Extensions of the Predictor-corrector Method for the Solution of Systems of ordinary Differential Equations

By C. V. D. Forrington

Two extensions of the predictor-corrector method for the solution of systems of ordinary differential equations are described. Both are aimed at reducing the number of derivative evaluations required to integrate over a given range. The combination of the two extensions to give an efficient general method is discussed.

1. Introduction

This article is concerned with the numerical solution of systems of equations of the form

$$\frac{dy_i}{dx} = f_i = f_i(x, y_1, \ldots, y_n) \quad (i = 1, 2, \ldots, n)$$

with particular reference to reducing the number of derivative evaluations required to integrate over a given range, while still retaining a check on the accuracy of the solution.

The advantages of predictor-corrector methods over Runge–Kutta processes in the above respects are well known and will not be reiterated here except to note that the formulae developed by Robertson (1959) and Hamming (1959) largely answer the charge that predictor-corrector methods may be unstable.

Methods of generalizing predictor-corrector formulae to allow different step-lengths to be used in each equation, and to allow arbitrary increments in the independent variable, are developed below, and applications of each extension are given. Each extension is considered in some detail and the combination of the two to give an efficient general method for the solution of systems of equations is discussed.

The methods are developed in terms of a particular form of fourth-order predictor-corrector pair, but the extensions to other forms and other orders will be obvious. For the purpose of this article, the step-length in an equation is defined as the increment in the independent variable between derivative evaluations. In the descriptions it is assumed that only a single application of the corrector is made.

2. Different Step-Lengths in each Equation

2.1 *Basis of method. The generalized predictor*

Consider the equations

$$\frac{dy}{dx} = f(x, y, z)$$

$$\frac{dz}{dx} = g(x, y, z)$$

and assume that the solution has reached the point x_0, using step-length h for the solution y and k for z, where $h = mk$ and m is an integer.

Let the values of y and f at the points $x_0 - ih$ be denoted by y_{-i} and f_{-i} and z, g at the points $x_0 - ik$ by z_{-i}, g_{-i}.

It is desired to progress to the point $x_0 + h\ (= x_0 + mk)$ by means of one predictor-corrector step in the first equation and m steps in the second. More explicitly, we do not wish to calculate any values of f at intermediate points of the range x_0 to $x_0 + h$.

It will be assumed that a predictor-corrector method of the form

$$y_1^* = a_0 y_0 + a_{-1} y_{-1} + h(b_0 f_0 + b_{-1} f_{-1} + b_{-2} f_{-2} + b_{-3} f_{-3}) \quad (1)$$

$$y_1 = c_0 y_0 + c_{-1} y_{-1} + h(d_1 f_1^* + d_0 f_0 + d_{-1} f_{-1} + d_{-2} f_{-2}) \quad (2)$$

where the asterisks denote predicted values and the coefficients are chosen to give fourth-order accuracy, is to be used.

In order to evaluate g at the appropriate points in the range x_0 to $x_0 + mk$, the values of y at these points are required, and they must be obtained without any evaluations of f, and hence the normal predictor-corrector sequence cannot be used for the first equation. Instead, the values of y must be obtained in some other way. A convenient method of doing this is to use a generalized predictor formula of the form

$$y_p = a_0(p) y_0 + a_{-1}(p) y_{-1} + h[b_0(p) f_0 + b_{-1}(p) f_{-1} + b_{-2}(p) f_{-2} + b_{-3}(p) f_{-3}].$$

This is of the same form as (1), but the coefficients are allowed to be functions of p, so as to give the value of y at a general point.

By expanding each term as a Taylor series about x_0, and equating powers of h up to h^4 and then solving for a_0, a_{-1} and b_0 to b_{-3}, we obtain

$$\left.\begin{aligned} a_0 &= 1 - a_{-1} \\ a_{-1} &= a_{-1} \\ 24\, b_0 &= (p^4 + 8p^3 + 22p^2 + 24p) + 9a_{-1} \\ 24\, b_{-1} &= -(3p^4 + 20p^3 + 36p^2) + 19a_{-1} \\ 24\, b_{-2} &= (3p^4 + 16p^3 + 18p^2) - 5a_{-1} \\ 24\, b_{-3} &= -(p^4 + 4p^3 + 4p^2) + a_{-1} \end{aligned}\right\} \quad (3)$$

and the truncation error is given by

$$K_p = (6p^5 + 45p^4 + 110p^3 + 90p^2 - 19a_{-1})\frac{h^5y^{(5)}}{720}.$$

The value of a_{-1} may be chosen arbitrarily, and it will be determined from considerations of stability, truncation error, and simplicity of formulae.

