i)^2 + \sum_{j=1}^{m_j}(x_j - l_i)^2 y_j \\ f_2 = -s_{i-1}^{'}(l_i)\sum_{j=1}^{m_j}(x_j - l_i)^4 - s_{i-1}(l_i)\sum_{j=1}^{m_j}(x_j - l_i)^3 + \sum_{j=1}^{m_j}(x_j - l_i)^3 y_j \end{cases} \tag{6}$$

In the above model because of having two degree of freedom, the element length should be adjusted to a suitable length in order to achieve a good approximation to data.

[15] developed a B-spline model for approximation of engineering data. The B-spline equation which is applied in this method was summation of $n+1$ B-spline functions as Eq. (7).

$$s(x) = \sum_{i=0}^{n} c_i\, b_i\left[u_i(x)\right] = c_0 b_0(u_0) + c_1 b_1(u_1) + \ldots c_n b_n(u_n) \tag{7}$$

Where $u \in [-2, 2],\ u_i(x) = (x - x_i)/h \quad and \quad h = (b - a)/h$

Each cubic B-spline function $b_i(u_i)$ is an exponential function as Eq. (8) and Fig. (1).

$$b_i(u) = a\, e^{-bu_i^2} - c \tag{8}$$

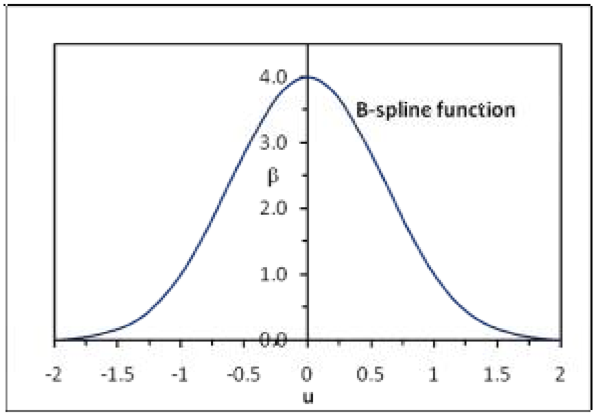

Figure 1. B-spline function.

Where $a = 4.016478$, $b = 1.3740615$ and $c = 0.016478$. The method of least square fitting is applied in the model.

$$SSE = \sum_{i=1}^{m} \left[y_i - s(x_i) \right]^2 \tag{9}$$

Where SSE is sum of square errors and m is the number of data pairs. By substitution of $s_i(x)$ in Eq. (9) it becomes,

$$SSE = \sum_{i=1}^{m} \left[y_i - \sum_{k=0}^{n} c_k \, b_k \right]^2 \tag{10}$$

Squaring Eq. (10) it results,

$$SSE = \sum_{i=1}^{m} y_i^2 + \sum_{i=1}^{m} \sum_{j=0}^{n} \sum_{k=0}^{n} c_j \, c_k \, b_j^i \, b_k^i - 2 \sum_{i=1}^{m} \sum_{k=0}^{n} c_k \, b_k^i \, y_i \tag{11}$$

Taking the derivatives of Eq. (11) respect to the coefficients c_k it results,

$$\frac{\partial SSE}{\partial c_j} = 2 \sum_{i=1}^{m} \sum_{k=0}^{n} c_k \, b_k^i \, b_j^i - 2 \sum_{i=1}^{m} b_j^i \, y_i = 0 \quad , \quad j = 0,1,2,\ldots,n \tag{12}$$

Eq. (12) forms a symmetric linear system of equations as,

$$\begin{bmatrix} \sum_{i=1}^{m} b_0^2 & \sum_{i=1}^{m} b_0 b_1 & \cdots & \sum_{i=1}^{m} b_0 b_n \\ \sum_{i=1}^{m} b_0 b_1 & \sum_{i=1}^{m} b_1^2 & \cdots & \sum_{i=1}^{m} b_1 b_n \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{m} b_0 b_n & \sum_{i=1}^{m} b_1 b_n & \cdots & \sum_{i=1}^{m} b_n^2 \end{bmatrix} \begin{Bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{m} b_0 y_i \\ \sum_{i=1}^{m} b_1 y_i \\ \vdots \\ \sum_{i=1}^{m} b_n y_i \end{Bmatrix} = [A]\{c\} = \{b\} \tag{13}$$

With regarding to the basic function characteristics Eq. (8),

$$\sum_{i=1}^{m} b_i(x) \, b_j(x) = 0 \quad for \quad j \geq i+4 \tag{14}$$

Therefore according to the conditions of Eq. (14) the linear system of equations (13) transforms to a heptadiagonal system of equations. This causes lots of zeroes for most of the components of matrix A. The problem of the model is the requirement of the first derivatives at the first and last control points. The above derivatives are usually not available and somehow should be estimated. Also the situation of control points in this model is important to obtain a satisfactory and reasonable result.

