D = molecular diffusion coefficient

D_r = total radial diffusion coefficient, $D_r = D + \epsilon_r$

D_x = total axial diffusion coefficient, $D_x = D + \epsilon_x$

D_1 = dimensionless total radial diffusion coefficient, D_r/D

D_2 = dimensionless total axial diffusion coefficient, D_x/D

f_k = functions defined by Equations (6) and (11)

F = function defined by Equation (12)

G = function defined by Equation (14)

J_0 = zero order Bessel function of first kind

J_1 = first order Bessel function of first kind

K = dispersion coefficient defined by Equation (8)

$\overline{K}$ = time average dispersion coefficient defined by Equation (25)

N_{Pe} = Peclet number, $U_R R/D$

N_{Sc} = ν/D

r = radial coordinate

R = tube radius

t = time

u = axial velocity

u_1 = u/U_R

U_R = reference velocity, $2U(\infty)$ where $U(\infty)$ is the bulk average velocity corresponding to fully developed laminar flow

v = radial velocity coordinate

v_1 = v/U_R

x = axial distance

X = $x/R\ N_{Pe}$

x_s = length of slug at $t = 0$

X_s = x_s/RN_{Pe}

y = r/R

α_n = roots of $J_0\ (\alpha_n) = 0$

γ_n = roots of $J_1\ (\alpha_n) = 0$

ϵ = eddy diffusion function

θ = dimensionless local concentration, C/C_R

θ_m = dimensionless average concentration

λ = auxilliary parameter replacing τ when applying Duhamel theorem

ν = kinematic viscosity

τ = dimensionless time, tD/R^2

σ = auxilliary parameter replacing X when applying Duhamel theorem

LITERATURE CITED

1. Ananthakrishnan, V., W. N. Gill, and A. J. Barduhn, *AIChE J.*, **11**, 1063 (1965).
2. Aris, Rutherford, *Proc. Roy. Soc. (London)*, **252A**, 538 (1959).
3. Bailey, H. R., and W. B. Gogarty, *ibid.*, **259A**, 352 (1962).
4. Bird, R. B., W. E. Stewart, and E. N. Lightfoot, "Transport Phenomena," John Wiley, New York (1960).
5. Carrier, G. F., *Quart. Appl. Math.*, **14**, 108 (1956).
6. Christiansen, E. B., and H. E. Lemmon, *AIChE J.*, **11**, 995, (1965).
7. Gill, W. N., *Proc. Roy. Soc. (London)*, **298A**, 335 (1967).
8. ———, and V. Ananthakrishnan, *AIChE J.*, **13**, 801 (1967).
9. Gill, W. N., *Chem. Eng. Sci.*, **22**, 1013 (1967).
10. Philip, J. R., *Australian J. Phys.*, **16**, 287 (1963).
11. Taylor, G. I., *Proc. Roy. Soc.*, *(London)*, **219A**, 186 (1953).
12. Fan, L. T., and C. L. Hwang, *Kansas State Univ., Bull., Spec. Rep.*, 67, **50**, No. 3 (1966).
13. Lighthill, M. J., *J. Inst. Math. Its Appl.*, **2**, 97 (1966).

Manuscript received September 11, 1967; revision received December 19, 1967; paper accepted January 8, 1968.

Stability of Numerical Integration Techniques

G. P. DISTEFANO

Monsanto Company, St. Louis, Missouri

This paper presents the cause of instabilities which arise during the numerical solution of ordinary differential equations. By using the numerical integration routines presently available, one actually approximates the differential equation by a difference equation. If the difference equation is of a higher order than the original differential equation, the approximate solution contains extraneous solutions which are not at all related to the true solution. It is the behavior of these extraneous solutions that one is usually concerned with in a stability analysis.

Also presented is a procedure for obtaining a bound on the largest allowable integration step size for a class of chemical engineering problems. A detailed explanation of the procedure is illustrated for unsteady state distillation calculations.

The purpose of this paper is to present a clear, concise discussion of the main causes of instabilities which arise during the numerical integration of ordinary differential equations on a digital computer. Ordinary differential equations are an integral part of chemical engineering calculations in such basic areas as chemcial kinetic studies, transient distillation, heat transfer, etc. The complexity of such systems vary from a single equation to a system of several hundred simultaneous differential equations. In practice, these equations are usually highly nonlinear, and one must resort to numerical techniques in order to obtain a solution.

Two main problems arise in the numerical solution of differential equations, truncation error and numerical instability. Truncation error usually lends itself nicely to rigorous error analysis, and procedures are available to contain truncation error within some prescribed limits. Numerical instability, on the other hand, is more complex, and only in the recent literature has there been any attempt at a rigorous mathematical analysis.

By using the numerical integration routines presently available, one actually approximates a differential equation by a difference equation which is solved in a step-by-step or marching fashion.

STABILITY

It is important to begin this paper with a clear definition between convergence and stability of finite difference techniques. By convergence one means that the finite difference solution approaches the true solution as the

G. P. Distefano is now with Electronic Associates, Inc. Princeton, New Jersey.

interval size approaches zero. The concept of stability on the other hand is associated with the propagation of errors of the numerical technique as the calculations progress with a finite interval size, that is, what the effects of errors made on one step will have on succeeding steps. It should be clear from the above definitions that convergence does not necessarily imply stability.

The problem of instability arises because in most instances the order of the approximating difference equation is higher than that of the original differential equation. Hence, the difference equation possesses extraneous solutions which in some instances can dominate the solution, so that the solution of the difference equations bears little, if any, resemblance to the true solution of the original differential equation. It happens frequently that the spurious solutions do not vanish even in the limits as the increment sizes approaches zero. This phenomenon is called *strong instability*, and implies lack of convergence as well as lack of stability. When a method possesses convergence but has unstable asymptotic behavior, the phenomenon is called *weak instability*. For example, a numerical routine that is stable for some finite increment size, h_1, but is unstable for some larger increment size, $h_2 > h_1$, is said to possess a weak instability.

For linear differential equations, the problem can be discussed qualitatively in the following manner (*1*). Suppose the original differential equation is of order q, it possesses a solution of the form,

$$x(t) = \sum_{i=1}^{q} c_i \, e^{\alpha_i t} + G(t) \tag{1}$$

where $e^{\alpha_i t}$ are the complementary functions, and $G(t)$ is the particular solution.

Suppose now that the approximating finite-difference equation is of the order m, so that its corresponding solution is of the form,

$$x(t_n) = \sum_{i=1}^{m} \overline{c}_i \, \beta_i^n + \overline{G}(t_n) \tag{2}$$

where β_i^n are the complementary functions, and $\overline{G}(t_n)$ is the particular solution.

Two cases require consideration. In the first place, let the possibility that $m = q$ exist (that is, the order of the finite-difference equation is the same as that of the differential equation). One can expect convergence in this case, as it is reasonable to expect that as h approaches zero, each β_i^n tends to a corresponding $e^{\alpha_i t}$, and convergence results. Therefore, in cases such as this, the method can possess at worst a weak instability.