Thus we have a method of finding values of y in the range x_0 to $x_0 + mk$ without any evaluations of f at intermediate points, and these values may be used in calculating $g_1, g_2, \ldots, g_m$.

2.2 *A practical procedure*

If the value $a_{-1} = 0$ is taken in the generalized predictor formula, we get

$$y_p = y_0 + h(b_0f_0 + b_{-1}f_{-1} + b_{-2}f_{-2} + b_{-3}f_{-3}) \quad (4)$$

where b_0 to b_{-3} are obtained from (3).
When $p = 1$, the formula reduces to

$$y_1 = y_0 + \frac{h}{24}(55f_0 - 59f_{-1} + 37f_{-2} - 9f_{-3}) \quad (5)$$

with
$$K_1 = \frac{251}{720}h^5y^{(5)}$$

which is the Adams–Bashforth fourth-order predictor.

A convenient corrector formula is the Adams–Bashforth fourth-order corrector

$$y_1 = y_0 + \frac{h}{24}(9f_1 + 19f_0 - 5f_{-1} + f_{-2}) \quad (6)$$

which has a truncation error

$$K_2 = -\frac{19}{720}h^5y^{(5)}.$$

Using the above formulae, the sequence of operations to advance from x_0 to $x_0 + h$ is given below:

1 Put $p = \frac{1}{m}$, $q = 1$
2 → Predict z_q using (5)
3 Predict y_p using (4)
4 Calculate g_q
5 Correct z_q using (6)
6 Calculate g_q using corrected value of z_q
7 Put $p = 2p$, $q = q + 1$

m times

8 Calculate f_1 using predicted value of y_1 and value of z_m
9 Correct y_1 using (6)
10 Calculate f_1 using corrected value of y_1

2.3 *Notes on the method*

(*a*) *Improvement of value given by generalized predictor*

The method used by Hamming (1959) to improve the value obtained from the ordinary predictor formula may be extended to the generalized predictor, when the improved value is given by

$$Y = y_p + \frac{K_p}{K_2 - K_1}(p_0 - c_0)$$

where K_1 is the value of K_p at $p = 1$, K_2 is the truncation error coefficient in the corrector used, and p_0 and c_0 are the predicted and corrected values obtained for the previous step.

For the case $a_{-1} = 0$, the quantity to be added to the predicted value becomes

$$-\frac{6p^5 + 45p^4 + 110p^3 + 90p^2}{270}(p_0 - c_0).$$

(*b*) *Accuracy and change of step-length*

With the second equation, using the shorter step-length, the criterion for changing the step-length is the usual one of using the value $\frac{K_2}{K_2 - K_1}(p_1 - c_1)$ as an estimate of the truncation error arising from the step, where p_1, c_1 are the predicted and corrected values obtained for that step. If this is greater than the maximum permissible error, then the step-length must be shortened, while if it is much smaller, a longer step-length could be used.

The accuracy requirement for the first equation is somewhat different, since it is necessary that the values obtained from the generalized predictor should be within the prescribed accuracy, since they are used without correction in obtaining the solution of the second equation. Only a predicted value is available at the intermediate points of a stage, and it is necessary to use the difference between the predicted and corrected values at the end of the stage as a criterion.

For the case $a_{-1} = 0$, the truncation error in the generalized predictor is maximum in the range 0 to 1 when $p = 1$. Therefore, if the value obtained at $p = 1$ is within the required accuracy, so will the values obtained at intermediate points. The estimated truncation error in the predictor at the end of a stage is $\frac{K_1}{K_2 - K_1}(p_1 - c_1)$ and therefore p_1 and c_1 should agree to the required number of places, otherwise the stage must be repeated using a shorter step-length in the first equation.

In practice, since the predicted values of the first solution must be within the prescribed accuracy without correction, a somewhat shorter step-length must be used than if the corrector was applied each time. With the case $a_{-1} = 0$, a reduction of about 40% in the step-length used in the first equation would compensate for the increased truncation error coefficient.

