THE COMPUTATION OF SPLINES AND THE SOLUTION OF RELATED EQUATION SYSTEMS

by

G. E. McMaster

A Thesis
Submitted in Partial Fulfilment
of the Requirement of

DOCTOR OF PHILOSOPHY

at

The University of Manitoba
Winnipeg, Manitoba

Department of Computer Science

"THE COMPUTATION OF SPLINES AND THE SOLUTION OF RELATED EQUATION SYSTEMS"

by

G. E. McMaster

A dissertation submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisor, Dr. W. D. Hoskins, without whom the production of this thesis would have been impossible. His vast knowledge in the area of mathematics and computation and his infectious enthusiasm proved invaluable. More important, perhaps, were his interest, guidance, encouragement, and availability when they were needed most.

I would like to thank the members of my committee: Dr. R. G. Stanton, Dr. D. W. Trim and Dr. D. D. Cowan.

I wish to thank Shirley Blackwood not only for her careful typing of the thesis but for her constant good nature and her willingness to always type that one extra page. I would also like to thank Miss Eileen Robbins for help in typing the final draft.

Anyone who has ever written a thesis knows what an enormous burden this task imposes on ones family. My wife, Beverly, suffered my constant attention to the thesis with the utmost of goodwill and patience. This thesis is Beverly's, Jeffrey's, Luke's and Jonathan's as much as it is mine.

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Goals of the Thesis

The problem of spline interpolation and smoothing falls quite naturally into five main sections:

(a) the formation of the system of linear equations defining the coefficients of the basis functions used in the spline representation,

(b) the examination of properties of the coefficient matrix of this linear system,

(c) the solution of this linear system,

(d) the evaluation of the spline for various values of the argument,

(e) applications that use a spline representation to advantage.

In this thesis, results are obtained primarily in areas (a), (b), (c), and (e).

1.2 The Numerical Evaluation of Splines

A large number of equivalent mathematical descriptions of the polynomial spline are extant in the literature. Representative of these different forms are the works of Greville [1969]; Cox, [1971, 1972, 1973] with the use of B-splines as basis functions for spline interpolation; Ahlberg, Nilson and Walsh [1967] and their use of both divided differences and Hermite Interpolation to derive so-called consistency equations between derivatives; Fyfe [1971] with cardinal spline forms; Späth [1970] with Lidstone polynomials; and Golomb [1968] employing Bernoulli polynomials. Mathematically, all forms give exactly the same spline; computationally there is a wide variation with respect to the properties possessed and the condition number of the resulting set of simultaneous equations. Numerically, it has always been distressing that the forms with the fewest number of parameters, e.g., the representation used by Curtis and Powell [1967], should always be the worst to use from the point of view of rounding errors (c.f. Cox [1971] and Cox [1972]) and that well-conditioned forms should involve too many redundant parameters and large equation systems (Späth [1969]).

The B-spline or basis spline, has been proposed by a number of authors (Anselone and Laurent [1968]; Cox [1971]; Herriot and Reinsch [1971]; Lafata and Rosen [1970]; and Schumaker [1969]) as a convenient basis for problems of interpolation and smoothing. In forming the linear algebraic equations defining the multipliers of the basis functions and in evaluating the subsequent approximating spline, it is necessary to employ an algorithm for evaluating the B-spline. Cox [1972] and de Boor [1973] have obtained,independently, methods for B-spline evaluation that are numerically stable and economical.

1.3 *Outline of the Thesis*

In Chapter 2, background results on the B-spline are presented along with recent theorems which permit in Chapter 3 the formulation of systems of linear equations for both smoothing and interpolating splines. As well, an economical method for the least squares evaluation of a multivariate spline is given. In Chapter 3, algorithms for smoothing and for interpolation are given. The format used for the presentation of algorithms in this thesis closely approximates that of Wilkinson and Reinsch [1971]. Very briefly, the format followed is:

(a) The computer programs that supplement the derived mathematical algorithms are presented in Algol W (Hoare et al [1966]).