3. FORMULATION OF MODEL

The presented model implies the optimization on cubic spline method. Therefore here it pays attention shortly to the cubic spline method. The final curve is sum of a series of piecewise cubic spline curves. Those local curves have continuity of c^2 at the internal and the boundary points. The local cubic spline function on each element and for example in element i is,

$$s_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad i = 0,1,\cdots,n-1 \tag{15}$$

The Eq. (15) consists of n cubic spline functions on n+1 pairs data with element length $h_i = x_{i+1} - x_i$. The boundary conditions at each node or element edge are,

$$\begin{cases} s_i(x_i) = f_i \\ s_i(x_i) = s_{i-1}(x_i) \\ s_i'(x_i) = s_{i-1}'(x_i) \\ s_i''(x_i) = s_{i-1}''(x_i) \end{cases} , \quad i = 0,1,\cdots,n-1 \tag{16}$$

With substitution of conditions of Eq. (16) in Eq. (15) it results,

$$\begin{cases} a_i = f_i \\ a_i = a_{i-1} + h_{i-1}b_{i-1} + h_{i-1}^2 c_{i-1} + h_{i-1}^3 d_{i-1} \\ b_i = b_{i-1} + 2h_{i-1}c_{i-1} + 3h_{i-1}^2 d_{i-1} \\ c_i = c_{i-1} + 3h_{i-1}d_{i-1} \end{cases} , \quad i = 0,1,\ \cdots,n-1 \tag{17}$$

Doing some manipulations on the above equations b_{i-1} can be determined as,

$$b_{i-1} = \frac{1}{h_{i-1}}(f_i - f_{i-1}) - \frac{h_{i-1}}{3}(2c_{i-1} + c_i) \tag{18}$$

If the indexes of Eq. (18) increase by one step and then it is substituted in equations (17) the following linear system of equations results.

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} = \frac{3}{h_i}(f_{i+1} - f_i) - \frac{3}{h_{i-1}}(f_i - f_{i-1}) \quad , \quad i = 0,1,2,\ \cdots,n-1 \tag{19}$$

The above linear system of equations is three diagonal and unknowns parameters are c_i's. For solving the linear system, the parameters c_{-1} and c_n should be determined for including the boundary conditions. The clamped or natural boundary conditions are usually applied for the above linear system. However for simplicity it is possible to assume $c_{-1} = c_n = 0$, $c_{-1} = c_0$, and $c_{n-1} = c_n$. The resulted final curve is smooth and suitable to the real function $f(x)$ because it has degree three on each element. If the variations of data follow a simple cubic spline curve a single spline function is sufficient for the approximation of data. For that simple example the governing approximation function is,

$$s = a + b\,x + c\,x^2 + d\,x^3 \tag{20}$$

Where coefficients a, b, c and d can be determined from the following linear system of equations.

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \sum x_i^5 \\ \sum x_i^3 & \sum x_i^4 & \sum x_i^5 & \sum x_i^6 \end{bmatrix} \begin{Bmatrix} a \\ b \\ c \\ d \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum x_i\, y_i \\ \sum x_i^2 y_i \\ \sum x_i^3\, y_i \end{Bmatrix} \tag{21}$$

In Eq. (21) n is the number of pairs data. The above model can be done easily by Excel worksheet program. A single cubic spline function cannot be applied for most cases of engineering problems because of variables complexity of data. For those data if the piecewise cubic spline model is used for the approximation, the final curve will be smooth and follows and coincides the variability of data. Suppose the data distribution be as Fig. (2), the cubic spline function for element i is,

$$s_i(x) = a_i + b_i(x-\ell_i) + c_i(x-\ell_i)^2 + d_i(x-\ell_i)^3 \tag{22}$$

Where ℓ_i the control point for element i and it equals x_i.

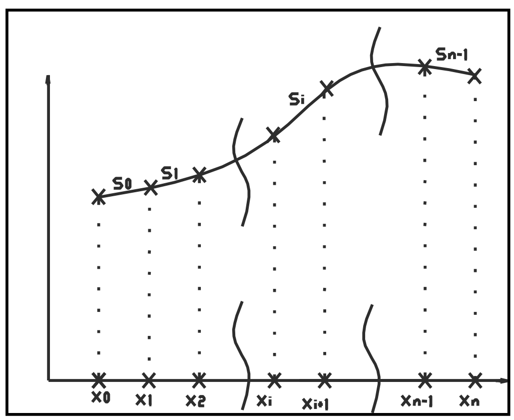

Figure 2. The nonuniform distribution of data.