The second case warranting discussion, $m > q$, is more serious than the first in that the finite-difference solution contains extraneous solutions bearing no resemblance to the solution of the original differential equation. If these parasitic solutions tend to vanish as h approaches zero, then the technique possesses convergence. If their asymptotic behavior does not affect that of the remaining complementary functions, the technique is said to be stable. It may be possible that the parasitic solutions do not vanish as h approaches zero, and/or their absolute magnitude tends to grow as the calculations proceed. In such cases the technique can lack converegence or stability, or both. It should be obvious, therefore, that techniques having this characteristic (that is, $m > q$) can possess a strong instability.

It is the growth of the extraneous solutions relative to the true solution that one is concerned with in a stability analysis. If the extraneous solutions tend to zero as the calculations progress, the solution is said to possess absolute stability; if the extraneous solutions tend to grow as the calculations progress, but at a rate much slower than the true solution, the solution is said to possess relative stability; if the extraneous solutions tend to grow as the calculations progress, at a rate faster than the true solution, the solution is said to be unstable.

DIFFERENTIAL EQUATIONS

In order to discuss stability in a quantative way, one must define a differential equation as well as the numerical routine used in the approximate solution. The classical equation studied in this connection is the simple linear first-order differential equation,

$$x' = Ax \tag{3}$$

where A is a constant. It can be shown that, to a first-order approximation, the results obtained from a stability analysis on the above linear equation can be exteneded to the nonlinear case

$$x' = f(x, t) \tag{4}$$

where $\partial f/\partial x$ from Equation (4) plays a role similar to the constant, A, in Equation (3).

The nonlinear function, $f(x, t)$, can be linearized by expansion of the function about the point (x_o, t_o) in a Taylor series truncated after the first-order terms. The resulting linearized form for Equation (4) is given in Equation (5):

$$x' = Ax + Bt + c \tag{5}$$

where

$$A = \left(\frac{\partial f}{\partial x}\right)_o \tag{6}$$

$$B = \left(\frac{\partial f}{\partial t}\right)_o \tag{7}$$

$$C = \left[f_o - x_o\left(\frac{\partial f}{\partial x}\right)_o - t_o\left(\frac{\partial f}{\partial t}\right)_o \right] \tag{8}$$

Hildebrand (*2*) has shown that the stability characteristics of the linear Equation (5) are very similar to the stability characteristics of equations of the form given by Equation (4). For example, it can be shown that the same extraneous solutions exist in both cases, even though the growth of these extraneous solutions relative to the true solution will be slightly different in each case. However, the basic method of analysis is the same for both cases. It is not felt, therefore, that the added complication of the additional linear terms in Equation (5) should be taken into consideration in this paper.

Therefore, the following analysis will be based on the equation

$$x' = f(x, t) \cong Ax \tag{9}$$

where

$$A \cong \left(\frac{\partial f}{\partial x}\right) \tag{10}$$

and it is assumed that $(\partial f/\partial x)$ is relatively invariant in the region of interest. Equation (9) has as its solution

$$x = c_1 \, e^{At} = c_1 \, e^{Ahn} = c_1 \, (e^{\overline{h}})^n \tag{11}$$

where

$$t = nh, \tag{12}$$

and

$$\overline{h} = \left(\frac{\partial f}{\partial x}\right) h \tag{13}$$

The constant, c_1, is evaluated from the initial conditions of the differential equation. Neither the constant, c_1, nor the particular solution affect the stability properties of the integration routine. Therefore, no reference to either of these quantities will be made in the following discussion.

SYSTEMS OF DIFFERENTIAL EQUATIONS

Any system of simultaneous nonlinear first-order differential equations such as

$$\frac{d\vec{x}}{dt} = \vec{f}(\vec{x}, t) \tag{14}$$

can be linearized about an operating point $(\vec{x}_o, t_o)$ and put into matrix form as

$$\frac{d\vec{x}}{dt} = \tilde{A}\,\vec{x} + t\,\vec{F} + \vec{f}_o \tag{15}$$

where $\tilde{A}$ denotes the Jacobian matrix, $\vec{F}$ is a vector containing the terms $(\partial f_i/\partial t)$, and $\vec{f}_o$ is the original function plus some first-order terms, all evaluated at the point $(\vec{x}_o, t_o)$. It can be shown (3) that the quantity in the system of equations that plays a role similar to $(\partial f/\partial x)$ in the case of the single differential equation is the largest negative eigenvalue of the $\tilde{A}$ matrix. Therefore, the equation which corresponds to Equation (13) for a system of simultaneous differential equations is

$$\bar{h} = \lambda_{\max}\, h \tag{16}$$

where $\lambda_{\max}$ represents the value of the largest negative eigenvalue of the Jacobian matrix.

NUMERICAL TECHNIQUES

Numerical integration techniques fall into two rather broad categories: single-step and multistep methods. Some common examples are:

1. Single-step methods (a) Explicit methods: Forward Euler, Runge-Kutta formulas. (b) Implicit methods: Backward and modified Euler formulas.

2. Multistep methods (a) Explicit methods: All open-end predictor formulas. (b) Implicit methods: All closed -end corrector formulas. (c) Predictor-corrector methods: One or more predictor application followed by one or more corrector applications or iterations.

Predictor-corrector methods can be considered as mixed explicit, implicit methods. Unless the corrector (implicit) is iterated until convergence results, the predictor formula (explicit) must enter into the stability analysis because any error in the predicted value will affect the corrected value. It should be mentioned here that any purely implicit technique requires an iterative method of solution for nonlinear differential equations.

In the above examples the single-step methods were considered separately because they represent first-order difference equations approximations to first-order differential equations. Therefore, these techniques do not introduce extraneous solutions and can possess at worst a weak instability. It can be generalized further by stating that any single step method or any multistep method that requires no history points of the independent variables (but may require history points of the derivative of the independent variable) can possess at worst a weak instability. For example, it can be shown that the general corrector equation,

$$x_{n+1} = x_{n-r} + h(a_{n+1}x'_{n+1} + a_n x'_n + a_{n-1}x'_{n-1} + \cdots + a_{n-m}x'_{n-m}) \tag{17}$$

can be strongly unstable if $r > 0$, but cannot be strongly unstable if $r = 0$.

SINGLE-STEP METHODS

Explicit Method: Forward Euler Formula

If one uses the forward Euler formula

$$x_{n+1} = x_n + hx'_n \tag{18}$$

to approximate Equation (9), Equation (18) becomes

$$x_{n+1} - (1 + \bar{h})\,x_n = 0 \tag{19}$$

Equation (19) is a first-order difference equation whose solution is of the form

$$x_n = \bar{c}_1\beta^n = \bar{c}_1(1 + \bar{h})^n \tag{20}$$

where $\bar{c}_1$ is evaluated from the initial conditions. In this case $\bar{c}_1$ should be equal to c_1. Since the true solution to Equation (9), given in Equation (11), tends to grow without bound for positive values of A (positive $\bar{h}$) one hopes that the approximate solution also grows without bound for positive $\bar{h}$. However, for negative A (negative $\bar{h}$), the true solution tends to zero. One wants the approximate solution to approach zero in the region of negative $\bar{h}$, a result which can be realized if, and only if, $|\beta|$ is less than one, as can be seen from Equation (20). Herein, we will be concerned only with investigating the nature of the approximate solution in the region of negative $\bar{h}$.

From the form of Equations (11) and (20) one can see that the numerical technique uses $(1 + \bar{h})$ to approximate $e^{\bar{h}}$. We can therefore conclude that the technique is stable when

$$|\beta| = |1 + \bar{h}| < 1 \tag{21}$$

A plot of $|\beta|$ vs. $\bar{h}$ (Figure 1) shows that Euler's forward formula gives stable results for $(-\bar{h}) < 2$.