(b) The theoretical development giving the mathematical basis for the algorithm is given first. If a competitive published routine exists to solve part of the problem, then it is used, and only the reference is given.

(c) The formal parameter list giving all the input and output parameters for the main procedures is given.

(d) Organizational and notational details explaining unusual features of the algorithm such as storage techniques used or interesting testing procedures are given where necessary.

Error analysis, in general, is not included since, for the methods of solution used, detailed error analyses already exist and the solution methods can be proven stable. In the case of B-spline evaluation, Cox [1972] presents a rigorous error analysis; for the solution of the system of band equations, Wilkinson [1963, 1965] and Wilkinson and Reinsch [1971] give a complete error analysis. In the testing of the smoothing spline in Chapter 3, an interesting forward error analysis (Cody [1973]) is used, and is described in detail.

In Chapter 3, the solution of second-order linear differential equations using cubic splines is examined. In Chapter 4, new decoupling techniques for the rapid solution of systems of band equations resulting from spline representations are presented. The algorithms are competitive in a serial computer system, but are more effective in a particular parallel processing environment. Different concepts relating to parallel processing have been investigated (Flynn [1966, 1972]). Previous direct methods for the solution of tridiagonal linear systems using parallel processing (Stone [1973a], Kogge and Stone [1972]) were directed to SIMD computer systems (single-instruction-stream multiple-data-streams) Traub [1973].

The methods in Chapter 4 adapt well to an MIMD or multiple-instruction-steam multiple-data-stream parallel processing system for which algorithms seem to be difficult to obtain (Stone [1973b]).

In order to determine the effectiveness of the decoupling algorithms in an MIMD environment, an MIMD speed-up coefficient α is defined: let n be the total member of arithmetic operations in the algorithm and m the number that can proceed in parallel, then

$$\alpha = n/(n-m/2)$$

The speed-up factor is evaluated for the variations in the decoupling algorithms for the solution of tridiagonal systems in Chapter 4 and is found in most cases to be approximately 2 for large n. The general polydiagonal decoupling routines employ an extension of the technique used in the tridiagonal case, and similar economies can be expected.

In Chapter 5, properties of some classes of coefficient matrices that arise in the solution of systems of equations determining spline parameters are investigated. The analysis used is then extended to obtain properties of related matrices.

Chapter 2

B-Spline Function

2.1 *Introduction*

The spline function is a piecewise polynomial function that has excellent approximating properties, tends to be smoother and more flexible to use than a polynomial and usually provides better approximating properties (Greville [1969], de Boor [1963]). If the function being approximated is smooth, then spline functions are likely to give better estimates of the low-order derivatives than polynomials (Späth [1974]).

In this thesis, the determination and the evaluation of polynomial splines of odd degree $2r+1$ is examined. A spline function of degree $2r+1$ defined on n given knots $x_1 < x_2 < \ldots < x_n$ is a function $S(x) \varepsilon C^{2r}$ such that $S(x) \varepsilon S(x_1, x_2, \ldots, x_n)$, the class of polynomials of degree at most $2r+1$ in each of the intervals in the set $I = \{(-\infty, x_1), (x_1, x_2), \ldots, (x_n, \infty)\}$. In the practical application of spline functions, a finite range $a \le x \le b$ is almost always used,and hence $x_1 > a$ and $x_n < b$. In order to determine $S(x)$, first note that $S(x)$ has $n + 2r + 2$ parameters. The $n+1$ polynomials defined on I contain $(2r+2).(n+1)$ undetermined constants; however, $n(2r+1)$ of these constants are determined by the continuity requirements on $S(x)$, i.e. $S(x) \varepsilon C^{2r}$. The additional constants can be determined either by various interpolatory requirements or in a least squares sense.

In this chapter, we consider the spline as a conventional interpolating function; spline interpolation using additional information, such as values of the derivatives at the ends of the interval of interest, is considered in Chapter 3.

There are many ways (cf. Chapter 1) for representing a polynomial spline; however, if the spline is expressed as a linear combination of B-splines, then stable and efficient computational algorithms can be generated (Greville [1972], Cox [1972], de Boor [1973]). The B-spline was first introduced for the uniform partition by Schoenberg [1946] and for the non-uniform partition by Curry and Schoenberg [1966].