At the edges of each element continuity of c^2 can be defined as equations (23) which are acceptable for most engineering applications.

$$\begin{cases} s_i(\ell_i) = s_{i-1}(\ell_i) \\ s_i^{'}(\ell_i) = s_{i-1}^{'}(\ell_i) \\ s_i^{''}(\ell_i) = s_{i-1}^{''}(\ell_i) \end{cases} \tag{23}$$

Applying the 1st and 2nd derivatives on Eq. (22) and then substituting them into Eq. (23) the following relationships between parameters a_i, b_i , c_i and d_i are obtained.

$$\begin{cases} a_i = a_{i-1} + h_{i-1}b_{i-1} + h_{i-1}^2 c_{i-1} + h_{i-1}^3 d_{i-1} \\ \quad b_i = b_{i-1} + 2h_{i-1}c_{i-1} + 3h_{i-1}^2 d_{i-1} \\ \qquad c_i = c_{i-1} + 3h_{i-1}d_{i-1} \end{cases} \tag{24}$$

Where $h_i = \ell_{i+1} - \ell_i$ is the length of element i . The above parameters on each element are calculated by using the minimization of sum of squares and applying the continuity conditions of Eq. (23) as follows,

$$SSE = \sum_{j=1}^{m_i} (y_j - Y_j)^2 = \sum_{j=1}^{m_i} \left[y_j - s_i(x_j) \right]^2 \quad , \quad x_j \in [\ell_i, \ell_{i+1}] \tag{25}$$

$$Min\ SSE = Min\ \sum_{j=1}^{m_i} \begin{bmatrix} y_j - a_i - b_i(x_j - \ell_i) - c_i(x_j - \ell_i)^2 \\ - d_i(x_j - \ell_i)^3 \end{bmatrix}^2 \tag{26}$$

In order to include the continuity conditions Eq. (24) the Lagrange coefficients l_i, m_i and n_i are considered corresponding to Eq. (26).

$$\begin{aligned} Min\ SSE = Min \Bigg\{ & \sum_{j=1}^{m_i} \left[y_j - s_i(x_j) \right]^2 + \\ & l_i(a_i - a_{i-1} - h_{i-1}b_{i-1} - h_{i-1}^2 c_{i-1} - h_{i-1}^3 d_{i-1}) + \\ & m_i(b_i - b_{i-1} - 2h_{i-1}c_{i-1} - 3h_{i-1}^2 d_{i-1}) + v_i(c_i - c_{i-1} - 3h_{i-1}d_{i-1}) \Bigg\} \end{aligned} \tag{27}$$

The minimization of equation (27) is possible by taking their derivatives respect to the parameters and Lagrange coefficients.

$$\begin{aligned} & \frac{\partial SSE}{\partial a_i} = 0 \quad , \quad \frac{\partial SSE}{\partial b_i} = 0 \quad , \quad \frac{\partial SSE}{\partial c_i} = 0 \quad , \quad \frac{\partial SSE}{\partial d_i} = 0 \\ & \frac{\partial SSE}{\partial l_i} = 0 \quad , \quad \frac{\partial SSE}{\partial m_i} = 0 \quad , \quad \frac{\partial SSE}{\partial v_i} = 0 \end{aligned} \tag{28}$$

The equations (28) form a linear system of equation of dimension 7×7 for element i.

$$\begin{bmatrix} 2m_i & 2\sum(x_j-\ell_i) & 2\sum(x_j-\ell_i)^2 & 2\sum(x_j-\ell_i)^3 & 1 & 0 & 0 \\ 2\sum(x_j-\ell_i) & 2\sum(x_j-\ell_i)^2 & 2\sum(x_j-\ell_i)^3 & 2\sum(x_j-\ell_i)^4 & 0 & 1 & 0 \\ 2\sum(x_j-\ell_i)^2 & 2\sum(x_j-\ell_i)^3 & 2\sum(x_j-\ell_i)^4 & 2\sum(x_j-\ell_i)^5 & 0 & 0 & 1 \\ 2\sum(x_j-\ell_i)^3 & 2\sum(x_j-\ell_i)^4 & 2\sum(x_j-\ell_i)^5 & 2\sum(x_j-\ell_i)^6 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} a_i \\ b_i \\ c_i \\ d_i \\ l_i \\ m_i \\ v_i \end{Bmatrix} = \begin{Bmatrix} 2\sum y_j \\ 2\sum(x_j-\ell_i)y_j \\ 2\sum(x_j-\ell_i)^2 y_j \\ 2\sum(x_j-\ell_i)^3 y_j \\ a_{i-1}+h_{i-1}b_{i-1}+h_{i-1}^2c_{i-1}+h_{i-1}^3d_{i-1} \\ b_{i-1}+2h_{i-1}c_{i-1}+3h_{i-1}^2d_{i-1} \\ c_{i-1}-3h_{i-1}d_{i-1} \end{Bmatrix} \tag{29}$$

The above system can be solved for vector {x} according to the following equation.

