NPL Report CMSC 13/02

Report to the National Measurement System Policy Unit, Department of Trade and Industry

From the Software Support for Metrology Programme

Fixed- and free-knot univariate least-squares data approximation by polynomial splines

By
M G Cox, P M Harris and P D Kenward

May 2002

Fixed- and free-knot univariate least-squares data approximation by polynomial splines

M G Cox, P M Harris and P D Kenward
Centre for Mathematics and Scientific Computing

May 2002

ABSTRACT

Fixed- and free-knot least-squares data approximation by polynomial splines is considered. Classes of knot-placement algorithms are discussed. A practical example of knot placement is presented, and future possibilities in free-knot spline approximation addressed.

ISSN 1471–0005

Authorised by Dr Dave Rayner,
Head of the Centre for Mathematics and Scientific Computing

National Physical Laboratory,
Queens Road, Teddington, Middlesex, United Kingdom TW11 0LW

Contents

1 Introduction

Empirical models are important to metrology in cases where the knowledge of the underlying physics for a measurement system is insufficient to characterise it completely. For empirical models depending on one variable, polynomial and particularly *polynomial spline curves*, when used with care, are generally very satisfactory for representing data. A polynomial spline curve is composed of a sequence of polynomial curves joined together at points called *knots* and in such a way as to ensure smoothness of the complete curve.

Spline curves provide a flexible class of functions that are effective for representing a wide variety of shapes. However, the knots generally have no physical meaning for the metrologist, and yet the effectiveness of a spline representation can depend critically on their number and positions. Consequently, metrology users require assistance with knot placement via appropriate algorithms and software.

Automating the choice of the number and positions of the knots is generally a very difficult problem. However, a number of knot placement strategies are available to help the user select a sensible, if not necessarily optimal, set of knots. The strategies are broadly classified as those that

- position knots according to the *distribution* of the data abscissae (and possibly data ordinates),
- sequentially *insert* knots in order to improve (maximally) the quality of the spline fit to the data (measured using some metric), and
- sequentially *delete* knots in order to (minimally) change the quality of the fit.

The aim of this report is to describe a number of these knot placement strategies and to illustrate their application to metrology data. Examples of strategies of each of the above types are considered. A software package[1], containing implementations of some of these strategies is available through MetroS[2] [2]. The software package acts as a pre-processor to NPLFit[3], software developed by NPL for modelling experimental data using polynomial and spline curves, and also available through MetroS.

The report is organised as follows. The representation of univariate polynomial splines in terms of B-splines is reviewed (Section 2), and the problem

[1] See http://www.npl.co.uk/ssfm/metros/packages/KnotPlacement/
[2] See http://www.npl.co.uk/ssfm/metros/
[3] See http://www.npl.co.uk/ssfm/metros/packages/nplfit/

of obtaining fixed- and free-knot least-squares spline approximations to data described (Section 3). The accepted approach to the fixed-knot case is recalled (Section 4) and the manner in which spline uncertainties can be evaluated is given (Section 5). The importance of families of spline approximants is emphasised (Section 6). The free-knot problem is formulated (Section 7) and several of the established and some lesser-known knot-placement strategies reviewed (Section 8). Conclusions are drawn and future possibilities indicated (Section 9).

2 Univariate polynomial splines

Let $I := [x_{\min}, x_{\max}]$ be an interval of the x-axis, and

$$x_{\min} = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1} \leq \lambda_N < \lambda_{N+1} = x_{\max}$$

a partition of I.

A spline $s(x)$ of order n (degree $n-1$) on I is a piecewise polynomial of order n on $(\lambda_j, \lambda_{j+1})$, $j = 0, \ldots, N$. The spline s is C^{n-k-1} at λ_j if $\mathrm{card}(\lambda_\ell = \lambda_j, \ell \in \{1, \ldots, n\}) = k$. So, for example, a spline $s(x)$ of order 4 for which the points λ_j are distinct is a piecewise cubic polynomial of continuity class C^2, i.e. continuous in value, first and second derivatives, at the points λ_j.

The partition points $\boldsymbol{\lambda} = \{\lambda_j\}_1^N$ are the (interior) *knots* of s. To specify the complete set of knots needed to define s on I in terms of B-splines, the knots $\{\lambda_j\}_1^N$ are augmented by knots $\{\lambda_j\}_{1-n}^{-1}$ and $\{\lambda_j\}_{N+2}^{q}$, $q = N+n$, satisfying

$$\lambda_{1-n} \leq \cdots \leq \lambda_0, \qquad \lambda_{N+1} \leq \cdots \leq \lambda_q.$$

For many purposes, a good choice [11] of additional knots is

$$\lambda_{1-n} = \cdots = \lambda_0, \qquad \lambda_{N+1} = \cdots = \lambda_q.$$

It readily permits derivative boundary conditions to be incorporated in spline approximants [8].

On I, $s(x)$ has the *B-spline representation* [6]

$$s(x) := s(\mathbf{c}, \boldsymbol{\lambda}; x) = \sum_{j=1}^{q} c_j N_{n,j}(\boldsymbol{\lambda}; x), \tag{1}$$

where $N_{n,j}(\boldsymbol{\lambda}; x)$ is the *B-spline* [6, 13] of order n with knots $\{\lambda_k\}_{j-n}^{j}$ and $\mathbf{c} = (c_1, \ldots, c_q)^{\mathrm{T}}$ are the *B-spline coefficients* of s. Each $N_{n,j}(\boldsymbol{\lambda}; x)$ is a spline with knots $\boldsymbol{\lambda}$, is non-negative and has compact support. Specifically,

$$N_{n,j}(\boldsymbol{\lambda}; x) > 0, \quad x \in (\lambda_{j-n}, \lambda_j), \qquad \mathrm{supp}(N_{n,j}(\boldsymbol{\lambda}; x)) = [\lambda_{j-n}, \lambda_j]. \tag{2}$$

The B-spline basis $\{N_{n,j}(\boldsymbol{\lambda}; x)\}_{j=1}^{q}$ for splines of order n with knots $\boldsymbol{\lambda}$ is generally very well-conditioned [11]. Moreover, the basis functions for any $x \in [x_{\min}, x_{\max}]$ can be formed in an unconditionally stable manner using a three-term recurrence relation [6, 13]. Specifically, the *relative errors* in the values $fl(N_{n,j}(\boldsymbol{\lambda}; x))$ of the basis function computed using IEEE floating-point arithmetic [19] satisfy

$$|fl(N_{n,j}(\boldsymbol{\lambda}; x)) - N_{n,j}(\boldsymbol{\lambda}; x)| \leq CnN_{n,j}(\boldsymbol{\lambda}; x)\eta,$$

where C is a constant that is a small multiple of unity and η is the unit roundoff of the floating point processor [6]. The B-spline basis for splines of order 3 with interior knots at $x = (1, 2, 5)^{\mathrm{T}}$ and coincident end knots at $x = 0$ and 10, is shown in Figure 1.

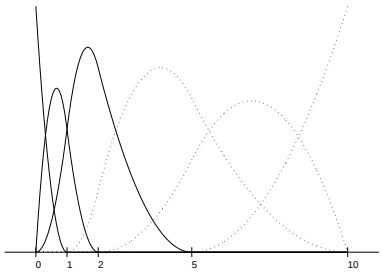

Figure 1: The B-spline basis for splines of order 3 for some nonuniformly spaced knots. The first three B-spline basis functions are shown as solid lines and the remaining three as dotted lines.