In Section 2.2, basic properties of the B-spline are given, along with requirements for the definition of the underlying knot set. An efficient algorithm (Cox [1972]), for B-spline evaluation is outlined which is used in an L_2 algorithm in Chapter 3. An integral result for the product of B-splines on a uniform mesh, useful for smoothing periodic data sets, is obtained in Section 2.3. In Section 2.4, an economical method for determining the coefficients of a multivariate B-spline representation for interpolation or for least squares curve fitting is obtained.

2.2 *The Basis Spline*

Most formulations of spline problems tend to give rise to ill-conditioned systems of linear equations (Greville [1969], Cox [1971]). Problems in solving the system are aggravated when the degree of the spline is increased and when there are many knots in the partition.

For example, it may be readily demonstrated that $S(x)$ is uniquely represented (Greville [1968]) by the two sets of parameters $P(\underline{x}) = (x_1, x_2, \ldots, x_n)$ and $C = (c_1, c_2, \ldots, c_{n+2r+2})$ where

$$S(x) = \sum_{i=1}^{n} c_i \, (x - x_i)_+^{2r+1} + \sum_{i=0}^{2r+1} c_{n+i+1} \, x^i$$

and

$$x_+ = \begin{cases} x & \text{when } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This representation, although useful for purposes of mathematical analysis, leads to an ill-conditioned system of equations for the determination of the c_i, and is unwieldy to evaluate.

A very desirable representation for the spline is one whose support is finite and whose basis functions require a minimal number of knots in their definition.

The basis spline or B-spline of degree $2r+1$ (order $2r+2$) is non-zero over $2r+2$ consecutive intervals between knots (hence the nomenclature, spline of minimum support); $2r+2$ is the smallest number of intervals over which a spline of degree $2r+1$ can be non-zero. The B-spline is local in the sense that at any point only k B-splines, where k is equal to the order, are non-zero. These properties permit the representation of a spline in terms of B-splines in a stable numerically compact form (Cox [1972]). The forward B-spline (Schoenberg [1973]) $M_{2r+2,i}(x, P(\underline{x}))$ is the spline of degree $2r+1$ specified by the knot set $P(\underline{x}) = \{x_1, x_2, \ldots, x_n\}$. The knot set is specified in the form $P(\underline{x})$ to emphasize that the knots are chosen within the range of the given data by the curve fitter using a general knowledge of the shape of the underlying curve as indicated by the data and by trial and error. In general, more knots are required in those regions where the behaviour of the curve is changing rapidly and fewer knots where it is changing slowly; however, the exact positioning of the knots is often not critical (Cox and Hayes [1973]). With a little experience, satisfactory knot positions can be found after one or two trials.

The B-spline $M_{2r+2,i}(x, P(\underline{x}))$ may be formally defined as follows (de Boor [1973]). Let

$$x_+^{2r+1} = \begin{cases} x^{2r+1} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

and

$$M_{2r+2}(x;y) = (y-x)_+^{2r+1} .$$

Then $M_{2r+2,i}(x, P(\underline{x}))$ is the divided difference of order $2r+2$ of $M_{2r+2}(x;y)$ with respect to the variable y based on the arguments $x_{i-(2r+2)}, x_{i-(2r+1)}, \ldots, x_i$. The evaluation of $M_{2r+2,i}(x, P(\underline{x}))$ through the use of divided differences leads to an unstable evaluation procedure and another technique (de Boor [1973], Cox [1972]) will be used. It is evident from the definition, however, that $M_{2r+2,i}(x, P(\underline{x}))$ is zero everywhere except in the $2r+2$ intervals in the range $R = \{x_{i-(2r+2)} < x < x_i\}$ and is uniquely determined (Cox [1972]), using the $2r+3$ knots defining R, except for a constant multiplier. The sign of the constant multiplier may be chosen to make the B-spline $M_{2r+2,i}(x, P(\underline{x}))$ positive on R. $M_{2r+2,i}(x, P(\underline{x}))$ may be shown to have a single maximum in R, and it and its derivatives up to the $2r$'th are zero at the end points of R, i.e., at $x = x_i$ and $x = x_{i-(2r+2)}$.