$$\{x\} = [P]^{-1}\{e\} = [Q]\{e\} \tag{30}$$

or

$$\begin{cases} a_i = a_{i-1} + h_{i-1}b_{i-1} + h_{i-1}^2 c_{i-1} + h_{i-1}^3 d_{i-1} \\ b_i = b_{i-1} + 2h_{i-1}c_{i-1} + 3h_{i-1}^2 d_{i-1} \\ c_i = c_{i-1} + 3h_{i-1}d_{i-1} \\ d_i = a_1^i + a_2^i a_{i-1} + a_3^i b_{i-1} + a_4^i c_{i-1} + a_5^i d_{i-1} \\ l_i = b_1^i + b_2^i a_{i-1} + b_3^i b_{i-1} + b_4^i c_{i-1} + b_5^i d_{i-1} \\ m_i = g_1^i + g_2^i a_{i-1} + g_3^i b_{i-1} + g_4^i c_{i-1} + g_5^i d_{i-1} \\ v_i = q_1^i + q_2^i a_{i-1} + q_3^i b_{i-1} + q_4^i c_{i-1} + q_5^i d_{i-1} \end{cases} \tag{31}$$

where

$$\begin{cases} a_1^i = q_{11}^i\, e_4^i \\ a_2^i = q_{12}^i \\ a_3^i = h_{i-1}\, q_{12}^i + q_{13}^i \\ a_4^i = h_{i-1}^2\, q_{12}^i + 2h_{i-1}\, q_{13}^i + q_{14}^i \\ a_5^i = h_{i-1}^3\, q_{12}^i + 3h_{i-1}^2\, q_{13}^i + 3h_{i-1} q_{14}^i \end{cases}, \quad \begin{cases} b_1^i = e_1^i + q_{21}^i\, e_4^i \\ b_2^i = q_{22}^i \\ b_3^i = h_{i-1}\, q_{22}^i + q_{23}^i \\ b_4^i = h_{i-1}^2\, q_{22}^i + 2h_{i-1}\, q_{23}^i + q_{24}^i \\ b_5^i = h_{i-1}^3\, q_{22}^i + 3h_{i-1}^2\, q_{23}^i + 3h_{i-1} q_{24}^i \end{cases} \tag{32}$$

$$\begin{cases} g_1^i = e_2^i + q_{31}^i\, e_4^i \\ g_2^i = q_{32}^i \\ g_3^i = h_{i-1}\, q_{32}^i + q_{33}^i \\ g_4^i = h_{i-1}^2\, q_{32}^i + 2h_{i-1}\, q_{33}^i + q_{34}^i \\ g_5^i = h_{i-1}^3\, q_{32}^i + 3h_{i-1}^2\, q_{33}^i + 3h_{i-1} q_{34}^i \end{cases}, \quad \begin{cases} q_1^i = e_3^i + q_{41}^i\, e_4^i \\ q_2^i = q_{42}^i \\ q_3^i = h_{i-1}\, q_{42}^i + q_{43}^i \\ q_4^i = h_{i-1}^2\, q_{42}^i + 2h_{i-1}\, q_{43}^i + q_{44}^i \\ q_5^i = h_{i-1}^3\, q_{42}^i + 3h_{i-1}^2\, q_{43}^i + 3h_{i-1} q_{44}^i \end{cases}$$

If the last two equations of Eq. (31) solve simultaneously for parameters a_{i-1} and b_{i-1} the Eq. (38) is resulted.

$$\begin{cases} a_{i-1} = m_1^i m_i + m_2^i v_i + m_3^i c_{i-1} + m_4^i c_i + m_5^i \\ b_{i-1} = n_1^i m_i + n_2^i v_i + n_3^i c_{i-1} + n_4^i c_i + n_5^i \end{cases} \tag{33}$$

where

$$\begin{cases} m_1^i = \dfrac{q_3^i}{T_i} \\ m_2^i = -\dfrac{g_3^i}{T_i} \\ m_3^i = \dfrac{3h_{i-1}(g_3^i q_4^i - g_4^i q_3^i) + g_5^i q_3^i - g_3^i q_5^i}{3h_{i-1}T_i} \\ m_4^i = \dfrac{g_3^i q_5^i - g_5^i q_3^i}{3h_{i-1}T_i} \\ m_5^i = \dfrac{g_3^i q_1^i - g_1^i q_3^i}{T_i} \end{cases},$$

and

$$\begin{cases} n_1^i = -\dfrac{q_2^i}{T_i} \\ n_2^i = \dfrac{g_2^i}{T_i} \\ n_3^i = \dfrac{3h_{i-1}(g_4^i q_2^i - g_2^i q_4^i) + g_2^i q_5^i - g_5^i q_2^i}{3h_{i-1}T_i} \\ n_4^i = \dfrac{g_5^i q_2^i - g_2^i q_5^i}{3h_{i-1}T_i} \\ n_5^i = \dfrac{g_1^i q_2^i - g_2^i q_1^i}{T_i} \end{cases}$$

$$T_i = g_2^i q_3^i - g_3^i q_2^i$$

The following linear system of equations is obtained by substitution equations (32) in equations (31), finally

$$\begin{aligned} &-l_i + (b_2^i m_1^i + b_3^i n_1^i)\, m_i + (b_2^i m_2^i + b_3^i n_2^i)\, v_i + (b_2^i m_3^i + b_3^i n_3^i + b_4^i - b_5^i t^{i-1})\, c_{i-1} \\ &\quad + (b_2^i m_4^i + b_3^i n_4^i + b_5^i t^{i-1})\, c_i = -(b_1^i + b_2^i m_5^i + b_3^i n_5^i) \end{aligned}$$

