Chapter 4 -- Numerical Methods

4.1 Overview

The numerical solution of the state vector equation is described in this chapter. The equation is a two-point boundary value problem in its simplest form, and becomes a multi-point BVP in its more general configuration. The equation may be linear or nonlinear, depending on the strain-displacement relations chosen. There are at least two ways to solve the problem: the "Field Method" as presented in Cohen (1974), or using a shooting technique [Ascher, et. al. (1988), Keller (1968), Kant and Ramesh (1981), Stoer and Bulirsch (1991), Kalnins (1964)].

The field method has the advantages of being numerically stable and relatively fast compared to other solution procedures but is difficult to implement because it requires special treatment of kinematic boundary conditions. The method is based upon a transformation (a Ricatti transformation; see Ascher, et. al. (1988), pp. 164-170, Jordan and Shelley (1966)) of the two-point BVP into two separate initial value problems to be integrated in succession. The general shooting technique is slower than the field method because numeric instability must be handled by segmentation of the solution region (i.e., *multiple* shooting), but is adaptable to the solution of nonlinear problems, and makes no distinction between kinematic and static variables.

There are any number of possible implementations of the multiple shooting method. We have chosen to use the so-called *stabilized marching method*, as presented by Ascher, Mattheij, and Russell (1988). This choice was made based on the ability of the stabilized marching technique to guarantee stability of the solution procedure -- other methods might not be able to make such a guarantee; those methods seem to require more user judgement in the selection of shooting points.

In addition to an exposition of the stabilized marching technique, the current chapter contains an explanation of cubic spline interpolation. The cubic spline is used in the current analysis in order to model the geometry of the meridian. It is felt that use of a spline function, wherein the entire meridian is described by the position of just a few points, makes the resulting analysis code easier to use than it would otherwise be. If the geometry were to be described exactly, then a new subroutine would need to be written for analysis of each new meridian shape. Furthermore, cubic

spline functions have been used successfully for structural shape optimization (Hinton, Rao and Sienz (1992), Hinton and Rao (1993), Weck and Steinke (1983-4)). Incorporation of the cubic spline here thus sets the code up for possible future use within an optimization program.

4.2 The stabilized marching method

4.2.1 Shooting for a linear, two-point boundary value problem

For the sake of exposition, assume that a two-point BVP with variable coefficients may be written in the form

$$\frac{dy}{ds} = A(s)y(s) + P(s) \qquad a < s < b$$

$$B_a y(a) = \beta_a \qquad B_b y(b) = \beta_b$$
(4.1)

where y is an n-vector, s is the independent variable, A(s) is the (n-by-n) coefficient matrix, P(s) is an n-vector of loading terms, B_a is (p-by-n), B_b is ((n-p)-by-n), β_a is a p-vector and β_b is a (n-p)-vector. That is, there are p conditions prescribed at the initial end, and (n-p) conditions prescribed at the final end, for the nth order system. The form of equation (4.1) is the same as that of equations (3.11) and (3.9), under the assumption of separated boundary conditions. Other boundary conditions are possible of course, and such situations are also solvable by shooting, but the solution procedure is made a bit more numerically challenging. Separated BC's frequently occur, and are the usual case for shell problems. More information on handling non-separated BC's may be found in Keller (1968) and Ascher, et. al. (1988).

The essence of a shooting procedure is that it solves a linear two-point boundary value problem by direct numerical integration as if it were an initial value problem. This is easily done -- many good integration routines exist -- once the initial values are properly chosen. The trick lies in proper choice of initial conditions (IC's). The IC vectors must be chosen in such a way as to satisfy exactly the known boundary conditions on the initial edge, without imposing unnecessary restrictions on the values of the unprescribed dependent variables.

