National Conference with International Participation
ENGINEERING MECHANICS 2006
Svratka, Czech Republic, May 15 – 18, 2006
paper no. **305**

AUTOMATIC GENERATION OF VARIATIONAL STATEMENT

T. Nánási*

Summary: *Variational statements corresponding to boundary value problems are usually computed by hand calculations. In this paper an algorithm is derived which enables to implement the computation of variational statements as a computer program in the enviroment of symbolic programming languages such as Mathematica or Maple.*

1. Introduction

In general, the variational statement for the weak solution of the two-point boundary value problem $L\ u(x) = f(x),\ B_s\ u(s) = h_s$, defined on interval $x \in (0,1)$ with end points s=0,1 for operator $L = (L, B_0, B_1)$ of order $2m$ has the following form: find $u \in U$ such that $W(u,v) = 0$ for all $v \in V$. One of the main ideas underlying the concept of the weak solution is that with properly chosen bilinear form $W(u,v)$ and suitable Hilbert (Sobolev, etc.) spaces U, V it is possible to reduce the solution of the original problem of order $n = 2m$ to the solution of the problem, in which only derivatives up to the m-th order appear. Except for this order reduction there is an additional benefit in the possibility to solve the problems on spaces with less demanding requirements on the continuity and differentiability of the admissible functions. In most of practically used formulations the homogeneous geometric boundary conditions enter the weak formulation in definition of the spaces U, V; while the dynamic boundary conditions enter the definition of the bilinear form $W(u,v)$. This form should be chosen so that the weak solutions automatically guarrant the fullfilment of the dynamic boundary conditions. However, as discussed by Rektorys (1985), the question of the construction of the actual bilinear form $W(u,v)$ corresponding to given boundary value problem is not trivial, especially when complicated boundary conditions are considered.

In this paper the Dirichlet's remainder is represented in matrix form. From two decompositions of this matrix an algorithm is deduced, which resolves the above problem of assigning a bilinear form for weak formulation to the given boundary value problem at least in the class of one-dimensional linear differential operators. The proposed method is based on rewriting of the Green's identity $(Lu,v) - (u, L^*v) - H[u,v] = 0$ to a form, to which the given data of the boundary value problem can be substituted directly.

* Ing. Tibor Nánási, CSc.: Department of Applied Mechanics, Slovak University of Technology, Faculty of Materials Science and Technology, Paulínska 16, 917 24 Trnava, Slovakia, e-mail: tibor.nanasi@stuba.sk

2. Matrix representation of the operator and of the Green`s identity

Consider two-point boundary value problem

$$Lu(x) = f(x), \quad x \in (0,1),$$
$$\mathbf{B}_0 u(0) = \mathbf{h}_0, \quad \mathbf{B}_1 u(0) = \mathbf{h}_1, \tag{1}$$

where the domain operator is the linear differential expression L of even order $n=2m$

$$L = a_n D^{(n)} + a_{n-1} D^{(n-1)} + a_{n-2} D^{(n-2)} + \ldots + a_1 D + a_{0,} \tag{2}$$

a_i are constant real coefficiens, $\mathbf{B}_0$, $\mathbf{B}_1$ are $m \times 2m$ matrices of rank m representing the coefficients of given boundary conditions and $D^{(k)} = d^k/dx^k$.

For simplicity we suppose $a_n=1$. Column vectors $\mathbf{h_0}$, $\mathbf{h_1}$ represent the nonhomogeneous part of boundary conditions and the state vector $\mathbf{u}$ is defined as

$$\mathbf{u}(x) = \left[u(x), u^{(1)}(x), \ldots, u^{(n-2)}(x), u^{(n-1)}(x)\right]^T \tag{3}$$

where $u^{(k)}(x)=D^{(k)}u(x)$. To study the properties of the above boundary value problem (1), let us consider the Green's identity

$$(Lu, v) = (u, L^* v) + H[u,v], \tag{4}$$

which reduces to the integration by parts formula in the case of one-dimensional domain. The H[u,v] is the Dirichlet's remainder term

$$H[u,v] = H_1[u(1), v(1)] - H_0[u(0), v(0)] = \sum_{s=0}^{s=1} (-1)^{s+1} \mathbf{u}^T(s) \mathbf{H}_s \mathbf{v}(s), \tag{5}$$

in which the Dirichlet's expressions H_s[u(s),v(s)], rewritten as bilinear forms with matrices $\mathbf{H_s}$ in the state vectors u(s) and v(s), are evaluated at the boundary points s=0 and s=1. L^* is the differential expression adjoint to the differential expression L.

