[9]Levi, M., "Geometric Phases in the Motion of Rigid Bodies," *Archives for Rational Mechanics and Analysis*, Vol. 122, No. 3, 1993, pp. 213–229.
[10]Junkins, J. L., and Shuster, M. D., "The Geometry of Euler Angles," *Journal of the Astronautical Sciences*, Vol. 41, No. 4, 1993, pp. 531–543.

Finite Element and Runge–Kutta Methods for Boundary-Value and Optimal Control Problems

Carlo L. Bottasso* and Andrea Ragazzi†
Politecnico di Milano, 20158 Milan, Italy

Introduction

WE consider a class of finite element in time (FET) and some implicit Runge–Kutta (RK) schemes for the solution of boundary-value problems. FETs have been recently advocated as a new way of solving this class of problems, with particular emphasis on optimal control problems (for example, see Ref. 1 and the references therein). The finite element method develops solutions discretizing appropriate weak forms of the equations that describe a given problem. On the other hand, the RK method directly discretizes the governing ordinary differential equations (ODEs). Although the two approaches look different, in this work we prove that the discrete equations arising from the class of FETs here considered are linear combinations of the equations generated by some implicit RK schemes and that the FET unknowns are linear combinations of the RK unknowns within each time step. Under these circumstances, this means that the two approaches are in reality the same method that yields identical numerical solutions and, hence, enjoys the same numerical properties. The analysis is valid for the p version of the method, that is, for arbitrarily high order. Similar results were derived in Ref. 2 for initial-value problems. That work is here extended to cover the case of boundary-value differential-algebraic problems.

RK methods are probably the best understood and most widely studied family of integration schemes, which makes them a mature and trusted technology with extensive applications to the class of problems here considered. FETs are less widely known, but their application to certain problems leads to strikingly simple solution procedures, for example, in the case of optimal control.[3] Our hope is that the proof of equivalence here offered might help close the gulf existing between practitioners using the two methods. We believe that additional insight can be usually obtained by a unified view, and we are, therefore, not suggesting to abandon one approach in favor of the other. Furthermore, some developments, for example a posteriori error estimation for adaptive mesh control, might be easier to accomplish in one framework rather than in the other.[4]

Boundary-Value Problems

We consider a generic boundary-value problem, defined by a system of first-order ODEs

$$\dot{\boldsymbol{y}} = \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{z}, t) \tag{1a}$$

If algebraic constraints of the form

$$\boldsymbol{a}(\boldsymbol{y}, \boldsymbol{z}, t) = 0 \tag{1b}$$

are also present, we have an index-one differential-algebraic equation (DAE) problem. In Eqs. (1a) and (1b), $(\dot{\cdot}) = \mathrm{d}\cdot/\mathrm{d}t$, $\boldsymbol{y} \in \mathbb{R}^Y$, $\boldsymbol{z} \in \mathbb{R}^Z$, $t \in \mathbb{R}$, $\boldsymbol{g}: \mathbb{R}^Y \times \mathbb{R}^Z \times \mathbb{R} \to \mathbb{R}^Y$, and $\boldsymbol{a}: \mathbb{R}^Y \times \mathbb{R}^Z \times \mathbb{R} \to \mathbb{R}^Z$. $I = [0, T]$ is the time domain. Suitable boundary and additional constraint conditions are given as

$$\boldsymbol{C}[\boldsymbol{y}(0), \boldsymbol{y}(T), T] = 0 \tag{1c}$$

This format covers a wide range of applications, in particular optimal control problems with or without unknown terminal time and path constraints. Minor modifications allow treatment of problems with interior point constraints, possibly at unknown intermediate times, and multiphase problems (see Ref. 4 and references therein for details).

We partition I according to $0 = t_1 < t_2 < \cdots < t_{N+1} = T$ and let $h_n := t_{n+1} - t_n$. We consider global methods, that is, methods that produce an approximate solution representation over the entire interval of interest. In the terminology of the finite element method, these procedures assemble the equations over the whole interval. Alternative strategies are based on the solution of corresponding initial-value problems, for example, shooting and related techniques.