The shooting procedure will involve a superposition of a single inhomogeneous solution with a number (n-p) of homogeneous solutions:

$$y(s) = Y(s)\xi + v(s) \tag{4.2}$$

where y(s) is the vector of dependent variables, s represents the independent variable, Y(s) is a (n-by-(n-p)) matrix whose columns are homogeneous solution vectors, ξ is a (n-p)-vector of superposition constants, and v(x) is the (n-by-1) particular solution vector. The homogeneous and particular equations will thus be given by

$$\frac{dY}{ds} = A(s)Y(s)$$

$$\frac{dv}{ds} = A(s)v(s) + P(s)$$
(4.3)

Then with the boundary conditions at the initial edge defined as in (4.1), the use of equation (4.2) yields

$$B_a Y(a) \xi + B_a v(a) = \beta_a, \tag{4.4}$$

which may be satisfied by taking either

$$B_a Y(a) = I$$

$$\xi = \beta_a$$

$$v(a) = 0$$
(4.5)

or

$$B_a Y(a) = 0$$

$$B_a v(a) = \beta_a$$
(4.6)

The choice of initial conditions given by equation (4.5) leads to standard shooting as used by, for example, Kalnins (1964). The stabilized marching method of Ascher, et. al. (1988) begins with BC's as in equation (4.6). We will here utilize (4.6).

The BC's (4.6) may be satisfied by use of QR factorization of the transpose of the initial condition coefficient matrix, B_a^T . The method is detailed by Ascher, et. al., but is repeated here. Dimensionally, B_a^T is (n-by-p); using QR factorization, it may be decomposed into a product of an orthogonal matrix H(n-by-n) and a (n-by-p) matrix R which is block upper triangular:

$$B_{a}^{T} = HR$$

$$HH^{T} = H^{T}H = I,$$

$$R = \begin{bmatrix} r^{T} \\ 0 \end{bmatrix},$$
(4.7)

where r is (p-by-p) lower triangular and the zero matrix is ((n-p)-by-p). Now, H is viewed as a concatenation of two matrices, set side by side: H = [H, H], where H is of dimension (n-by-p) and H is (n-by-(n-p)). We thus get from (4.6), (4.7)

$$[r,0] \begin{bmatrix} \widehat{H}^T \\ \widehat{H}^T \end{bmatrix} Y(a) = 0,$$

which is satisfied by

$$Y(a) = H (4.8)$$

We also get from (4.6), (4.7)

$$[r,0] \begin{bmatrix} \widehat{H}^T \\ \widehat{H}^T \end{bmatrix} v(a) = \beta_a,$$

which is satisfied by

$$v(a) = \widehat{H} r^{-1} \beta_a \tag{4.9}$$

With the initial condition vectors found by equations (4.8) and (4.9), we may now proceed to integrate the initial value problems found by inserting the IC vectors into the equation (4.1). The integration limits will be from a to b. This will entail a single nonhomogeneous integration with IC

$$y(a) = v(a)$$

and (n-p) homogeneous integrations with the columns of Y(a) as IC vectors. The integration process yields the particular solution vector v(b), and the homogeneous solution matrix Y(b).

At the end of the integrations, we must find the superposition constant vector ξ . This is

done by requiring the satisfaction of the final end BC's: $B_b y(b) = \beta_b$. Using the superposition (4.2), we get

$$B_h[Y(b)\xi + v(b)] = \beta_h$$

or

$$\xi = [B_b Y(b)]^{-1} [\beta_b - B_b v(b)]$$
 (4.10)

Following the above detailed procedure, we have found all of the necessary bits, and we now find the values of y at the two shooting points by use of equation (4.2).

4.2.2 Multiple shooting for linear boundary value problems

It often happens that the solution to an ordinary differential equation has terms which, left unchecked, grow rapidly as the independent variable is increased. In such a case, any attempt to numerically integrate the equation over a large region quickly results in a loss of accuracy. This loss of accuracy occurs as the growing modes quickly overflow the memory register or lead to the necessity of taking small differences between large terms, thus overwhelming the more "stable" modes. Such a problem does not occur when the BVP is solved exactly, because consideration of the boundary conditions leads to the logical conclusion that the coefficients of rapidly growing terms must vanish. For shooting, the problem is handled by the technique of *multiple* shooting.

The multiple shooting technique works in much the same way as the standard shooting method; the difference is that in multiple shooting the region of solution, a < s < b, is partitioned into a number of smaller regions or segments, over each of which the standard shooting technique is applied. The partitioning, when followed by a rescaling of integrated solutions, prohibits excessive growth of the "unstable" modes. The final solution is then found by enforcing certain conditions of transition at the common shooting points between segments. These transition conditions may allow for a good deal of flexibility in the problem: for example, they might reflect the presence of discontinuities in the dependent variable vector (e.g., concentrated loading) or in the constituent functions which make up the coefficient matrix A(s) of equation (4.1).

