ACCURATE ELEMENT METHOD STRATEGY FOR THE INTEGRATION OF FIRST ORDER ORDINARY DIFFERENTIAL EQUATIONS

M. BLUMENFELD[1]

Metoda intitulată "The Accurate Element Method" dezvoltată recent de către autor poate rezolva in acelaşi mod toate problemele legate de integrarea ecuaţiilor diferenţiale ordinare: Problema cu valori iniţiale (IVP), cea bilocală (BVP) si cea a valorilor proprii (EVP) [1,2]. Pentru problema cu valori initiale (IVP) mai multi specialişti în domeniul analizei numerice au considerat necesar un studiu mai dezvoltat privind două aspecte esenţiale: stabilitatea metodei şi capacitatea sa de a înlocui soluţia exactă a unei ecuaţii diferenţiale ordinare prin polinoame. Articolul de faţa este rezultatul acestui studiu.

The Accurate Element Method (AEM) developed recently by the author can solve in the same way all the problems connected to the Ordinary Differential Equations (ODE) namely the Initial Value Problem (IVP), the Boundary Value Problem (BVP) and the Eigenvalue Problem (EVP) [1,2]. Connected to the IVP problem several academics considered necessary a more elaborate study concerning two essential problems: the stability of the method and its capacities to replace the exact solution of an ODE by several polynomials. The paper is the result of this study.

Keywords: Ordinary Differential Equation, Accurate Element Method, Implicit Methods, Target Value Problem, Field Polynomial Solution

1. Introduction: Problems to be solved

Consider a first order Ordinary Differential Equation (ODE)

$$E_1(x)\frac{d\phi}{dx}+E_0(x)\,\phi+E_F(x)=E_1(x)\,\phi^{(1)}+E_0(x)\,\phi^{(0)}+E_F(x)=0 \quad (1.1)$$

where $E_1(x)$, $E_0(x)$, $E_F(x)$ are three known functions. This ODE has to be integrated between a **start** (*initial*) point x_S and a **target** (*final*) point x_T, leading to two different problems that will be analyzed below:

1. Supposing the initial value $\phi_S=\phi_S^{(0)}=\phi(x=x_S)$ as known one has to calculate the value of the function $\phi_T=\phi_T^{(0)}=\phi(x=x_T)$ at the target point, no

[1] Professor Emeritus, Strength of Materials Department, University "Politehnica" of Bucharest, ROMANIA

matter how far this point is. This will be considered as a **Target Value Problem (**TVP**)**, being in fact the trivial approach of an Initial Value Problem (IVP).

2. Supposing the integration field $x_S - x_T$ divided in a small number *NE* of sub-intervals (elements), find *NE* polynomials $\phi_n(x)$ (*n=1,2…NE*) that can be considered as *accurate solutions of the ODE on each element*. This problem that can have an important use for a wide class of real time simulation programs like SIMULINK and AMESim[2], will be referred as a **Field Polynomial Solution** (FPS).

It will result that the strategy for solving these two problems leads to different approaches that can be nevertheless controlled and optimized. The Accurate Element Method (AEM), which can usually give good answers to both *TVP* and *FPS* will be compared with the "classic" fourth–order Runge-Kutta method, which is not always able to solve a *TVP* and whose results are only discrete (non-continuous) values.

2. ODEs to be integrated

Three linear ODEs will be integrated below. For all of them the functions $E_1(x)$, $E_0(x)$, $E_F(x)$ are represented by polynomials.

2.1 **ODE1**

The ODE1 is a "build-up" problem for which a solution has been chosen as a nineteen-degree polynomial given by the product

$$\phi(x) = (FA)\,(FB) \qquad (2.1)$$

where FA = (x-1) (x-1.6) (x-1.7) (x-1.8) (x-2.02)(x-3.05) (x-3.2) (x-3.6) (x-4.3)

FB = (x-4.55) (x-4.86) (x-5.2) (x-5.64) (x-6.02)(x-7.03)(x-8)(x-9) (x-9.78) (x-10)

If $E_1 = 1 + 2x - x^2 + 3x^3 + x^4$ and $E_0 = 5 - 2x + 3x^2 + 4x^3 + x^4$, the free term $E_F(x)$ results as a *twenty-three degree polynomial*. The starting value ϕ_S of the function results from (2.1) by replacing $x \Rightarrow x_S$.

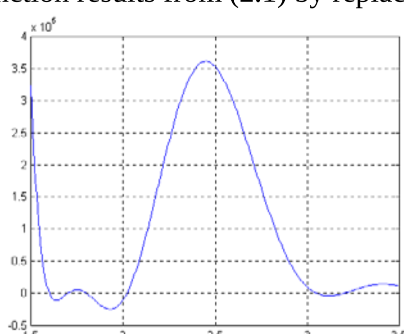

Fig.2.1 From *x*=1.5 to *x*=3.5

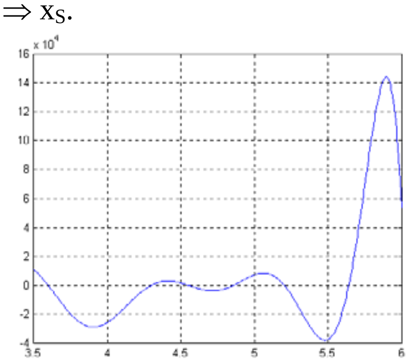

Fig.2.2 From *x*=3.5 to *x*=6

[2] Advanced Modeling and Simulation language for Engineers, Imagines, Roanne, 2006.

The function $\phi(x)$ that has 18 extremes is represented between $x=1.5 - 3.5$ in Fig 2.1 and between $x=3.5 - 6$ in Fig 2.2. Because the solution (2.1) is known, it possible in this case to compare the *TVP* result ($\phi_{computed}$) to the accurate one ($\phi_{accurate}$), by calculating an *actual relative error* given by

$$actual\ error = \frac{\phi_{computed} - \phi_{accurate}}{\phi_{accurate}} \tag{2.2}$$

2.2 ODE2

The second ODE is taken at random being given by

$$(1.1 - 0.1x)\,\phi^{(1)} + (5 - 2x + 3x^2 + 4x^3 + x^4)\,\phi^{(0)} + 2 + 3\,x + 2x^2 + x^3 = 0 \tag{2.3}$$

The integration will start from $x_S=0$, for which ϕ_S is chosen as

$$\phi_S = \phi(x = x_S) = \phi(0) = 1 \tag{2.4}$$

No closed solution is known by the author for ODE2 (2.3). The graphic of the solution obtained by using the strategy described later on is given in Fig.2.3.

