NAG Library Chapter Introduction

E01 − Interpolation

Contents

1 Scope of the Chapter

This chapter is concerned with the interpolation of a function of one or more variables. When provided with the value of the function (and possibly one or more of its lowest-order derivatives) at each of a number of values of the variable(s), the NAG Library routines provide either an interpolating function or an interpolated value. For some of the interpolating functions, there are supporting NAG Library routines to evaluate, differentiate or integrate them.

2 Background to the Problems

In motivation and in some of its numerical processes, this chapter has much in common with Chapter E02 (Curve and Surface Fitting). For this reason, we shall adopt the same terminology and refer to dependent variable and independent variable(s) instead of function and variable(s). Where there is only one independent variable, we shall denote it by x and the dependent variable by y. Thus, in the basic problem considered in this chapter, we are given a set of distinct values $x_1, x_2, \ldots, x_m$ of x and a corresponding set of values $y_1, y_2, \ldots, y_m$ of y, and we shall describe the problem as being one of interpolating the data points (x_r, y_r), rather than interpolating a function. In modern usage, however, **interpolation** can have either of two rather different meanings, both relevant to routines in this chapter. They are

(a) the determination of a function of x which takes the value y_r at $x = x_r$, for $r = 1, 2, \ldots, m$ (an **interpolating function** or **interpolant**),

(b) the determination of the value (**interpolated value** or **interpolate**) of an interpolating function at any given value, say $\hat{x}$, of x within the range of the x_r (so as to estimate the value at $\hat{x}$ of the function underlying the data).

The latter is the older meaning, associated particularly with the use of mathematical tables. The term 'function underlying the data', like the other terminology described above, is used so as to cover situations additional to those in which the data points have been computed from a known function, as with a mathematical table. In some contexts, the function may be unknown, perhaps representing the dependency of one physical variable on another, say temperature upon time.

Whether the underlying function is known or unknown, the object of interpolation will usually be to approximate it to acceptable accuracy by a function which is easy to evaluate anywhere in some range of interest. Polynomials, rational functions (ratios of two polynomials) and piecewise polynomials, such as cubic splines (see Section 2.2 in the E02 Chapter Introduction for definitions of terms in the latter case), being easy to evaluate and also capable of approximating a wide variety of functions, are the types of function mostly used in this chapter as interpolating functions. An interpolating polynomial is taken to have degree $m - 1$ when there are m data points, and so it is unique. It is called the **Lagrange interpolating polynomial**. The rational function, in the special form used, is also unique. An interpolating spline, on the other hand, depends on the choice made for the knots.

One way of achieving the objective in (b) above is, of course, through (a), but there are also methods which do not involve the explicit computation of the interpolating function. Everett's formula and Aitken's successive linear interpolation (see Dahlquist and Björck (1974)) provide two such methods. Both are used in this chapter and determine a value of the Lagrange interpolating polynomial.

It is important to appreciate, however, that the Lagrange interpolating polynomial often exhibits unwanted fluctuations between the data points. These tend to occur particularly towards the ends of the data range, and to get larger with increasing number of data points. In severe cases, such as with 30 or 40 equally spaced values of x, the polynomial can take on values several orders of magnitude larger than the data values. (Closer spacing near the ends of the range tends to improve the situation, and wider spacing tends to make it worse.) Clearly, therefore, the Lagrange polynomial often gives a very poor approximation to the function underlying the data. On the other hand, it can be perfectly satisfactory when its use is restricted to providing interpolated values away from the ends of the data range from a reasonably small number of data values.

In contrast, a cubic spline which interpolates a large number of data points can often be used satisfactorily over the whole of the data range. Unwanted fluctuations can still arise but much less frequently and much less severely than with polynomials. Rational functions, when appropriate, would also be used over the whole data range. The main danger with these functions is that their polynomial

denominators may take zero values within that range. Unwanted fluctuations are avoided altogether by a routine using piecewise cubic polynomials having only first derivative continuity. It is designed especially for monotonic data, but for other data still provides an interpolant which increases, or decreases, over the same intervals as the data.

The concept of interpolation can be generalized in a number of ways. Firstly, at each x, the interpolating function may be required to take on not only a given value but also given values for all its derivatives up to some specified order (which can vary with r). This is the Hermite–Birkoff interpolation problem. Secondly, we may be required to estimate the value of the underlying function at a value $\hat{x}$ outside the range of the data. This is the process of **extrapolation**. In general, it is a good deal less accurate than interpolation and is to be avoided whenever possible.