There are two subclasses of the multiple shooting technique. If all of the segments can be

handled independently, then the integrations on all segments may be done simultaneously. This leads to the technique of *parallel shooting*. If the segments must be handled in succession, then the technique proceeds segment-by-segment, according to the value of the independent variable. Such techniques are termed *marching techniques*. The choice of whether to use parallel shooting or a marching technique depends in part upon whether there will be any discontinuities in the region of solution; for the current work, a marching technique has been chosen in order to accommodate possible geometric and load discontinuities.

The first step in the process of multiple shooting is the partitioning of the region of integration into a number (N) of segments, whose endpoints are denoted by s_i , i=1,2,...,N+1:

$$a = s_1 < s_2 < \dots < s_{N+1} = b (4.11)$$

Note that, strictly speaking, the *a priori* partitioning of equation (4.11) is only necessary for parallel shooting. For marching techniques, it is possible to set the shooting points s_i on an as-you-go basis. For instance, Ascher, et al suggest dynamically setting shooting points based upon the growth of unstable modes as indicated by the degree of distinctness between the (initially orthonormal) homogeneous solution vectors; Stoer and Bulirsch recommend setting shooting points by comparison of the integrated solutions on each segment to a predefined comparison function. Regardless of whether the shooting points are predetermined or chosen dynamically, we will have on the i^{th} segment the ODE

$$\frac{dy_i}{ds} = A(s)y_i(s) + P(s) s_i < s < s_{i+1} (4.12)$$

In addition to equation (4.12), the problem is defined by the initial and final end conditions

$$B_a y_1(a) = \beta_a$$

$$B_b y_N(b) = \beta_b$$
(4.13)

and the transition conditions

$$K_i y_i(s_{i+1}) = y_{i+1}(s_{i+1}) - \Delta_{i+1}$$
 (4.14)

In equation (4.14), the term K_i represents a connectivity matrix for the dependent variable vector,

and the term Δ_{i+1} is a discontinuity vector. For example, full continuity is described by $K_i = I$, and the matrix $\Delta_{i+1} = 0$. It may be noted that as a consequence of the segmentation, the two-point BVP has been transformed into a multi-point BVP, with internal boundary conditions defined by equation (4.14).

Superposition is assumed to apply on all segments, so that we have

$$y_i(s) = Y_i(s)\xi_i + v_i(s)$$
 $s_i \le s \le s_{i+1},$ (4.15)

leading to the homogeneous and particular equations

$$\frac{dY_i}{ds} = A(s)Y_i(s)$$

$$s_i < s < s_{i+1}$$

$$\frac{dv_i}{ds} = A(s)v_i(s) + P(s)$$
(4.16)

The process continues by the selection of IC's on the first segment. These IC's may be chosen just as they were in section (4.2.1), i.e., using equations (4.8), (4.9). Hereafter, the particular initial condition vector on the i^{th} segment will be denoted by α_i , and the matrix of initial condition vectors for homogeneous integrations will be known as F_i . That is, we take $F_i = Y_i(s_i)$, and $\alpha_i = v_i(s_i)$. By equations (4.7), (4.8), (4.9), we get

$$B_{a}^{T} = \widehat{H}, \overline{H} \begin{bmatrix} r^{T} \\ 0 \end{bmatrix}$$

$$F_{1} = \overline{H}$$

$$\alpha_{1} = \widehat{H} r^{-1} \beta_{a}$$

$$(4.17)$$

We now may input the initial conditions F_1 and α_1 into equation (4.16), and integrate to get $Y_1(s_2)$ and $v_1(s_2)$.

For all segments following the first, the initial condition vectors must be chosen based upon the integrated solutions on the previous segment, taking into account the presence of any known discontinuities. In the stabilized marching method, the choice of new IC's is made based upon a reorthogonalization of the homogeneous solutions of the previous segment.