For L defined by (2), the explicit form of the Green's identity reads

$$\left(\sum_{i=0}^{n} a_i u^{(i)}, v\right) = \left(u, \sum_{i=0}^{n} (-1)^i a_i v^{(i)}\right) + \sum_{s=1}^{s=1} (-1)^{s+1} \sum_{i=1}^{n} a_i \sum_{j=0}^{i-1} (-1)^j u^{(i-1-j)}(s) v^{(j)}(s). \tag{6}$$

From (6) we get directly the adjoint differential expression as

$$L^* = \sum_{i=0}^{n} (-1)^i a_i D^{(i)}.$$

Due to the unfriendly form of the Dirichlet's remainder term in (6), it is not quite clear how to substitute the given boundary conditions for functions $u(x)$ to the Green's formula and how to compute the adjoint boundary conditions for functions $v(x)$.We recall, that adjoint boundary conditions $\mathbf{B}_s^* \mathbf{v}(s) = \mathbf{0}$ together with given boundary conditions (considered now as homogeneous) are minimal conditions to eliminate the Dirichlet's remainder term. To achieve the aim of this paper, first we rewrite the Dirichlet's remainder to enable the direct substitution of the given data, i.e. of the L, $\mathbf{B}_0$, $\mathbf{B}_1$, $f(x)$, $\mathbf{h_0}$ and $\mathbf{h_1}$ into the Green's identity.

Following Nánási (1994) the domain equation $Lu = f$ can be written as an equivalent system of first order differential equations

$$\mathbf{L}u(x) = \mathbf{A}D(u) + \mathbf{B}u(x) = \mathbf{f}(x), \tag{7}$$

where

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-2} & a_{n-1} & 1 \\ -a_2 & -a_3 & -a_4 & \cdots & -a_{n-1} & -1 & 0 \\ a_3 & a_4 & a_5 & \cdots & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{n-1} & 1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} a_0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & a_2 & a_3 & a_4 & \cdots & a_{n-2} & a_{n-1} & 1 \\ 0 & -a_3 & -a_4 & -a_5 & \cdots & -a_{n-1} & -1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & -a_{n-1} & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}, \qquad (8)$$

$$\mathbf{f}(x) = [f(x), 0, 0, \ldots, 0]^T, \quad D = \frac{d}{dx}, \quad \mathbf{L} = \mathbf{A}D + \mathbf{B}.$$

The Green's identity in the state space has now the form

$$((\mathbf{A}D + \mathbf{B})\mathbf{u}(x), \mathbf{v}(x)) = (\mathbf{u}(x), (-\mathbf{A}^T D + \mathbf{B}^T)\mathbf{v}(x)) + \sum_{s=0}^{1} (-1)^{s+1} \mathbf{u}^T(s)\mathbf{A}^T \mathbf{v}(s). \qquad (9)$$

Inspection of (6) shows, that the Dirichlet's remainder in (6) can be expressed by the matrix $\mathbf{A}^T$, i.e.

$$\mathbf{H}_0 = \mathbf{H}_1 = \mathbf{A}^T.$$

From (9) the advantages of the formulation at the state space in terms of the special matrix **A** are obvious:

1) Both the adjoint expression $\mathbf{L}^* = -\mathbf{A}^T D + \mathbf{B}^T$ and the Dirichlet's remainder term are given by the same matrices **A** and **B**. Moreover, the matrix **B** is closely related to the matrix **A**, as easily checked by inspection.
2) Actually it is not necessary to carry out the process of integration by parts, as according to (9) its result is expressed directly in terms of input matrices **A**, **B**.