RK Methods

The ε-embedding method[5] for an s-stage RK scheme applied to the solution of Eqs. (1a) and (1b) results in the relations

$$\hat{\boldsymbol{y}} = (\boldsymbol{1} \otimes \mathbb{I}_Y)\boldsymbol{y}^n + h_n(\boldsymbol{A} \otimes \mathbb{I}_Y)\hat{\boldsymbol{g}} \tag{2a}$$

$$0 = \hat{\boldsymbol{a}} \tag{2b}$$

$$\boldsymbol{y}^{n+1} = \boldsymbol{y}^n + h_n(\boldsymbol{b}^T \otimes \mathbb{I}_Y)\hat{\boldsymbol{g}} \tag{2c}$$

where $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^s$ and $\boldsymbol{A}$ is $s \times s$. The sY-dimensional vectors are defined as $\hat{\boldsymbol{y}} := (\boldsymbol{y}^1, \ldots, \boldsymbol{y}^s)$ and $\hat{\boldsymbol{g}} := [\boldsymbol{g}(\boldsymbol{y}^1, \boldsymbol{z}^1, t^1), \ldots, \boldsymbol{g}(\boldsymbol{y}^s, \boldsymbol{z}^s, t^s)]$ and the sZ-dimensional vectors as $\hat{\boldsymbol{z}} := (\boldsymbol{z}^1, \ldots, \boldsymbol{z}^s)$ and $\hat{\boldsymbol{a}} := [\boldsymbol{a}(\boldsymbol{y}^1, \boldsymbol{z}^1, t^1), \ldots, \boldsymbol{a}(\boldsymbol{y}^s, \boldsymbol{z}^s, t^s)]$, where $t^i := t_n + c_i h_n$. Finally, the symbol $\otimes$ denotes the tensor (Kronecker) product of matrices, whereas $\boldsymbol{1}$ is the s-dimensional vector $\boldsymbol{1} := (1, \ldots, 1)$ and $\mathbb{I}_Y$ is the unit $Y \times Y$ matrix.

Equations (1c) do not require discretization and, therefore, can be dropped from the subsequent discussion. The values $\boldsymbol{z}^1, \boldsymbol{z}^2, \ldots, \boldsymbol{z}^{N+1}$ can be computed a posteriori, using the equations $\boldsymbol{a}(\boldsymbol{y}^{n+1}, \boldsymbol{z}^{n+1}, t_{n+1}) = 0$, so that the solution $(\boldsymbol{y}^{n+1}, \boldsymbol{z}^{n+1})$ lies on the manifold.

FET Methods

FET methods are based on the approximation of a weak form of Eqs. (1a–1c):

$$\int_I [\boldsymbol{v} \cdot (\dot{\boldsymbol{y}} - \boldsymbol{g}) + \boldsymbol{w} \cdot \boldsymbol{a}]\,\mathrm{d}t + \boldsymbol{\theta} \cdot \boldsymbol{C} = 0 \tag{3}$$

In practice, one seeks a solution that satisfies in a weak sense Eqs. (1a) and (1b) and the boundary conditions (1c). See Refs. 2 and 6 for a unified approach to FETs and the references cited in 4 for additional background and information. Integration by parts of the first term of Eq. (3) yields

$$\int_I (\dot{\boldsymbol{v}} \cdot \boldsymbol{y} + \boldsymbol{v} \cdot \boldsymbol{g} - \boldsymbol{w} \cdot \boldsymbol{a})\,\mathrm{d}t = \boldsymbol{v} \cdot \boldsymbol{y}|_{\partial I} + \boldsymbol{\theta} \cdot \boldsymbol{C} \tag{4}$$

Under certain circumstances, weak forms (3) and (4) can be given a variational interpretation (for example, see Ref. 2). This is also the point of view of Ref. 3.