(c) Starting the solution

This may be achieved by starting with equal step-lengths k in the equations and integrating over $3m$ steps to get the necessary starting values. It was found convenient to use a Runge-Kutta method for the first $3m$ steps, although it is possible to change over to the predictor-corrector method, using equal step-lengths k after only three steps. This might be worthwhile if m is large.

2.4 *More than two equations*

The extension to systems of more than two equations presents no special difficulty. The number of coefficients to be calculated for each step forward in the equation with the longest step-length depends on the number of different step-lengths and the ratios between them. If there are r different step-lengths $\frac{h}{m_i}$ $(i = 1, 2, \ldots, r)$ where h is the longest, it would probably be convenient to calculate each coefficient as it is required rather than store all the necessary values. In this case, $4(r - 1)m$ evaluations must be made, where $m = \max(m_i)$.

3. Unequal Increments in the Independent Variable

3.1 *Generalized predictor-corrector formulae*

We now consider the case of equal step-lengths in the equations at any point, but with successive step-lengths which are unequal. The method is given for a single equation, but the extension to a system is straightforward.

Assume that y and f are known at the four points x_0, $x_0 + p_1$, $x_0 + p_2$, $x_0 + p_3$ and are denoted at those points by y_0, f_0; y_1, f_1; y_2, f_2; y_3, f_3. It is desired to find formulae of similar forms to (1) and (2) to give the value of y at the point $x_0 + p$.

First consider the predictor and write it in the form

$$y^* = a_0 y_0 + a_1 y_1 + b_0 f_0 + b_1 f_1 + b_2 f_2 + b_3 f_3 \quad (7)$$

where a_0, a_1 and b_0 to b_3 are functions of p_1, p_2, p_3 and p. Each of the y's and f's may be expanded in a Taylor series about x_0, and the coefficients of the successive derivatives, up to the fourth, equated to give equations for a_0, a_1 and b_0 to b_3.

The equations obtained are

$$\begin{aligned} &a_0 + a_1 = 1 \\ &a_1 p_1 + b_0 + b_1 + b_2 + b_3 = p \\ &a_1 p_1^2 + 2b_1 p_1 + 2b_2 p_2 + 2b_3 p_3 = p^2 \\ &a_1 p_1^3 + 3b_1 p_1^2 + 3b_2 p_2^2 + 3b_3 p_3^2 = p^3 \\ &a_1 p_1^4 + 4b_1 p_1^3 + 4b_2 p_2^3 + 4b_3 p_3^3 = p^4. \end{aligned}$$

As before, one coefficient may be chosen arbitrarily and is determined by considerations of stability, truncation error, and simplicity of formulae.

A similar procedure with the corrector formula

$$y = c_0 y_0 + c_1 y_1 + df^* + d_0 f_0 + d_1 f_1 + d_2 f_2 \quad (8)$$

gives

$$\begin{aligned} &c_0 + c_1 = 1 \\ &c_1 p_1 + d + d_0 + d_1 + d_2 = p \\ &c_1 p_1^2 + 2dp + 2d_1 p_1 + 2d_2 p_2 = p^2 \\ &c_1 p_1^3 + 3dp^2 + 3d_1 p_1^2 + 3d_2 p_2^2 = p^3 \\ &c_1 p_1^4 + 4dp^3 + 4d_1 p_1^3 + 4d_2 p_2^3 = p^4. \end{aligned}$$

Again one coefficient may be chosen arbitrarily.

3.2 *A practical procedure*

Taking $a_1 = 0$ in the generalized predictor, corresponding to the fourth-order Adams–Bashforth predictor, we get

$$\left.\begin{aligned} a_0 &= 1 \\ a_1 &= 0 \\ b_3 &= \frac{p^2(6p_1p_2 - 4pp_2 - 4pp_1 + 3p^2)}{12p_3(p_1 - p_3)(p_2 - p_3)} \\ b_2 &= \frac{p^2(6p_1p_3 - 4pp_3 - 4pp_1 + 3p^2)}{12p_2(p_1 - p_2)(p_3 - p_2)} \\ b_1 &= \frac{p^2 - 2p_2b_2 - 2p_3b_3}{2p_1} \\ b_0 &= p - (b_1 + b_2 + b_3). \end{aligned}\right\} \quad (9)$$

$c_1 = 0$ in the generalized corrector gives

$$\left.\begin{aligned} c_0 &= 1 \\ c_1 &= 0 \\ d_2 &= \frac{p^3(2p_1 - p)}{12p_2(p - p_2)(p_1 - p_2)} \\ d_1 &= \frac{p^3(2p_2 - p)}{12p_1(p - p_1)(p_2 - p_1)} \\ d &= \frac{p^2 - 2p_1d_1 - 2p_2d_2}{2p} \\ d_0 &= p - (d + d_1 + d_2) \end{aligned}\right\} \quad (10)$$

which, when equal step-lengths are used, gives the Adams–Bashforth fourth-order corrector.