3. The first decomposition of the matrix $\mathbf{A}^T$

Let us decompose the matrix $\mathbf{A}^T$ at both end points to a product of two regular $2m\times 2m$ matrices

$$\mathbf{A}^T = \mathbf{A}_{L0}\mathbf{A}_{R0}, \quad \mathbf{A}^T = \mathbf{A}_{L1}\mathbf{A}_{R1}. \qquad (10)$$

To incorporate the given boundary conditions, we chose the half of the columns of the matrices $\mathbf{A}_{L0}$, $\mathbf{A}_{L1}$ to be identical with the rows of the boundary matrices $\mathbf{B}_0$ and $\mathbf{B}_1$, respectively. The rest of the columns of the matrices $\mathbf{A}_{Ls}$ can be chosen arbitrarily under the condition that the matrix $\mathbf{A}_{Ls}$ is regular. Then evidently the complementary rows of the matrices

$$\mathbf{A}_{Rs} = (\mathbf{A}_{Ls})^{-1}\mathbf{A}^T, \quad s = 0,1 \qquad (11)$$

can be identified with the matrices of adjoint boundary conditions $\mathbf{B}_1^*$ .By the term complementary rows we mean rows with row indices $2m+1-i_k$, where i_k are the column indices corresponding to the given boundary conditions in matrices $\mathbf{A}_{L0}$, $\mathbf{A}_{L1}$. For details see Nánási (1994). The decompositions (11) are considerably simplified, if the given boundary conditions are first transformed to canonical form, i.e. they are resolved with respect to the highest derivatives and all highest derivatives in each boundary conditions are eliminated from the rest of the boundary conditions. Then the given boundary conditions can be coded to the matrix $\mathbf{A}_{Ls}$ in such order, that the resulting matrix is triangular. The most simple form is achieved if the dummy columns contain single non-zero element, namely unity, on the diagonal. The matrix $\mathbf{A}^T$ has also triangular structure, therefore it follows directly from (11), that the resulting matrix $\mathbf{A}_{Rs}$ has triangular structure.

For convenience we introduce auxiliary matrices which are binary description of the given and adjoint boundary conditions. Let i_k, $k=1,2,...,m$, $1 \le i_k \le 2m$ are the column indices of those columns in $\mathbf{A}_{Ls}$,which correspond to given boundary conditions. Let us construct a binary vector $\boldsymbol{i}_{Ls}$ of length $2m$ such that $\mathbf{i}_{Ls,j} = 1$ for $j = i_k$ and $\mathbf{i}_{Ls,j} = 0$ for $j \ne i_k$. Let $\mathbf{i}_A$ is a vector of length $2m$, which has all components equal to unity. The binary description of adjoint boundary conditions is given by the auxiliary vector $\mathbf{i}_{Rs} = \mathbf{i}_A - \mathbf{i}_{Ls}$. Later we make use of substitution matrices defined by relations

$$\mathbf{I}_{us} = \mathbf{I}_{2m} - \mathrm{diag}\{\mathbf{i}_{Ls}\}, \quad \mathbf{I}_{vs} = \mathbf{I}_{2m} - \mathrm{diag}\{\mathbf{i}_{Rs}\} = \mathbf{I}_{2m} - \mathbf{I}_{us}, \tag{12}$$

where $\mathbf{I}_{2m}$ is $2m \times 2m$ identity matrix. The first decomposition thus provides an algebraic tool for direct computation of adjoint boundary conditions via relation (11).