Valuable properties of s can be deduced [13] from those of the B-splines. A useful property is that, for any $x \in I$, $s(x)$ is a convex combination of the coefficients of the B-splines whose support contains x. Thus, local bounds for s can readily be found:

$$\min_{j<k\leq j+n} c_k \leq s(x) \leq \max_{j<k\leq j+n} c_k, \qquad x \in [\lambda_j, \lambda_{j+1}].$$

These bounds imply a mimicking property for s, viz., that the elements of $\mathbf{c}$ tend to vary in much the same way that s varies. Figure 2 depicts a spline curve s of order 4 with "non-polynomial" shape having interior knots at $x = (1, 2, 5)^{\mathrm{T}}$, coincident end knots at $x = 0$ and 10, and B-spline coefficients $(0.00, 0.20, 0.60, 0.22, 0.18, 0.14, 0.12)^{\mathrm{T}}$. To reproduce this shape to visual accuracy with a polynomial would require a high degree and hence many more defining coefficients. The mimicking property is evident: successive elements of $\mathbf{c}$ rise, fall sharply and then gently, behaving in a similar way to s.

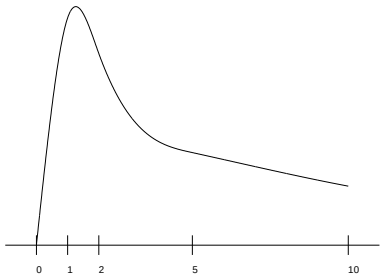

Figure 2: A spline curve with "non-polynomial" shape illustrating the mimicking property.

3 Fixed- and free-knot approximation

Two types of least-squares data approximation (or data modelling) by splines are regularly considered. One is the determination of the B-spline coefficients $\mathbf{c}$ for given data, a prescribed order n and prescribed knots $\boldsymbol{\lambda}$. The other is the determination of $\mathbf{c}$ *and* $\boldsymbol{\lambda}$ for given data and spline order n. The former problem is linear with respect to the parameters of the spline, just $\mathbf{c}$ being regarded as unknown. The latter is nonlinear, both $\mathbf{c}$ and $\boldsymbol{\lambda}$ being unknown.

The linear case is well understood, with highly satisfactory algorithms [11] and software implementations [1, 17] available. The nonlinear case remains a research problem, although useful algorithms (Section 8) have been proposed, implemented and used. Many of these algorithms "iterate" with respect to $\boldsymbol{\lambda}$, where for each choice of knots the resulting linear problem is solved for $\mathbf{c}$. Thus, the linear problem (Section 4) is important in its own right and as part of the solution strategy for knot-placement algorithms.

4 Least-squares data approximation by splines with fixed knots

The least-squares data approximation problem for splines with fixed knots can be posed as follows. Given are data points $\{(x_i, y_i)\}_1^m$, with $x_1 \leq \cdots \leq x_m$, and corresponding *weights* $\{w_i\}_1^m$ or *standard uncertainties* $\{u_i\}_1^m$. The w_i reflect the *relative* quality of the y_i,[4] u_i is the standard uncertainty of y_i and corresponds to the standard deviation of possible "measurements" at $x = x_i$ of the function underlying the data, y_i being one realisation. Given also are the N knots $\boldsymbol{\lambda} = \{\lambda_j\}_1^N$ and the order n of the spline s.

[4]The x_i are taken as exact for the treatment here. A generalised treatment is possible, in which the x_i are also regarded as inexact. The problem becomes nonlinear (in $\mathbf{c}$).

When *weights* are specified, the problem is to determine the spline $s(x)$ of order n, with knots $\boldsymbol{\lambda}$, such that the sum of squares of the elements $\{w_i e_i\}_1^m$ is minimised with respect to $\mathbf{c}$. When *standard uncertainties* are specified, the sum of squares of the elements $\{u_i^{-1} e_i\}_1^m$ is minimised with respect to $\mathbf{c}$. If $w_i = u_i^{-1}$, $i = 1, \ldots, m$, the two formulations are identical in terms of the spline produced. When weights are specified, s is referred to as a *spline approximant*. When uncertainties are prescribed, s is known as a *spline model*. There are differences (Section 5) in interpretation in terms of the statistical uncertainties associated with the solution and in terms of validating the spline model so obtained.

The use of a formulation in terms of standard uncertainties, together with the B-spline representation (1) of s, gives the linear algebraic formulation[5]

$$\min_{\mathbf{c}} \mathbf{e}^{\mathrm{T}} V_{\mathbf{y}}^{-1} \mathbf{e}, \qquad \mathbf{e} = \mathbf{y} - A\mathbf{c}, \tag{3}$$

where $\mathbf{y} = (y_1, \ldots, y_m)^{\mathrm{T}}$, A is an $m \times q$ matrix with $a_{i,j} = N_{n,j}(x_i)$, and

$$V_{\mathbf{y}} = \mathrm{diag}(u_1^2, \ldots, u_m^2).$$

Matrix computational methods can be applied to this formulation. As a consequence of property (2) of the B-splines, A is a rectangular banded matrix of bandwidth n [9].

The linear algebraic solution can be effected using Givens rotations to triangularise the system, back-solution then yielding the coefficients $\mathbf{c}$ [7]. The number of floating-point operations (flops) required is to first order $O(mn^2)$, i.e., independent of the number of knots. Hence computing a spline model for many knots is hardly more expensive than one for a few knots. Moreover, since for many problems cubic splines ($n = 4$) yield a good balance between approximation properties and smoothness (continuity class C^2), regarding the order as *fixed* gives a flop count $O(m)$.

$\mathbf{c}$ is unique [12] if there is a strictly ordered subset $\mathbf{t} = \{t_j\}_1^q$ of $\mathbf{x}$ such that the Schoenberg–Whitney conditions [25]

$$t_j \in \mathrm{supp}(N_{n,j}(\boldsymbol{\lambda}; x)), \qquad j = 1, \ldots, q, \tag{4}$$

hold. In a case where the conditions (4) do not hold[6], an appropriate member can be selected from the space of possible solutions. Such a selection is also advisable if the conditions are in a practical sense "close" to being violated. A particular solution can be determined by augmenting the least-squares

[5] A further generalisation is possible in which mutual dependencies are permitted among the measurement errors. In this case, $V_{\mathbf{y}}$ is non-diagonal.

[6] A set of knots giving rise to this circumstance may be a consequence of an automatic knot-placement procedure.

formulation by a minimal number of *equality* constraints for $\mathbf{c}$ such that A has full column rank [11].

An instance of the type of data set to which the algorithms of this paper are addressed is shown in Figure 3. Such a data set (cf. Section 2) has the variety of behaviour that cannot readily be reproduced by some other classes of approximating functions.

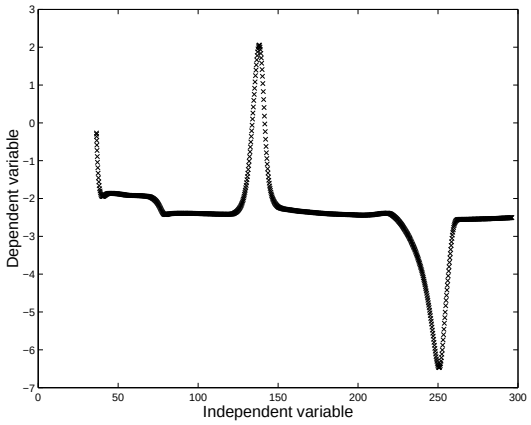

Figure 3: A data set representing heat flow as a function of temperature. Such data forms the basis of the determination of thermophysical properties of materials under test. The data set comprises 980 data points.