Implicit Method: Modified Euler Formula

If one uses the modified Euler formula

$$x_{n+1} = x_n + \frac{h}{2}(x'_{n+1} + x'_n) \tag{22}$$

to approximate Equation (9), Equation (22) becomes

$$(2 - \bar{h})\,x_{n+1} - (2 + \bar{h})\,x_n = 0 \tag{23}$$

The solution to Equation (23) is given by

$$x_n = \bar{c}_1\,\beta^n = \bar{c}_1\left(\frac{2 + \bar{h}}{2 - \bar{h}}\right)^n \tag{24}$$

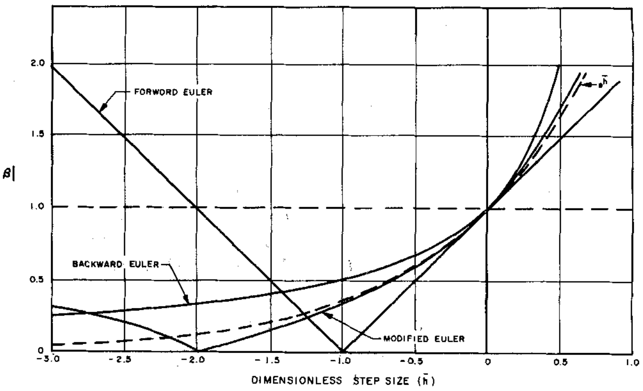

Fig. 1. Euler methods.

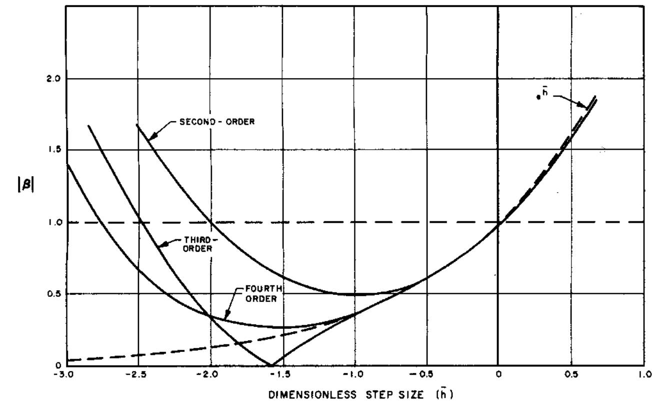

Fig. 2. Runge-Kutta methods.

For a plot of $|\beta|$ vs. $\bar{h}$ one can see the modified Euler technique is stable for all values of $\bar{h}$ (see Figure 1).

RUNGE-KUTTA FORMULAS

The Runge Kutta formulas of mth order approximate the differential equation with a Taylor's series expansion of order m, but are derived in such a way that all higher-order terms are expressed in terms of first-order derivatives.

It can be easily shown (*4*) that the mth-order Runge-Kutta approximation to the solution of the simple differential equation given in Equation (9) is

$$x_{n+1} = \left(1 + \bar{h} + \frac{\bar{h}^2}{2!} + \frac{\bar{h}^3}{3!} + \cdots \frac{\bar{h}^m}{m!}\right) x_n \quad (25)$$

which is an exact approximation of $e^{\bar{h}}$ up to order m. This function is plotted against $e^{\bar{h}}$ in Figure 2 for the second, third, and fourth-order Runge-Kutta formulas. It can be seen that the higher-order Runge-Kutta formulas have a greater stability range than the lower-order formulas, but only at the expense of more derivative evaluations per increment. The second-order formula requires two derivative evaluations per increment, the third-order requires three, the fourth-order requires four, the fifth-order requires six, and so on. Since the level of effort in computation required to move one increment in time is directly proportional to the number of derivative evaluations per increment, the performance of a method should be evaluated in terms of the stability range and the number of derivative evaluations required per increment.

MULTISTEP METHODS

The true solution of Equation (9) is given by Equation (11). If the approximate solution is obtained by using a multistep method of either an explicit (open integration) type or a mixed (predictor-corrector) type of order m, the finite-difference solution is given by

$$x_n = \bar{c}_1 \beta_1{}^n + \bar{c}_2 \beta_2{}^n + \bar{c}_3 \beta_3{}^n + \cdots + \bar{c}_m \beta_m{}^n \quad (26)$$

provided all the roots, β_1, are real and distinct. The form of Equation (26) for complex roots, repeated roots, etc., is given by Hildebrand (*5*).

It can be shown that if the numerical technique possess consistency (*6*), one of the solutions to the finite-difference equation (say, $\bar{c}_1\beta_1{}^n$) will approximate the true solution $[c_1(e^{\bar{h}})^n]$ in the region of small $\bar{h}$. In other words, β_1 will tend to $e^{\bar{h}}$ in the limit as $\bar{h}$ approaches zero, and all other roots or solutions will be extraneous solutions. Therefore, an explicit mth-order method will contain $(m - 1)$ extraneous solutions. It can be shown that purely implicit mth-order method will contain only $(m - 2)$ extraneous solutions. This result can be seen in the case of the second-order modified Euler, as well as in the case of the third-order Adams-Moulton corrector.

Explicit Method: Third-Order Adams-Moulton Predictor

By using the third-order Adams-Moulton predictor formula

$$x_{n+1} = x_n + \frac{h}{12}\left(23x'_n - 16x'_{n-1} + 5x'_{n-2}\right) \quad (27)$$

in the solution of Equation (9) results in the difference equation approximation

$$x_{n+1} - \left(1 + \frac{23}{12}\bar{h}\right)x_n + \left(\frac{16}{12}\bar{h}\right)x_{n-1} - \left(\frac{5}{12}\bar{h}\right)x_{n-2} = 0 \quad (28)$$

Equation (28) is a third-order difference equation which possesses two extraneous solutions when used to approximate a first-order differential equation. These extraneous solutions are obtained by finding the roots of the characteristic equation of Equation (28), which is given below as

$$\beta^3 - \left(1 + \frac{23}{12}\bar{h}\right)\beta^2 + \left(\frac{4}{3}\bar{h}\right)\beta - \left(\frac{5}{12}\bar{h}\right) = 0 \quad (29)$$

The solution of Equation (28) will be of the form,

$$x_n = \bar{c}_1\beta_1{}^n + \bar{c}_2\beta_2{}^n + \bar{c}_3\beta_3{}^n \quad (30)$$

provided that the three roots (β_1, β_2, β_3) are all real and distinct. For a discussion on the form of solving Equation (28) for the case of complex or repeated roots see the literature (*5*).

The roots of Equation (29) are plotted in Figure 3. In all of the graphs the roots are displayed in the following fashion: for real roots the absolute value of the root is plotted, and for conjugate complex roots the modulus of the pair is plotted as a single quantity (thus conjugate complex pairs of roots show as a single curve).