4. The second decomposition of the matrix $\mathbf{A}^T$

To construct the variational statement it is necessary to integrate the constituent expressions in the bilinear expression (Lu,v) so that in the resulting integral part only derivatives of the order lower than $m+1$ appear. Essentially it would be sufficient to integrate only the terms $a_i u^{(i)}(x)\ v(x)$ with $i \ge m+1$. However, this type of integration would destroy the symmetry of originally symmetric problems. To preserve the symmetry, we integrate each of the expressions $a_i u^{(i)}(x)\ v(x)$ up to the half of its degree i. Full integration by parts of the terms with even order provides even number of boundary terms, so the notion of the half degree is intuitively clear. Contrary, the integration by parts of the terms with odd degree provides odd number of boundary terms and the notion of the half degree becomes unambigious. Optionally we may decide to a compromise - the critical derivatives may be split into equal halfs. One of them is included to integration by parts of the whole expression from the order $2m$ down to the order m and the second identical half-term is included to the integration of the whole expression from the order m down to the order 0. This gives rise to the half-terms on the diagonals of both matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ in formulas (17).

From detailed analysis of the full, partial or backward integration of the bilinear expressions $(Lu(x),v(x))$ resp. $(u(x), L^*v(x))$ we have the relations

$$\begin{aligned}
(Lu,v) &= (u,L^*v) + \mathbf{u}^T(s)(\mathbf{A}_1 + \mathbf{A}_2)\mathbf{v}(s)\big|_0^1,\\
(Lu,v) &= (L_1u,L_2v) + \mathbf{u}^T(s)\mathbf{A}_1\mathbf{v}(s)\big|_0^1,\\
(v,L^*v) &= (L_1u,L_2v) - \mathbf{u}^T(s)\mathbf{A}_2\mathbf{v}(s)\big|_0^1,
\end{aligned} \tag{13}$$

where

$$\mathbf{a}_0 = \begin{bmatrix} \mathbf{A}_0 & 0_{m+1,m} \\ 0_{m,m+1} & 0_{m,m} \end{bmatrix} \quad \text{and} \quad (L_1 u, L_2 v) = \int_0^1 \mathbf{u}^T(x)\,\mathbf{a}_0 \mathbf{v}(x)\,dx.$$

The actual form of the matrices $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{A}_2$ depends on the strategy chosen to integrate the terms with odd order of the derivative. The above mentioned one leads to matrices

$$\mathbf{A}_0 = \begin{bmatrix} a_0 & -\frac{1}{2}a_1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{2}a_1 & -a_2 & \frac{1}{2}a_3 & 0 & \cdots & 0 & 0 \\ 0 & -\frac{1}{2}a_3 & a_4 & -\frac{1}{2}a_5 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (-1)^{m-1}a_{m-2} & \frac{1}{2}(-1)^m a_{m-1} \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2}(-1)^{m-1}a_{m-1} & (-1)^m \end{bmatrix},$$

$$\mathbf{A}_1 = \begin{bmatrix} \frac{1}{2}a_1 & 0 & 0 & \cdots & 0 \\ a_2 & -\frac{1}{2}a_3 & 0 & \ldots & 0 \\ a_3 & -a_4 & \frac{1}{2}a_5 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & -1 & 0 & \vdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} \frac{1}{2}a_1 & -a_2 & a_3 & \cdots & a_{n-1} & -1 \\ 0 & -\frac{1}{2}a_3 & a_4 & \cdots & 1 & 0 \\ 0 & 0 & \frac{1}{2}a_5 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad (14)$$

i.e. with the boundary terms expressed as combination of symmetric and antisymmetric bilinear forms. Whatever the strategy of the actual integration is, we always end up with decomposition

$$\mathbf{A}^T = \mathbf{A}_1 + \mathbf{A}_2, \quad (15)$$

in which the matrices $\mathbf{A}_1$, $\mathbf{A}_2$ represent the partial contributions to the Dirichlet's remainder resulting from integration by parts of the operator on the left side of the scalar product *(.,.)* from order $n=2m$ to the half-order m and from the order m to the order 0, respectively.