5 Spline uncertainties

Once a valid spline model has been obtained, the uncertainties associated with the spline can be evaluated [10]. Uncertainty evaluations are essential in metrology, where all measurement results are to be accompanied by a quantification of their reliability [3], and important in other fields. The key entity is the covariance matrix $V_{\mathbf{c}}$ of the spline coefficients $\mathbf{c}$. Using recognised procedures of linear algebra,

$$V_{\mathbf{c}} = (A^{\mathrm{T}} V_{\mathbf{y}}^{-1} A)^{-1}. \tag{5}$$

From this result, the standard uncertainty of any quantity that depends on $\mathbf{c}$ can be evaluated. Specifically, for a given constant vector $\mathbf{p}$, the standard uncertainty $u(\mathbf{p}^{\mathrm{T}}\mathbf{c})$ of $\mathbf{p}^{\mathrm{T}}\mathbf{c}$ is given by

$$u^2(\mathbf{p}^{\mathrm{T}}\mathbf{c}) = \mathbf{p}^{\mathrm{T}} V_{\mathbf{c}} \mathbf{p}.$$

By setting $\mathbf{p}$ to contain the values of the B-spline basis at a point $x \in I$, the standard uncertainty of $s(x)$ can be formed. The standard uncertainty of a

nonlinear function of $\mathbf{c}$ can be estimated by first linearising the expression about the solution value of $\mathbf{c}$.

If weights rather than uncertainties are specified for the data, (5) takes the form

$$V_{\mathbf{c}} = \hat{\sigma}^2 (A^{\mathrm{T}} W^2 A)^{-1},$$

where $\hat{\sigma}$ estimates the standard deviation of the weighted residuals $\{w_i e_i\}_1^m$,

$$W = \mathrm{diag}(w_1, \ldots, w_m)$$

and

$$\hat{\sigma}^2 = \mathbf{e}^{\mathrm{T}} W^2 \mathbf{e} \,/ (m - q)$$

evaluated at the solution.

6 Families of approximants

When dealing with certain classes of approximating function it is natural and useful to consider *families of approximants*. A simple example is polynomial approximation, for polynomials $p_j(x)$ of order $j = 1, 2, \ldots, N$, for some maximum order N. Each member of the family "contains" the previous member. It is then meaningful to consider the approximation measure, here

$$S = \sqrt{\mathbf{e}^{\mathrm{T}} V_{\mathbf{y}}^{-1} \mathbf{e}},$$

with respect to indices denoting members. Thus, the value of S for the polynomial approximant of order j can be inspected with respect to index j for $j = 1, 2, \ldots, N$. For data approximation, it is more meaningful to use as the measure the *root-mean-square residual* given by dividing S by $(m-j)^{1/2}$. For representative data, the expectation is that as j increases this quantity should stabilise to an essentially constant value. This property provides a useful validation procedure. If weights u_i^{-1} are used as in Section 4 this measure should settle to the value unity. Thus the approximant with index j (normally the smallest such) that achieves the value one is sought.

Within most of the strategies outlined in Section 8 it is possible to produce results for $N = 1, 2, \ldots$ knots, and thus to study the effect of the number of knots on the quality of the approximant. From such information it may be possible to select an acceptable solution. If for each number of knots, the knots contain those for the previous number, and a least-squares approximant is determined, the sequence of approximants for $N = 1, 2, \ldots$ knots forms a *family*. A family has the property that the sequence of values of the least-squares measure S is monotonically decreasing.

7 Least-squares data approximation by splines with free knots

The problem of least-squares data approximation by splines with free knots can be formulated in the same way as that for fixed knots (Section 4), except that the knots are not specified *a priori*, either in location or number. The formulation (3) no longer yields a linear problem, since the matrix A of B-spline values is now a function of $\boldsymbol{\lambda}$. Instead,

$$\mathbf{e}(\boldsymbol{\lambda}) = \mathbf{y} - A(\boldsymbol{\lambda})\mathbf{c},$$

and it is required to solve

$$\min_{\boldsymbol{\lambda};\mathbf{c}} \mathbf{e}^{\mathrm{T}}(\boldsymbol{\lambda}) V_{\mathbf{y}}^{-1} \mathbf{e}(\boldsymbol{\lambda}). \tag{6}$$

In order to reflect the fact that for any given knot set the B-spline coefficients are given by solving a relatively simple, linear problem, formulation (6) can be expressed as

$$\min_{\boldsymbol{\lambda}} \left(\min_{\mathbf{c}} \mathbf{e}^{\mathrm{T}}(\boldsymbol{\lambda}) V_{\mathbf{y}}^{-1} \mathbf{e}(\boldsymbol{\lambda}) \right). \tag{7}$$

Extensive use is made of this elementary result.

There are theoretical difficulties associated with existence, uniqueness and characterisation of best free-knot least-squares spline approximants, which influence practical considerations. A best spline in the class of splines required may not exist. Take as $\{x_i\}_1^m$, $m = 21$ uniformly spaced values in $[-1, 1]$ and $y_i = |x_i|^3$. To see that a best least-squares spline s of order 4 with three interior knots for this data may not exist, consider the choice $\lambda_1 = -\epsilon$, $\lambda_2 = 0$ and $\lambda_3 = \epsilon$. The least-squares error can be made smaller than any given $\delta > 0$ for some $\epsilon > 0$. However, if the least-squares error is made zero by the choice $\epsilon = 0$, the resulting three coincident knots at $x = 0$ mean that s has lower continuity than the class of splines considered. In practice, allowing knots to come "too close" together can introduce undesirable "sharpness" into the approximant. Buffering of knots [16], to ensure a minimal separation helps in this regard. The use of a candidate knot set introduces a form of buffering. In some circumstances the coalescing of knots would be ideal in terms of the resulting closeness of s to the data. In some applications the loss of smoothness would be unacceptable. Therefore, whether buffering is appropriate depends on the use to be made of s.

The solution may not be unique. Figure 4 shows a set of 201 uniformly spaced points in $[-1, 1]$ taken from $f(x) = \mathrm{sign}(x) \min(x, 1/2)$. Figure 5 shows the root-mean-square residual as a function of knot location for least-squares splines of order 4 with one interior knot. There are two best approximants, one with its knot at $x = -0.63$ and the other at $x = +0.63$.

One of the two approximants is shown in Figure 4. The other spline is its skew-symmetric counterpart.

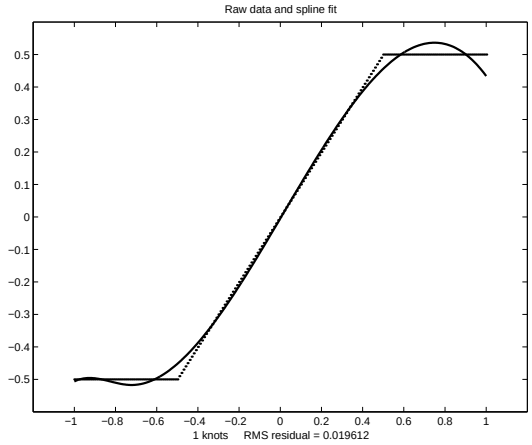

Figure 4: 201 uniformly spaced points in $[-1, 1]$ taken from $f(x) = \text{sign}(x)\min(x, 1/2)$ and a best least-squares spline approximant with one knot.

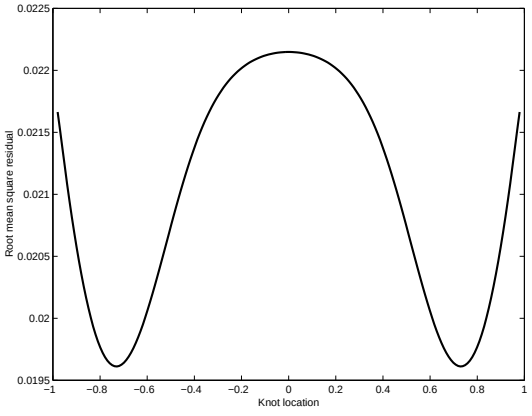

Figure 5: The root-mean-square residual as a function of knot location for least-squares spline approximants with one knot to the data of Figure 4.