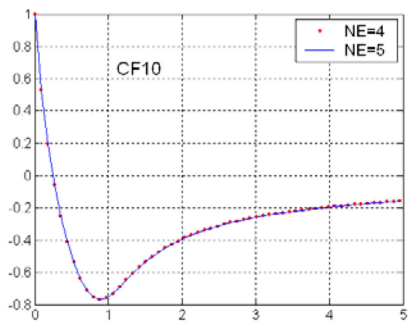

Fig.2.3. ODE2 from x=0 to x=5

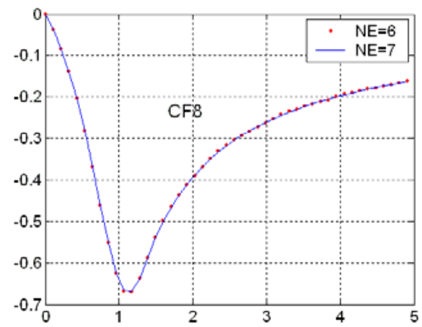

Fig.2.4.ODE3 from x=*0* to x=5

2.3 ODE3

The ODE3 is similar to ODE2 (2.4), but with a *modified first term* $E_1(x)$

$$(6 - 5x + x^2)\,\phi^{(1)} + (5 - 2x + 3x^2 + 4x^3 + x^4)\,\phi^{(0)} + 2 + 3x + 2x^2 + x^3 = 0 \tag{2.5}$$

The integration starts from $x_S=0$, for which ϕ_S is chosen as

$$\phi_S = \phi(x = x_S) = \phi(0) = 0 \tag{2.6}$$

No closed solution is known by the author for ODE2 (2.5). The graphic of the solution obtained by using the strategy described later on is given in Fig.2.4.

3. The complete transfer relation and concordant functions

Since 1768 the one-step "classical" methods used for the integration of *ODEs* (like Euler, Heun, Runge-Kutta) *carefully avoid integration*. On the contrary, the *Accurate Element Method* (AEM) starts by an **accurate integration of the ODE**, being based on two concepts:

1. The **COMPLETE TRANSFER RELATION** (CTR) , represents the result of an accurate integration that leads to one or more *integral equations*. For instance, the first order ODE (2.1) can be integrated between x_S and x_T leading to

$$\int_{xS}^{xT} E_1(x)\frac{d\phi}{dx}dx + \int_{xS}^{xT} E_0(x)\phi(x)\ dx + \int_{xS}^{xT} E_F(x)dx = 0 \qquad (3.1)$$

If the coefficients E_1=constant and E_0=constant, the first integral is performed straightforward[3]

$$E_1\phi_T = E_1\phi_S - E_0\int_{xS}^{xT}\phi(x)\ dx - \int_{xS}^{xT} E_F(x)dx \qquad (3.2)$$

The only difficulty (typical for an integral equation) is the first integral that includes the *unknown* function $\phi(x)$ under the integral sign. This integral can be performed by replacing $\phi(x)$ by an approximation function $\tilde{\phi}(x)$. Seemingly, for a first order ODE the only possible approximation is the linear interpolation function

$$\tilde{\phi}(x) = K_1 + K_2\ x \qquad (3.3)$$

where $K_1 = (\phi_S x_T - \phi_T x_S)/(x_T - x_S)\ ;\ K_2 = (\phi_T - \phi_S)/(x_T - x_S)$ (3.4)

If such function is replaced in (3.1) it will lead to poor and unacceptable result, similar to those obtained by using Euler's method.

2. The **CONCORDANT FUNCTION** (CF) is a higher order polynomial that can be used as a replacing function $\tilde{\phi}(x)$ instead of (3.3), **without modifying the number of the end unknowns**, which remain ϕ_S and ϕ_T. Suppose, for instance, that the ODE2 (2.3) has to be integrated between $x_S=1$ and $x_T=2$, for which it will be used a Concordant Function represented by a *third-degree* polynomial referred as CF4

$$\tilde{\phi}(x) = C_1 + C_2\,x + C_3\,x^2 + C_4\,x^3 \qquad (3.5)$$

whose derivative is $d\tilde{\phi}(x)/dx = C_2 + 2\ C_3\,x + 3C_4\,x^2$ (3.6)

In order to obtain the four unknown constants C_i (i = *1,2,3,4*), two end conditions are obvious:

$x=x_S=1 \Rightarrow \phi_S = C_1 + C_2 + C_3 + C_4\ (a)$; $x = x_T = 2 \Rightarrow \phi_T = C_1 + 2C_2 + 4C_3 + 8C_4\ (b)$

[3] For the general case when $E_1=E_1(x)$ and $E_0=E_0(x)$ the relation (3.2) becomes [1]

$$E_{1T}\,\phi_T = E_{1S}\,\phi_S - \int_{xS}^{xT}\left(E_0(x) - \frac{dE_1(x)}{dx}\right)\phi(x)\ dx - \int_{xS}^{xT} E_F(x)dx \qquad (3.2a)$$

where $E_{1S}=E_1(x=x_S)$ and $E_{1T}=E_1(x=x_T)$ are two constants resulting from $E_1(x)$.

The two other conditions are obtained by *AEM* from the derivative (3.6), which is also transferred at both ends by replacing in (3.6) the two end abscissas

$x{=}x_S \Rightarrow \left(d\tilde{\phi}(x)/dx\right)_S = C_2 + 2C_3 + 3C_4\,(c)$; $x{=}x_T \Rightarrow \left(d\tilde{\phi}(x)/dx\right)_T = C_2 + 4C_3 + 12C_4 \quad (d)$

The four conditions (*a*), (*b*), (*c*), (*d*), represent a system of 4 equations from which it result the 4 constants C_i. They depend at this stage not only on the initial unknowns ϕ_S and ϕ_T but *also on the two unknown end derivatives.* The Accurate Element Method eliminates **accurately** these new unknowns **by using the governing equation itself**. The equation ODE2 (2.3) considered here applied at both ends leads to

$$x{=}x_S{=}1 \Rightarrow \left(d\tilde{\phi}(x)/dx\right)_S + 11\phi_S + 8 = 0 \quad \Rightarrow \quad \left(d\tilde{\phi}(x)/dx\right)_S = -11\,\phi_S - 8 \quad (3.7a)$$

$$x{=}x_T \Rightarrow \quad 0.9\left(d\tilde{\phi}(x)/dx\right)_T + 61\phi_T + 24 = 0 \Rightarrow \left(d\tilde{\phi}(x)/dx\right)_T = -(61/0.9)\,\phi_T - (24/0.9) \quad (3.7b)$$