In Figure 3 it can be seen that β_1 is the true root, and β_2 and β_3 are extraneous roots. At $(-\bar{h}) = 0.6$, $|\beta_2|$ is greater than one, and in Equation (30) the term containing $|\beta_2|$ will grow without bound as n gets large, thus dominating the solution and swamping the true root which tends to zero since $|\beta_1|$ is less than one. The degree

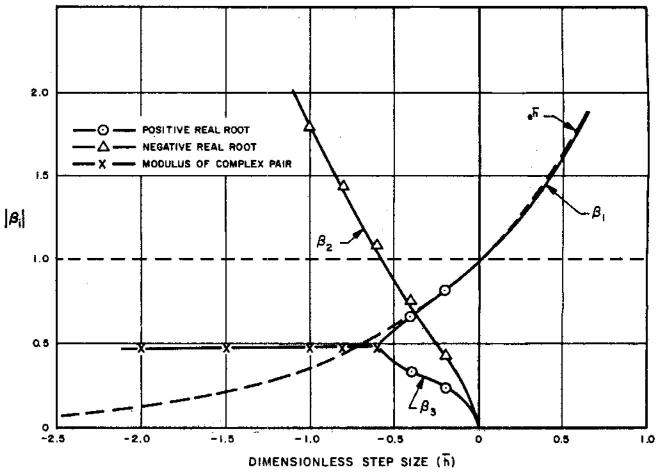

Fig. 3. Third-order Adams-Moulton predictor.

to which the extraneous solutions affect the overall solution depends upon the values of the constants $\bar{c}_1$, $\bar{c}_2$, and $\bar{c}_3$, as well as the value of n. However, as can be seen from Equation (30) the term $\bar{c}_2\beta_2^n$ will eventually dominate the solution as n gets large and $\bar{c}_2$ is finite. The three constants, $\bar{c}_i$, are evaluated from three previously stored history points at t_n, t_{n-1} and t_{n-2}.

Implicit Method: Adams-Moulton Third-Order Corrector

If one uses the third-order Adams-Moulton corrector formula,

$$x_{n+1} = x_n + \frac{h}{12}\,(5x'_{n+1} + 8x'_n - x'_{n-1}) \qquad (31)$$

to approximate Equation (9), Equation (31) becomes

$$\left(1 - \frac{5}{12}\bar{h}\right)x_{n+1} - \left(1 + \frac{2}{3}\bar{h}\right)x_n + \left(\frac{\bar{h}}{12}\right)x_{n-1} = 0 \qquad (32)$$

Equation (32) is a second-order difference equation which will yield one extraneous solution. One is concerned with the rate of growth of this extraneous solution as the increment size gets large. The third-order Adams-Moulton predictor has been shown to contain two extraneous solutions; therefore, the predictor-corrector pair will also contain two extraneous solutions. The solution of Equation (32) is of the form,

$$x_n = \bar{c}_1\beta_1^n + \bar{c}_2\beta_2^n \qquad (33)$$

where

$$\beta_{1,2} = \frac{(12 + 8\bar{h}) \pm \sqrt{144 + 144\bar{h} + 84\bar{h}^2}}{2(12 - 5\bar{h})} \qquad (34)$$

Figure 4 shows that β_1 is the true root and β_2 is an extraneous root. At $(-\bar{h}) = 6.0$, $|\beta_2|$ is greater than one, and in Equation (33) the term containing β_2^n will overwhelm the true solution as n gets large. Thus, the method is unstable for $(-\bar{h}) > 6$.

The region to the right of the point $\bar{h} = -1.5$ in Figure 4 is the region of relative stability, and the region to the right of point $\bar{h} = -6.0$ is the region of absolute stability. The region between two points is, therefore, a stable region, that is, the asymptotic behavior of the approximate solution will behave like that of the true solution. However, in many cases, a large error may exist because the extraneous root dominates the approximate solution for large n. This area has received relatively little attention in the technical literature, and is therefore an important area for further investigation.

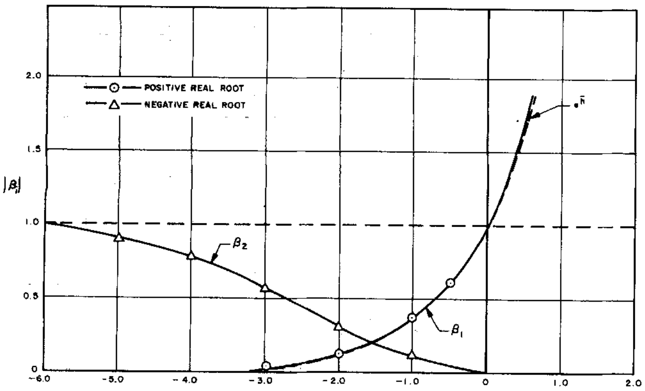

Fig. 4. Third-order Adams-Moulton corrector.

PREDICTOR-CORRECTOR METHODS

Methods of Application

Before discussing any specific technique, it is necessary to discuss the manner in which a predictor-corrector method is applied. In essence, the open-end predictor equation is used to extrapolate from the point x_n to the point x_{n+1}. By using the predicted x_{n+1}, a closed-end corrector equation is then applied to interpolate for an improved value of x_{n+1}. A rigorous truncation error analysis can then follow to modify the corrected value or both the predicted and corrected values. Predictor-corrector methods applied in this manner require but two functional (derivative) evaluations per increment, independent of the order of the method. Predictor-corrector methods can also be applied with several repeated corrector applications at the end of the interval, in which case the number of functional evaluations per increment is directly proportional to the number of such corrector iterations.

Letting P represent a predicted value, C a corrected value, M_p modified predicted value, and M_c a modified corrected value, some typical predictor-corrector schemes might be:

1. P
2. P-C
3. P-C-M_c
4. P-M_p-C-M_c
5. P-C-C (denoted P-$2C$)
6. P-C-M_c-C-M_c
7. P-M_p-C-M_c-C-M_c

*8. P-nC

†9. P-∞C

‡10. C

For purpose of illustration, the fourth-order Adams-Moulton equations applied in the manner of the fourth scheme is shown below:

Predict:

$$x_{n+1}^{(P)} = x_n + \frac{h}{24}\,(55x'_n - 59x'_{n+1} + 37x'_{n-1} - 9x'_{n-2}) \qquad (35)$$

Modify:

$$x_{n+1}^{(M_p)} = x_{n+1}^{(P)} + \frac{251}{270}\,(x_n^{(C)} - x_n^{(P)}) \qquad (36)$$

Correct:

$$x_{n+1}^{(C)} = x_n + \frac{h}{24}\,(9x'^{(M_p)}_{n+1} + 19x'_n - 5x'_{n-1} + x'_{n-2}) \qquad (37)$$

Modify:

$$x_{n+1}^{(M_c)} = x_{n+1}^{(C)} - \frac{19}{270}\,(x_{n+1}^{(C)} - x_{n+1}^{(P)}) \qquad (38)$$

In a rigorous sense, Equation (36) should contain the difference between the corrected and predicted value at the $(n + 1)$ step, as in Equation (38). Since the corrected value of the dependent variable, $x_{n+1}^{(C)}$, will not be available at this stage, it is recommended that one use the difference between the corrected and predicted value at the previous increment. Applied thusly, an iterative procedure is avoided.

* Denotes one predictor application followed by a fixed number "n" of corrector iterations.

† Means that the corrector is iterated until it converges.

‡ Denotes the stability of the corrector equation alone (implicit solution).