8 Knot-placement strategies

Many knot-placement strategies have been proposed and used. Some of these strategies are outlined and their properties indicated. Several of the strategies generate a *family* of candidate spline approximants, with advantages for model validation.

8.1 Manual methods

Manual methods can be classed as those methods for which the user examines the general "shape" of the function underpinning the data, selecting the number and location of the knots on this basis. With practice and visual aids, acceptable solutions can often be obtained [7]. Naturally, knots are chosen to be more concentrated where "things are happening" in contrast to regions where the underpinning behaviour is innocuous.

8.2 Strategies that depend on the distribution of the data

The strategies described in this section are "simple" in the sense that the only information used is that given by the *distribution* of the data points and represented by the abscissa (x_i) and ordinate (y_i) values. Strategies described in later sections make use of more comprehensive information derived from spline approximants for the data based on (existing) candidate knot sets. The strategies are divided into those that make use only of the abscissa values (Section 8.2.1), and those depending on both abscissa and ordinate values (Section 8.2.2). None of the strategies described generate a family of spline approximants. This is because, for any N, the knot set of size N does not contain that of size $N-1$.

8.2.1 Strategies that depend only on abscissa values

Strategies based on the manner in which the values of the independent variable are distributed may be used to place the knots (at points that are not necessarily the data abscissae themselves). A facility in DASL (the NPL Data Approximation Subroutine Library) [1] and NPLFit provides one such strategy, based on the Schoenberg–Whitney conditions (4) in the following way. Intuitively, these conditions imply that there is no region where there are "too many" knots compared with the number of data points. *Mathematically*, these conditions guarantee uniqueness. *Numerically*, their satisfaction does not ensure that the solution is well-defined. If the conditions are "close" to being violated, $\mathbf{c}$ will be sensitive to perturbations in the data. In particular, since the behaviour of $\mathbf{c}$ "controls" that of s (Section 2), the spline is likely to exhibit spurious behaviour such as large undesirable oscillations if $\|\mathbf{c}\|_2 \gg \|\mathbf{y}\|_2$.

It follows that a sensible choice of knots would be such that the Schoenberg–Whitney conditions are satisfied "as well as possible" for a data subset. Such a choice is made in DASL [1] for spline approximation of arbitrary order. It is also made in a cubic spline *interpolation* routine in the NAG Library

[17], regarding spline interpolation as a special case of spline approximation in which $q = m$ and $N = m - n$. The choice made is seen most simply by first applying it to spline interpolation. Consider the choice

$$\lambda_j = \frac{1}{2}(x_{j+\lfloor n/2 \rfloor} + x_{j+\lfloor (n+1)/2 \rfloor}), \qquad j = 1, \ldots, m-n,$$

where $\lfloor v \rfloor$ is the largest integer no larger than v. For n even, $\lambda_j = x_{j+n/2}$. Thus, the choice $t_j = \lambda_{j-n/2}$ would be made. However (Section 2), $\mathrm{supp}(N_{n,j}) = [\lambda_{j-n}, \lambda_j]$. Thus, *index-wise*, the Schoenberg–Whitney conditions are satisfied as well as possible in the sense that the index of $\lambda_{j-n/2}$ falls halfway between the indices of the support endpoints λ_{j-n} and λ_j. Comparable considerations apply for n odd. Precisely this choice is recommended [15, 17] in the context of cubic spline interpolation. It is the "not a knot" criterion, as a practical alternative to the classical use of boundary derivatives. A knot is placed at each "interior" data value x_i apart from x_2 and x_{m-1}.

The above choice can be interpreted as follows. Consider the graph $x = F(\ell)$ given by the join of the points $\{(i, x_i)\}_1^m$. The jth interior knot, λ_j, for $j = 1, \ldots, m-n$, is given by $F(j + n/2)$. The successive spacings between the index arguments of F for $j = 0, \ldots, N+1$, using $F(0) = x_{\min}$ and $F(N+1) = x_{\max}$, are therefore

$$1 + n/2, \underbrace{1, \ldots, 1}_{N-1}, 1 + n/2.$$

For *approximation*, these successive spacings are proportionally increased to account for the fact that there are fewer knots. The resulting expression for the jth interior knot is

$$\lambda_j = F(1 + (m-1)(j + n/2 - 1)/(q-1)), \qquad j = 1, \ldots, N.$$

The choice can be interpreted as placing the interior knots such that there is an approximately equal number of data points in each knot interval (interval between adjacent knots), except that in the first and the last interval there are approximately $n/2$ times as many points. The strategy [1] has the property that when N is such that the data is interpolated, the choice of knots agrees with one of the recommended choices for spline interpolation.[7]

Figure 6 illustrates the above strategy for a spline interpolant and approximant of order 4 to data with abscissae

$$\mathbf{x} = (0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 3, 4, 5, 7.5, 10)^{\mathrm{T}}.$$

[7] The approach tends to give better knot locations if the data is gathered in a manner which ensures that the local density of the data is greater in regions where the behaviour of y is more marked.

Each figure shows the graph $x = F(\ell)$. For the interpolant (left-hand graph), ten knots are chosen to coincide with the abscissa values $x_3, \ldots, x_{12}$. For the approximant (right-hand graph), four knots are chosen such that there are two points in each interval, excepting the first and last interval where there are four points, i.e., $n/2 = 2$ times as many. The distribution of the knots reflects that of the abscissa values.

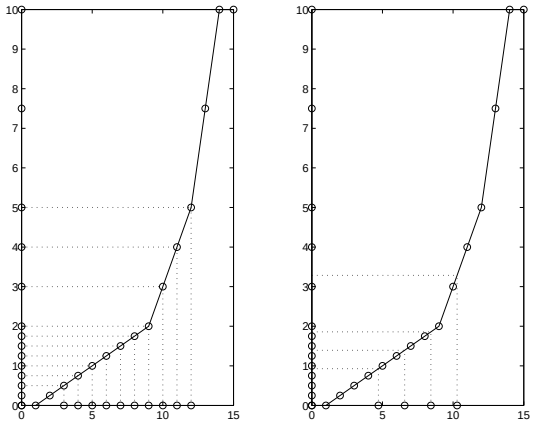

Figure 6: A knot placement strategy depending only on the abscissa values.

Figure 7 shows the results of applying the above strategy to the thermophysical data of Figure 3. The root-mean-square residual is plotted as a function of the number of knots N for $N = 0, \ldots, 200$. The root-mean-square residual values (shown on a logarithmic scale) vary by a factor of (approximately) 1000 for this range of N, from 0.96 for $N = 0$ to 0.62×10^{-3} for $N = 200$. The "non-smooth" behaviour of the root-mean-square residual values as a function of the number of knots is a consequence of the set of spline approximants not constituting a family. However, even though the approximants do not form a family, the saturation of the the root-mean-square residual values for $N > 180$ to a constant (of, approximately, 0.65×10^{-3}) suggests that an acceptable fit to the data is generated.

A simpler strategy to that described above is to select uniformly spaced knots between x_{min} and x_{max}. Such a choice makes minimal use of the available knowledge about the data, only that the data abscissa values are contained within the interval $[x_{min},\ x_{max}]$. The Schoenberg–Whitney conditions will not necessarily automatically be satisfied by such a choice, and the spline approximant would therefore not be unique, although the approach indicated at the end of Section 4 could be applied.

Figure 8 shows the results of selecting $N = 0, \ldots, 200$ uniformly spaced knots for the thermophysical data. The behaviour of the root-mean-square residual as a function of N is very similar to that shown in Figure 7, excepting that

the root-mean-square residual values are still decreasing over the complete range of N.