If (3.7a) and (3.7b) are replaced in (*c)* and (*d)*, respectively, all the four constants C_i **will depend only on the initial unknowns ϕ_S and ϕ_T** being given by

$C_1 = -59\,\phi_S + (1184/9)\,\phi_T + (40/3)$; $C_2 = 144\,\phi_S - (2978/9)\,\phi_T - (112/3)$ (3.8)

$C_3 = -108\,\phi_S + (2395/9)\,\phi_T + (104/3)$; $C_4 = 24\,\phi_S - (601/9)\,\phi_T - (32/3)$

Now (3.5) is replaced in (3.1), leading finally to *a single equation with two unknowns* (ϕ_S, ϕ_T). Because the initial (start) value ϕ_S is known the target value ϕ_T results immediately.

This methodology allows the use of higher degree polynomials by applying a similar methodology [1]. For instance for CF6 (a five-degree polynomial with 6 constants) it is necessary to use besides the first derivative of the ODE, also its second derivative. This procedure (increasing by two units the degree of the polynomial and adding simultaneously a higher order derivative, which applied at both ends give two new accurate conditions) may continue similarly for polynomials of any degree. Below will be analyzed the influence on the precision of *7 different Concordant Functions* (CF4, CF6, CF8, CF10, CF12, CF14, CF16). The smallest degree of the replacing polynomial corresponds to CF4 (4 constants, third degree) and the highest to CF16 (16 C_i, fifteen-degree).

Remark. The derivatives mentioned above concern only the functions $E_1(x)$, $E_0(x)$, $E_F(x)$, *independently of each other*. This task is processed by a program without any intervention of the user [1].

4. The Accurate element method is an implicit method thus generally stable

The *stability* of the procedure for solving numerically an ODE is a term not always clearly specified. Usually a procedure is considered as *unstable* if errors introduced at some stage of the calculation are propagated without bound throughout subsequent calculations. A large number of studies have been dedicated to the methodologies establishing the conditions that secure the stability of the Runge-Kutta method, which is an *explicit method*. Usually they lead to the limitation of the length of each step, possibly imposing a great number of steps.

The Accurate Element Method is an *implicit method*. In fact the integral equation (3.2) shows that the target value ϕ_T depends on the first integral *that includes $\phi(x)$*. The function $\phi(x)$ is replaced by (3.5), where the constants C_i that depend on ϕ_S and ϕ_T are given by (3.8). After performing the integration *the result will finally include only three terms* : $\mathbf{K_1\, \phi_S + K_2\, \phi_T + K_3}$, where K_1, K_2, K_3, are three constants depending on each specific problem[4]. Consequently (3.2) becomes

$$E_{1T}\,\phi_T = E_{1S}\,\phi_S - (K_1\phi_S + K_2\phi_T + K_3) - \int_{xS}^{xT} E_F(x)\,dx \tag{4.1}$$

It results that the target value ϕ_T **is included on both left and right sides of (4.1)**. The target value results by solving (4.1)

$$\phi_T = \frac{\left[E_{1S} - K_1\right]\phi_S + \left[K_3 - \int_{xS}^{xT} E_F(x)\,dx\right]}{E_{1T} + K_2} \tag{4.2}$$

This represents a specific form of an *implicit* solution that is **generally or unconditionally stable**[5]. For instance an integral *using a single element* starting from $x_S=0$ and having as target $\mathbf{x_T=10\ 000}$ was perfectly stable **(**see[1], page 118). It is important to underline that for a linear ODE the Accurate Element Method **obtains ϕ_T directly, without any iterative approach or any procedure for solving a system of equations** usually specific for the implicit methods.

5. Preliminary step for solving an ODE: finding the roots of $E_1(x)$

The procedure for obtaining a CF is conditioned by the computation of the end derivatives as it results from (3.7a) and (3.7b) [see [1], page178]. This remark is also valid for Euler or Runge-Kutta methods. Consequently, it is necessary to know before starting to solve an ODE the roots of $E_1(x)$ in order to avoid them during the computation. For the three ODEs analyzed here it results:
ODE1: two real roots, $x_1 = -3.43$, $x_2 = -3.67$, **outside** the integration interval;
ODE2: one real root, $x_1 = 11$, **outside** the integration interval;
ODE3: two real roots, x1= 2, $x_2 = 3$, both of them **inside** the integration interval.

---

[4] Here:

$$\int_{xS}^{xT} \phi(x)\;dx = \int_1^2 \left(C_1 + C_2\,x + C_3\,x^2 + C_4\,x^3\right)dx = -5\phi_S + (619/108)\phi_T - 16/9$$

[5] S.C.Chapra, R.P.Canale, Numerical Methods for Engineers, McGraw-Hill, 2002

It results that this aspect has to be taken into consideration **only for ODE3.**

6. Parameters that can be chosen: Concordant function and/or number of elements(NE)

In order to improve the results the user that tries to integrate an ODE can modify the CF and/or the number of sub-intervals (elements) on which the integration is performed. This double possible choice is benefactor, but raises a dilemma: *which parameter has to be modified and how?* The short time since AEM has been developed [2] and the relatively small number of examples analyzed by the author are not enough to give a definite answer. Nevertheless below is outlined a strategy that can be a guide for the users (or for possible researchers) in order to find a good answer to the above question. Before presenting the strategy, it deserves to underline some aspects, which will simplify the search:

1. Though the number of CFs is here limited to 7, there is a huge difference between the behavior of a third degree (CF4) and a fifteen-degree (CF16) polynomial. The lower value CF4 leads to results that are better but not far from those obtained by using the fourth-order Runge-Kutta [1]**.** The results improve usually when a higher CF is used, but not always steadily so that a good estimation of the precision **during the computation** is compulsory.

2. The estimation of the precision is based on *the comparison between two successive computations*: the new and the previous. This approach is not time-consuming *due to the small number of elements involved*. As it was shown in §4 *AEM is unconditionally stable* so that the computation starts usually **by using a single element (NE=1),** no matter the length of integration interval. The number of elements that have to be used in order to obtain good results (if possible accurate) is quite small: *usually under 10 elements, seldom more than 30*.