Since each B-spline spans $2r+2$ adjacent intervals (the order of the B-spline), then the knot set $P(\underline{x})$ determines $n-(2r+2)$ different

B-splines provided that $n > 2r+2$. The spline representation defined on the knot set $P(\underline{x})$ involves $n+2r+2$ degrees of freedom and requires $n+2r+2$ independent B-splines. It is then necessary to add $2r+2$ artificial knots to augment the given knot set at or outside each end of the given range of interest $[a,b]$ giving a total of $n+2r+2$ knots. For computational convenience, it is possible (Cavasso and Laurent [1969]) to place these extra knots at the appropriate end points, namely,

$$x_{-2r+1} = x_{-2r+2} = \dots = x_0 = a$$

(2.2-1)

$$x_{n+1} = x_{n+2} = \dots = x_{n+2r+2} = b$$

giving knots of multiplicity $2r+2$ at both a and b . The discontinuities that this arrangement introduces are at the end points of the range of interest,and so are of no concern. The $n+2r+2$ B-splines are then non-zero only in the range $a < x < b$.

If the spline is to represent a periodic data set $y = (y_v)$ of period r where

(2.2-2) $$y_m = y_k \quad \text{if} \quad m \equiv k \bmod (r) ,$$

then the knot set may be extended in an obvious manner using the spacing of the original x_i . This method for extending the knot set is assumed in Chapter 3 in order to obtain a periodic L_2 B-spline for smoothing purposes.

If the spline $S(x)$ of order $2r+2$ with the prescribed knot set $P(\underline{x}) = \{x_1, x_2, \dots, x_n\}$ is to interpolate to the function $f(x)$ at $x = t_1, t_2, \dots, t_p$, then it is assumed that the elements of the given set of nodes and the user-defined knot set $P(\underline{x})$ are strictly ordered, that is:

$$t_1 < t_2 < \dots < t_p$$

(2.2-3)

$$t_1 < x_1 < x_2 < \dots < x_n < t_p \ .$$

It is usual to assume that $a = t_1$ and $b = t_p$.

The $n+2r+2$ degrees of freedom remaining in $S(x)$ may be reduced by applying the interpolation conditions

(2.2-4) $$S(t_i) = f(t_i); \quad i = 1, 2, \dots, p \ .$$

To ensure that $S(x)$ be determined uniquely, we require that the number of given nodes, the number of selected knots, and the order of the spline be related by

(2.2-5) $$p = n + 2r + 2 \ .$$

If the number of given nodes p is greater than $n+2r+2$, then the spline $S(x)$ may be determined using the method of least squares.

In order to ensure a unique $S(x)$, the specified knots $P(\underline{x})$ must be chosen to satisfy the Schoenberg-Whitney [1953] conditions

(2.2-6)
$$\begin{aligned} t_1 &< x_1 < t_{1+2r+2} \\ t_2 &< x_2 < t_{2+2r+2} \\ &\cdot \quad \cdot \quad \cdot \\ t_n &< x_n < t_p \end{aligned}$$

which ensures that each B-spline in the representation $S(x)$ has one node in its range of definition.

The evaluation of a B-spline $M_{2r+2,i}(x, P(\underline{x}))$ may be effected using a stable recurrence relation given in detail in Cox [1972] or de Boor [1973], namely

$$M_{r,i}(x, P(\underline{x})) = \frac{(x-x_{i-r})\,M_{r-1,i-1}(x, P(\underline{x})) + (x_i - x).M_{r-1,i}(x, P(\underline{x}))}{(x_i - x_{i-r})}$$

(2.2-7)

commencing with

$$M_{1,i}(x, P(\underline{x})) = \begin{cases} 1/(x_i - x_{i-1}), & x_{i-1} \le x < x_i \\ 0 & \text{otherwise.} \end{cases}$$

The expression on the right-hand side of (2.2-7) is the convex sum of two positive values which gives, in the main, the stability to the B-spline evaluation.