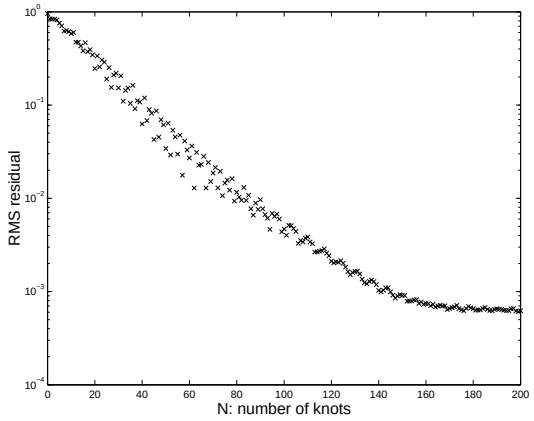

Figure 7: The root-mean-square residual as a function of the number of knots for the application of a knot-insertion strategy to the thermophysical data of Figure 3. The knots are positioned so that there is (approximately) the same number of abscissa values within each knot interval with the exception that the first and last intervals contain $n/2$ times as many. The figure depicts the root-mean-square residual on a logarithmic scale, so its value varies by a factor of 1000 from 0 to 200 knots.

8.2.2 Strategies that depend on abscissa and ordinate values

The strategies described in Section 8.2.1 essentially locate a number of knots uniformly with respect to point index or, for the simpler strategy, with respect to abscissa value. One way of generalising these strategies is to generate a number of knots that are uniformly spaced but along (an approximation to) the underlying curve represented by the data. An approximation to the underlying curve is constructed as the piecewise linear curve obtained by joining the data points by straight-line segments.

Let d_i denote cumulative Euclidean distance from the first to the ith data point, i.e.,

$$d_1 = 0, \quad d_i = d_{i-1} + \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}, \quad i = 2, \ldots, m.$$

The parameter values

$$t_i = \frac{d_i}{d_m}, \quad i = 1, \ldots, m,$$

provide a scaled measure of the distance of the ith data point from the first point. The scaling is such that the total length from first to last data points

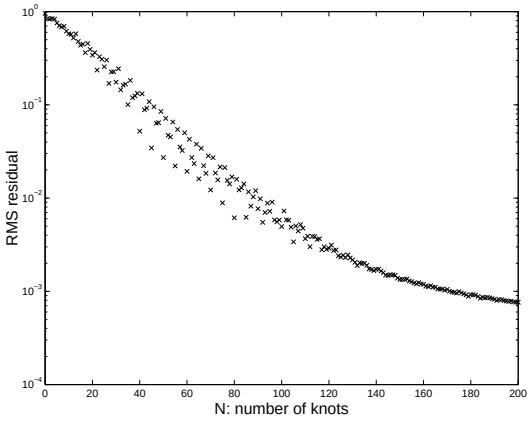

Figure 8: The root-mean-square residual as a function of the number of knots for the application of a knot-insertion strategy to the thermophysical data of Figure 3. The knots are positioned uniformly within the interval defined by the smallest and largest abscissa values.

is unity. Let p be the smallest value of i such that $t_i > j/(N+1)$. The jth knot λ_j is then located at

$$\lambda_j = x_{p-1} + \frac{j/(N+1) - t_{p-1}}{t_p - t_{p-1}}(x_p - x_{p-1}).$$

The knots $\lambda_j,\ j = 1, \ldots, N$, have the property that the distance between adjacent knots measured along the straight-line segments between the data points is the constant value $1/(N+1)$.

Figure 9 shows the results of applying the above strategy to the thermophysical data. The behaviour of the root-mean-square residual as a function of N is very similar to that shown in Figures 7 and 8. One reason for this is that the range of ordinate values is small compared to that of the abscissa values. Unlike the strategies of Section 8.2.1 that locate the knots relative to the abscissa values, the strategy described here is very dependent on the relative *scaling* of the abscissa and ordinate values. Thus, by considering data $(\alpha x_i, y_i),\ i = 1, \ldots, m$, for $0 < \alpha < 1$, the strategy can be expected to give different results according to the value of α.

8.3 Sequential knot-insertion strategies

In a sequential knot-insertion strategy, a succession of approximants is obtained, in which for each approximant a knot is inserted in the knot interval that gives rise to the greatest contribution to the least-squares error. A knot interval is an interval between adjacent knots, where the endpoints of

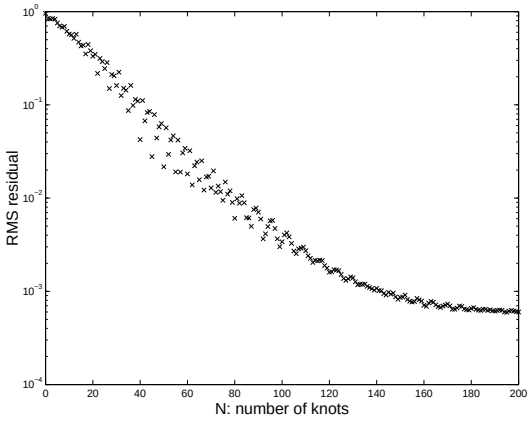

Figure 9: The root-mean-square residual as a function of the number of knots for the application of a knot-insertion strategy to the thermophysical data of Figure 3. The knots are positioned uniformly with respect to chord length along a piecewise linear curve obtained by joining the data points by straight-line segments.

I count as knots for this purpose. Previously inserted knots are retained undisturbed. Several variants are possible (also see Section 8.11), e.g.:

- Start the process with a number of knots already in place, perhaps obtained from information specific to the application.
- Candidate positions for a new knot are
  - The continuum of points within the interval. The approach gives rise to the minimisation of a univariate function that may possess local minima.
  - The subset within the interval of a discrete set of points chosen *a priori*, e.g., the data abscissa themselves or a uniformly spaced set of x-values. The approach gives rise to a finite computation for the globally-best choice of knot, relative to the discretisation, with respect to previous knots.
- More than one knot can be inserted at a time. Doing so gives an approach that is intermediate between full optimisation (Section 8.7) and sequential (single) knot insertion. Computation times rise rapidly with the number of "simultaneous" knots so inserted, so in practice only a small number, say two or three, might be feasible.
- Use a merit function other than the least-squares error (Section 8.5).

Figure 10 shows the root-mean-square residual as a function of the number of knots for the application of this strategy to the thermophysical data of Figure 3. Figure 11 show the positions of the eighty knots generated by the knot insertion strategy corresponding to the right-most point in Figure 10. Compared with the results presented in Section 8.2 the root-mean-square residual values have saturated to the same essentially constant value but over a much smaller range of N. The strategy has generated a spline approximant defined by significantly fewer knots. The distribution of the knots is far from uniform with knots concentrated where the slope of the underlying curve represented by the data changes most rapidly. The strategy has generated a distribution of knots that is intuitively good for this data set.

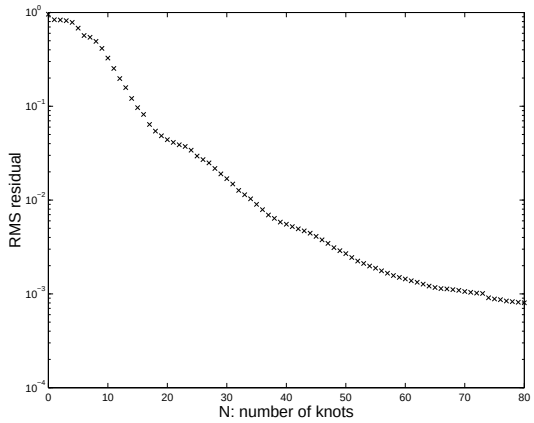

Figure 10: The root-mean-square residual as a function of the number of knots for the application of a knot-insertion strategy to the thermophysical data of Figure 3. At each iteration a single knot is inserted so as to maximise the reduction of the root-mean-square residual. Candidate positions for each new knot are the data abscissa values.

