1 SUMMARY

To calculate a **weighted least squares fit to given data values by a cubic spline** which has knots assigned by the user. Let the given data values be y_i, i=1,2,...,m with weights w_i, i=1,2,...,m. The subroutine calculates a cubic spline $s(x)$, with knots ξ_j, j=1,2,...,n and smoothing factors θ_j, j=2,3,...,n−1 assigned by the user, which minimizes

$$\sum_{i=1}^{m}(w_i(y_i-s(x_i)))^2+\sum_{j=2}^{n-1}(\theta_j d_j)^2$$

where d_j is the discontinuity in the third derivative of $s(x)$ at the jth knot ξ_j.

The subroutine returns the values $s(\xi_j)$, and first derivative values $s'(\xi_j)$, j=1,2,...,n which together with the knots ξ_j, j=1,2,...,n define the spline.

ATTRIBUTES — Version: 1.0.0. **Types:** `VB06A`, `VB06AD` **Calls:** `FD05`. **Original date:** November 1973. **Origin:** M.J.D. Powell.

2 HOW TO USE THE PACKAGE

2.1 The argument list and calling sequence

The single precision version

```
CALL VB06A(M,N,XD,YD,WD,RD,XN,FN,GN,DN,THETA,IPRINT,W)
```

The double precision version

```
CALL VB06AD(M,N,XD,YD,WD,RD,XN,FN,GN,DN,THETA,IPRINT,W)
```

`M` is an `INTEGER` which must be set by the user to m, the number of data points. Note that M ≥ 1, and may be decreased by the routine if certain of the restrictions are violated, see section 3.

`N` is an `INTEGER` which must be set by the user to n, the number of knot points. Note that N ≥ 2, and may be decreased by the routine if certain of the restrictions are violated, see section 3.

`XD` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ m which must be set by the user to the positions x_i, i=1,2,...,m. These must be ordered such that $\xi_1 \le x_1 \le x_2 \le \ldots x_m \le \xi_n$.

`YD` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ m which must be set by the user to the data values y_i, i=1,2,...,m.

`WD` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ m which must be set by the user to the weights w_i, i=1,2,...,m. If an unweighted fit is required the user must set `WD(I)=1` for `I=1,M`.

`RD` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ m which will be set by the subroutine to the residuals $r_i=y_i-s(x_i)$, i=1,2,...,m.

`XN` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ n set by the user to the knot values ξ_j, j=1,2,...,n. These must be ordered such that $\xi_1<\xi_2<\ldots<\xi_n$.

`FN` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ n set by the subroutine to the values of the spline at the knot points, i.e. $s(\xi_j)$, j=1,2,...,n.

`GN` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ n set by the subroutine to the values of the first derivative of the spline at the points, i.e. $s'(\xi_j)$, j=1,2,...,n.

`DN` is a `REAL` (`DOUBLE PRECISION` in the D version) array of length ≥ n set by the subroutine to the values of the third derivative discontinuities at the knots $s'''(\xi_j+0)-s'''(\xi_j-0)$, j=2,3,...,n−1. The numbers `DN(1)` and `DN(N)` are used as working space.

THETA is a REAL (DOUBLE PRECISION in the D version) array of length ≥ n which must be set by the user to the parameters θ_j in the smoothing term in the sum of squares given in section 1. THETA(1) and THETA(N) are ignored by the subroutine but the user must set THETA(J), J=2, N-1 to θ_j, j=2,3,...,n–1. Choosing values for the θ_j's must be done with care and the topic is discussed in some detail by M.J.D. Powell in AERE report TP308 (1967). We give here a few general points. If θ_j=0, j=2,...,n–1 a straightforward least squares fit is obtained equivalent to the result produced by VB05B/BD. Making the θ_j's uniformly very large will force all the third derivative discontinuities to zero and the spline fit will tend towards a cubic polynomial fit. Varying the magnitudes of the θ_j's relative to each other will make the fit less flexible where they are large and more flexible where they are small. Apart from these considerations one should also aim to get a smooth set of d_j's and to keep the two parts of the sum of squares in section 1 about the same size. Note that the library routine VC03A/AD is available for solving the spline fitting problem and it chooses knots and smoothing factors automatically.

IPRINT is an INTEGER set by the user to control the printing. If IPRINT > 0 details of the fit will be printed giving values of ξ_j, $s(\xi_j)$, $s'(\xi_j)$, j=1,2,...,n, d_j and θ_j, j=2,3,...,n–1 and x_i, y_i, w_i, $s(x_i)$ and r_i, i=1,2,...,m. If IPRINT ≤ 0 no printing will be done except possibly diagnostic messages in the event of an error.

W is a REAL (DOUBLE PRECISION in the D version) array of length ≥ 8n+36 which must be provided by the user for the subroutine to use for workspace.

3 GENERAL INFORMATION

Use of common: : None

Workspace: : 8n+36 words provided by the user in the argument W. **Calls:** : None

Input/output: : Results are printed when IPRINT > 0 and diagnostic messages are printed in the event of an error. All printing is done on FORTRAN stream 6 and each set of results starts on a new page.

Restrictions:

$m \geq 1$

$n \geq 2$

$\xi_1 < \xi_2 < ... < \xi_n$

$\xi_1 < x_1 < x_2 ... < x_m < \xi_n$

If the restrictions are not satisfied the subroutine usually deletes some data and/or knots so that it can provide a fit, in which case the value of M and/or N is decreased and a diagnostic message is printed. Also the information in the one-dimensional arrays is compressed to match the new value of M and/or N. Note that a fit is defined unambiguously only if M ≥ 4. **Original date:** : July 1967, completely re-written, November 1973.

4 METHOD

The method is similar to that of VB05B/BD, but the smoothing term gives n–2 additional equations, each depending on 5 unknowns. The smoothing term is incorporated because the unsmoothed least squares spline fit tends to exhibit spurious oscillations. Also a modification is made to reduce large errors which tend to occur at the ends of the range. This modification consists of minimizing expression (A) in section 1 subject to fixed values of $s'''(\xi_1)$ and $s'''(\xi_n)$. These fixed values are calculated to make the first term of expression (A) (i.e. the weighted sum of squares of residuals), least. The smoothing term is discussed in M.J.D. Powell *The local dependence of least squares cubic splines,* AERE Report TP308 (1967).