Using equations (7) and (8) as predictor and corrector respectively, assume that a point x_n has been reached and that the previous values of y and f have been obtained at $x_n - h_1$, $x_n - h_2$, $x_n - h_3$. The values of y and f at $x_n + h$ are required. The necessary coefficients in (7) and (8) are obtained by putting $p_1 = -h_1$, $p_2 = -h_2$, $p_3 = -h_3$, $p = h$ in (9) and (10). This enables y at $x_n + h$ to be calculated. Thus values of y at unequal intervals in x may be calculated.

3.3 *Notes on the method*

(*a*) *Improvement of predicted value*

The trucation error in the generalized predictor is given by

$$\frac{K_1 y^{(5)}}{5!}$$

where $K_1 = (p^5 - a_1 p_1^5 - 5b_1 p_1^4 - 5b_2 p_2^4 - 5b_3 p_3^4)$, and in the generalized corrector by

$$\frac{K_2 y^{(5)}}{5!}$$

where $K_2 = (p^5 - c_1 p_1^5 - 5dp^4 - 5d_1 p_1^4 - 5d_2 p_2^4)$.

An estimate of the truncation error in the predictor is given by

$$\frac{K_1}{K_2' - K_1'}(p_0 - c_0)$$

where p_0, c_0 are the predicted and corrected values obtained for the previous step, K_1', K_2' the truncation error coefficients for the previous step, and K_1 the truncation error coefficient in the predictor for the current step. This may be added to the predicted value to give an improved approximation to y.

(*b*) *Accuracy and change of step-length*

An estimate of the truncation error in the corrector is given by $\frac{K_2}{K_2 - K_1}(p_1 - c_1)$, where all the quantities refer to the current step. This may be used as a criterion for determining the suitability of the step-length in the same way as with the ordinary predictor-corrector method.

(*c*) *Starting the solution*

The first three steps in the process must be performed by some other method, not depending on earlier values. It is convenient to use a fourth-order Runge–Kutta method for this purpose.

(*d*) *More than one equation*

The method extends immediately to systems of equations, the coefficients in the formulae being the same for each equation. Thus as the number of equations increases, the proportion of the computation time spent in evaluating the coefficients at each stage will decrease.

4. Applications of Above Methods

The first method is of use when some solutions of a system are varying much more rapidly with the independent variable than others. Such systems arise in some physical problems where ratios of a hundred to one in the time constants involved are not uncommon. If equal step-lengths were used in the equations of such systems the computing time might be prohibitively long, whereas using different step-lengths, allowing a uniform accuracy bound for all the equations, a considerable reduction is possible. The method has been demonstrated for various equations on the R.A.E. Mercury computer, and satisfactory results have been obtained.

The second method may be used where it is desired to change the step-length frequently, or make it a function of some other variables in the problem. Alternatively, it can be used as an efficient method of changing the step-length in ordinary predictor-corrector methods which is both fast and requires very little extra storage space. The coefficients in the formulae need only be evaluated when a change of step-length is about to take place, or has taken place in the previous three steps.

The possibility of choosing an optimum step-length at each stage for use with the second method is being investigated, and one idea is for successive step-lengths to be determined by the formula $h_2 = k\left(\frac{e}{e'}\right)^{1/5} h_1$, where e is the maximum permissible truncation error in a step, e' is the estimated truncation error for the previous step, and k is some constant less than one.

5. Combining the two Methods

In theory there is no difficulty in combining the two methods to give a general method for the solution of systems of equations in which the step-length in any equation at any stage depends only on the rate of variation of the solution of the particular equation in the neighbourhood of the current value of the independent variable. However, there are several practical disadvantages with a completely general procedure of the above form. For example, for each step forward in the equation using the shortest step-length, up to $4n + 4$ different coefficients would have to be calculated, and the time taken to do this might be a significant proportion of the total computing time. Also up to $4n$ quantities, representing the previous three step-lengths and the current step-length in each equation, must be stored, and this, together with a longer program, would represent a considerable increase in the storage space required, compared with ordinary methods.