8.4 Sequential knot-removal strategies

In a sequential knot-removal strategy, the starting point is an initial spline approximant having a "large" number of knots that typically would be regarded as an acceptable approximant to the data and that contains (perhaps many) more knots than desired. Also see Section 8.11. Each successive approximant is obtained from the previous approximant by deleting one (or more) knots. The knot selected for removal is chosen as that having least effect in terms of the change in the least-squares error. The process is continued until an acceptable approximant is no longer obtained.

The initially large number of knots (Section 8.11) provides an appreciable number of candidate knots for removal and thus greater flexibility. The

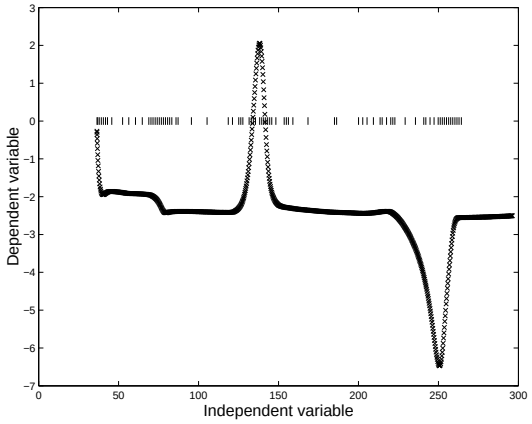

Figure 11: The thermophysical data shown in Figure 3 and the positions of the 80 knots returned by the application of a knot insertion strategy.

rationale is that in contrast to successive knot insertion a succession of acceptable approximants is obtained as opposed to a succession of unacceptable approximants, until the final "solution" is provided. There are variants, as with sequential knot insertion. For example, several knots can be removed at each stage.

A different class of knot removal algorithms [21] is based on a general class of ℓ_p norms. It is not concerned specifically with data approximation, but with replacing an initial spline approximant (that may correspond to an approximant) by one that is acceptably close according to the measure.

The two sets of crosses in Figure 12 correspond to the values of the root-mean-square residual as a function of the number of knots for the application of the knot-insertion strategy followed by the knot-removal strategy for the thermophysical data of Figure 3. The two sets, where the "progress" takes place from left to right along the "top set", followed by right to left along "the bottom set", constitutes a form of hysteresis. The behaviour in the two directions is distinctly different. In particular, the figure indicates that once an acceptable approximation has been obtained by knot insertion, the use of knot removal can deliver an approximation of comparable quality with many fewer knots or alternatively for the same number of knots an appreciably better approximation can be obtained. In this case, with 30 knots, knot removal gives a least-squares error that is one quarter of that for knot insertion. For a least-squares error of 0.005, 30 knots are required using knot removal and 43 using knot insertion.

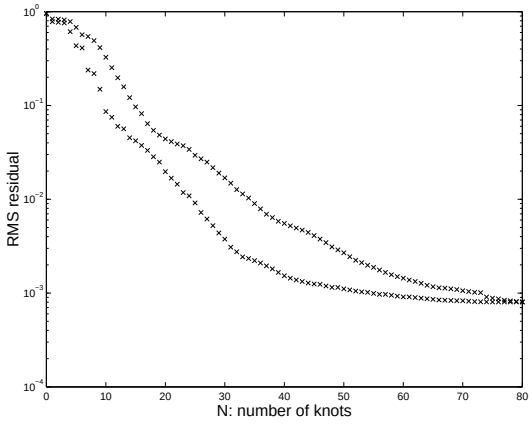

Figure 12: The root-mean-square residual as a function of the number of knots for the application of knot-insertion and knot-removal strategies to the thermophysical data of Figure 3. The "upper set" of crosses indicates the values obtained for knot insertion (Figure 10) and the lower for knot removal. The knot-removal strategy starts with the knot set provided by the knot-insertion strategy, which was terminated after 80 knots had been placed, and removes at each iteration a single knot so as to minimise the increase in the root-mean-square residual.

8.5 Approaches based on removing trends

A variant of the sequential knot-insertion strategies described in Section 8.3 is to replace the least-squares measure of fit by a test of the "randomness" of the residuals associated with the spline fit based on a current set of knots.

One implementation of this approach is in terms of a "trend test" [24]. Suppose

$$\{x_{p_j}, x_{p_j+1}, \ldots, x_{p_j+q_j}\}$$

is the set of data abscissae values contained in the jth knot interval, and $s(x)$ is the spline fit based on the current knots. Compute the value of T_j, given by

$$T_j = \frac{\sum_{k=p_j}^{p_j+q_j-1} r_k r_{k+1}}{\sum_{k=p_j}^{p_j+q_j} r_k^2},$$

where

$$r_k = u_k^{-1} e_k = w_k e_k$$

are the (weighted) residuals associated with $s(x)$ at the data abscissa values in the jth knot interval. A value of T_j exceeding $1/\sqrt{q_j}$ indicates that there is a trend in the residuals within the corresponding interval. A knot is inserted at the midpoint of any knot interval in which a trend is detected.

Another way of detecting a trend in the residual values r_k is to identify regions where consecutive values of r_k have the same sign. A "long" run of such residuals indicates a trend in the residuals and hence a region where the fit based on the current knots is unacceptable. A knot is inserted within the region given by the longest run of residuals of the same sign. The new knot is chosen to coincide with the data abscissa value corresponding to the central residual of this run (if the number of residuals is odd) or the midpoint of the two central residuals (if the number is even).

Figures 13 and 14 show the results of applying these strategies to the thermophysical data of Figure 3. The strategy based on the "trend test" is different from other strategies considered because at each iteration a number (typically greater than one) of knots is inserted (one into each knot interval for which a trend in the residual values is detected). Consequently, the graph of root-mean-square residual values comprises only a few points, and it is difficult to judge whether these values have saturated to an essentially constant value. To overvoce this disadvantage, a variant of the strategy based on inserting one knot at a time, indicated by the greatest value of T_j, could be used.

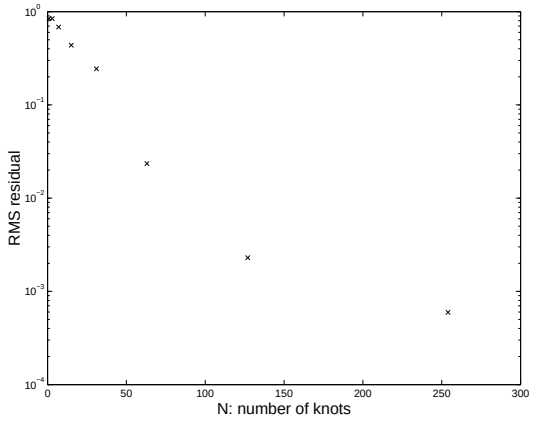

Figure 13: The root-mean-square residual as a function of the number of knots for the application of a knot-insertion strategy to the thermophysical data of Figure 3. A knot is inserted at the midpoint of any knot interval in which a trend in the residual values is detected [24].