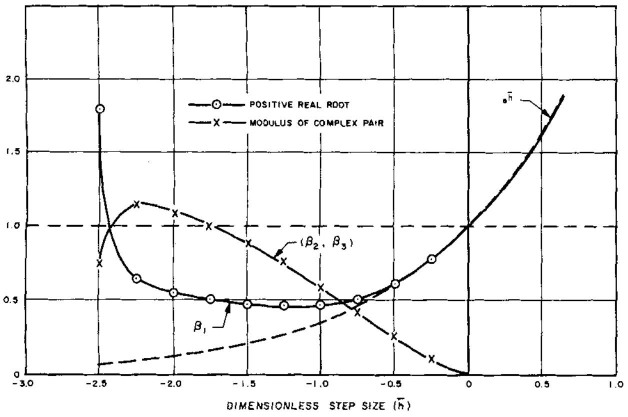

Fig. 5. Third-order Adams-Moulton predictor-corrector.

Adams-Moulton Third-Order Predictor-Corrector Pair

The value of x at t_{n+1} calculated from the Adams-Moulton third-order predictor is given as $x_{n+1}{}^{(P)}$ by Equation (27). The derivative of x at t_{n+1} is therefore,

$$x'^{(P)}_{n+1} = Ax^{(P)}_{n+1} = Ax_n + \frac{\bar h}{12}(23Ax_n - 16Ax_{n-1} + 5Ax_{n-2}) \quad (39)$$

A corrected value of x at time, t_{n+1}, is obtained by using the corrector equation in the following fashion:

$$x_{n+1} = x_n + \frac{h}{12}(5x'^{(P)}_{n+1} + 8x'_n - x'_{n-1}) \quad (40)$$

By substituting Equation (39) into Equation (40) and rearranging results in the difference equation approximation, we obtain

$$x_{n+1} - \left(1 + \frac{13}{12}\bar h + \frac{115}{144}\bar h^2\right)x_n + \left(\frac{\bar h}{12} + \frac{80}{144}\bar h^2\right)x_{n-1} - \left(\frac{25}{144}\bar h^2\right)x_{n-2} = 0 \quad (41)$$

Equation (41) is a third-order difference equation, the three roots of which are shown in Figure 5. Equation (41) has a pair of conjugate complex roots throughout the entire $-\bar h$ region, thus resulting in only two curves in Figure 5 for the three roots. It can be seen that the predictor-corrector pair remains stable out to $(-\bar h) < 1.8$.

Since the primary concern in this paper is the value of $(-\bar h)$ at which any of the roots gets larger than one in absolute value (and not in a critical study of the behavior of the roots), the following graphs will present only the absolute value of the largest root as a function of $(-\bar h)$. For purpose of illustration, the third-order Adams-Moulton equations discussed previously and displayed in Figures 3, 4, and 5 are summarized in the new format in Figure 6.

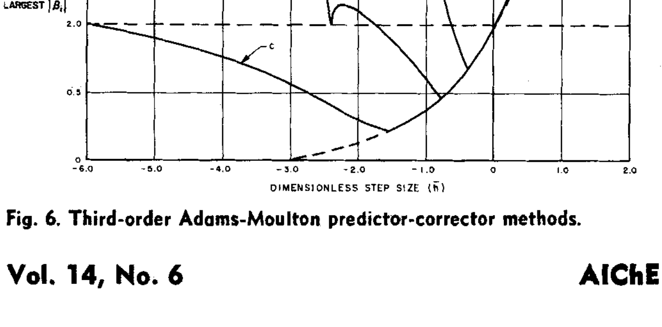

Fig. 6. Third-order Adams-Moulton predictor-corrector methods.

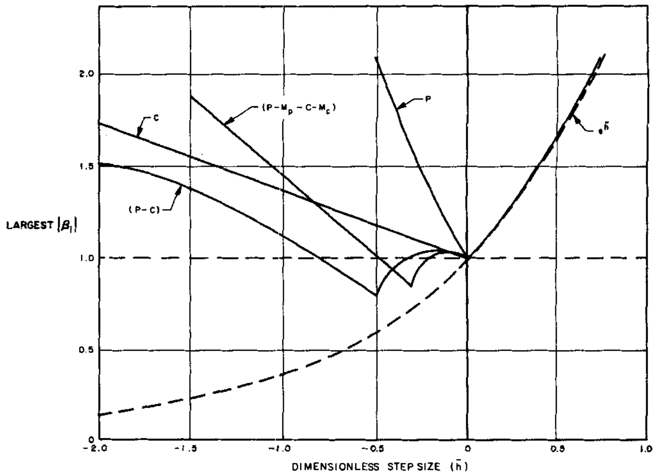

Fig. 7. Fourth-order Milne predictor-corrector methods.

Behavior of Other Predictor-Corrector Methods

Figure 7 illustrates schemes one, two, four and ten for the fourth-order Milne equations (7). It can be seen that for negative $\bar h$ there is no region of relative stability, and only a small region of absolute stability. The most stable technique results from scheme two, *P-C*, and the least stable scheme ten, *C*. From this plot, one can see that the Milne method is not very useful for differential equation possessing the property

$$\frac{\partial f}{\partial x} < 0 \quad (42)$$

that is, in the region of negative $\bar h$.

Figure 8 illustrates schemes one, two, four, and ten for the fourth-order Milne-Hamming predictor-corrector equations (7), (8). For this technique scheme ten, *C*, is the most stable, and scheme two, *P-C*, is the least stable. The most practical approach is that of scheme four which is stable out to $(-\bar h) < 0.8$.

Figure 9 displays the results of schemes one, two,[§] five, six, seven and ten for the fourth-order Adams-Moulton equations (9). From this plot the following conclusions can be drawn:

1. scheme ten is the most stable,

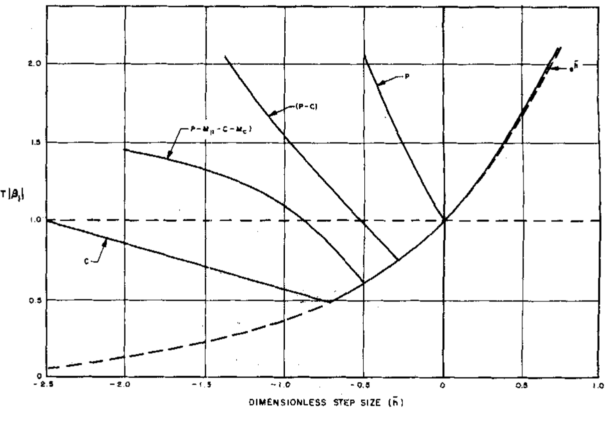

Fig. 8. Fourth-order Milne-Hamming predictor-corrector methods.

[§] The figure for scheme two, *P-C*, given elsewhere (9) is incorrect. This has been corrected in this paper.

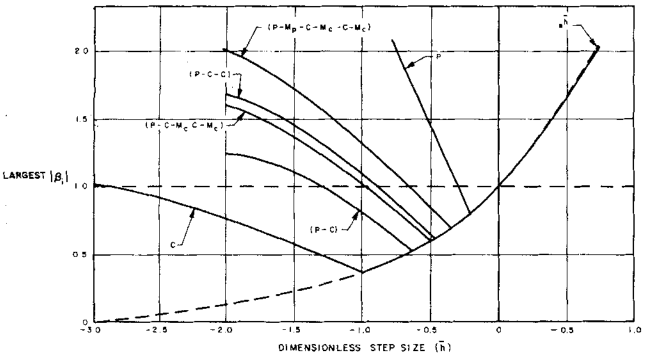

Fig. 9. Fourth-order Adams-Moulton predictor-corrector methods.