7. The target value problem (TVP)

Our experience has shown that for the *Target Value Problem* the results improve (usually but not always) when the order of the CF increases. In order to appreciate the efficiency of each CF, below are presented the values of ϕ_T obtained by using *all the seven-CFs mentioned above*. Obviously, such an approach represents only an "academic" approach, because for practical problems only some of the *CF*s are considered. The user has to be sustained – in order to choose an appropriate CF and the smallest number of elements – by some quantitative criteria that can be moreover included in a the program: *using a small number of criteria is good; using a single criterion is for the best*.

7.1 Criteria for choosing an efficient concordant function and a small number of elements (NE)

As it will result below from all the numerical experiments the results of TVP usually improve when the degree of CF increases. Consequently, based on our experience it is advisable to start the computation by using a higher order CF (for instance CF16) and *NE=1*.

Obviously no criterion is available in order to appreciate how good is the value of the ϕ_T obtained with *NE=1*. Consequently, it is necessary to proceed to a second computation by increasing NE with an increment that can be *1 element* or *many elements*. The computation can be continued in a first attempt by using *NE=2, 3, 4...* elements. Starting from *N=2* one can have a first appreciation of the precision of the whole procedure by calculating a *relative error* similar to (2.2). This time the comparison is made between the new computed value (ϕ_{new}) and the value already known from the previous computation ($\phi_{previous}$). This criterion will be referred as **the estimated error** [1,2]

$$estimated\ error = \frac{\phi_{new} - \phi_{previous}}{\phi_{new}} \tag{7.1}$$

It is useful to observe that the value resulted from (7.1) gives a good estimation of the number of digits that can be considered as accurate [1]. Usually an estimated error having as exponent *e-6* or *e-7* can be considered as satisfactory because 6 or 7 digits of the result are expected to be accurate. For a given *Imposed Estimated Error* (IEE) the program is able to stop automatically the computation at a desired level of precision (for a first attempt one can use *IEE = 1e-6* (10^{-6}) or *IEE = 1e-7* (10^{-7}), but other values can also lead to good results). This procedure is based on the presumption that the computed value of ϕ_T converges towards an accurate value, which means that the *estimated error decreases when the number of elements NE increases.* Or, as it resulted from some numerical experiments, this tendency is true only up to a certain number of elements, then the values of the *estimated error* become to oscillate [1]. In such case the computation with a chosen CF can be stopped by the program, the result retained being for instance the last value of ϕ_T before the starting of the oscillations.

Being used for a single CF the criterion (7.1) can be considered as being an *internal evaluation of the precision*. It is possible to use an *external criterion*, by choosing another CF and performing the same procedure as that described above. In this case one can on one side *compare the results obtained with the two CFs* (consequently to verify the first computation) and on another side try to *overcome the oscillation of the results*.

7.2 The ODE1 solved as a target value problem

7.2.1 Accurate element method

For ODE1 two analysis will be performed starting from x_S=1 and having as targets x_T=6 (Table 1) and x_T=11 (Table 2), respectively. The results include: the *estimated error* (7.1), the *actual error* (2.2) for *NE = 1,3,5* and the target value ϕ_T obtained for *NE=5* given in the last column. It is useful to observe:

1. The *actual error* (2.2) is given only for ODE1 because the solution (2.1) is known.

2. The *estimated error* decreases when the degree of CF increases.

3. For CF4, CF6, CF8 the results obtained for a small number o elements NE are quite far from the accurate values.

Table 1

ODE1 (x_S=1 ; x_T=6); $\phi_{T,exact}$= 5.833807525634766 e+4

NE⇒	NE=1	NE=3		NE=5		
CF⇓	Act.err.	Est.err.	Act.err.	Est.err.	Act.err.	ϕ_T(x=6)
CF4	-488	27.52	-8.89e-1	2.87e-1	-7.18e-1	1.8462353339e+4
CF6	3426	-32	-8.62e-1	2.68e-1	-3.12e-1	3.3250465205e+4
CF8	-10836	8.90e-1	2.53	-3.98e-1	1.38e-1	6.5017232134e+4
CF10	19148	3.53	-6.61e-1	3.58e-2	-5.48e-3	**5.8**121358160e+4
CF12	-19497	7.69e-1	1.89e-2	-2.13e-3	2.75e-5	**5.833**9409129e+4
CF14	10973	1.00e-2	-6.54e-3	-6.23e-4	-1.30e-4	**5.833**1752647e+4
CF16	-2914	-2.78e-3	-2.75e-3	-6.72e-4	-1.26e-4	**5.833**1974901e+4

Table 2

ODE1 **(x_S=1 ; x_T=11); $\phi_{T,exact}$= 3.9563530203 e+13**

NE⇒	NE=1	NE=3		NE=5		
CF⇓	Act.err.	Est.err.	Act.err.	Est.err.	Act.err.	ϕ_T(x=11)
CF4	1.44	-2.16e-1	6.58e-1	-9.85e-2	2.36e-1	5.893854401e+13
CF6	-2.91	1.11	-5.14e-1	1.04e-1	-8.84e-2	3.645999924e+13
CF8	5.16	-6.10e-1	1.55e-1	-2.69e-2	1.07e-2	**3.9**98864283e+13
CF10	-5.15	2.75e-1	-2.81e-2	3.02e-3	-7.22e-4	**3.95**3891953e+13
CF12	2.79	-6.92e-2	2.31e-3	-1.35e-4	1.51e-5	**3.956**412986e+13
CF14	-1.20	6.94e-3	-8.90e-5	2.35e-6	-1.43e-7	**3.95635**2454e+13
CF16	2.93e-1	-2.20e-4	9.40e-7	-9.62e-9	5.57e-9	**3.9563530**38e+13

7.3 The ODE2 solved as a target value problem

7.3.1 Accurate element method

The ODE2 (2.4) will be integrated on two different intervals: between $x_S = 0$, $x_T = 5$ (Table 3) and $x_S = 0$, $x_T = 10$ (Table 4). The results given in both two tables represent the *estimated error* (7.1) for *NE = 2,3,4,5*. The target values ϕ_T given in the second column are obtained by using *NE=2*.

For the case $\mathbf{x_T = 5}$ (Table 3) it is useful to observe:

1.Based on the *estimated errors* one can divide the *CF*s in two groups: those with poor or medium errors (CF4, CF6, a little better CF8) and those errors indicating accurate results (CF10 to CF16).

2. The value that can be considered as accurate is ϕ_T = – **0.1606841854**, which is confirmed by all the CFs between CF10 and CF16.