8.6 Theory-based approaches

The distance of a spline $s(x)$ with knots $\boldsymbol{\lambda}$ from a sufficiently differentiable function $f(x)$ is proportional to

$$h^n |f^{(n)}(\xi)|,$$

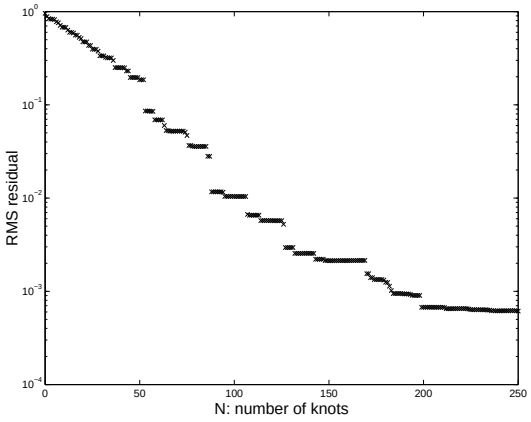

Figure 14: The root-mean-square residual as a function of the number of knots for the application of a knot-insertion strategy to the thermophysical data of Figure 3. A knot is inserted at the data abscissa value corresponding to the central residual of the longest run of residuals of the same sign.

where h is the local knot spacing and ξ is a value of x [15]. Consider inverting this expression in order approximately to equalise the error with respect to x. The lengths of the knot intervals should consequently be chosen to be proportional to $|f^{(n)}(\xi)|^{-1/n}$, where ξ is a value in the neighbourhood of the respective knot interval. Consider the function

$$F(x) = \int_{x_{\min}}^{x} |f^{(n)}(t)|^{1/n} dt \Big/ \int_{x_{\min}}^{x_{\max}} |f^{(n)}(t)|^{1/n} dt \; . \tag{8}$$

Take knots given by

$$F(\lambda_j) = \frac{j}{N+1}, \qquad j = 1, \ldots, N. \tag{9}$$

This result corresponds to dividing the range of the monotonically increasing function $F(x)$, for $x \in I$, into $N+1$ contiguous subranges of equal length, taking the values of x corresponding to the subrange endpoints as the knots.

In practice f, let alone F, is unknown. Various efforts have been made to estimate f and hence F from the data points. For instance, if the data is approximated by a spline of order $n+1$, its nth derivative, a piecewise-constant function, can be used to estimate F [4]. It is then straightforward to form the required knots. The approach begs the question in the case of data. In order to estimate knots for a spline of order n, it is first necessary to construct a spline approximant of order $n+1$ for the data, the construction of which itself requires a choice of knots.

Alternatively [14], a spline approximant of order n for the data can be constructed for some *convenient* choice of knots. Its nth derivative is of

course zero (except at the knots). However, its $(n-1)$th derivative is piecewise constant, a function that can be approximated by the join of the mean values at the knots of the constant pieces to the immediate right and left, with special consideration at the endpoints of I. The derivative of this piecewise-linear function then provides a piecewise-constant representation of the nth derivative, that can be used as before. Knots can then be deduced from this form as above. The advantage of this approach is that it can be iterated [14]. If the process "converges", the result can be used to provide the required knot set. The process can work well, but is capable of producing disappointing results. Several variants of the basic concept are possible. The approach warrants careful revisiting.

8.7 "Overall" optimisation approaches

For any given value of N, the problem is regarded as an optimisation problem with respect to the overall error measure. It is necessary to provide a sensible initial estimate of the knot positions. Local solutions which may be grossly inferior to the global solution are possible [5]. At an optimal solution, knots may coalesce, thus reducing the continuity of the spline at such points [20]; the same comment applies to the sequential-knot-insertion and optimisation approach (Section 8.8).

8.8 Sequential knot insertion and optimisation

Sequential knot insertion with optimisation is identical to the sequential knot-insertion strategy (Section 8.3) except that, after each knot is inserted, all previously-inserted knots are adjusted such that the complete set of knots at that stage are (locally) optimal with respect to the overall error measure. One such strategy [16] carries out the optimisation at each stage by adjusting in turn each knot in the current knot set in order to achieve satisfactory reduction in the least-squares error, and repeating the complete adjustment as necessary. This strategy is not as poor as the traditional one-variable-at-a-time strategy for nonlinear optimisation because knots far from the newly-inserted knot tend to have little effect on the error measure.

Buffering to prevent knots coalescing and reducing the continuity of the approximant can be used. Various features can be incorporated to improve computational efficiency, including the use of contemporary nonlinear least-squares optimisation. It is emphasised that for each choice of knots the problem is linear (cf. Section 7).

8.9 Optimal discontinuous piecewise-polynomial approximation

Consider the class S_N of splines having N interior knots of multiplicity n (i.e., nN interior knots in all, counting coincidences). An $s \in S_N$ will in general be discontinuous at these knots. It is possible to determine the globally optimal locations of such knots, using the principle of dynamic programming [5]. The approach is based on the fact that the best approximant $s_N \in S_N$ to the leading p ($\geq nN$) data points is given by the best over $q = nN - n + 1, nN - n + 2, \ldots, p - N$ of $s_{N-1} \in S_{N-1}$ for the leading $q \leq p - N$ points, together with a *polynomial piece* of order n over points $q+1$ to p. By this simple recursive means the globally best knots for splines of any order that are discontinuous at any number of knots can be computed.

Such a solution may not be suitable as the final result in an application. However, it can be useful as part of a knot placement strategy. For example, suppose good knots for a spline of order n are required. An approach would be to determine an optimal discontinuous spline of order $n + 1$. Use this spline to estimate f in expression (8). The integral in the numerator of (8) will be continuous piecewise linear and estimates of the optimal knots for a $C^{(n-2)}$ spline readily obtained from (9). Mixed results have informally been obtained by the authors with an implementation of this approach. It is suggested that it be revisited.

8.10 Knot dispersion

A set of knots of multiplicity n is positioned using an appropriate strategy, such as that in Section (8.9) and a $C^{(-1)}(I)$ spline with these knots determined. Each of these multiple knots is "dispersed", viz., replaced by n nearby simple knots, and a replacement $C^{(n-2)}(I)$ spline computed. A careful strategy for knot dispersion is required. Again, informal experiments have been made by the authors and mixed results obtained.

8.11 Knot initialisation and candidate knot locations

Several of the above procedures require or can benefit from an initial placement of the knots. Some make use of "candidate knot locations".

The solution to the free-knot spline approximation problem returned by iterative algorithms typically depends on the starting set of knots. Although an algorithm may return a result that satisfies the necessary and sufficient conditions for a solution [18], this result may be locally rather than globally

optimal. There is no known characterisation of a globally optimal solution. The careful interpretation of solutions is therefore important.

The use of candidate knot positions can be helpful. For instance, it may be decided that for splines of even order, only knots that coincide with data abscissae are in the candidate set, or, for splines of odd order, knots only at points mid-way between adjacent data abscissae may be so regarded. Such criteria are consistent with the choice for interpolating splines and the generalisation covered in Section 8.2.1. The Lyche-Mørken knot removal algorithms [21, 22] use data abscissae as candidate knots. The use of a finite number of candidate knot locations helps to reduce the dimensionality of the problem: there can then only be a *finite* number of possible knot sets. For large N this number can be extremely large, making it prohibitive to examine all possibilities. However, for small N, e.g., 1, 2 and 3, it may indeed be possible, and can pay dividends. Knot insertion and knot removal algorithms can also implement the concept. For example, at each stage of a knot insertion strategy, two or three knots can be inserted "simultaneously". By the method of their introduction these new knots will be optimal relative to the knots previously used and the available candidate knot locations.

Another aspect of a candidate knot set is that if it is sufficiently dense it will contain, to a degree of approximation dictated by its "spacing", the optimal knots for the given data set [20]. For instance, consider a set of $m \gg 100$ data points specified over an interval I normalised to $[-1, 1]$. Take 100 uniformly spaced points spanning this interval. This set will contain, to approximately two figures, each globally optimal knot set having $N \leq 98$ knots[8] (assuming all knots are simple). If a spline based on these 98 candidate interior knots provided a valid model, a suitable knot removal algorithm might be expected to be able to identify reasonably closely the optimal knot sets. Work is required to determine the degree of success in this regard.