$$\begin{aligned} &(a_2^i m_1^i + a_3^i n_1^i)\, m_i + (a_2^i m_2^i + a_3^i n_2^i)\, v_i + (a_2^i m_3^i + a_3^i n_3^i + a_4^i - a_5^i t^{i-1})\, c_{i-1} \\ &\quad + (a_2^i m_4^i + a_3^i n_4^i + a_5^i t^{i-1} + t^i)\, c_i - t^i c_{i+1} = -(a_1^i + a_2^i m_5^i + a_3^i n_5^i) \end{aligned}$$

$$\begin{aligned} &n_1^{i-1} m_{i-1} - n_1^i m_i + n_2^{i-1} v_{i-1} - n_2^i v_i + (n_3^{i-1} + h_{i-2}) c_{i-2} + (n_4^{i-1} - n_3^i + h_{i-2}) c_{i-1} \\ &\quad - n_4^i c_i = -n_5^i - n_5^{i-1} \end{aligned} \tag{34}$$

$$\begin{aligned} &(m_1^{i-1} + h_{i-2} n_1^{i-1}) m_{i-1} - m_1^i m_i + (m_2^{i-1} + h_{i-2} n_2^{i-1}) v_{i-1} - m_2^i v_i + (m_3^{i-1} + h_{i-2} n_3^{i-1} \\ &\quad + \frac{2}{3} h_{i-1}^2) c_{i-2} + (m_4^{i-1} - m_3^i + h_{i-2} n_4^{i-1} + \frac{1}{3} h_{i-2}^2) c_{i-1} - m_4^i c_i = -h_{i-2} n_5^{i-1} + m_5^i - m_5^{i-1} \end{aligned}$$

where $t^i = 1/3h_i$.

The equations (34) form a linear system of equations with dimension 4×4 which is a local system for element i. Therefore, for problem with n elements the final or global linear system of equations has dimension 4n×4n. The total unknown parameters are 4n and they consist c_i , μ_i , u_i and λ_i for i=0, 1, 2, …, n-1. The situation of entries in local linear system is shown in Fig. (3).

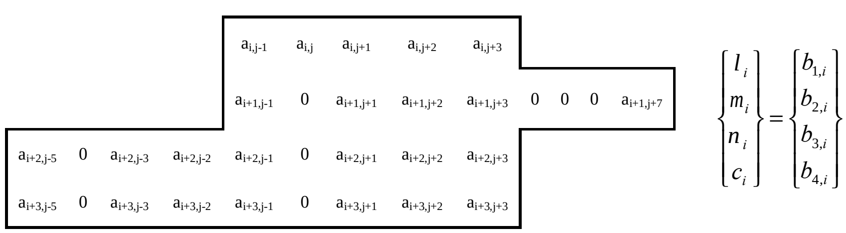

Figure 3. The situation of local linear system entries.

The governing final linear system of equations for the following boundary conditions relationships has unique solution.

$$m_{-1}=0 \quad , \quad c_{-2}=0 \quad , \quad c_{-1}=c_n=0.5 \quad , \quad v_{-1}=0$$
$$n_5^{-1}=m_5^{-1}=n_4^{-1}=m_4^{-1}=0 \tag{35}$$

Where c_{-1} and c_n are equivalent to second derivatives of cubic spline functions $s_{-1}(x)$ and $s_n(x)$ at control points ℓ_{-1} and ℓ_n , respectively. Fig. (4) shows the situations of local linear systems in the global or final linear system of equations. The final matrix is bounded and has a lot of zeros entries.

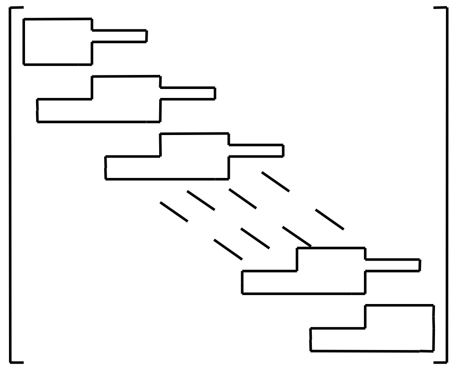

Figure 4. The situations of local system entries.

4. CREATING THE FINAL MATRIX

The domain of data [a,b] is divided into n elements by selecting n+1 control points. The element length could be uniform with length h or variable as h_k where $h_k=\ell_{k+1}-\ell_k$. The cubic spline function on element k is $s_k(x)$ where $k=0,1,2,\cdots,n-1$ and $k+1$ is the element number. The nonzero components of local matrix for element $k+1$ are determined from the following equations where $0\le k\le n-1$ and $i,j=4k+1$.