3. This value is obtained **starting with NE=2** for CF10 to CF16.

4. *In this case the precision does not increase with the rise of the degree of the CF*, because seemingly the best value corresponds to CF12 (the smallest *estimated error*).

Table 3

ODE2 (x_S=0 ; x_T=5)

CF	ϕ_T	Estimated errors			
	NE=2	NE=2	NE=3	NE=4	NE=5
CF4	-0.160 1813334619398	3.25 e-3	1.76 e-3	7.31 e-4	1.91 e-4
CF6	-0.1606 830756433705	-1.97 e-5	1.11 e-5	-6.09 e-6	8.53 e-7
CF8	-0.16068418 29643502	2.13 e-8	1.37 e-8	1.21 e-9	9.20 e-10
CF10	-0.1606841854 438049	-9.02 e-11	2.37 e-11	-8.63 e-12	-1.28 e-13
CF12	-0.1606841854 464673	-1.82 e-12	-3.45 e-12	3.44 e-12	-8.23 e-13
CF14	-0.1606841854 524707	-5.78 e-11	-6.81 e-10	1.69 e-11	5.01 e-12
CF16	-0.1606841854 007920	7.09 e-11	1.43 e-10	1.77 e-10	-2.04 e-11

For the case $\mathbf{x_T}$ **= 10** (Table 4) the conclusions are somehow different:

1.The smaller degree CFs that give poor results are CF4 and CF6 (better)

2.The value that can be considered as accurate is ϕ_T = **– 0.086244366353** (CF8,CF10,CF12)

Table 4

ODE2 (x_S=0 ; x_T=10)

CF	ϕ_T	Estimated errors			
	NE=2	NE=2	NE=3	NE=4	NE=5
CF4	-0.0862 3535064991550	7.91 e-5	3.26 e-5	2.19 e-5	1.65 e-5
CF6	-0.086244366 06473081	-9.03 e-9	6.63 e-9	-5.05 e-9	3.19 e-9
CF8	-0.086244366353 17155	7.33 e-14	2.39 e-14	7.21 e-14	8.24 e-15
CF10	-0.086244366353 20805	1.08 e-13	2.59 e-13	-6.27 e-13	1.54 e-13
CF12	-0.086244366353 60698	6.52 e-12	-2.23 e-12	-1.71 e-12	-1.59 e-12
CF14	-0.08624436635 569120	-5.42 e-12	-3.53 e-11	2.93 e-12	3.10 e-12
CF16	-0.0862443663 4221143	-8.31 e-11	1.81 e-10	6.85 e-14	-1.25 e-10

3.The results given by CF14 and CF16 are a little worse (the accuracy does not increase steadily in this case together with the degree of CF)

4. The above value of ϕ_T corresponds to NE=2. *For nearly all CFs the accuracy does not increase significantly when the number of elements increases.*

7.3.2 Runge – Kutta method

This time the values given in Tables 3 and 4 will be compared to those given by the Runge – Kutta method with constant steps. In the last case the results are disappointing (if not useless) due probably to the instability. In fact, for the case $\mathbf{x_T = 5}$**,** for NE=500 and NE=1000 the code MATLAB gives a harsh verdict: NaN (Not a Number), resulting from operations which have undefined numerical results. If one increases NE the answers continue to be wrong: for NE=2000, ϕ_T = **–** 1.874 e+141 (?); for NE=3000, ϕ_T = **–** 2.835 e+13 (?). Only starting from NE=4000 the result becomes credible ϕ_T = **–** 0.1606841935, being not far from NE=5000 for which ϕ_T = **–** 0.1606841868. This last result that coincides with 8 digits to the value given in Table 3 needs duration **80 times greater then that corresponding to AEM**.

For the case $\mathbf{x_T = 10}$ the Runge – Kutta method with constant steps **gives no answer**, because for NE=5000, NE=10000, NE=15000 steps the result is invariably **NaN**.

7.3.3 Comparison between CFs and Runge-Kutta for accurate solutions

Two integration of ODE2 have been performed both starting from $x_S=0$ but with targets having smaller values $x_T=1$ and $x_T=2$, respectively (Table 5).

Table 5

ODE2

	$\mathbf{x_S = 0}$; $\mathbf{x_T = 1}$		$\mathbf{x_S = 0}$; $\mathbf{x_T = 2}$	
	NE	ϕ_T(x=1)	NE	ϕ_T(x=2)
CF16	4	-0.7529609952 579179	3	-0.3965673643 553024
CF14	4	-0.7529609952 655538	4	-0.3965673643 558060
CF12	5	-0.7529609952 683732	5	-0.3965673643 549966
CF10	7	-0.7529609952 746303	6	-0.3965673643 793142
CF8	11	-0.7529609952 928821	9	-0.3965673643 194347
CF6	29	-0.7529609952 965156	18	-0.3965673643 988034
CF4	82	-0.752960995 9757841	131	-0.3965673643 954343
Runge-Kutta	320	-0.752960995 0214358	1500	-0.3965673643 917458

These integration intervals allow a normal behavior of the Runge-Kutta method, so that a comparison between this last method and AEM with different CFs is possible. The comparison is based on the **number of elements *NE* necessary to obtain accurate results.** For each case the results with 10 digits have been considered as representing the accurate value:

for $x_T=1$, ϕ_T(x=1) = -0.7529609952 ; for $x_T=2$, ϕ_T(x=2) = -0.3965673643

7.4 The ODE3 solved as a target value problem

The ODE3 will be integrated between x_S=0 and x_T=5. As it was shown in §4 solving ODE3 rises a special problem: E_1(x) has two real roots ($x_1= 2$, $x_2=3$), both of them **inside** the integration interval. The Runge-Kutta method tries to find the value of ϕ at the end of each step, by using the value of the derivative at the beginning of the same step. Or, when the abscissa *x* tends to the value of the smallest root (in our case $x = 2$), $E_1(x)$ tends to zero. In such case the value of the first derivative – that results from a computation where $E_1(x)$ is the denominator [see (3.5a)] – tends to infinity. Consequently *the Runge-Kutta method stops at the smallest root*. This can be observed quite clearly from Table 6 where, in order to obtain credible results when x increases towards 2, it was necessary to rise the number of steps up to $NE = 5000$. For $x = 1.998$ no credible value of ϕ has been obtained even for $NE=10000$.