9 Conclusions and future possibilities

It is rarely required to determine a least-squares spline approximant that is globally or even locally optimal with respect to its knots. An approximant that met some closeness requirement with the smallest possible number of knots is an academic rather than a pragmatic objective. Today, the more important consideration is to obtain an approximant that represents the data in that its smoothness is consistent with that of the function underlying the data and the uncertainties in the data. (This statement must be qualified for situations where the continuity class of splines is a consideration as discussed

[8] The two endpoints do not constitute interior knots.

above.) These ends *may* be achieved by seeking an approximant with a reasonable but not necessarily optimal number of knots.

The use of knot removal strategies is likely to attract research effort in the future. One reason for this statement is that the need to work with large initial knot sets is not as computationally prohibitive with today's powerful personal and other computers. Another reason is that the approach can be expected to produce better approximants, i.e., smaller least-squares errors for the same number of knots. A study [23] of knot removal algorithms [21, 22] applied to typical metrology data has shown that promising results may be obtained using such algorithms.

Large data sets, as are now frequently being produced in metrology from computer-controlled measuring systems, are ideal for the purpose of obtaining a sound initial approximant in the form of a valid model containing possibly many more knots than the minimum possible. Their size permits initial approximants to be obtained, even with large numbers of uniformly spaced knots, that provide valid but highly redundant models for the data. The fact that such sets do not contain "appreciable gaps", because of the manner in which they are gathered, means that this fact together with the quantity of data far outweighing this initial number of knots goes a long way towards ensuring that this initial approximant is valid. There is much scope for an appreciable number of knots to be removed. The initial large number of knots may also have been obtained by the use of a knot *insertion* strategy. It is the experience of the authors that knot insertion can introduce appreciably more knots than given by the optimal choice.

Because the early approximants may be far from optimal, an insertion algorithm can produce knots that are totally different from those in an optimal approximant. In contrast, a knot removal algorithm has a possibility to obtain good knots. (See Section 8.11.) For instance, because of the sequential manner in which knots are inserted, there may be two or more close or even coincident knots, although a good knot set might not have this property. It is also possible that such knots, although not part of an optimal set, are influential in their effect on a knot removal algorithm, with the result that they appear in the "final" approximant.

The problem of data containing wild points is not addressed satisfactorily by existing knot placement algorithms. Because such points are responsible for a large contribution to the least-squares error, more knots would be placed in the neighbourhood of such a point than would otherwise had been done. The knot placement strategy can then be influenced more by the errors in the data than by the properties of the underlying function. Formulations and hence algorithms are needed that have greater resilience to such effects.

In solving the fixed-knot spline approximation problem as part of the free-knot problem, a knot set differs from a previous knot set only by the addition or removal of a small number of knots. In linear algebraic terms the "new" matrix $A(\boldsymbol{\lambda}')$, say, differs in only a few rows from the previous matrix $A(\boldsymbol{\lambda})$. Considerable gains in computational efficiency can be obtained by accounting for this fact. This paper has not addressed this issue, concentrating more on the *concepts* in the area. There is much scope, however, for the application of the recognised stable updating and downdating techniques of linear algebra [18]. Their application will not reduce the *computational complexity* of a procedure, but could reduce computation times for large problems by an appreciable factor.

10 Acknowledgements

The work described here was supported by the National Measurement System Policy Unit of the UK Department of Trade and Industry as part of its NMS Software Support for Metrology programme.

The authors acknowledge the contribution [23] made by John Mason and Colin Ross of the University of Huddersfield to the work described in this report.

References

[1] G. T. Anthony and M. G. Cox. The National Physical Laboratory's Data Approximation Subroutine Library. In J. C. Mason and M. G. Cox, editors, *Algorithms for Approximation*, pages 669–687, Oxford, 1987. Clarendon Press.

[2] R. M. Barker. Best Practice Guide No. 5. Software Re-use: Guide to MetroS. Technical report, National Physical Laboratory, Teddington, UK, 2000.

[3] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. Guide to the Expression of Uncertainty in Measurement, 1995. ISBN 92-67-10188-9, Second Edition.

[4] H. G. Burchard. On the degree of convergence of piecewise polynomial approximations on optimal meshes. *Amer. Math. Soc.*, 234:531–559, 1977.

[5] M. G. Cox. Curve fitting with piecewise polynomials. *J. Inst. Math. Appl.*, 8:36–52, 1971.

[6] M. G. Cox. The numerical evaluation of B-splines. *J. Inst. Math. Appl.*, 10:134–149, 1972.

[7] M. G. Cox. A survey of numerical methods for data and function approximation. In D. A. H. Jacobs, editor, *The State of the Art in Numerical Analysis*, pages 627–668, London, 1977. Academic Press.

[8] M. G. Cox. The incorporation of boundary conditions in spline approximation problems. In G. A. Watson, editor, *Lecture Notes in Numerical Analysis 630: Numerical Analysis*, pages 51–63, Berlin, 1978. Springer-Verlag.

[9] M. G. Cox. The least squares solution of overdetermined linear equations having band or augmented band structure. *IMA J. Numer. Anal.*, 1:3–22, 1981.

[10] M. G. Cox. The NPL Data Approximation Subroutine Library: current and planned facilities. *NAG Newsletter*, 2/87:3–16, 1987.

[11] M. G. Cox. Algorithms for spline curves and surfaces. In L. Piegl, editor, *Fundamental Developments of Computer-Aided Geometric Modelling*, pages 51–76, London, 1993. Academic Press.

[12] M. G. Cox and J. G. Hayes. Curve fitting: a guide and suite of algorithms for the non-specialist user. Technical Report NAC 26, National Physical Laboratory, Teddington, UK, 1973.

[13] C. de Boor. On calculating with B-splines. *J. Approx. Theory*, 6:50–62, 1972.

[14] C. de Boor. Good approximation by splines with variable knots II. In G. A. Watson, editor, *Numerical Solution of Differential Equations, Lecture Notes in Mathematics No. 363*, pages 12–20. Springer-Verlag, Berlin, 1974.

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

[16] C. de Boor and J. R. Rice. Least squares cubic spline approximation II - variable knots. Technical Report CSD TR 21, Purdue University, 1968.

[17] B. Ford, J. Bentley, J. J. du Croz, and S. J. Hague. The NAG Library 'machine'. *Software - Practice and Experience*, 9:56–72, 1979.

[18] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, London, 1981.

[19] IEEE. IEEE standard for binary floating-point arithmetic. Technical Report ANSI/IEEE standard 754-1985, IEEE, IEEE Computer Society, New York, USA, 1985.

[20] D. Jupp. Non-linear least square spline approximation. Technical report, Flinders University, Australia, 1971.

[21] T. Lyche and K. Mørken. A discrete approach to knot removal and degree reduction for splines. In J. C. Mason and M. G. Cox, editors, *Algorithms for Approximation*, pages 67–82, Oxford, 1987. Clarendon Press.

[22] T. Lyche and K. Mørken. A data reduction strategy for splines with applications to the approximation of functions and data. *IMA Journal of Numerical Analysis*, 8:185–208, 1988.

[23] J.C. Mason and C. Ross. Report on an investigation into the Lyche and Mørken spline knot reduction algorithm. University of Huddersfield, to be published, 2002.

[24] M.J.D. Powell. Curve fitting by splines in one variable. In J.G. Hayes, editor, *Numerical approximation to functions and data*, pages 65–83, London, 1970. The Athlone Press.

[25] I. J. Schoenberg and Anne Whitney. On Pólya frequency functions III. *Trans. Am. Math.*, 74:246–259, 1953.