1962

Stability and local accuracy of numerical methods for ordinary differential equations

Roger Lyle Crane
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Mathematics Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/2123

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
This dissertation has been microfilmed exactly as received

CRANE, Roger Lyle, 1933—
STABILITY AND LOCAL ACCURACY OF
NUMERICAL METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS.

Iowa State University of Science and Technology
Ph.D., 1962
Mathematics
University Microfilms, Inc., Ann Arbor, Michigan
STABILITY AND LOCAL ACCURACY OF NUMERICAL METHODS
FOR ORDINARY DIFFERENTIAL EQUATIONS

by

Roger Lyle Crane

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Applied Mathematics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1962
TABLE OF CONTENTS

CHAPTER I:  INTRODUCTION 1

CHAPTER II:  LINEAR MULTISTEP METHODS 6

CHAPTER III:  STABILITY OF MODIFIED PREDICTOR - CORRECTOR METHODS 23

CHAPTER IV:  LOCAL ACCURACY OF MODIFIED PREDICTOR - CORRECTOR METHODS 39

CHAPTER V:  CONVERGENCE OF MODIFIED PREDICTOR - CORRECTOR METHODS 52

CHAPTER VI:  STRONGLY STABLE PREDICTOR - CORRECTOR METHODS 64

BIBLIOGRAPHY 77

ACKNOWLEDGMENTS 78
CHAPTER I: INTRODUCTION

Many numerical techniques for the approximate solution of ordinary differential equations have been developed. Quite often, the solution of the equation $y' = f(x,y)$ is approximated by use of a linear difference equation.

To illustrate two types of such difference equations, consider the equation

$$y_{n+k} = \sum_{j=0}^{k-1} a_j y_{n+j} + h \sum_{j=0}^{k} b_j y'_{n+j},$$

where $x_{n+j} = x_n + jh$, $y_{n+j} = y(x_{n+j})$, $y'_{n+j} = f(x_{n+j}, y_{n+j})$, and the $a_j$ and $b_j$ are constants chosen so that $y_{n+k}$ is a good approximation of the solution of the differential equation. Note that if $b_k = 0$, equation (1.1) provides an explicit method for computing $y_{n+k}$ when the values of $y_{n+j}$ for $j = 0, 1, \cdots, k - 1$ are known. This is true since the corresponding values $y'_{n+j}$ can be computed from $y'_{n+j} = f(x_{n+j}, y_{n+j})$. Equation (1.1), with $b_k = 0$, is sometimes called an open formula and is often called a predictor formula. The latter terminology results because, in some sense, one is able to predict the value $y_{n+k}$ when one knows the $y$ values at the previous points. When $b_k \neq 0$, equation (1.1) provides an implicit relation for the determination of $y_{n+k}$, since $y_{n+k}$ appears in the term $y'_{n+k} = f(x_{n+k}, y_{n+k})$. Such a formula is called a closed formula, or, in many cases, a corrector formula. The term
corrector arises since it is often true that the presence of the term \( f(x_{n+k}, y_{n+k}) \) results in a more accurate determination of the value \( y_{n+k} \). In general, it is necessary to use an iterative procedure in order to solve (1.1) when \( b_k \neq 0 \). This iterative procedure can be illustrated by rewriting (1.1) in the form

\[
y^{(i+1)}_{n+k} = \sum_{j=0}^{k-1} (a_j y_{n+j} + h b_j y'_{n+j}) + h b_k f(x_{n+k}, y^{(i)}_{n+k}).
\]

Henrici [5, p. 216] shows that if \( f(x,y) \) is continuous and satisfies a Lipschitz condition with respect to \( y \), equation (1.1) has a unique solution and, further, the iterative process defined by (1.2) converges to this unique solution independent of the choice of \( y^{(0)}_{n+k} \), provided that \( h \) is sufficiently small.

One commonly recommended technique for solving the equation \( y' = f(x, y) \) is to use a predictor formula to generate the value \( y^{(0)}_{n+k} \) and, subsequently, to use a corrector formula iteratively. In practice, of course, the iterative process is not continued until convergence is attained, but until the difference between two successive iterates is smaller than some prescribed tolerance. Such a procedure will be called an iterative predictor-corrector method in order to distinguish it from other tech-
niques which will be described below.