The approximation of Eq. (4) is developed as follows. We construct finite-dimensional trial function spaces Y^h and Z^h as $Y^h := \{\boldsymbol{y}^h \in [\mathcal{P}^{k_y}(I_n)]^Y\}$ and $Z^h := \{\boldsymbol{z}^h \in [\mathcal{P}^{k_z}(I_n)]^Z\}$, and test function spaces V^h and W^h as $V^h := \{\boldsymbol{v}^h \in [\mathcal{P}^{k_v}(I_n)]^Y\}$ and $W^h := \{\boldsymbol{w}^h \in [\mathcal{P}^{k_w}(I_n)]^Z\}$. $\mathcal{P}^k$ denotes the space of the kth-order polynomials. Within the nth time element, the finite element trial solutions are defined as follows:

$$\boldsymbol{y}^h(t) := \boldsymbol{y}^n$$

for $\tau = 0$,

$$\boldsymbol{y}^h(t) := \sum_{i=1}^{k_y+1} s_i^{k_y}(\tau)\boldsymbol{y}_i$$

Received 21 June 1999; revision received 30 March 2000; accepted for publication 2 April 2000.

*Ricercatore, Dipartimento di Ingegneria Aerospaziale, Via La Masa 34; Carlo.Bottasso@polimi.it.
†Graduate Student, Dipartimento di Ingegneria Aerospaziale, Via La Masa 34.

for $0 < \tau < 1$,

$$y^h(t) := y^{n+1}$$

for $\tau = 1$, and

$$z^h(t) := \sum_{i=1}^{k_z+1} s_i^{k_z}(\tau)\, z_i$$

for $0 < \tau < 1$. The weighting functions are defined as

$$v^h(t) := \sum_{i=1}^{k_v+1} s_i^{k_v}(\tau) v_i, \qquad w^h(t) := \sum_{i=1}^{k_w+1} s_i^{k_w}(\tau) w_i$$

for $0 < \tau < 1$ and $0 \le \tau \le 1$, respectively, with $\tau = (t - t_n)/(t_{n+1} - t_n)$. Here, y_i, z_i, v_i, and w_i are the vectors of nodal unknowns and test functions for node i. Clearly, y_i and z_i depend on n, but this dependence is not reflected in the notation. Also $s_i^{k_y}(\tau)$, $s_i^{k_z}(\tau)$, $s_i^{k_v}(\tau)$, and $s_i^{k_w}(\tau)$ are the finite element shape functions of order k_y, k_z, k_v, and k_w, respectively, for node i. Shape functions can be constructed in different ways. For our discussion, we only require from the shape functions the standard property

$$\sum_{i=1}^{k_v+1} s_i^{k_v} = 1$$

that implies

$$\sum_{i=1}^{k_v+1} s_i^{k_v}(\tau)' = 0$$

From the definitions, note that the trial solutions y^h are continuous within each time element, but are discontinuous across the interface of the elements, namely, at times $t_1, t_2, \ldots, t_{N+1}$, where we have the discrete boundary values $y^1, y^2, \ldots, y^{N+1}$. Furthermore, the trial solutions z^h are discontinuous across the element interfaces, but in this case there are no discrete boundary values at the boundary times.

In conclusion, we are looking for solutions $y^h \in Y^h$ and $z^h \in Z^h$ that satisfy

$$\int_{I_n} [\dot{v}^h \cdot y^h + v^h \cdot g(y^h, z^h, t) - w^h \cdot a(y^h, z^h, t)]\, dt = v^h \cdot y^h|_{\partial I_n} \qquad n = 1, \ldots, N \quad (5)$$

for all $v^h \in V^h$ and $w^h \in W^h$. We choose for simplicity to drop the boundary condition term $\theta \cdot C$ from the following discussion because it does not require discretization. The order of the test and trial finite element shape functions is then chosen according to the relation $k_y = k_z = k_w = k_v - 1$.