Interpolation can also be extended to the case of two or more independent variables. If the data values are given at the intersections of a regular two-dimensional mesh bicubic splines (see Section 2.3.2 in the E02 Chapter Introduction) are very suitable and usually very effective for the problem. For other cases, perhaps where the data values are quite arbitrarily scattered, polynomials and splines are not at all appropriate and special forms of interpolating function have to be employed. Many such forms have been devised and two of the most successful are in routines in this chapter. They both have continuity in first, but not higher, derivatives.

3 Recommendations on Choice and Use of Available Routines

3.1 General

Before undertaking interpolation, in other than the simplest cases, you should seriously consider the alternative of using a routine from Chapter E02 to approximate the data by a polynomial or spline containing significantly fewer coefficients than the corresponding interpolating function. This approach is much less liable to produce unwanted fluctuations and so can often provide a better approximation to the function underlying the data.

When interpolation is employed to approximate either an underlying function or its values, you will need to be satisfied that the accuracy of approximation achieved is adequate. There may be a means for doing this which is particular to the application, or the routine used may itself provide a means. In other cases, one possibility is to repeat the interpolation using one or more extra data points, if they are available, or otherwise one or more fewer, and to compare the results. Other possibilities, if it is an interpolating function which is determined, are to examine the function graphically, if that gives sufficient accuracy, or to observe the behaviour of the differences in a finite difference table, formed from evaluations of the interpolating function at equally-spaced values of x over the range of interest. The spacing should be small enough to cause the typical size of the differences to decrease as the order of difference increases.

3.2 One Independent Variable

3.2.1 Interpolated values: data without derivatives

When the underlying function is well represented by data points on both sides of the value, $\hat{x}$, at which an interpolated value is required, E01ABF should be tried first if the data points are equally spaced, E01AAF if they are not. Both compute a value of the Lagrange interpolating polynomial, the first using Everett's formula, the second Aitken's successive linear interpolation. The first routine requires an equal (or nearly equal) number of data points on each side of $\hat{x}$; such a distribution of points is preferable also for the second routine. If there are many data points, this will be achieved simply by using only an appropriate subset for each value of $\hat{x}$. Ten to twelve data points are the most that would be required for many problems. Both routines provide a means of assessing the accuracy of an interpolated value, with E01ABF by examination of the size of the finite differences supplied, with E01AAF by intercomparison of the set of interpolated values obtained from polynomials of increasing degree.

In other cases, or when the above routines fail to produce a satisfactory result, one of the routines discussed in the next section should be used. The spline and other piecewise polynomial routines are the most generally applicable. They are particularly appropriate when interpolated values towards the ends of the range are required. They are also likely to be preferable, for reasons of economy, when many interpolated values are required.

E01AAF above, and three of the routines discussed in the next section, can be used to compute extrapolated values. These three are E01AEF, E01BEF and E01RAF based on polynomials, piecewise polynomials and rational functions respectively. Extrapolation is not recommended in general, but can sometimes give acceptable results if it is to a point not far outside the data range, and only the few nearest data points are used in the process. E01RAF is most likely to be successful.

3.2.2 Interpolating function: data without derivatives

E01AEF computes the Lagrange interpolating polynomial by a method (based on **Newton's formula with divided differences** (see Fröberg (1970)) which has proved numerically very stable. Thus, it can sometimes be used to provide interpolated values in more difficult cases than can E01AAF (see the previous section). However, the likelihood of the polynomial having unwanted fluctuations, particularly near the ends of the data range when a moderate or large number of data points are used, should be remembered.

Such fluctuations of the polynomial can be avoided if you are at liberty to choose the x values at which to provide data points. In this case, a routine from Chapter E02, namely E02AFF, should be used in the manner and with the x values discussed in Section 3.2.2 in the E02 Chapter Introduction.

Usually however, when the whole of the data range is of interest, it is preferable to use a cubic spline as the interpolating function. E01BAF computes an interpolating cubic spline, using a particular choice for the set of knots which has proved generally satisfactory in practice. If you wish to choose a different set, a cubic spline routine from Chapter E02, namely E02BAF, may be used in its interpolating mode, setting NCAP7 $= M + 4$ and all elements of the parameter W to unity.

The cubic spline does not always avoid unwanted fluctuations, especially when the data shows a steep slope close to a region of small slope, or when the data inadequately represents the underlying curve. In such cases, E01BEF can be very useful. It derives a piecewise cubic polynomial (with first derivative continuity) which, between any adjacent pair of data points, either increases all the way, or decreases all the way (or stays constant). It is especially suited to data which is monotonic over the whole range.

In this routine, the interpolating function is represented simply by its value and first derivative at the data points. Supporting routines compute its value and first derivative elsewhere, as well as its definite integral over an arbitrary interval. The other routines mentioned, namely E01AEF and E01BAF, provide the interpolating function either in Chebyshev series form or in B-spline form (see Sections 2.2.1 and 2.2.2 in the E02 Chapter Introduction). Routines for evaluating, differentiating and integrating these forms are discussed in Section 3.7 in the E02 Chapter Introduction. The splines and other piecewise cubics will normally provide better estimates of the derivatives of the underlying function than will interpolating polynomials, at any rate away from the central part of the data range.