$$\begin{cases} a_{i+2,j-5}=n_3^{k-1}+h_{k-2} \\ a_{i+3,j-5}=m_3^{k-1}+h_{k-2}n_3^{k-1}+\dfrac{2}{3}h_{k-2}^2 \end{cases}$$

$$\begin{cases} a_{i+2,j-3}=n_1^{k-1} \\ a_{i+3,j-3}=m_1^{k-1}+h_{k-2}n_1^{k-1} \end{cases} , \quad \begin{cases} a_{i+2,j-2}=n_2^{k-1} \\ a_{i+3,j-2}=m_2^{k-1}+h_{k-2}n_2^{k-1} \end{cases}$$

$$\begin{cases} a_{i,j-1}=b_2^k m_3^k+b_3^k n_3^k+b_4^k-b_5^k t^{k-1} \\ a_{i+1,j-1}=a_2^k m_3^k+a_3^k n_3^k+a_4^k-a_5^k t^{k-1} \\ a_{i+2,j-1}=n_4^{k-1}-n_3^k+h_{k-2} \\ a_{i+3,j-1}=m_4^{k-1}-m_3^k+h_{k-2}n_4^{k-1}+\dfrac{1}{3}h_{k-2}^2 \end{cases} , \quad a_{ij}=-1 \tag{36}$$

$$\begin{cases} a_{i,j+1}=b_2^k m_1^k+b_3^k n_1^k \\ a_{i+1,j+1}=a_2^k m_1^k+a_3^k n_1^k \\ a_{i+2,j+1}=-n_1^k \\ a_{i+3,j+1}=-m_1^k \end{cases} , \quad \begin{cases} a_{i,j+2}=b_2^k m_2^k+b_3^k n_2^k \\ a_{i+1,j+2}=a_2^k m_2^k+a_3^k n_2^k \\ a_{i+2,j+2}=-n_2^k \\ a_{i+3,j+2}=-m_2^k \end{cases}$$

$$\begin{cases} a_{i,j+3} = b_2^k m_4^k + b_3^k n_4^k + b_5^k t^{k-1} \\ a_{i+1,j+3} = a_2^k m_4^k + a_3^k n_4^k + a_5^k t^{k-1} + t^k \\ a_{i+2,j+3} = -n_4^k \\ a_{i+3,j+3} = -m_4^k \end{cases}, \quad a_{i+1,j+7} = -t^k$$

The entries of right-hand side vector b_i for element $k+1$ is obtained from equations (37).

$$\begin{cases} b_i = -(b_1^k + b_2^k m_5^k + b_3^k n_5^k) \\ b_{i+1} = -(a_1^k + a_2^k m_5^k + a_3^k n_5^k) \\ b_{i+2} = n_5^k - n_5^{k-1} \\ b_{i+3} = m_5^k - m_5^{k-1} - h_{k-2} n_5^{k-1} \end{cases} \tag{37}$$

The effects of boundary conditions Eq. (35) should be included in the 1st, 2nd and nth elements. Solving the resulted global linear system of equations by LU method, the unknown parameters l_k , m_k , n_k and c_k for $k = 0,1,2,\cdots,n-1$ will be obtained. Then by applying Eqs. (31) and (32) the cubic spline coefficients a_k , b_k and d_k can be calculated. The resulted final spline curve has the continuity c^2 and the maximum optimization respect to the sum of square errors. For testing and checking the formulation obtained above about three problems are chosen and explained in the following.

5. PROBLEM 1

In this example about 41 data pairs in the domain $x \in [1,5]$ are generated based on the following equation.

$$f(x) = 3 + 7\sin\sqrt{x}\,\frac{(\ln x)^x}{x^{\ln x}} + e \tag{38}$$

Where e is random number generation in the interval [-1, 1]. For first try two elements with uniform length $h_0=h_1=2$ are considered with control points $\ell_0 = 1$, $\ell_1 = 3$ and $\ell_2 = 5$. The coefficients c_{-1} and c_3 for boundary elements assume 0.50. The final linear system of equations has dimension 8×8. The parameters and cubic spline coefficients are brought in Table (1).

Table 1. The parameters and coefficients.

λ_0	-1.258	a_0	3.67
μ_0	1.339	b_0	1.503
u_0	1.021	d_0	0.014
c_0	-0.247	a_1	5.806
λ_1	-0.877	b_1	0.689
μ_1	-0.001	c_1	-0.160
ν_1	0.090	d_1	0.110

The two spline functions from this problem are,

$$\begin{cases} s_0(x) = a_0 + b_0(x-1) + c_0(x-1)^2 + d_0(x-1)^3 \\ s_1(x) = a_1 + b_1(x-3) + c_1(x-3)^2 + d_1(x-3)^3 \end{cases} \tag{39}$$

Fig. (5) Shows the piecewise cubic spline curve for this problem.

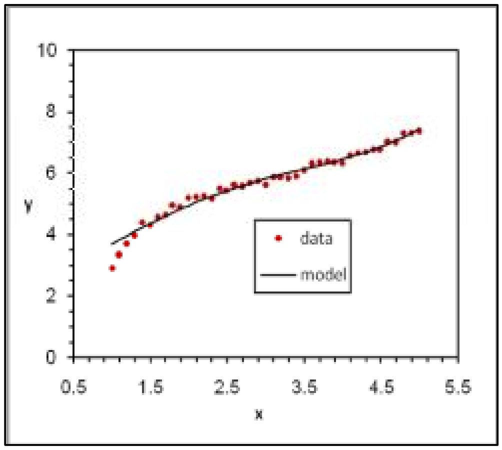

Figure 5. Cubic spline curve for n=2.