Table 6

ODE3

x	AEM		Runge-Kutta		
	ϕ_T (NE=2)	Estim.er.	NE=1000	NE=5000	NE=10000
1.9	-0.41533884308	1.98 e-10	-0.4153388	*	*
1.95	-0.40407866493	1.41 e-10	-0.4040787	*	*
1.98	-0.39762555610	2.75 e-11	-0.19122 (?)	-0.39762555695	*
1.99	-0.39552244408	-6.10 e-11	-2.77 e+10 (?)	-0.39552245265	*
1.995	-0.39447964673	1.35 e-10	*	-0.39535726520	*
1.998	-0.39385674271	1.42 e-10	*	-5.2695 e+10 (?)	-0.92980 (?)

The *AEM* leads to good estimated errors up $x = 1.998$, but when x tends to 2 the computation fails due to a division by $E_1=0$. For instance if $x_T=5$ and $NE=5$, the right end abscissa of the second element will be $x=2$, so that for this case no answer can be obtained. The problem of avoiding $x=2$ (or $x=3$) can be solved by modifying slightly the "target" abscissa (for instance $x_T=5.23$). If not, one can use the *Non-symmetrical Concordant Functions* [1], which are not affected by the coincidence with the roots of $E_1(x)$.

The computation for finding the target value $\phi_T(x=5)$ has been performed for $NE=1,\ 2,\ 3,\ 4$ then stopped in order to avoid $NE=5$ (see above). From the values given in Table 7 it results:

1.The higher order *CF*s (CF16,14,12,10) lead to very good values of the estimated errors.

Table 7

ODE3 (solved by AEM)

	NE=3		NE=4		Convergence criterion	
	$\phi_T(x=5)$	Estim.err.	$\phi_T(x=5)$	Estim.err.	NE=3	NE=4
CF16	-0.1608168891606	1.78 e-7	-0.1608168493921	-2.47 e-7	1.36 e+3	3.01 e+2
CF14	-0.1608168378795	-7.50 e-8	-0.1608168502207	8.67 e-8	1.41	1.01
CF12	-0.1608168491291	-6.20 e-9	-0.1608168475431	-9.86 e-9	1.34 e-2	6.12 e-3
CF10	-0.1608168552608	9.20 e-8	-0.1608168259850	-1.82 e-7	1.70 e-2	6.09 e-3
CF8	-0.1608172920844	2.62 e-7	-0.1608172320923	-3.73 e-7	2.53 e-2	8.16 e-3
CF6	-0.1608399995519	2.70 e-4	-0.1608374059234	-1.61 e-5	5.36 e-2	1.35 e-2
CF4	-0.1621552942969	1.00 e-4	-0.1602284414990	-1.20 e-2	7.48 e-2	3.23 e-3

2.It can be considered as accurate value $\boldsymbol{\phi_T(x=5)= -\ 0.1608168}$, value for which all the above mentioned *CF*s coincide for *NE=4.*

Because none of the *NE*s used lead to end abscissas $x=2$ or $x=3$, *AEM* "jumps" beyond them without any trouble.

8. The field polynomial solution (FPS)

Though finding polynomial solutions for an ODE is not always necessary, the analysis of the procedure is very instructive making clearer the behavior of *AEM*. The problem has already been formulated in §1: finding **Polynomial Solutions** $\phi_n(x)$ ($n=1,2...NE$), each one being valid on a single element from the integration field.

8.1 A single polynomial function cannot usually be a solution for a longer field

It is seldom possible to find a single polynomial solution of an ODE valid on a great integration field. Suppose for instance ODE1 having as accurate solution the 19th degree polynomial (2.2). It is obvious that the solution obtained – for instance – by using a third-degree polynomial (CF4) cannot accurately replace (2.2). Both functions have been drawn between x = 3.92 and x = 5.54 in Fig.8.1: the exact solution (2.2) as a continuous line and CF4 as a dotted line. Any comment concerning their coincidence is useless.

Nevertheless it is remarkable the fact that though CF4 has a completely different trajectory than the exact solution, they both tend to meet (with a certain error) at the target point. This is due to the **stock of information detained by CF4 that includes not only the target function ϕ_T but also its first derivative**[6]**.**

8.2 The CF curves converge towards the exact solution

When *NE=2* the trajectory of CF4 becomes to tend towards the accurate solution. This is due to the fact that now the first step calculates ϕ at the right end of the first element ($x=4.73$) so that CF4 **forcibly comes near to the exact solution**. As it results from Fig.8.2 the error in the middle point becomes smaller than that corresponding to *NE=1*, having as consequence the reducing of the error at the target point as compared to Fig.8.1.