E01RAF computes an interpolating rational function. It is intended mainly for those cases where you know that this form of function is appropriate. However, it is also worth trying in cases where the other routines have proved unsatisfactory. E01RBF is available to compute values of the function provided by E01RAF.

3.2.3 Data containing derivatives

E01AEF (see Section 3.2.2) can also compute the polynomial which, at each x_r, has not only a specified value y_r but also a specified value of each derivative up to order p_r.

3.3   Two Independent Variables

3.3.1 Data on a rectangular mesh

Given the value f_{qr} of the dependent variable f at the point (x_q, y_r) in the plane of the independent variables x and y, for each $q = 1, 2, \ldots, m$ and $r = 1, 2, \ldots, n$ (so that the points (x_q, y_r) lie at the $m \times n$ intersections of a rectangular mesh), E01DAF computes an interpolating bicubic spline, using a particular choice for each of the spline's knot-set. This choice, the same as in E01BAF, has proved generally satisfactory in practice. If, instead, you wish to specify your own knots, a routine from Chapter E02, namely E02DAF, may be used (it is more cumbersome for the purpose, however, and much slower for larger problems). Using m and n in the above sense, the parameter M must be set to $m \times n$, PX and

PY must be set to $m + 4$ and $n + 4$ respectively and all elements of W should be set to unity. The recommended value for EPS is zero.

3.3.2 Arbitrary data

As remarked at the end of Section 2, specific methods of interpolating are required for this problem, which can often be difficult to solve satisfactorily. Two of the most successful are employed in E01SAF and E01SGF, the two routines which (with their respective evaluation routines E01SBF and E01SHF) are provided for the problem. Definitions can be found in the routine documents. Both interpolants have first derivative continuity and are 'local', in that their value at any point depends only on data in the immediate neighbourhood of the point. This latter feature is necessary for large sets of data to avoid prohibitive computing time. E01SHF allows evaluation of the interpolant and its first partial derivatives.

The relative merits of the two methods vary with the data and it is not possible to predict which will be the better in any particular case.

E01SAF performs a triangulation of the scattered data points and then calculates a bicubic interpolant based on this triangulation and on the function values at the scattered points (which can be evaluated by E01SBF). Where derivative continuity is not essential and where bi-linear interpolated values are sufficient, E01EAF (which performs the same triangulation as E01SAF) and E01EBF (which performs barycentric interpolation using the set of function values) may be used.

3.4 Three Independent Variables

3.4.1 Arbitrary data

The routine E01TGF and its evaluation routine E01THF are provided for interpolation of three-dimensional scattered data. As in the case of two independent variables, the method is local, and produces an interpolant with first derivative continuity. E01THF allows evaluation of the interpolant and its first partial derivatives.