2. scheme two is the most stable practical method,
3. repeated corrector applications lowers the stability limit, and
4. modification of the predicted value lowers the stability limits considerably.

By using the simple *P-C* scheme, the method remains stable out to $(-\bar{h}) < 1.3$. The third-order Adams-Moulton equations have a stability limit of $(-\bar{h}) < 1.8$, whereas for the fifth-order method, the stability limit drops to approximately $(-\bar{h}) < 0.9$.

The fourth-order Milne predictor was used in conjunction with the fourth-order Fehlberg (*10*) corrector to generate the results shown in Figure 10 for the Milne-Fehlberg equations. It can be seen that scheme two, *P-C*, is the most practical means of applying these equations. Also of interest is the fact that (as in the Adams-Moulton equations) the second iteration of the corrector lowers the maximum allowable step size.

The behavior of the stability characteristics of the various methods presented in this paper are considerably different (see Table 1). For example, the characteristics of the Adams-Moulton method for the various schemes discussed above are almost directly opposite to that of the Milne-Hamming methods. The characteristics of the Adams-Moulton and the Milne-Fehlberg methods on the other hand are very similar in nature. In general it is practically impossible to predict in advance which of the above schemes of predictor-corrector applications is the most practical technique to apply unless one first investigates the roots of the difference equation for the particular differential equation in question. It can be concluded, however, that the use of the corrector equation in some fashion will usually improve the stability range over that of the predictor alone.

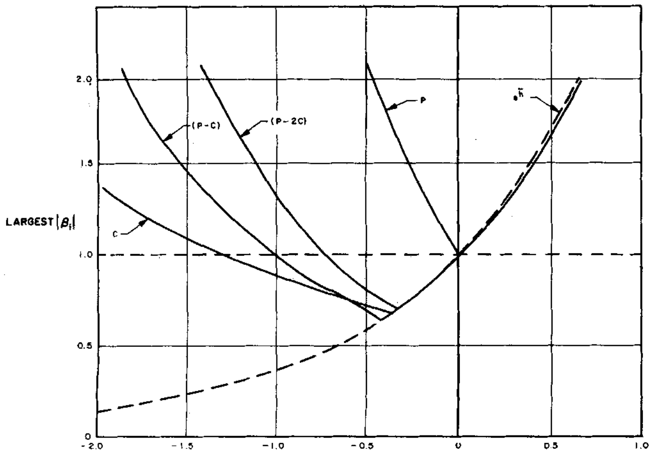

Fig. 10. Fourth-order Milne-Fehlberg predictor-corrector methods.

TABLE 1. STABILITY LIMITS OF VARIOUS NUMERICAL INTEGRATION TECHNIQUES

	Limiting Increment Size $(-\bar{h})_{(\max)}$ (dimensionless)	
Numerical Routine	Theoretical	Empirical
1. forward Euler	2.0	—
2. modified Euler		
a) *P-C*	2.0	1.1
b) *P-3C*	2.0	1.3
c) *P-6C*	2.0	1.5
d) *P-10C*	2.0	1.6
e) *P-∞C**	2.0	—
f) *C*†	unlimited	—
3. backward Euler		
a) *P-C*	1.0	0.7
b) *P-2C*	1.4	0.8
c) *P-3C*	1.0	0.7
d) *P-4C*	1.2	—
e) *P-∞C**	1.0	—
f) *C*†	unlimited	—
4. Runge-Kutta		
a) second-order	2.0	1.1
b) third-order	2.5	2.0 unstable‡
c) fourth-order	2.7	2.5
5. third-order Adams-Moulton		
a) *P*	0.6	—
b) *P-C*	1.8	1.5
c) *P-C-M$_c$*	2.0	1.5
d) *P-M$_p$-C-M$_c$*	1.6	1.1
e) *P-2C*	1.2	1.0
f) *P-3C*	—	1.3
g) *P-6C*	—	1.4
h) *P-10C*	—	1.6
i) *P-∞C**	2.4	—
j) *C*†	6.0	—
6. fourth-order Adams-Moulton		
a) *P*	0.3	—
b) *P-C*	1.3	—
c) *P-M$_p$-C-M$_c$-C-M$_c$*	0.6	—
d) *P-C-M$_c$-C-M$_c$*	1.0	—
e) *P-2C*	0.9	—
f) *C*†	2.9	—
7. fourth-order Milne		
a) *P*	0.0	—
b) *P-C*	0.8	—
c) *P-M$_p$-C-M$_c$*	0.5	—
d) *P-2C*	0.0	—
e) *P-3C*	0.0	—
f) *C*†	0.0	—
8. fourth-order Milne-Hamming		
a) *P*	0.0	—
b) *P-C*	0.5	—
c) *P-M$_p$-C-M$_c$*	0.8	—
d) *P-2C*	0.9	—
e) *C*†	2.7	—
9. fourth-order Milne-Fehlberg		
a) *P*	0.0	—
b) *P-C*	1.0	—
c) *P-2C*	0.7	—
d) *C*†	1.3	—
10. fourth-order Adams-Fehlberg		
a) *P*	0.3	—
b) *P-C*	—	1.5 unstable‡
c) *P-10C*	—	1.5 unstable‡
d) *C*†	1.3	—

* Means that the corrector is iterated until it converges.

† Denotes the stability limit of the closed-end corrector equation alone (implicit solution).

‡ Only one value was tried and at this value the method was unstable. No attempt was made to find the correct limiting increment size.

$$\tilde{A} = \begin{bmatrix} -\left(\frac{L_0 + D}{H_0}\right) & \frac{V_1K_1}{H_0} & \cdots & \cdots & \cdots & 0 \\ \frac{L_0}{H_1} & -\left(\frac{L_1 + K_1V_1}{H_1}\right) & \frac{V_2K_2}{H_1} & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \frac{L_{n-1}}{H_n} & -\left(\frac{L_n + K_nV_n}{H_n}\right) & \frac{V_{n+1}K_{n+1}}{H_n} & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \frac{L_N}{H_{N+1}} & -\left(\frac{L_{N+1} + K_{N+1}V_{N+1}}{H_{N+1}}\right) \end{bmatrix} \qquad (46)$$

One note concerning repeated applications of the corrector equation should be interjected here. With regard to scheme nine, which is one predictor application and repeated applications of the corrector until it converges, in general, one cannot estimate how many iterations of the corrector are necessary for convergence, or even if convergence will result from repeated applications of the corrector. In addition, the net effect of repeated corrector iterations in some cases is to reduce the allowable step size, because of the nature of the convergence of the iterative scheme. For these reasons, this scheme is rarely used in practice.

APPLICATION TO PROCESS DYNAMICS

Process dynamic studies by their very nature require the solution of large systems of simultaneous nonlinear differential equations. Needless to say, one must often resort to numerical solution on a digital computer. The following analysis shows how the results presented in this paper can be used to obtain an approximate estimate of the maximum integration step size allowable from a stability standpoint.