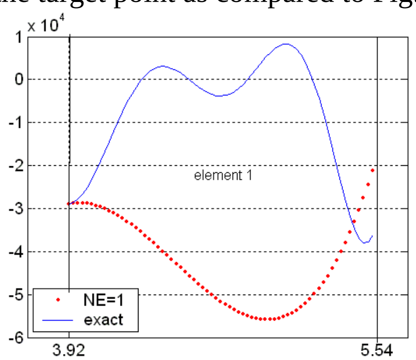

Fig.8.1

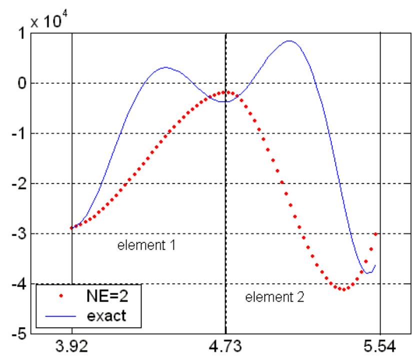

Fig. 8.2

[6] For the higher CFs also the second, third or higher order derivatives

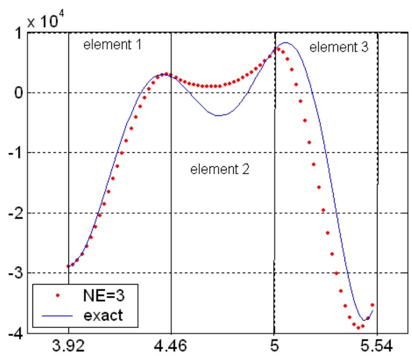

Fig. 8.3

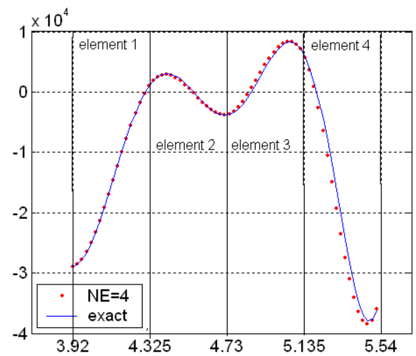

Fig. 8.4

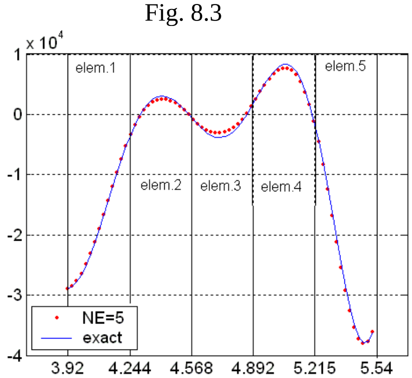

Fig. 8.5

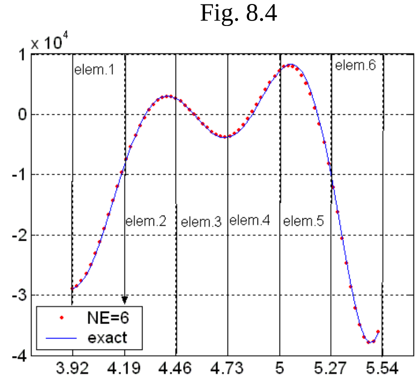

Fig. 8.6

For *NE=3* (Fig.8.3) the errors of CF4 at the right ends of *element 1* (x=4.46) and *element 2* (x=5) are quite small, so that the target value ϕ_T(x=5.54) becomes more credible. This tendency continues for *NE=4* (Fig.8.4), *NE=5* (Fig.8.5), *NE=6* (Fig.8.6). What is important is that each time when *NE* is increased there is *a new point that comes near to the exact solution*, which seemingly makes that for *NE=6* (Fig.8.6) the two curves coincide. From this analysis it results that **when the number of elements increases the CF curve converges towards the exact solution.** But the visual examination of a graphic is not enough to decide if a CF curve is satisfactory convergent, because the qualitative appreciation can be roughly delusive. A numerical criterion becomes necessary in order to decide if the convergence process has reached a satisfactory level.

8.3 The function $\tilde{\phi}$ in natural coordinates

In order to simplify the approach the Concordant Function has been given in §3 as (3.3), which is a function of *x*. Because according to the methodology of AEM the constants C_i are obtained *by calculating the inverse of a square matrix*, this procedure becomes difficult (even inaccurate) when the degree of CF increases beyond CF18 [1]. This problem is eliminated if $\tilde{\phi}$ is expressed in *natural (dimensionless) abscissas* given by

$$\eta = \frac{x - x_L}{x_R - x_L} = \frac{x - x_L}{h} \quad (8.1) \quad , \qquad \text{where} \qquad h = x_R - x_L \quad (8.2)$$

x_L being the *Left end abscissa*, respectively x_R the *Right end abscissa* of the element. It is obvious that for $x = x_L$, $\eta = 0$, while for $x = x_R$, $\eta = 1$. By using (8.1), the function $\tilde{\phi}$ (3.3) becomes [1]

$$\tilde{\phi}(\eta) = \overline{C}_1 + \overline{C}_2\,\eta + \overline{C}_3\,\eta^2 + \overline{C}_4\,\eta^3 \quad (8.3)$$

The use of the natural abscissas is necessary because the inverse of the matrix that allows obtaining the coefficients C_i *is always the same regardless of the Cartesian end abscissas of the elements*. Consequently the inverse has been calculated once and for all[7], being given (for all CFs between CF4 and CF16) in the Appendix A of [1]. The coefficients included in (8.3) being known, $\tilde{\phi}$ can be written as function of *x* by using (8.1)

$$\tilde{\phi}(x) = \overline{C}_1 + \overline{C}_2\left(\frac{x - x_L}{h}\right) + \overline{C}_3\left(\frac{x - x_L}{h}\right)^2 + \overline{C}_4\left(\frac{x - x_L}{h}\right)^3 \quad (8.4)$$

This function (used for drawing all the graphics) can also be written as

$$\tilde{\phi}(x) = \left[\overline{X}\right]\left[CxT\right] \quad (8.5), \quad \text{where}$$

$$\left[\overline{X}\right] = \left[1 \quad (x - x_L) \quad (x - x_L)^2 \quad (x - x_L)^3\right] (8.6); \left[CxT\right] = \left[\overline{C}_1 \quad \frac{\overline{C}_2}{h} \quad \frac{\overline{C}_3}{h^2} \quad \frac{\overline{C}_4}{h^3}\right]^T \quad (8.7)$$

8.4 A numerical criterion for establishing the level of convergence

The convergence process of the CF curves towards the exact solution can be established by comparing two curves based on an increasing number of elements. If one notes $\phi_{previous}$ and ϕ_{new} the two solutions of two successive curves, there are many possibilities to establish a criterion of convergence. Though the number of elements is different between the two cases, one can calculate the ordinates of the two curves *at the same abscissa* x_{test} by using (8.5). Here will be accepted the criterion used with good results in [2, page 160]

$$convergence\,criterion = \frac{1}{NP}\sqrt{\sum_{n=0}^{n=NP}\left(\frac{\phi_{new} - \phi_{previous}}{\left(\phi_{new} + \phi_{previous}\right)/2}\right)^2} \quad (8.9)$$

where *NP* is the number of test points x_{test} and $\phi_{mean}=(\phi_{previous} + \phi_{new}) / 2$.

The most difficult problem is to choose the value of the criterion (8.9), which can be considered as a conventional frontier between "unacceptable" and

[7] for a CF16 the inverted matrix is [16 × 16]

"acceptable". This value chosen by the author based on some numerical experiments is

$$\text{allowable convergence criterion} = 9.9 \times 10^{-3} \qquad (8.10)$$

Obviously this value is disputable.

Remark. It is important to observe that the continuity between two elements is secured (regardless the number of elements) as it follows: for CF4, continuity C_1 (function and its first derivative), for CF6, continuity C_2 (function and the derivatives *1* and *2*), for CF8, continuity C_3 (function and the derivatives *1, 2* and *3*) and so on [2].

8.5 A strategy for finding a field polynomial solution

As it was shown in §4 there are two parameters that can be chosen by the user: CF and/or NE. The strategy used for TVP in §7 was simple: because the precision resulted by using a low degree polynomial is usually unsatisfactory, one starts by using CF16 (or other high order CF), the parameter to be modified being *the number of elements*. The same procedure is repeated (if necessary) for smaller degrees CF. Usually two such attempts are enough.