The recurrence relation (2.2-7) is valid for coincident knots, provided that there are no more than 2r+1 coincidences (the degree of the spline) at any knot. This permits the spline $S(x)$ to have reduced continuity at one or more points in the range of interest $[a,b]$. This form for the spline is called a deficient spline.

For improved numerical stability in the B-spline evaluation, it is preferable (Hayes [1974b]) to use the normalized B-spline (de Boor [1972]) defined as

(2.2-8) $$N_{2r+2,i}(x, P(\underline{x})) = (x_i - x_{i-(2r+2)})\,M_{2r+2,i}(x, P(\underline{x})) \quad .$$

The normalized B-spline may be computed from the given recurrence relation for the M's by omitting the final division by $x_i - x_{i-(2r+2)}$.

The spline $S(x)$ on $[a,b]$ may then be expressed uniquely using the $n+2r+2$ normalized B-splines defined on the augmented knot set as

(2.2-9) $$S(x) = \sum_{i=1}^{n+2r+2} c_i \cdot N_{2r+2,i}(x, P(\underline{x})) ,$$

the c_i being constant. In order to determine the c_i, the interpolation condition $S(t_j) = f(t_j)$, where $j = 1, \ldots, p$, may be applied to give the linear system of equations

(2.2-10) $$\sum_{i=1}^{n+2r+2} c_i \, N_{2r+2,i}(t_j, P(\underline{x})) = f(t_j),$$

where $j = 1, 2, \ldots, p$.

The linear independence of the B-spline functions $N_{2r+2,i}(x, P(\underline{x}))$ and the restriction that the user-specified knot set $P(\underline{x})$ satisfy the Schoenberg-Whitney [1953] conditions, ensures a unique solution to the system of linear equations (2.2-10) (Cox [1974]).

If $p > n+2r+2$, then the coefficients c_i in the system of equations (2.2-10) may be obtained in a least squares manner by way of the normal equations. The system of equations to be solved may be represented as

(2.2-11) $$NN^*\underline{C} = \underline{F}$$

where $F = \{f(t_j)\}$, $C^T = \{c_1, c_2, \ldots, c_{n+2r+2}\}$ and the elements n_{ij} of N are given by

(2.2-12) $$n_{ij} = N_{2r+2,i}(t_j, P(\underline{x})) .$$

The system of equations (2.2-11) may then be solved by Gaussian elimination. To ensure a unique solution to (2.2-11), at least one of the p given knots must be in the range of definition of each of the $N_{2r+2,i}(x, P(\underline{x}))$ (Hayes and Halliday [1974]).

The least squares solution to (2.2-10) may be obtained more stably by the use of Householder reductions of the matrix NN^* (Bunsinger and Golub [1965]). This is at the cost of nearly doubling the amount of computation.

If the interpolatory spline of degree $2r+1$ defined on the knot set $P(\underline{x})$ is given by some polynomial of degree r or less in each of the intervals $(-\infty, x_1)$, $(x_n, +\infty)$ and the knot set is taken as the given nodes, then a natural spline (Greville [1969]) is obtained. In this circumstance, the coefficient matrix of the linear system to be solved is of strict band style with the non-zero elements appearing on the diagonal band of width $2r+1$. Algorithms for the solution of such linear systems are developed in Chapter 4 to obtain the parameters of this frequently used spline representation.

Finally, we mention Marsden's identity (Marsden [1970]) which permits a polynomial of degree n to be expressed in terms of B-splines. This identity is

$$(2.2\text{-}13) \qquad (u-x)^{k-1} = \sum_i \phi_{i,k}(u) \cdot N_{k,i}(x, P(\underline{x}))$$

where

$$\phi_{i,k}(u) = \prod_{r=1}^{k-1} (u-i-r) \quad .$$

This result is employed in Chapter 3 to inexpensively generate test data to validate a given B-spline representation, since a B-spline of degree m must exactly represent polynomials of degrees $0, 1, 2, \ldots, m$.