FET Methods as RK Processes

We shall now show that the FET formulation (5) can be written as (and, therefore, is equivalent to) an RK process of the type of Eqs. (2a–2c). We begin by selecting a quadrature rule for the evaluation of the integrals in Eq. (5). The rule is defined by q abscissas c_i and weights b_i. We consider the case $q = k_y + 1$; therefore, the number of quadrature points is the same as the number of finite element nodes. This yields

$$\sum_{i=1}^{k_v+1} v_i \cdot \left[\sum_{r=1}^{q} b_r \frac{d\left(s_i^{k_v}\right)}{d\tau}(c_r)\, y^h(t^r) + h_n \sum_{r=1}^{q} b_r s_i^{k_v}(c_r)\, g(y^h(t^r), z^h(t^r), t^r) \right] + h_n \sum_{i=1}^{k_w+1} w_i \cdot \sum_{r=1}^{q} b_r s_i^{k_w}(c_r)\, a[y^h(t^r), z^h(t^r), t^r] = v_{k_v+1} \cdot y^{n+1} - v_1 \cdot y^n \quad (6)$$

The usual finite element method practice consists in expressing the discrete equations in terms of the nodal values. We shall take a different approach and assume as unknowns the values of the finite element trial functions at the quadrature points. This can be done because we are using a quadrature rule that uses as many integration points as there are finite element nodes, as already mentioned. See Ref. 2 for a discussion on some interesting consequences of this assumption.

Define the $q \times q$ matrices $V := [s_i^{k_v}(c_j)]$, $V' := [s_i^{k_v}(c_j)']$, $W := [s_i^{k_w}(c_j)]$, and $B := [\diagdown b_{i\diagup}]$, with $i, j = 1, \ldots, q$. Considering that the w_i are arbitrary, from Eq. (6) we get $[(WB) \otimes \mathbb{I}_M]\hat{a} = 0$. Because WB is nonsingular, this implies $\hat{a} = 0$. Consider now that the v_i are arbitrary; the corresponding first $k_v (= q)$ of the $k_v + 1$ independent equation sets of Eq. (6) can be written concisely:

$$[(V'B) \otimes \mathbb{I}_N]\hat{y} = -(\mathbf{1}_1 \otimes \mathbb{I}_N)\, y^n - h_n[(VB) \otimes \mathbb{I}_N]\hat{g} \quad (7)$$

where $\mathbf{1}_1$ is the q-dimensional vector $\mathbf{1}_1 := (1, 0, \ldots, 0)$ that has the first entry equal to one and all the others equal to zero. Solving for $\hat{y}$, we have

$$\hat{y} = -[(B^{-1}V'^{-1}\mathbf{1}_1) \otimes \mathbb{I}_N]y^n - h_n[(B^{-1}V'^{-1}VB) \otimes \mathbb{I}_N]\hat{g} \quad (8)$$

It is easy to show that $-B^{-1}V'^{-1}\mathbf{1}_1 = \mathbf{1}$, which is analogous to the so-called B simplifying assumption in the theory of RK methods. Then, taking the aforementioned properties of the shape functions into account, we get from Eq. (6) that $y^{n+1} = y^n + h_n(b^T \otimes \mathbb{I}_N)\hat{g}$.

By comparison with Eqs. (2a) and (2b), we conclude that Eq. (6) is formally equivalent to a q-stage RK method with Butcher tableau

$$\begin{array}{c|c} c & -B^{-1}V'^{-1}VB \\ \hline & b^T \end{array} \quad (9)$$

This result, together with the fact that $\hat{a} = 0$ as shown, states that the FET equations, with the proper quadrature, are linear combinations of the RK equations. Furthermore, we recall that the internal unknowns are the values of the shape functions at the quadrature points, which are linear combinations of the finite element nodal values, whereas the boundary terms coincide in the two cases. This means that, under the stated conditions, the two methods will yield identical results and are, therefore, in this sense equivalent.

As said, this formulation assumes discontinuous trial solutions z^h at the element interfaces and does not directly define the values $z^1, z^2, \ldots, z^{N+1}$. However, these values can be computed a posteriori so that the solution lies on the manifold, using the equations $a(y^{n+1}, z^{n+1}, t_{n+1}) = 0$. This is the same procedure used for correcting the ε-embedding method.

It remains to be seen whether the numerical values assumed by Eq. (9) for different choices of the quadrature rule correspond to existing RK schemes or not. To this purpose, we tested the Gauss, Lobatto, and Radau–Left rules. In the first case we obtained the Kuntzmann–Butcher method, in the second the Lobatto IIIB method, and in the third the Radau IA method.

