Optimum Choice of Integration Step

Pavel Pospíšil, Milena Hanzálková, Karel Mikulášek

Department of Computer Science and Engineering,
Faculty of Electrical Engineering and Computer Science,
Technical University of Brno,
Božetěchova 2, 612 66 BRNO, The Czech Republic,
Telephone: +42-5-7275277 (+42-5-41212219),
Fax:+42-5-41211141
E-Mail `pospispa@dcse.fee.vutbr.cz`

1 Abstract

The best-known and most accurate method of calculating a new value of a numerical solution of a differential equation is to construct the Taylor series.

Modern Taylor Series Method [1] has proved to be both very accurate and fast. It is based on a direct use of the Taylor series.

The main idea behind the Modern Taylor Series Method is an automatic integration method order setting, i.e. using as many Taylor series terms for computing as needed to achieve the required accuracy.

It is the aim of this paper to show the advantage of the method when multiple arithmetic is used. Two examples of homogenous differential equations are presented. The rather detailed analysis can be applied to further types of problems.

Keywords: homogenous differential equations, multiple arithmetic, Modern Taylor Series Method.

2 Homogenous Equation - Example 1

Let us solve the differential equation

$$y' = y, \quad y(0) = 1. \tag{1}$$

The numerical solution of (1) will depend on the number of Taylor series terms used.

Note : The analytical solution of (1) is

$$y = e^t. \tag{2}$$

Results illustrating the use of the Taylor series for applying a numerical integration method are shown in Tab.1. Tab.1 demonstrates the results of a numerical solution of the differential equation (1) after one computation step (with the integration step h=1s).

In each line of Tab.1 the "Reduced value y(1)" of the numerical solution of the differential equation (1) and the corresponding value of the "Absolute error" - for the method order ORD (number of Taylor series terms used) are printed.

In the column "Absolute error" we can read the difference between the exact and numerical solutions.

Reduced value y(1)	ORD	Absolute error
2.7182818	10	-0.000000027312660911
2.71828182	11	-0.000000002260552523
2.718281828	12	-0.000000000172876824
2.7182818284	13	-0.000000000012286394
2.71828182845	14	-0.000000000000815681
2.71828182845	15	-0.000000000000050959
2.71828182845904	16	-0.000000000000003220
2.71828182845904	17	-0.000000000000000444
2.71828182845904	18	-0.000000000000000333
2.71828182845904	19	-0.000000000000000333
2.71828182845904	20	-0.000000000000000333

Tab.1: Long real arithmetic

In order to make the results of the numerical solution more clear and illustrative only the digits tallying the digits of the exact solution are shown.

It follows from Tab.1 that the requirement of a higher method order is justifiable - with the same integration step h a higher method order (i. e. with more terms of the Taylor series) can yield a higher accuracy (it approximates better the exact solution).

However, the increase in the accuracy of the result is not unlimited. In Tab.1 the accuracy stops increasing when the ORD reaches the value of 18. This is caused by an underflow during the computation of the higher order Taylor series terms. The addition of these terms changes neither the resulting value of y_{n+1}

nor the absolute error - the absolute error has reached its saturated value ESAT (the value of ESAT depends on the word width of the arithmetic of the computer used).

2.1 Evaluation of the Computation Speed

Further important results can be achieved by experimental calculations in solving numerically the differential equation (1) with an explicit use of the Taylor series.

The exact solution $y(1) = e^1$ can also be approximated by solving the equation (1) with the integration step $h = 0.5s$ (Tab.2) or with the integration step $h = 0.1s$ (Tab.3).

Reduced value y(1)	ORD	Absolute error
2.718281828459045	14	-0.000000000000000222

Tab.2: Long real arithmetic

Reduced value y(1)	ORD	Absolute error
2.718281828459045	10	0.000000000000000222

Tab.3: Long real arithmetic

The results in Tab.2 and Tab.3 are again characterized by the saturated absolute errors ESAT (the computation is terminated when three consecutive absolute error values have not changed).

The relation between the number of operations (required for reaching the saturated absolute error ESAT) and the integration step h is given in Tab.4.

h	ORD	NRCS	Addition	Multipl.	DIV
1	18	1	18	18	17
0.5	14	2	28	28	26
0.1	10	10	100	100	90

Tab.4

In the column "DIV" ve have the number of division operations, in the column "Multipl." we have the number of product operations and in the column "Addition" the number of addition operations necessary for computing the results according to Tab.1, Tab.2 and Tab.3. In the column "ORD" in Tab.4 there is always the value of the method order with which the absolute error (according to Tab.1, Tab.2, Tab.3) reached its saturated value ESAT.

In the column "NRCS" , to complete the picture, the number of computation steps is shown after which, given the integration step h, the point $t_n = 1s$ has been reached (i. e. the point at which the exact solution is $y_n = e^1$).

Regardless of the practical construction of addition, multiplication and division it is clear from Tab.4 that the number of operations (required for reaching the highest accuracy - with the corresponding integration step h) increases as the integration step h shortens.

3 Homogenous Equation - Example 2

Let us solve the differential equation

$$x' = -x, \quad x(0) = 1. \tag{3}$$

Note : The exact solution of (3) is

$$z = e^{-t}. \tag{4}$$

In combination with (1) the equation (3) can be called a "check function". The reason for this is that for the product of the analytical solution of equations (1) and (3) we have

$$z = y * x = e^{t} * e^{-t} = 1 \tag{5}$$

and thus (5) might be used for testing the accuracy of the numerical solution.

The Example 2 has been solved by means of a multiple arithmetic (Tab.5).

3.1 Multiple arithmetic

The numbers over which the arithmetic operations are performed are in a floating-point representation:

$$x = M.z^{E}$$

Their accuracy is determined by the number of bits used for encoding the mantissa. A module has been developed enabling to select an arbitrarily great number of bits for encoding the mantissa. It is virtually limited only by the segment size so that, with a small reserve, the maximum length of the mantissa is 64500 x 8 bits. It is easy to see how accurate the calculations with such a tool are especially if we consider that with a mathematical co-processor the maximum length of the mantissa is only 64 bits. On the other hand it must also be taken into account that the increased accuracy is traded off by the time needed to perform the mathematical operations.

To use the multiple arithmetic for calculating the solution of homogenous differential equations it is sufficient to implement the following operations: addition, subtraction, multiplication, division.

Arithmetic	Number of terms	Valid digits	Time[s]	h_{max}
16B	121	20	0.027	25
16B	169	5	0.039	38
32B	266	20	0.170	65
32B	331	5	0.215	83
64B	590	20	1.291	155
64B	651	5	1.427	172
128B	1230	20	9.830	333
128B	1287	5	10.270	349

Tab.5: Maximum integration step h_{max} for the arithmetic used

4 CONCLUSION

The principal idea of an extremely accurate and fast method for numerical solutions is presented in the paper.

The numerical solutions are most exact when the method order ORD is accordingly high for the given integration step h. The accuracy of the result can be influenced in a considerable way by the word width.

The Example 2 has been solved by means of a multiple arithmetic (Tab.5).

5 REFERENCES

[1] Kunovský,J.: Modern Taylor Series Method, Habilitation work, TU Brno, 1995