5. The extended Geen's identity

Instead of the direct use of the Green's identity (4) we modify it by backward integration of L^* from the order $2m$ of derivatives of functions $v(s)$ to the half order m in (4), which leads to the extended integral identity

$$(Lu, v) - (L_1 u, L_2 v) - \sum_{s=0}^{1} (-1)^{s+1} \mathbf{u}^T(s)\mathbf{A}_{Ls}\mathbf{A}_{Rs}\mathbf{v}(s) + \sum_{s=0}^{1} (-1)^{s+1} \mathbf{u}^T(s)\mathbf{A}_2\mathbf{v}(s) = 0. \quad (16)$$

This extended integral identity still holds true for arbitrary pair of sufficiently smooth functions $u(x)$, $v(x)$. Not specified differential expressions L_1 and L_2 are the results of partial integration by parts of the domain operator L, for formaly selfadjoint problems $L_1 = L_2$. The integral identity (16) can be interpreted as transformation of the extended Green's identity (4) with respect to the data of the boundary value problem (1). Matrices $\mathbf{A}_{Ls}$ and $\mathbf{A}_{Rs}$ contain information on the homogeneous parts of given and adjoint boundary conditions, respectively.

With respect to (10) and (11) it is obvious that the last two terms in (16) are equal to

$$-\sum_{s=0}^{1}(-1)^{s+1}\mathbf{u}^T(s)\mathbf{A}_1\mathbf{v}(s)$$

which is the result of the conventional integration by parts "up to the half order". In developing the extended identity (16), the trivial identities have been used to "unfold" the matrix $\mathbf{A}_1$ so that the given data of the boundary value problem (1) could be encoded for arbitrary set of given homogeneous or nonhomogeneous boundary conditions. This unfolding is necessary, as the columns *m+1,...,2m* of the matrix $\mathbf{A}_1$ consist of only zero elements and can give no information on the geometrical boundary conditions.

6. Variational statements for boundary problems

The key results of this paper are based on observation that the formula (16) essentialy represents unconstrained variational statement corresponding to the given boundary value problem. The final variational statement can be derived from (16) by substitution of the given data.

To keep the final result, which is a scalar value, still in the form of a bilinear form in state vectors *u(s), v(s),* instead of the usual direct substitutions we insert properly chosen "substitution" matrices introduced in (12), which render equivalence with the classical substitutions. The substitution of the homogeneous part of given boundary conditions for functions *u(s)* is achieved with right multiplication of the matrix $\mathbf{A}_{Ls}$ by the matrix $\mathbf{I}_{us}$, while the substitution of the nonhomogeneous part of given boundary conditions is achieved by replacing the term $\mathbf{u}^T(s)\ \mathbf{A}_{Ls}$ with term $\mathbf{h}^e_s\ \mathbf{I}_{vs}$, where the vector $\mathbf{h}^e_s$ of length *2m* is the extension of the vector $\mathbf{h}_s$ of the nonhomeogeneous part of the given boundary conditions, obtained by adding zero elements to the dummy positions. In this way the eventual nonzero elements of the extended vector $\mathbf{h}^e_s$ take the same positions in $\mathbf{h}^e_s$ as are the column positions of the corresponding homogeneous parts of given boundary conditions in the extended matrix $\mathbf{A}_{Ls}$. Here we used the obvious identity

$$\mathbf{h}^e_s\mathbf{I} = \mathbf{h}^e_s\mathbf{I}_{vs}$$

to get rid of terms multiplied by zero components of the extended vector $\mathbf{h}^e_s$ With these substitutions Eq. (16) becomes after some rearrangements

$$\begin{aligned} W(u,v) = (f,v) - (L_1u, L_2v) - \sum_{s=0}^{1}(-1)^{s+1}\mathbf{h}^e_s\mathbf{I}_{vs}\mathbf{A}_{Rs}\mathbf{v}(s) + \\ + \sum_{s=0}^{1}(-1)^{s+1}\mathbf{u}^T(s)(-\mathbf{A}_{Ls}\mathbf{I}_{vs}\mathbf{A}_{Rs} + \mathbf{A}_2)\mathbf{v}(s) = 0. \end{aligned} \qquad (17)$$