Reorthogonalization of the homogeneous solution matrix provides for stabilization of the method, giving rise to the name "stabilized marching." Stabilization occurs as a result of two effects: first, by rendering the IC vectors normal, any large modes are forced to take on reasonable values prior to integration. This first effect is not unique to the stabilized marching method; it is a key aspect of all marching techniques. The second stabilization effect has to do with the eventual calculation of the superposition constants, ξ_i . More will be said on this later.

A (n-by-(n-p)) column-orthonormal matrix G_{i+1} and a ((n-p)-by-(n-p)) upper triangular matrix Γ_i are defined by the equation

$$K_i Y_i(s_{i+1}) = G_{i+1} \Gamma_i (4.18)$$

These new matrices may be found using a combination of singular value decomposition (SVD) and QR factorization. The process goes as follows: by SVD,

$$K_{i}Y_{i}(s_{i+1}) = U_{i}\Sigma_{i}V_{i},$$
 (4.19)

in which U_i is (n-by-(n-p)) and column-orthonormal, Σ_i is ((n-p)-by-(n-p)) diagonal, and V_i is ((n-p)-by-(n-p)) and orthogonal. QR factorization next yields

$$(\Sigma_i V_i) = Q_i R_i, (4.20)$$

where Q_i is ((n-p)-by-(n-p)) orthogonal and R_i is ((n-p)-by-(n-p)) upper triangular. We then get the form of (4.18) from (4.19), (4.20) by taking

$$G_{i+1} = U_i Q_i$$

$$\Gamma_i = R_i$$
(4.21)

The equation (4.18) is now used in (4.15), the result substituted into (4.14) and some rearrangement is performed to get

$$\xi_i = \Gamma_i^{-1} G_{i+1}^T [F_{i+1} \xi_{i+1} + \alpha_{i+1} - \Delta_{i+1} - K_i v_i(s_{i+1})]$$
(4.22)

Next, using (4.15), (4.18) and (4.22) in (4.14) and rearranging, we get

$$(I - G_{i+1}G_{i+1}^T)[F_{i+1}\xi_{i+1}] + (I - G_{i+1}G_{i+1}^T)\alpha_{i+1}$$

$$= (I - G_{i+1}G_{i+1}^T)[K_i v_i(s_{i+1}) + \Delta_{i+1}]$$
(4.23)

In view of the column-orthonormality of G_{i+1} , i.e., $G_{i+1}^T G_{i+1} = I$, we may take

$$F_{i+1} = G_{i+1}, (4.24)$$

which will then ensure that satisfaction of the conditions of transition will not depend upon the calculated value of ξ_{i+1} . This leaves the equation

$$(I - G_{i+1}G_{i+1}^T)\alpha_{i+1} = (I - G_{i+1}G_{i+1}^T)[K_i v_i(s_{i+1}) + \Delta_{i+1}]$$

Clearly, one solution to this latest equation is given by $\alpha_{i+1} = [K_i v_i(s_{i+1}) + \Delta_{i+1}]$, but again considering the column-orthonormality of G_{i+1} , we may more generally take

$$\alpha_{i+1} = (I - G_{i+1}G_{i+1}^T)^r [K_i v_i (s_{i+1}) + \Delta_{i+1}]$$
(4.25)

where r is any non-negative integer. It may be noted that if $K_i = I$, $\Delta_{i+1} = 0$, we have the situation of continuity of the dependent variable vector at a shooting point. If, additionally, the integer r is set equal to one, then we obtain the initial condition vector suggested by Ascher, et al, and by Keller (1968). We have arrived at the choice for α_{t+1} based only upon consideration of connectivity, but it is pointed out by Keller (for the simpler case where there are no geometric or load discontinuities) that the formulation of equation (4.25) with r=1 yields a particular solution IC vector which is mutually orthogonal to the homogeneous IC vectors. This orthogonality property is retained in the generalization presented here, for r > 0. The initial conditions (4.24) and (4.25) thus represent a generalization of the initial conditions of the given references to situations where the dependent variable is non-continuous at a shooting point.