The small deviation of the cubic spline curve from the first element data is because of the effect of boundary condition on the coefficient c_{-1}. However with increasing one element to the above elements that deviation can be removed. This is shown in Fig. (6). Hence with increasing the number of control points by one the approximation model becomes more compatible to the data.

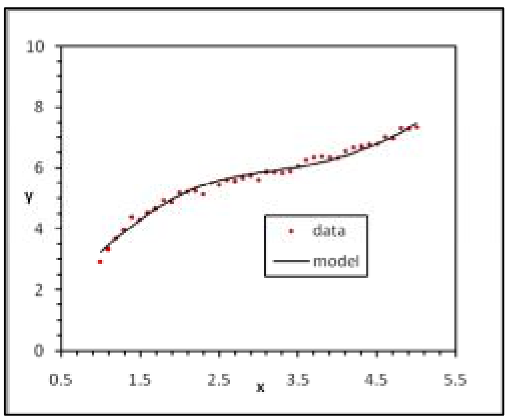

Figure 6. Cubic spline curve for n=3.

6. PROBLEM 2

In this example about 81 data pairs are generated by Eq. (40) in the interval $x \in [0,8]$.

$$f(x) = 0.05x^3 e^{-(x^2-7x+10)} + 0.5x^2 + e \tag{40}$$

Where e is the uniform random number which is obtained similar to the previous problem for each x_i. About 4 elements or 5 control points as $\ell_0 = 0$, $\ell_1 = 2$, $\ell_2 = 3.4$, $\ell_3 = 6$ and $\ell_4 = 8$ are considered. The elements lengths are nearly uniform with sizes $h_0 = 2$, $h_1 = 1.4$, $h_2 = 2.1$ and $h_3 = 2$. The values of 0.5 and 0.7 are assumed for c_{-1} and c_4 , respectively. Fig. (7) shows the governing piecewise cubic spline curve for the above control points. As it can be seen from the figure the approximation solution to the data is not accurate and suitable. This is because of the highly variations of data, also insufficient number of control points. The sum of square errors for this case is 776. If the problem for the second case is solved for 8 elements and 9 control points the approximated model will be satisfactory and follows completely the variations of data. The element length for the latter case is uniform and it equals 1.0. The piecewise cubic spline curve for this case is in Fig. (8).

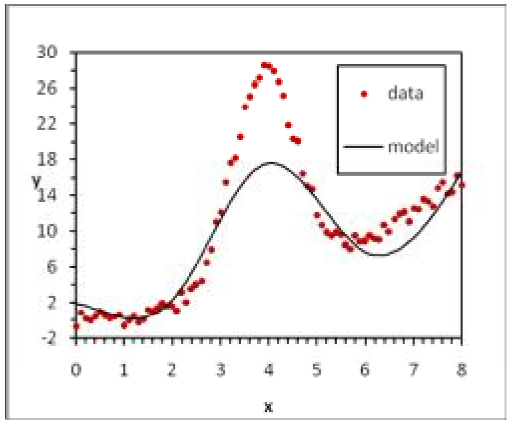

Figure 7. Cubic spline method for n=4.

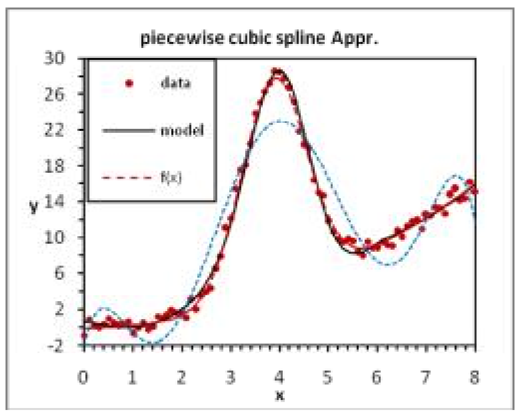

Figure 8. Cubic spline method for n=8.

As it can be seen on Fig. (8) the presented model defines a suitable and acceptable approximation curve for data of high variability. The piecewise cubic spline curve has a close relationship to the real function $f(x)$. The SSE value for 8 elements decreases from 776 to 47. The dashed line curve in Fig. (8) is generated by the polynomial of degree 6 approximation. It is not a suitable model for those data. The polynomial curve has more oscillations respect to the data and the presented model.