For purpose of illustration, the case of unsteady state distillation is treated in detail. Transient distillation was chosen because it is representative of all stage-by-stage processes, and the tridiagonal form of the differential equations is amenable to rigorous mathematical analysis. The linearized set of differential equations for dynamic distillation is given (for each component) by

$$\frac{d}{dt}\vec{x} = \tilde{A}\vec{x} + \vec{g} \qquad (43)$$

where

$$\vec{x} = \begin{bmatrix} x_o \\ x_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_{N+1} \end{bmatrix} \qquad (44)$$

$$\vec{g} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ Fx_F \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \end{bmatrix} \qquad (45)$$

From the form of Equations (43) through (46) it can be seen that external inputs (feeds) do not alter the form of the $\tilde{A}$ matrix. A liquid drawoff would show up in a corresponding diagonal term, whereas a vapor drawoff would show up in one of the first superdiagonal terms of the $\tilde{A}$ matrix.

Following a stability analysis, one is interested in obtaining a bound on the largest negative eigenvalue of the $\tilde{A}$ matrix for use in Equation (16). This in turn will yield a bound on the largest allowable integration step size, $h_{\max}$, from

$$h_{\max} \leq (-\bar{h})_{\max}/|\lambda|_{\max} \qquad (47)$$

where $|\lambda|_{\max}$ is the absolute value of the largest negative eigenvalue of the $\tilde{A}$ matrix.

By using the Gerschgorin Theorem (*11*) on Equation (46) it can be easily shown that the greatest upper bound on $|\lambda|_{\max}$ is given by

$$|\lambda|_{\max} \leq \left(\frac{L_n + K_nV_n}{H_n}\right) + \left(\frac{L_{n-1}}{H_n}\right) + \left(\frac{K_nV_n}{H_n}\right)$$

$$\cong 2\left(\frac{L_n + K_nV_n}{H_n}\right)_{\max} \qquad (48)$$

that is, the upper limit on absolute value of the largest negative eigenvalue is approximately two times the largest diagonal element of the $\tilde{A}$ matrix. Typical values of $(L + KV)/H$ factors for distillation columns range between (100 to 1,000) hr.$^{-1}$. Since Equation (48) represents the greatest upper bound on $|\lambda|_{\max}$, use of this expression in Equation (47) will yield the minimum value of maximum allowable step size, $h_{\max}$.

A more exact relationship for $|\lambda|_{\max}$ can be obtained due to the tridiagonal nature of the $\tilde{A}$ matrix. In order to obtain an analytical expression for the relationship between $|\lambda|_{\max}$ and $(L_n + K_nV_n)/H_n$ it is assumed that L_n, V_n, K_n, and H_n are approximately constant (denoted by L, V, K, and H) throughout the entire column. In this case, the $\tilde{A}$ matrix assumes the form,

$$\tilde{A} = \begin{bmatrix} -b & c & 0 & - & - & - & - & 0 \\ a & -b & c & - & - & - & - & 0 \\ 0 & a & -b & c & - & - & - & 0 \\ & \ddots & \ddots & \ddots & \ddots & & \ddots & \\ 0 & - & - & - & - & - & a & -b \end{bmatrix} \qquad (49)$$

where approximately

$$b \cong a + c \qquad (50)$$

The largest negative eigenvalue of the $\tilde{A}$ matrix is then the largest negative root of the polynomial generated by expanding the characteristic equation,

$$P^M(\lambda) = \det [\tilde{A} - \lambda I]^M = 0 \qquad (51)$$

where M is the order of the $\tilde{A}$ matrix.

An analytical expression for the eigenvalues of the above Jacobian matrix is given by Amundson (*12*) as

$$\lambda_k = -b - \sqrt{4ac} \cos \left(\frac{\pi k}{M+1} \right); \quad k = 1, 2, 3, \text{- - -} M \qquad (52)$$

where M is the order of the $\tilde{A}$ matrix.

In terms of distillation parameters, the absolute value of the largest negative eigenvalue [as obtained from Equation (52) with $k = 1$] is given by

$$|\lambda|_{\max} = \left(\frac{L + KV}{H} \right) \left[1 + \left(\frac{2\sqrt{\alpha}}{\alpha + 1} \right) \cos \left(\frac{\pi}{N+3} \right) \right] \qquad (53)$$

where $\alpha = KV/L$ and N = number of plates. For large N, $|\lambda|_{\max}$ varies between $(L + KV)/H$ and $2(L + KV)/H$ for $\alpha = 0$ and $\alpha = 1$, respectively, and again approaches $(L + KV)/H$ as $\alpha \rightarrow \infty$.

Equation (53) used in conjunction with Equation (47) provides a means of estimating the maximum allowable step size for the integration of transient distillation equations. The resulting expression is

$$h_{\max} = \frac{(-\bar{h})_{\max} \left(\dfrac{L + KV}{H} \right)}{\left(\dfrac{L + KV}{H} \right) \left[1 + \left(\dfrac{2\sqrt{\alpha}}{\alpha + 1} \right) \cos \left(\dfrac{\pi}{N+3} \right) \right]} \qquad (54)$$

where $\alpha = KV/L$ and N = number of plates.

Table 1 presents both theoretical and empirical values for the limiting increment size, $(-\bar{h})_{\max}$, for the various techniques discussed in this work. It can be seen that, for an equivalent number of derivative evaluations, the third-order Adams-Moulton equations possess the most favorable stability characteristics, remaining stable out to $(-\bar{h})_{\max} \leqq 1.5$ with only two derivative evaluations per increment. By contrast, the fourth-order Runge-Kutta remains stable to $(-\bar{h})_{\max} \leqq 2.5$, but requires four derivative evaluations per increment.

Table 2 presents a comparison between the maximum increment size calculated using theoretical values of $(-\bar{h})_{\max}$ given in Table 1 in conjunction with Equation (54) and that found empirically by several investigators. From this table, one can see the remarkable agreement between the actual limiting increment size and that predicted by the simple relationship given in Equation (54).

Table 2. Comparison Between Calculated and Actual Maximum Step Sizes

Investigators	Technique	Scheme	$(L + KV)/H$ Factor (hr.$^{-1}$)	KV/L Ratio	N No. of Plates	$(h)_{\max}$* calc. (hr.)	$(h)_{\max}$ Actual (hr.)
1. Distefano, et al. (*21*) numbers represent average values over all runs.	Adams-Moulton (third-order)	P-C-M_c	525	1.2	12	0.0019	0.0020
2. Frank and Lapidus (*22*) binary system.	Runge-Kutta (fourth-order)	—	110	1.2	5	0.0130	0.0167
3. Frank and Lapidus (*22*) multicomponent system.	Runge-Kutta (fourth-order)	—	170	7.5	5	0.0098	0.0061
4. Holland (*13*, p. 116)	Runge-Kutta (fourth-order)	—	390	1.1	3	0.0037	0.0041
5. Distefano (*23*) run no. 4.	Adams-Moulton (third-order)	P-C-M_c	250	1.8	12	0.0041	0.0030
6. Distefano (*23*) run no. 4.	backward Euler	P-3C	250	1.8	12	0.0020	0.0015
7. Gamer (*24*) numbers represent average values.	Runge-Kutta-Gill (fourth-order)	—	1200	3.0	8	0.0012	0.0013
8. Huckaba and Danly (*25*) case reported.	modified Euler	P-∞C	310	2.3	12	0.0035	0.0040

* Calculated from Equation (54) using the theoretical value of $(-\bar{h})_{\max}$ from Table 1.