Following completion of the integration steps over all of the N segments, the shooting procedure is completed by solution for the superposition constants. This last step is done by first enforcing the final-end BC $B_b y(b) = \beta_b$. The final segment superposition constants are found from this BC by using the superposition equation (4.15) and rearranging to get (c.f. equation

(4.10)

$$\xi_N = [B_b Y_N(b)]^{-1} [\beta_b - v_N(b)]$$
 (4.26)

After solution of equation (4.26), all of the other superposition constants are found by use of equations (4.15), (4.18), (4.24), (4.25) in (4.15) to get

$$\xi_{i} = \Gamma_{i}^{-1} [\xi_{i+1} - G_{i+1}^{T} (K_{i} \nu_{i} (s_{i+1}) + \Delta_{i+1})]$$
(4.27)

It was stated earlier that there was an additional stabilization feature due to the reorthogonalization procedure. This additional feature may be seen by looking at equation (4.27): because the matrix Γ_i is upper triangular, the inversion of Γ_i is not really necessary. Instead, a simple backsubstitution procedure may be used. As a result, the effects of growth of unstable modes are not felt when solving for the superposition constants. That is, the reorthogonalization improves the conditioning of the equations to be solved for the superposition constants. Ascher, et al argue that it is quite often in this step that other multiple shooting methods fail.

The stabilized marching method for multiple shooting thus proceeds as follows:

- Partition the region of solution according to equation (4.11).
- Choose initial segment IC's by use of equations (4.18).
- Integrate equation (4.16) over the first segment to get $Y_1(s_2)$ and $v_1(s_2)$.
- For segments number 2 through N, choose new segment IC's according to equations (4.18), (4.24), (4.25). Integrate from $s = s_i$ to $s = s_{i+1}$.
- Repeat the last step until the final end is reached, i.e., until $s = s_{N+1} = b$.
- Solve for the final segment superposition constants, by equation (4.26).
- Use the recursion relation (4.27) to find the superposition constants for all segments preceding the last.

• Superpose the homogeneous and particular solutions on all segments and at all shooting points by use of equation (4.15).

4.3 Cubic spline interpolation

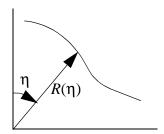


Fig. 4.1 Spline Coordinates

Suppose that the location of a curve in two-dimensional space is to be described by two coordinates, say $(\eta, R(\eta))$ as shown in Fig. 4.1, and suppose also that a number (N+1) of discrete pairs, $(\eta_i, R_i(\eta_i))$, i=0,1,...,N are known to exist on the curve. We refer to these discrete points as "knots" (or "nodes" or "support points") of the curve. If the functional form of $R(\eta)$ is not known then it will be necessary to interpolate $R(\eta)$, if values of R are desired at positions between the knots.

Many techniques exist for performance of numerical interpolation; most rely upon the use of polynomial expressions. It is possible, for example, to write a single polynomial whose plot passes through all of the knots. Unfortunately, if the number of knots is large, then the polynomial will have a high order and will thus be prone to exhibit much oscillatory behavior. As an alternative, we may interpolate using a piecewise polynomial expression, in which each piece is valid only over a single subinterval $[\eta_i, \eta_{i+1}]$. The piecewise polynomial function is known as a spline function, and we will refer to the polynomial over a single subinterval as a spline polynomial. That is, the spline function consists of a set of N spline polynomials. The order of the spline polynomials may be taken to be of a low order, depending upon the required continuity; the spline function is thus less likely to exhibit unwanted oscillatory behavior. Here we will describe the

cubic spline, in which all of the spline polynomials are of third order; the derivation is like that of Stoer and Bulirsch (1991), pp. 97-101.

Define

$$Y = \{R_i\}$$
 $i = 0, 1, ..., N$
 $h_{i+1} = \eta_{i+1} - \eta_i$ $i = 0, 1, ..., N-1$

then we seek a function $S[Y;\eta]$ which interpolates $R(\eta)$ and has the property

$$S[Y;\eta_i] = R_i (4.28)$$

The second derivative of $S[Y;\eta]$ with respect to η is known as the moment of S, and we denote

$$S''[Y;\eta_i] = M_i$$
 $i = 0, 1, ..., N$

Cubic spline polynomials provide continuity of the moments at the knots; we assume a linear variation of $S''[Y;\eta]$ over each subinterval. Thus,

$$S''[Y;\eta] = M_i \frac{(\eta_{i+1} - \eta)}{h_{i+1}} + M_{i+1} \frac{(\eta - \eta_i)}{h_{i+1}}$$