The strategy for solving a FPS is totally different because it is not obvious which CF will lead to better results. Apparently the problem is to find two successive numbers of elements that leads to a good convergence criterion for a given CF. In fact this is not enough, because the solution has to be **the minimum set of polynomials** giving an analytic form valid on the integration field. Consequently the strategy presented here includes the following steps:

1. Choosing a value of the *allowable convergence criterion* that is considered as suitable.

2. Making a **"transverse cut" throughout all the CFs**, by testing the behavior of each CF based on the criterion (8.9).

3. Selecting the minimum number of elements N_{min} that satisfies the chosen convergence criterion.

Remark. The last decision has to take also into account the *estimated error* of the computed target value.

8.6 Field polynomial solutions for ODE2 and ODE3

The strategy sketched above has been applied to ODE2 (2.3) and ODE3 (2.5) for an integration interval between x_S=0 and x_T=5. The results given in Table 8 and Table 9 include for each CF: the pair of elements for which the (8.10) *allowable convergence criterion* is reached {NE(conv)}, the computed value of the (8.9) *convergence criterion* {Conv.} and the (7.1) estimated error {Est.er.}. As it results for both cases the better results are those corresponding to the middle values of the concordant functions (*CF10* with N_{min}=4 for ODE2, *CF8* with N_{min}=6 for ODE3). The value of the *estimated error* is very good for ODE2

(3.69 e-13**)** but only satisfactory for ODE3 (2.61 e-6**).** The graphics given in Fig.2.3 and Fig.2.4 correspond to these two cases.

ODE2 *Table 8*

CF	NE(conv)	Conv.	Est.er.
CF4	14/15	7.60 e-4	5.73 e-7
CF6	7/8	7.81 e-4	-4.89 e-8
CF8	5/6	7.59 e-4	-5.52 e-10
CF10	**4/5**	**9.79 e-5**	**3.69 e-13**
CF12	5/6	3.99 e-4	1.25 e-13
CF14	17/18	1.02 e-4	-5.53 e-13
CF16	29/30	1.91 e-3	5.13 e-13

ODE3 *Table 9*

CF	NE(conv)	Er.pat.	Est.er.
CF4	12/13	7.33 e-4	-2.39 e-4
CF6	8/9	6.21e-4	5.22 e-7
CF8	**6/7**	**8.60 e-4**	**2.61 e-6**
CF10	8/9	1.46 e-5	3.32 e-9
CF12	15/16	7.41 e-4	-3.46 e-8
CF14	10/11	4.72 e-2	-3.92e-7
CF16	*	*	*

It is interesting to observe from Table 9 that for NE<30 the value (8.10) has not been reached for CF14, while for CF16 no result has been obtained though the value of the allowable convergence criterion has been reduced to 9.9×10^{-1}. Or the value of the *convergence criterion = 4.72 e-2* for CF14 (see Table 9) **is not enough to lead to a good convergence**, as it results clearly from Fig.8.7. On the other hand the *estimated error* for the target value is very good (-3.92e-7).

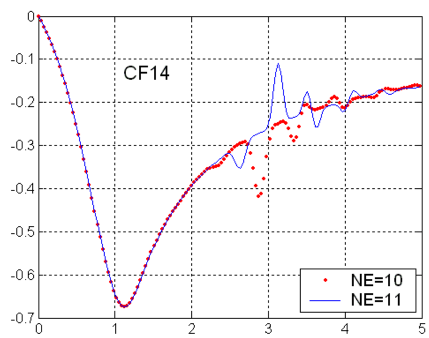

Fig.8.7

9. The analytic form of the solution

Obtaining a graphic that confirms the convergence tendency by the superposed curves is not the goal of the Field Polynomial Solution approach. After finding the minimum number of elements N_{min} from which on the convergence tendency is considered as satisfactory for a chosen CF, the answer has to be the **polynomial function** based on which the graphic has been drawn. At this stage the problem is solved because the constants included in [CxT] (8.7) are already obtained during the computation procedure.

Suppose it is necessary to describe the solution of ODE2. In order to simplify the exposure the integration interval is considered only **between x_S=0 and x_T=0.8**. If CF10 is used for this reduced interval it results N_{min}=1, because the convergence criterion between *NE=1* and *NE=2* is 2.74×10^{-5}. In this particular case the *Left end abscissa* used in (8.2) is $\mathbf{x_L = x_S = 0}$, so that $\boldsymbol{\eta} = \mathbf{x} / h$. Consequently, the analytic function results directly as a polynomial given by (3.3), but obviously including *10 terms* that corresponds to CF10

$$\tilde{\phi}(x) = 1 - 6.363636363636363\,\mathbf{x} + 13.71900826446281\,\mathbf{x^2} - 25.32682193839218\,\mathbf{x^3} + 36.49204289324499\,\mathbf{x^4} - 42.19755051295832\,\mathbf{x^5} + 40.21378331540018\,\mathbf{x^6} - 26.78179129972301\,\mathbf{x^7} + 10.03215529655342\,\mathbf{x^8} - 1.540291191858282\,\mathbf{x^9}$$

How accurate this solution is? A quick answer can be obtained by calculating the function at the middle of the integration interval, where the possible error is supposed to have a great value. If one replaces $x_M=0.4$ in $\tilde{\phi}(x)$ it results $\phi_M = \tilde{\phi}(0.4) = -0.3422291815$. A better result is obtained if the integration is performed directly between $x_S=0$ and $x_T=0.4$ in which case it results $\phi_T = -0.3423036660$. The error of ϕ_M is -2.17 e-4 , which is satisfactory.

A better answer can be obtained by using a more general method: verify if the polynomial solution (8.11) satisfies the ODE2 (2.3). This problem will be analyzed elsewhere.

Acknowledgements. The author is grateful to prof.S.Danaila and reader.B.Radu both for their objections and also for their suggestions that lead to the improvement of the analysis.

10. Conclusions

The paper identifies two types of problems connected to the integration of the Ordinary Differential Equations: the Target Value Problem (TVP) and the Field Polynomial Solution (FPS). The Accurate Element Method is an implicit method thus generally stable, which makes possible the integration over long intervals leading to accurate solutions with a small number of elements.

R E F E R E N C E S

1. Blumenfeld M., *The Accurate Element Method for solving Ordinary Differential Equations*, Editura JIF, Bucharest 2005 (in English). Large excerpts from this book can be found on the site www.blumenfeld.ro
2. Blumenfeld M., *A New Method for Accurate Solving of Ordinary Differential Equations*, Editura Tehnica, Bucharest 2001 (in English).