7. PROBLEM 3

This example consists of 46 data pairs. They are generated by the Eq. (41) in the interval $x \in [0.8, 3.97]$.

$$f(x) = -\frac{100}{(x-5)^2} Sin \frac{10}{x-5} \tag{41}$$

About 8 elements and 9 control points at $\ell_0 = 1.032$, $\ell_1 = 1.414$, $\ell_2 = 1.861$, $\ell_3 = 2.534$, $\ell_4 = 2.78$, $\ell_5 = 3.185$, $\ell_6 = 1.861$ and $\ell_7 = 3.967$ are considered. The element lengths are not uniform and vary from 0.245 to 0.673.

The variations of data are complex and for solution of this problem it requires the smaller and nonuniform elements. The obtained global matrix has dimension 28×28 and for including the boundary conditions effects c_{-1}=1.4 and c_8=0.7 are considered. Fig. (9) shows the cubic spline curve for this example. As the figure illustrates by applying only 7 elements to the problems, a satisfactory and an acceptable solution is obtained. The slight deviations of the final curve from the data can be omitted by adding one or two more elements to the problem similar to the previous examples. The dotted line curve in the figure is the approximation of data by Polynomial function of degree 5. Hence the polynomial estimation for this problem is inapplicable and unsatisfactory.

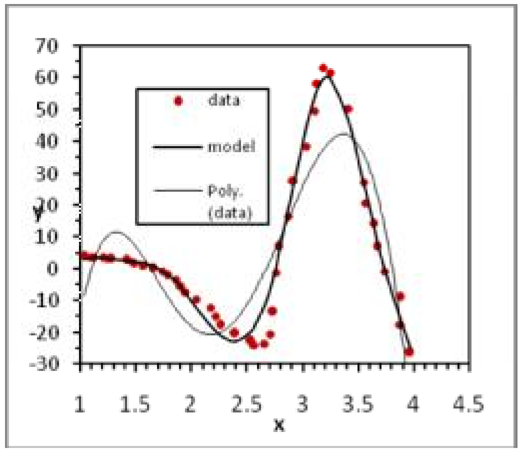

Figure 9. Piecewise cubic spline method for n=7.

8. CONCLUSIONS

The model which is developed in this research is the combination of two concepts, piecewise cubic spline interpolation and optimization. It has the possibility, applicability and suitable efficiency for approximation of engineering data which are obtained from laboratory and field tests or statistical concepts. The implementation of boundary conditions in the model is simple. By using the suitable elements lengths and element number, the model can be applied for estimation of complex data variations. The final curves obtained in this method are smooth and follow the data variations. The calculation of local matrices and collecting them into the global linear system of equations is not straightforward and it needs to transfer the formulation into a powerful programming like FORTRAN or C language. The sum of square errors in the model decreases with increasing the element number and choosing suitable positions of the control points. For the future studies in this subject it is recommended to use optimization technique for choosing the position of control points. Also the model and its formulation are expandable to two-dimensional approximation.

REFERENCES

[1] Burden, R. and Fairs, J. (2010). Numerical Analysis, 9th Ed. Brooks-Cole Pub. Comp. Boston.
[2] Collins, G. W. (2003). II Fundamental Numerical Methods and Data Analysis, (internet pdf file).
[3] Dierckx, P. (1993.) Curve and Surface Fitting with splines, Clarendon Press, NY.
[4] Dubin, D. (2003). Numerical and Analytical Methods for Scientists and Engineers Using Mathematica, John Wiley & Sons, Inc., Pub., Hokoben, New Jersy
[5] Fomel, S. (1997b). On the General Theory of Data Interpolation, SEP. 94, 165-179.
[6] Kiusalaas, J. (2005). Numerical Methods In Engineering with MATLAB, Cambridge Univ. Press,UK.
[7] Lobo, N. (1995). Curve fitting using spline sections of different orders, Proceedings of the First International Mathematics Symposium, P. 267-274.
[8] Nocedal, J. and Wright, S. J. (2006). Numerical Optimization, 2nd Ed., Springer, Science+Business Media, LLC, USA.
[9] Salamon, D. (2006). Curves and Surfaces for computer Graphics, Springer Science+Business Media, Inc. USA.
[10] Saxena, A. and Sahay B. (2005). Computer Aided Engineering Design, Springer, Anamaya Publishers, New Delhi, India.
[11] Zamani, M. (2009 a). Three simple spline methods for approximation and interpolation of data, Contemporary Engineering Science, Vol.2, No.8, P. 373-381.
[12] Zamani, M. (2009 b). An investigation of B-spline and Bezier methods for interpolation of data, Contemporary Engineering Science, Vol.2, No.8, P. 361-371.

[13] Zamani, M. (2010). A simple 2-D interpolation model for analysis of nonlinear data, Natural Science, Vol. 2, No. 6, June 2010, P. 641-646.
[14] Zamani, M. (2011). Approximation by a kind of B-spline Method, Indian Journal of Science and Technology, Vol. 4. No.2, P. 19-22.
[15] Zamani, M. (2012). A simple Piecewise Cubic Spline method for Approximation of Highly nonlinear Data, Natural Science, Vol. 4. No. 1, P 79-83.

BIBLIOGRAPHY OF AUTHORS

First author's Photo (3x4cm)	Xxxx (9 pt)