$$S'[Y;\eta] = -M_i \frac{(\eta_{i+1} - \eta)^2}{2h_{i+1}} + M_{i+1} \frac{(\eta - \eta_i)^2}{2h_{i+1}} + A_i \qquad \eta_i \le \eta \le \eta_{i+1} \qquad (4.29)$$

$$S[Y;\eta] = M_i \frac{(\eta_{i+1} - \eta)^3}{6h_{i+1}} + M_{i+1} \frac{(\eta - \eta_i)^3}{6h_{i+1}} + A_i (\eta - \eta_i) + B_i$$

where A_i , B_i are constants of integration.

Using (4.28) in (4.29), applied at η_i , η_{i+1} , A_i and B_i are found to be given by

$$A_{i} = \frac{R_{i+1} - R_{i}}{h_{i+1}} + \frac{h_{i+1}}{6} (M_{i} - M_{i+1})$$

$$B_{i} = R_{i} - M_{i} \frac{h_{i+1}^{2}}{6}$$
(4.30)

The equations (4.30) may be used in the last of equations (4.29), allowing us to write

$$S[Y;\eta] = \alpha_i + \beta_i(\eta - \eta_i) + \gamma_i(\eta - \eta_i)^2 + \delta_i(\eta - \eta_i)^3$$
 (4.31)

with

$$\alpha_{i} = R_{i}$$
 $\beta_{i} = \frac{R_{i+1} - R_{i}}{h_{i+1}} - \frac{2M_{i} + M_{i+1}}{6}h_{i+1}$

$$\gamma_{i} = \frac{1}{2}M_{i}$$
 $\delta_{i} = \frac{M_{i+1} - M_{i}}{6h_{i+1}}$

which is valid on the interval $[\eta_{i-1}, \eta_i]$, for i = 0, 1, ..., N-1

The spline function is completely defined by (4.31) once the values of M_i are known, for i=0,1,...,N. We thus have (N+1) unknowns to find. We have already enforced the required values of R at the knots, and we have asserted continuity of the moments at the interior knots. This leaves only continuity of $S'[Y;\eta]$ to be evaluated. Using (4.30) in (4.29), we get

$$S'[Y;\eta] = -M_i \frac{(\eta_{i+1} - \eta)^2}{2h_{i+1}} + M_{i+1} \frac{(\eta - \eta_i)^2}{2h_{i+1}} + \frac{R_{i+1} - R_i}{h_{i+1}} - \frac{h_{i+1}}{6} (M_{i+1} - M_i)$$
 (4.32)

Continuity of $S'[Y;\eta]$ is applied by asserting single-valuedness of this term at the i^{th} knot, regardless of whether the knot is viewed as a part of the i^{th} or $(i+1)^{th}$ subinterval, i=1,2,...,N-1. Thus it may be seen that the following condition must be satisfied at all of the (N-1) interior knots:

$$\frac{h_{i+1}}{6}M_i + \frac{(h_{i+1} + h_{i+2})}{3}M_{i+1} + \frac{h_{i+2}}{6}M_{i+2} = \frac{R_{i+2} - R_{i+1}}{h_{i+2}} - \frac{R_{i+1} - R_i}{h_{i+1}}$$
(4.33.a)

for i=0,1,...,N-2. We now have (N-1) equations for (N+1) unknowns; splining may be completed only with the addition of two more conditions. These conditions may be any two conditions on $S'[Y;\eta]$ or $S''[Y;\eta]$ at the end knots. For our purposes, we assume the slope R' is known at each end:

$$S'[Y;\eta_0] = R'_0$$

$$S'[Y;\eta_N] = R'_N$$

which yield

$$-\frac{h_1}{3}M_0 - \frac{h_1}{6}M_1 = R'_0 - \frac{R_1 - R_0}{h_1}$$

$$\frac{h_N}{6}M_{N-1} + \frac{h_N}{3}M_N = R'_N - \frac{R_N - R_{N-1}}{h_N}$$
(4.33.b)

Spline interpolation is thus completed using the spline polynomials of (4.31), with values of M_i given by solution of the linear algebraic system of (4.33.a), (4.33.b).