2.3 *Integral of the Product of Two B-Splines on a Uniform Mesh*

If the given knot set is assumed to be uniform, then there is no loss in generality in assuming that the B-spline is defined on the integer knots. One formulation for a B-spline of degree N or order $N+1$ is in terms of the backward difference of a truncated power function (Schoenberg and Curry [1966]). This definition gives a forward B-spline (Schoenberg [1973]) and, using the notation of Meek [1974], may be expressed as

(2.3-1) $$Q_{N+1}(x) = \frac{1}{N!} \nabla^{N+1} x_+^N$$

where

$$x_+ = \begin{cases} x & \text{for} \quad x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and where ∇ is the usual backward difference operator defined as

$$\nabla f_x = f_x - f_{x-h} \; .$$

The value h is the interval of differencing and, in this case, is assumed to be 1. Many of the useful computational properties of $Q_{N+1}(x)$ are summarized in Meek [1974]. A further result that enables an L_2 computational technique to be expressed in terms of the general consistency equations obtained by Fyfe [1971] concerns the integral of the product of two forward B-splines.

Theorem 2.3.1

(2.3-2) $$\int_{-\infty}^{\infty} Q_{N+1}(x-j)\, Q_k(x-\ell)\, dx = Q_{k+N+1}(N+1-\ell+j)$$

where j and ℓ are both integers and $N+1 > k$.

Proof:

The left hand side of equation (2.3-2) may be written as

$$L = \int_{-\infty}^{\infty} Q_{N+1}(t)\, Q_k\,(t-s)\, dt$$

with $s = \ell - j$. On using the definition of Equation (2.3-1),

$$L = \sum_{j=0}^{N} \int_{j}^{j+1} Q_{N+1}(t)\, Q_k\,(t-s)\, dt \quad . \tag{2.3-3}$$

Since $Q_{N+1}(t)$ is a polynomial of degree N in $[j, j+1]$, namely,

$$Q_{N+1}(t) = \frac{1}{N!}\, \nabla^{N+1}\, t_+^N$$

$$= \frac{1}{N!} \sum_{p=0}^{j} (-1)^p \binom{N+1}{p} (t-p)^N \ ,$$

where $t \in [j, j+1]$

it follows that the N^{th} derivative of $Q_{N+1}(t)$ is

$$Q_{N+1}^{(N)}\,(t) = \sum_{p=0}^{j} (-1)^p \binom{N+1}{p} \ , \quad t \in [j, j+1] \quad . \tag{2.3-4}$$

Equation (2.3-3) can be integrated $N+1$ times by parts to give

$$L = (-1)^N \sum_{j=0}^{N} \left[\frac{1}{(N+k)!}\, \nabla_t^k\, (t-s)_+^{N+k}\, Q_{N+1}^{(N)}\,(t) \right]_{t=j}^{t=j+1}$$

where ∇_t is the backward difference operator acting on the variable t . From Equation (2.3-1),it may be seen that $Q_{N+1}^{(N)}\,(t)$ is a constant in the interval $[j, j+1]$; so it is convenient to denote it by $Q_{N+1}^{(N)}\,(j)$ where $Q_{N+1}^{(N)}\,(-1)$ is defined as zero. Then the above expression may be rewritten in the form

$$L = \frac{(-1)^N}{(N+k)!} \sum_{j=0}^{N} Q_{N+1}^{(N)} (j) \left[\nabla_t^k (t-s)_+^{N+k} \right]_{t=j}^{t=j+1}$$

and rearranged as

$$L = \frac{(-1)^N}{(N+k)!} \sum_{j=0}^{N+1} \left[Q_{N+1}^{(N)} (j-1) - Q_{N+1}^{(N)} (j) \right] \nabla_j^k (j-s)_+^{N+k} \quad .$$