It holds information on the domain part of operator in terms of *(f,v)-(L_1u,L_2v),* on the given nonhomogeneous boundary conditions in terms of the linear form

$$-\sum_{s=0}^{1}(-1)^{s+1}\mathbf{h}^e_s\mathbf{I}_{vs}\mathbf{A}_{Rs}\mathbf{v}(s)$$

and on the boundary part of the operator in terms of the bilinear form

$$\sum_{s=0}^{1}(-1)^{s+1}\mathbf{u}^T(s)(-\mathbf{A}_{Ls}\mathbf{I}_{vs}\mathbf{A}_{Rs} + \mathbf{A}_2)\mathbf{v}(s)$$

The functionality of the above approach can be assured by further integration of the term (L_1u, L_2v), which leads to recovering of the original boundary value problem from bilinear identity (17). Backward integration in (17) gives

$$(f,v)-(Lu,v)-\sum_{s=0}^{1}(-1)^{s+1}\mathbf{h}_s^e\mathbf{I}_{vs}\mathbf{A}_{Rs}\mathbf{v}(s)+ \\ +\sum_{s=0}^{1}(-1)^{s+1}\mathbf{u}^T(s)(-\mathbf{A}_{Ls}\mathbf{I}_{vs}\mathbf{A}_{Rs}+\mathbf{A}_2+\mathbf{A}_1)\mathbf{v}(s)=0. \tag{18}$$

As $\mathbf{A}_1+\mathbf{A}_2=\mathbf{A}^T=\mathbf{A}_{Ls}\,\mathbf{I}\,\mathbf{A}_{Rs}$ and $\mathbf{I}=\mathbf{I}_{us}+\mathbf{I}_{vs}$, we end up with statement

$$(f-Lu,v)+\sum_{s=0}^{1}(-1)^{s+1}(\mathbf{u}^T(s)\mathbf{A}_{Ls}\mathbf{I}_{vs}-\mathbf{h}_s^e\mathbf{I}_{vs})\mathbf{A}_{Rs}\mathbf{v}(s)=0. \tag{19}$$

As (18) holds for arbitrary function *v(s)* and $\mathbf{A}_{Rs}$ is regular, we conclude that

$$Lu=f \qquad \text{and} \qquad \mathbf{u}^T\mathbf{A}_{Ls}\mathbf{I}_{vs}=\mathbf{h}_s^e\mathbf{I}_{vs}.$$

Thus, the first relation recovers the domain equation (1) and the second recovers the given boundary conditions .

The left side of (17) can be identified with the "unconstrained" form of the variational statement. The adjective unconstrained expresses the fact, that both the testing functions *v(s)* and the functions *u(s)* are not subject to any conditions at the boundary points of the definition domain. The selection of Hilbert spaces U_k and V_k is equivalent to the appropriate restriction of *W(u,v)*. In most cases the restriction to the spaces of functions conforming the homogeneous geometrical boundary conditions of given and adjoint problem is the best solution. Moreover, in the majority of practical problems the boundary problems are symmetric (selfadjoint), so even more simplicity is gained from the identity of the spaces U_k and V_k of admissible functions.

The formula (17) has been developed as an algorithm implemented in the enviroment of the symbolic programming language *Mathematica* and has been widely tested on many examples of boundary value problems.

7. Conclusions

The Dirichlet's remainder term appearing in the Green's formula associated with given two-point boundary value problem is written in a matrix form. Two suitable trivial decompositions of this matrix are introduced, which enable to deduce formal algorithms to compute symbolically both the data of adjoint boundary value problem and variational functionals (bilinear forms for weak solution) corresponding to the given boundary value problem in both unconstrained and constrained version.

8. Acknowledgement

This work was conducted under the support of the Agency VEGA MŠ SR for project 1/2076/05.

9. References

Rektorys, K. (1985) *Time discretization and partial differential equations.* SNTL, Prague.

Nánási, T. (1994) Matrix method for computation of the adjoint boundary conditions. *Strojnícky Časopis,* Vol. 45, No. 5, pp. 391-408.