Using standard results from RK theory, we can now conclude that the Kuntzmann–Butcher (or Gauss) RK methods are algebraically stable and also symplectic when applied to ODEs that describe the evolution of general (unconstrained) Hamiltonian systems. From our results, the considered FET method with Gauss quadrature is also algebraically stable and symplectic (and, hence, in general not Hamiltonian preserving), a result that has not been previously reported in the FET literature. Again using results from RK theory, we can also conclude that the Lobatto FETs are A stable, whereas the Radau–Left FETs are L and algebraically stable.

Conclusions

We have shown the existing relation between some implicit RK formulas and a class of FET methods for boundary-value problems governed by ODEs or index-one DAEs. Note that, although the two methods start from rather different points of view, they end up developing the same discretization of the equations governing the problem. This is not a new fact in the history of numerical methods: For example, it is well known that some finite difference, finite volume, and finite element methods are completely equivalent for certain classes of problems. We believe that additional insight can usually be gained by looking at a discretization process from different points of view, and this is exactly what happens in this case.

References

[1]Warner, M. S., and Hodges, D. H., "Solving Optimal Control Problems Using hp-Version Finite Elements in Time," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 1, 2000, pp. 86–94.

[2]Bottasso, C. L., "A New Look at Finite Elements in Time: A Variational Interpretation of Runge–Kutta Methods," *Applied Numerical Mathematics*, Vol. 25, 1997, pp. 355–368.

[3]Hodges, D. H., and Bless, R. R., "Weak Hamiltonian Finite Element Method for Optimal Control Problems," *Journal of Guidance, Control, and Dynamics*, Vol. 14, No. 1, 1991, pp. 148–156.

[4]Estep, D., Hodges, D. H., and Warner, M. S., "Computational Error Estimation and Adaptive Error Control for a Finite Element Solution of Launch Vehicle Trajectory Problems," *SIAM Journal on Numerical Analysis*, (to be published).

[5]Hairer, E., and Wanner, G., *Solving Ordinary Differential Equations II. Stiff and Differential Algebraic Problems*, Springer-Verlag, Berlin, 1996, p. 374.

[6]Borri, M., and Bottasso, C. L., "A General Framework for Interpreting Time Finite Element Formulations," *Computational Mechanics*, Vol. 13, No. 3, 1993, pp. 133–142.

Constrained Motions of Systems with Elastic Members

Shlomo Djerassi*
RAFAEL, Ministry of Defense, 31021 Haifa, Israel

Introduction

ANALYSES of systems with elastic members usually start with the generation of motion equations.[1–3] Authors use a variety of procedures to generate these equations. Common to procedures based on Lagrange's formulation are the time-consuming steps such as generation of kinetic energy expressions for generic particles of the elastic members and integration of these expressions over the entire volume of the bodies.

It is the contention of this Note that motion equations of systems with elastic members can be obtained more expeditiously if use is made of motion equations for constrained systems. This contention is discussed after review of Kane's equations for constrained systems, chosen here to be the working tool.

Accordingly, let S be a simple nonholonomic system of ν particles $P_i (i = 1, \ldots, \nu)$ of mass m_i possessing $\bar{n}$ generalized coordinates $q_1, \ldots, q_{\bar{n}}$ and n (where $n \leq \bar{n}$) generalized speeds $u_1, \ldots, u_n$ in N, a Newtonian reference frame. Suppose that the motion of S in N is defined as unconstrained and is governed by n dynamical equations, namely,

$$F_r + F_r^* = 0 \qquad (r = 1, \ldots, n) \tag{1}$$

where F_r and F_r^* are the rth generalized active force and the rth generalized inertia force for S, respectively. Moreover, suppose that m constraints of the form

$$u_k = \sum_{r=1}^{p} C_{kr} u_r + D_k \qquad (k = p+1, \ldots, n) \tag{2}$$

are imposed on the motion of S, where

$$p \triangleq n - m \tag{3}$$

and C_{kr} and D_k are functions of $q_1, \ldots, q_{\bar{n}}$ and time t. Then the motion of S in N is defined as constrained and is governed by p dynamical equations, namely,

$$F_r + F_r^* + \sum_{k=p+1}^{n} C_{kr}\left(F_k + F_k^*\right) = 0 \qquad (r = 1, \ldots, p) \tag{4}$$

These equations were first presented by Wampler et al.[4] in a matrix form.