3.5 Four and Five Independent Variables

3.5.1 Arbitrary data

The routine E01TKF and its evaluation routine E01TLF allow interpolation of four-dimensional scattered data, while the routine E01TMF and its evaluation routine E01TNF allow interpolation of five-dimensional scattered data. E01TKF and E01TMF are higher dimensional analogues to the routines E01SGF and E01TGF, while E01TLF and E01TNF are analogous to E01SHF and E01THF.
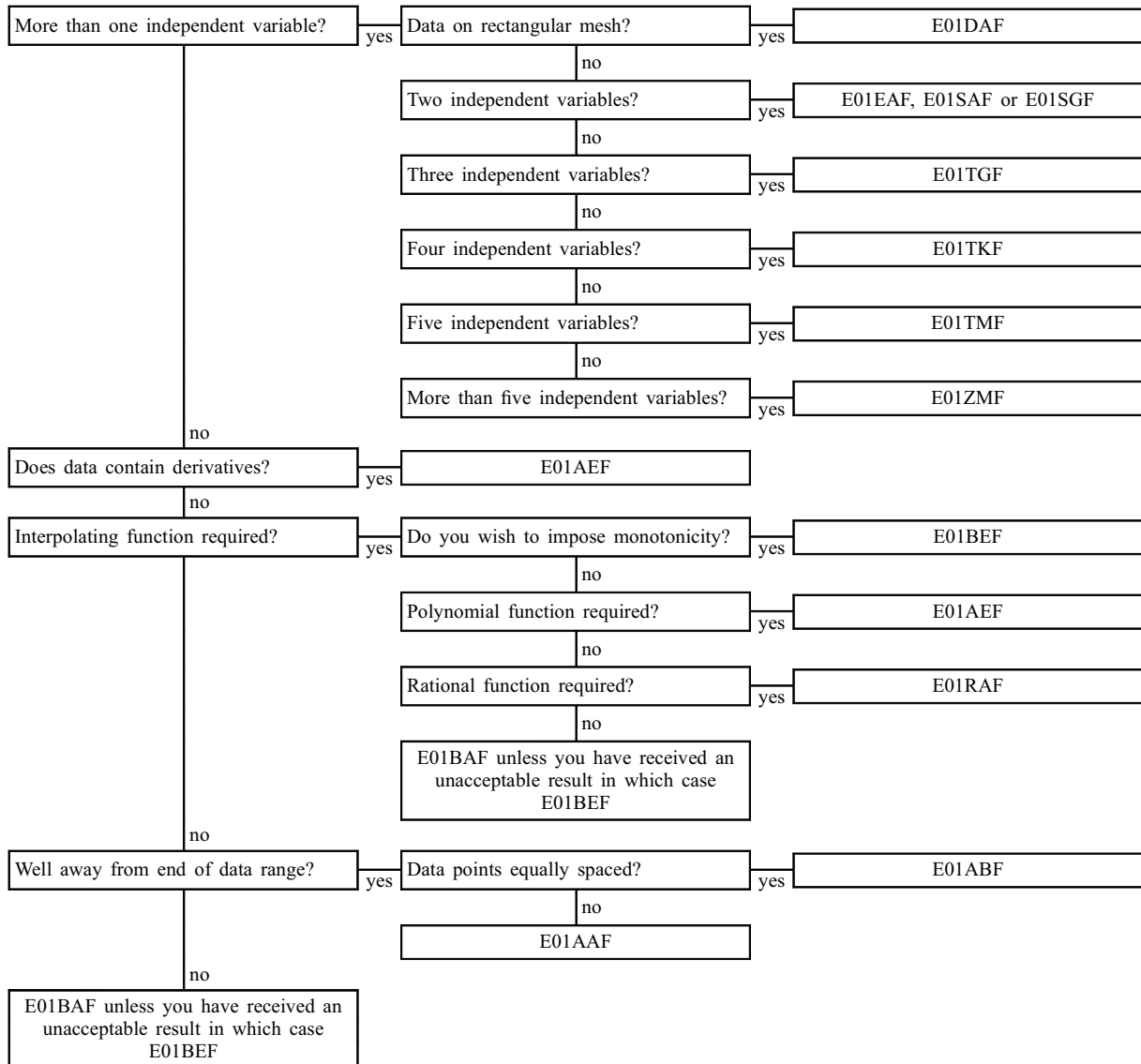
3.6 Multi-dimensional interpolation

3.6.1 Arbitrary data

Interpolation of scattered data in d-dimensions, where $d > 2$, is provided by routine E01ZMF. This extends the local method of E01TGF and E01TKF to higher dimensions. Evaluation of the interpolant, which has continuous first derivatives, is carried out by routine E01ZNF.

4   Decision Tree

Tree 1

More than one independent variable?	yes →	Data on rectangular mesh?	yes →	E01DAF

no ↓

Two independent variables?	yes →	E01EAF, E01SAF or E01SGF

no ↓

Three independent variables?	yes →	E01TGF

no ↓

Four independent variables?	yes →	E01TKF

no ↓

Five independent variables?	yes →	E01TMF

no ↓

More than five independent variables?	yes →	E01ZMF

no ↓

Does data contain derivatives?	yes →	E01AEF

no ↓

Interpolating function required?	yes →	Do you wish to impose monotonicity?	yes →	E01BEF

no ↓

Polynomial function required?	yes →	E01AEF

no ↓

Rational function required?	yes →	E01RAF

no ↓

E01BAF unless you have received an unacceptable result in which case E01BEF

no ↓

Well away from end of data range?	yes →	Data points equally spaced?	yes →	E01ABF

no ↓

E01AAF

no ↓

E01BAF unless you have received an unacceptable result in which case E01BEF

5   Functionality Index

Derivative,
   of interpolant,

Evaluation,
   of interpolant,

6 Auxiliary Routines Associated with Library Routine Parameters

None.

7   Routines Withdrawn or Scheduled for Withdrawal

The following lists all those routines that have been withdrawn since Mark 18 of the Library or are scheduled for withdrawal at one of the next two marks.

Withdrawn Routine	Mark of Withdrawal	Replacement Routine(s)
E01SEF	20	E01SGF
E01SFF	20	E01SHF

8   References

Dahlquist G and Björck Å (1974) *Numerical Methods* Prentice–Hall

Fröberg C E (1970) *Introduction to Numerical Analysis* Addison–Wesley