CONCLUSIONS AND RECOMMENDATIONS

The stability characteristics of several of the most widely used numerical integration routines were investigated theoretically. Although the stability limits varied from one technique to another, the variation was small compared to what is really needed to speed up digital calculations significantly.

Use of the stability results presented in this paper in the estimation of limiting integration step sizes was illustrated for the case of transient distillation calculations. A comparison was made between the maximum increment size predicted from Equation (54) and that found by computer trial and error by several integrators. In all cases, the actual results agreed quite closely with the predicted results. Since transient distillation is representative of all stage-by-stage processes, the procedures presented here for distillation can be easily extended to many other processes having similar equations.

From the results presented in Tables 1 and 2, it can be seen that even the most stable numerical integration techniques are not entirely satisfactory for the solution of problems in process dynamics. In most cases, one is restricted to using an extremely small integration step size, thus resulting in excessive computation time even on large present day digital computers. There exists a definite need for techniques that will allow for larger step sizes (from stability considerations) without significant loss in accuracy.

For process dynamics calculations on a digital computer, the following areas are recommended for future research:

1. Use of implicit methods (correctors) in the solution of nonlinear equations (*13*, *14*),

2. Use of methods which do not depend upon polynomial fits to past history points (*15* to *17*), and

3. Use of methods which were developed for systems that possess characteristics similar to those of physical systems, such as highly damped systems (*18*, *19*), and coupled systems with greatly different time constants (*20*).

ACKNOWLEDGMENT

The author wishes to thank the Monsanto Company, St. Louis, Missouri for permission to publish this work, and Electronic Associates, Inc., Princeton, New Jersey for providing the time for the author to transform isolated ideas into analytical treatment suitable for publication.

NOTATION

A = defined by Equation (3) and used in Equations (9), (10) and (13)

$\tilde{A}$ = denotes a Jacobian matrix, a matrix of elements $a_{ij} = \partial f_i / \partial x_j$

a = constant given by $a = L/H$

b = constant given by $b = (L + KV)/H$

c = constant given by $c = KV/H$

c_i = coefficients in the homogeneous solution of a differential equation

$\overline{c_i}$ = coefficients in the homogeneous solution of a difference equation

C = denotes a value from a corrector equation

$e^{\alpha_i t}$ = denotes the complementary solutions to a differential equation, (where e is the base of the natural logarithm)

$f(x, t)$ = arbitrary functional representation of a single nonlinear differential equation

$\vec{f}(\vec{x}, t)$ = arbitrary functional representation of a system of nonlinear differential equations

$\vec{F}$ = vector containing the terms $\partial f_i / \partial t$

$G(t)$ = denotes the particular solution to a differential equation

$\overline{G}(t_n)$ = denotes the particular solution to a difference equation

h = a small increment in the independent variable used as the step size in numerical integration

$\overline{h}$ = a dimensionless step size in the independent variable [defined in Equation (13) for a single equation and in Equation (16) for a system of equations]

H_n = liquid holdup on plate n of a distillation column (constant holdup denoted by H)

I = unit matrix

K_n = equilibrium value for a component on plate n of a distillation column (constant K_n denoted by K)

L_n = liquid flow rate from plate n of a distillation column (constant flow rate denoted by L)

m = order of a difference equation

M_c = modification to the value from a corrector equation

M = order of a Jacobian matrix

M_p = denotes a modification to the value from a predictor equation

n = nth increment in a step-by-step numerical solution, or the nth plate of a distillation column (use is clear from usage in the text)

N = total number of plates in a distillation column

$N+1$ = reboiler of a distillation column

q = the order of a differential equation

P = value from a predictor equation

$P^M(\lambda)$ = Mth order polynomial generated by expanding the characteristic determinant of a Jacobian matrix

t = independent variable

V_n = vapor flow rate from plate n in a distillation column (constant flow rate, V)

x = a single dependent variable

x' = the derivative of x with respect to t

$\vec{x}$ = a dependent variable vector

Greek Letters

α = KV/L ratio

β_i^n = complementary solutions to a difference equation (or the roots of the characteristic polynomial)

λ = eigenvalue of a Jacobian matrix

$|\lambda|_{\max}$ = absolute value of the largest negative eigenvalue of a Jacobian matrix

LITERATURE CITED

1. Fox, L., *"Numerical Solution of Ordinary and Partial Differential Equations,"* Addison-Wesley, Mass. (1962).
2. Hildebrand, F. B., *"Introduction to Numerical Analysis,"* McGraw-Hill, New York (1956).
3. Crane, R. L., and R. W. Klopfenstein, *J. Assoc. Computing Machinery*, **12**, 227 (1965).
4. Ceschino, F., and J. Kuntzmann, *"Numerical Solution of Initial Value Problems,"* Prentice-Hall, Englewood Cliffs, N.J. (1963).
5. Hildebrand, F. B., *"Methods of Applied Mathematics,"* Prentice-Hall, Englewood Cliffs, N.J. (1961).
6. Prater, S. V., *Nat. Aeron. Space Admin. Tech. Rep. No. N67-14407* (1967).
7. Chase, P. E., *J. Assoc. Computing Machinery*, **9**, 457 (1962).
8. Hamming, R. W., *ibid.*, **6**, 37 (1959).
9. Brown, R. R., J. D. Riley, and M. M. Bennett, *Math. Computation*, **1**, 90 (1965).
10. Fehlberg, E., *Nat. Aeron. Space Admin. Tech. Note D-599*, (March, 1959).
11. Varga, R. S., *"Matrix Iterative Analysis,"* Princeton-Hall, Englewood Cliffs, N.J. (1962).
12. Amundson, N. R., *"Mathematical Methods in Chemical Engineering,"* Prentice-Hall, Englewood Cliffs, N.J. (1966).
13. Waggoner, R. C., and C. D. Holland, *AIChE J.*, **11**, 113 (1965).
14. Holland, C. D., *"Unsteady State Processes with Applications in Multicomponent Distillation,"* Prentice-Hall, Englewood Cliffs, N.J. (1966).
15. Delves, L. M., *Math. Computation*, **20**, No. 94, 246 (April, 1966).
16. Brock, P., and F. J. Murray, *Math. Tables Aids Computation*, **6**, 63 (1952).
17. *Ibid*, 138 (1952).
18. Ralston, A., and H. S. Wilf, eds., *"Mathematical Methods for Digital Computers,"* p. 128, John Wiley, New York, (1960).
19. Richards, P. I., W. D. Lanning, and M. D. Torrey, *SIAM Review*, **7**, No. 3, 376 (July, 1965).
20. Treanor, C. E., *Math. Computation*, **20**, No. 93, 39 (January, 1966).
21. Distefano, G. P., F. P. May, and C. E. Huckaba, *AIChE J.*, **13**, 125 (1967).
22. Frank, A., and L. Lapidus, *Chem. Eng. Progr.*, **60**, No. 4, 61 (1964).
23. Distefano, G. P., *AIChE J.*, **14**, 190 (1968).
24. Gamer, J. D., MS thesis, U.S. Naval Post-graduate School, (1966).
25. Huckaba, C. E., and D. E. Danly, *AIChE J.*, **6**, 335 (1960).

Manuscript received August 22, 1967; revision received December 1, 1967; paper accepted December 4, 1967.