Rather more attractive is a restricted scheme which allows only a limited number of different step-lengths to be used at any stage, and which makes a change of step-length a relatively infrequent occurrence. For example, if the equations were split into two groups, representing two different step-lengths, and no change of step-length was occurring, only four coefficients would have to be calculated for each step forward in the shorter step-length, and these would be of the simple form given by (3). When a change of step-length was required, the second method could be used to obtain the appropriate coefficients, and very little extra storage space, independent of the total number of equations, would be required.

A scheme like the above would be applicable to physical systems involving two widely different time constants, where the equations for the rapidly varying solutions depend on the slowly varying solutions.

Other ideas for dealing with systems with widely different time constants are being developed by R. H. Merson and are based on his own variations of the Runge–Kutta method (Lance, 1960).

6. Stability of the Methods

A complete stability analysis for the above processes is beyond the scope of the present paper, but the following points are noted.

(*a*) By restricting the formulae to fourth-order accuracy when sufficient information is used to give fifth-order accuracy, the coefficients are obtained in terms of a single arbitrary parameter. Robertson (1959) considers the stability of the corrector applied to a single equation and using equal step-lengths, and obtains a condition on the value of the parameter for stable formulae. The value used in the formulae developed above satisfies that condition and gives a reasonable compromise between accuracy and radius of stability.

(*b*) The uncorrected values given by the generalized predictor and used at intermediate points of a stage will not themselves become unstable over the stage, since they are independently obtained from the corrected values of the solution at earlier pivotal points.

(*c*) The value taken above for the parameter gives the Adams–Bashforth fourth-order method when equal step-lengths are used, and this is well known as a stable process.

7. Examples

The following examples illustrate the economies that may be obtained by using different step-lengths in the equations.

(i) It was required to obtain the solution of the equations

$$\frac{dy_1}{dx} = f_1 = \cos x$$

$$\frac{dy_2}{dx} = f_2 = 100y_1 \cos 100x + \cos x \sin 100x$$

at $x = 1$, given that $y_1 = y_2 = 0$ at $x = 0$. The results obtained were checked against the formal solutions

$$y_1 = \sin x$$

$$y_2 = y_1 \sin 100x.$$

Using equal step-lengths h for each equation, the value $h = 0{\cdot}05$ gave six-figure accuracy in the first equation, while a value $h = 0{\cdot}0005$ was necessary to obtain the same accuracy in the second. Using different step-lengths h_1 and h_2 in the two equations, six-figure accuracy was obtained with $h_1 = 0{\cdot}025$ and $h_2 = 0{\cdot}0005$. With equal step-lengths it was necessary to evaluate f_1 at 4,000 points in [0, 1], whereas using different step-lengths it was calculated at only 80 points.

(ii) An example for which the solutions were not known is now given.

$$\frac{dy_1}{dx} = f_1 = -y_1\sqrt{(1 + x^2)}e^{-x \cos x}$$

$$\frac{dy_2}{dx} = f_2 = y_1 + \cos (20y_2).$$

Given $y_1 = 2, y_2 = 0$ at $x = 0$, the solutions at $x = 1$ were required.

Using equal step-lengths in the two equations, the following values of h were tried: $h = 0{\cdot}00125, 0{\cdot}0025, 0{\cdot}005, 0{\cdot}01$.

Results consistent to six figures were obtained with $h = 0{\cdot}00125$ and $h = 0{\cdot}0025$, and so the value $h = 0{\cdot}0025$ was taken as a suitable step-length. With this value, 800 evaluations of f_1 were necessary.

Using different step-lengths h_1 and h_2, consistent results were obtained with $h_1 = 0{\cdot}025$ and $h_2 = 0{\cdot}0025$. In this case only 80 evaluations of f_1 were required.

References

Hamming, R. W. (1959). "Stable Predictor-Corrector Methods for Ordinary Differential Equations," *J. Assoc. Comp. Mach.*, Vol. 6, p. 37.

Lance, G. N. (1960). *Numerical Methods for High Speed Computers*, London, Iliffe and Sons Ltd.

Robertson, H. H. (1959). "Some New Predictor-Corrector Methods," Int. Rep. No. Ma/45/1231. National Physical Laboratory.