Suppose that the members comprising the system in question are temporarily regarded as undergoing an unconstrained motion and that motion equations represented by Eqs. (1) are written for each member. Moreover, suppose that the kinematical constraint equations are formulated and cast in the form of Eqs. (2). Then substitutions in Eqs. (4) lead to the requisite motion equations. This procedure underlies works by Buffinton and Kane[5] and by Djerassi and Kane.[6] Using the elastic properties of the unconstrained members in the generation of Eqs. (1), these authors assumed that Eqs. (4)—where expressions from Eqs. (1) are used—automatically reproduce the elastic properties of the constrained system. A similar assumption was made by Thomson,[7] discussing the elastic properties of an elastic beam on three supports. The validity of this assumption cannot be proved rigorously; however, it can be illustrated, a task undertaken here in connection with two examples. These examples demonstrate the benefits associated with the exploitation of this assumption.

Cantilever Beam with a Massive Endpoint

Consider the system S comprising an elastic cantilever beam B and a particle P of mass m attached to B at its endpoint E. Suppose that B and P are temporarily regarded as undergoing an unconstrained motion. Then equations of motion of B and P can be generated independently. Accordingly, B is assumed to behave as a uniform Euler–Bernoulli beam, whose motion is governed by the equations

$$-M\dot{u}_i - EJL\lambda_i^4 q_i = 0 \qquad (i = 1, \ldots, \mu) \tag{5}$$

$$\dot{q}_i = u_i \qquad (i = 1, \ldots, \mu) \tag{6}$$

where EJ, L, and M are, respectively, the bending rigidity, the length, and the mass of the beam; and q_i, u_i, λ_i, and μ are, respectively, the ith generalized coordinate, the rth generalized speed, the ith eigenvalue of the equation

$$1 + \cos\lambda_i L \cdot \cosh\lambda_i L = 0 \tag{7}$$

and the number of modes used to describe the elastic deflection $y(x, t)$ of points of the beam. The latter can be expressed as

$$y(x,t) = \sum_{i=1}^{\mu} \phi_i(x) q_i(t) \tag{8}$$

where $\phi_i(x)$, the ith modal function, is given by

$$\phi_i(x) = -\cos\lambda_i x + \cosh\lambda_i x + k_i(\sin\lambda_i x - \sinh\lambda_i x)$$
$$k_i = \frac{\cos\lambda_i L + \cosh\lambda_i L}{\sin\lambda_i L + \sinh\lambda_i L} \tag{9}$$

if the boundary conditions are

$$y(0,t) = y'(0,t) = y''(L,t) = y'''(L,t) = 0 \tag{10}$$

In connection with the motion of P, $u_{\mu+1}$ is defined so that the velocity of P in N is expressed as ${}^N\boldsymbol{v}^P = u_{\mu+1}\boldsymbol{n}$. Then the dynamical equations governing the motion of S in N are Eqs. (5), and, in addition

$$-m\dot{u}_{\mu+1} = 0 \tag{11}$$

Consider a constrained motion of S with P attached to E. Then ${}^N\boldsymbol{v}^P = {}^N\boldsymbol{v}^E$, and because ${}^N\boldsymbol{v}^E = \dot{y}(L,t)\boldsymbol{n}$, it follows that $u_{\mu+1}(t) = \dot{y}(L,t)$, or

$$u_{\mu+1}(t) = \phi_1(L)u_1(t) + \cdots + \phi_\mu(L)u_\mu(t) \tag{12}$$

Received 4 March 1998; revision received 23 November 1999; accepted for publication 3 February 2000.

*Research Fellow, P.O. Box 2250.