However, Equation (2.3-4) gives, on substitution in the above expression,

$$L = \frac{(-1)^N}{(N+k)!} \sum_{j=0}^{N+1} (-1)^{j+1} \binom{N+1}{j} \nabla_j^k (j-s)_+^{N+k}$$

$$= \frac{1}{(N+k)!} \nabla_t^{N+k+1} (t-s)_+^{N+k} \Big|_{t=N+1} \quad .$$

From the definition (2.3-1), it follows that

$$L = Q_{N+k+1} (N+1-s) \quad . \qquad \text{Q.E.D.}$$

2.4 A Method for Obtaining the Coefficients of a Multivariate Spline

In this section, a general economical method for solving the system of equations defining a multivariate spline for interpolation or for surface representation is presented. A summary of recent advances in surface representation to which the method applies is given. The derived solution technique possesses definite computational savings over previous methods (Hayes and Halliday [1974], Hayes [1974a, 1974b], Späth [1974], Ahlberg et al [1967]) provided that the system of equations defining the spline parameters is not ill-conditioned or the coefficient matrix is not deficient.

We first examine methods for obtaining a bivariate spline representation where varying assumptions are made concerning the defining knot set. It is assumed that discrete data are given, that they may or may not contain random errors, and that these data are to be smoothed or fitted exactly. If the given data do contain errors, then these errors are assumed to be contained in the dependent variable. The surface representation methods considered here do not deal with the very different question of approximating mathematical functions where the value of the function for any values of the argument can be made available to any desired accuracy. Many publications dealing with cubic splines defined on two variables have appeared; however, these papers have largely concentrated on those interpolation problems in which the given data are known at the nodes of a rectangular mesh. This case is considered initially.

The bivariate spline is defined over a rectangular grid R specified by the partitions $P(\underline{x}) = \{x_1, x_2, \ldots, x_n\}$ and $Q(\underline{y}) = \{y_1, y_2, \ldots, y_m\}$. One of the rectangles R_{ij} in the grid may be defined as

(2.4-1) $$R_{ij} = \left\{ \begin{matrix} x_i \le x \le x_{i+1} \\ \\ y_j \le y \le y_{j+1} \end{matrix} \right\}$$

It is usual to treat a finite domain where $a < x < b$ and $c < y < d$.

To compute the coefficients of the bivariate spline, assume that the following data points are given,

(2.4-2) $$f(t_i, q_j) \quad (i = 1, \ldots, p; \; j = 1, \ldots, v)$$

where the t_i are defined in the x direction, the q_j in the y direction.

We require a method for the computation of a surface $s(x,y)$, defined on R, that either interpolates the values $f(t_i, q_j)$ or represents these values in a least squares sense and is such that $s(x,y) \varepsilon C^{2r,2r}$. To represent the general bivariate spline, a set of basis functions is required as in the case when the B-splines are used to represent the one-dimensional case. Such a set for the bivariate spline may be constructed mathematically from the tensor product of two sets of independent B-splines (de Boor [1962]), one in the x-direction, the other in the y-direction. The set of all cross products formed using functions from each set provides the basis functions for the bivariate spline. Thus it is necessary to augment the partition in the y direction as was done in the x direction (2.2-1). The augmented partition for the y variate is

(2.4-3)
$$y_{-2r+1} = y_{-2r+2} = \dots = y_0 = c$$
$$y_{m+1} = y_{m+2} = \dots = y_{m+2r+2} = d \quad .$$

The given nodes and the user-defined knots must satisfy the Schoenberg-Whitney [1953] conditions in both the x and y directions, that is, each B-spline defined on either the x or the y variate must have a node within its non-zero range of definition. The bivariate spline may then be defined uniquely on R (Hayes [1974b]) as

(2.4-4)
$$S(x,y) = \sum_{i=1}^{n+2r+2} \sum_{j=1}^{m+2r+2} a_{ij} N_{2r+2,i} (x,P(\underline{x})) . N_{2r+2,j} (y,Q(\underline{y})) \quad .$$

In the interpolatory case, $p = n+2r+2$ and $v = m+2r+2$ and the coefficients a_{ij} are determined by the $p.v$ equations

(2.4-5)
$$(t_i, q_j) = f(t_i, q_j) \qquad (i = 1, \dots, p; \; j = 1, \dots, v) \quad .$$