Each evaluation of $y_{n+k}^{(i+1)}$ in equation (1.2) takes additional computation time. Hence, it is common practice to use the predictor formula to determine $y_{n+k}^{(0)}$ and, subsequently, to use the corrector formula only once. In this case, the value $y_{n+k}^{(1)}$ so obtained is accepted as an approximation of $y_{n+k}$ and is used in succeeding computations. Such a procedure will be called simply a predictor-corrector method. It is described by the equations

$$p_{n+k} = \sum_{j=0}^{k-1} (a_j y_{n+j} + h b_j y'_{n+j}),$$

[(1.3.1)]

$$y_{n+k} = c_{n+k} = \sum_{j=0}^{k-1} (c_j y_{n+j} + h d_j y'_{n+j}) + h d_k f(x_{n+k}, p_{n+k}).$$

[(1.3.2)]

Here, the notation indicates that the coefficients of $y_{n+j}$ and $y'_{n+j}$ in the predictor do not necessarily coincide with those in the corrector. Also, the definitions $p_{n+k} \equiv y_{n+k}^{(0)}$ and $c_{n+k} \equiv y_{n+k}^{(1)} = y_{n+k}$ are used.

In order to motivate the consideration of a third algorithm, it is helpful to discuss the concept of accuracy of the difference equation (1.1). This concept will be considered in detail in later chapters. For the present, however, it is sufficient to note the following. Assume that the differential equation $y' = f(x, y)$ which is to be solved has a unique solution. In general, the value
$y_{n+k}$ obtained from (1.1) will not be the same as that of the unique solution even though the values $y_{n+j}$ for $j = 0, 1, \cdots, k-1$ are exact. This difference between $y_{n+k}$ and the true value of $y$ is called the local truncation error. It is, of course, desirable to reduce this error. That is the prime reason for consideration of the method which is defined by the equations

\begin{align*}
(1.4.1) & \quad p_{n+k} = \sum_{j=0}^{k-1} (a_j y_{n+j} + h b_j y_{n+j}'), \\
(1.4.2) & \quad m_{n+k} = p_{n+k} + K(p_{n+k-1} - c_{n+k-1}), \\
(1.4.3) & \quad c_{n+k} = \sum_{j=0}^{k-1} (c_j y_{n+j} + h d_j y_{n+j}') \\
& \quad \quad \quad + h d_k f(x_{n+k}, m_{n+k}), \text{ and} \\
(1.4.4) & \quad y_{n+k} = c_{n+k} + L(p_{n+k} - c_{n+k}).
\end{align*}

The quantities $p_{n+k-1}$ and $c_{n+k-1}$ are those obtained by use of the predictor and corrector, respectively, at the preceding point. $K$ and $L$ are constants which may be chosen to increase the accuracy of $y_{n+k}$.

The procedure defined by equations (1.4) will be called a modified predictor-corrector method. In particular, if $K \neq 0$ and $L = 0$, equations (1.4) define a $P$-modified predictor-corrector method. If $K = 0$ and $L \neq 0$, they define a $C$-modified predictor-corrector method, and if $K \neq 0$ and $L \neq 0$, they define a $PC$-modified predictor-corrector method. Note that when $K = L = 0$, the method
reduces to a predictor-corrector method as defined previously; since, in this case, \( M_{n+k} = P_{n+k} \) and \( y_{n+k} = C_{n+k} \). The specific manner in which the choice of \( K \) and \( L \) affects the accuracy of \( y_{n+k} \) is investigated in detail in Chapter IV.

Three numerical methods, the iterative predictor-corrector, the predictor-corrector, and the modified predictor-corrector, have been introduced in the order of their complexity. The behavior of the iterative method can be determined by analysis of the corrector equation since the convergence of the iterates does not depend upon the predicted value. However, the predictor cannot be ignored when the predictor-corrector method is studied. The modified method is even more complicated since the predictor, the corrector, and the constants \( K \) and \( L \) all affect the value \( y_{n+k} \). This study is primarily concerned with results pertaining to the latter two methods. In order to place these results in perspective, however, it is necessary to present some of the theory related to equations of the form (1.1). This work is, of course, applicable to the iterative method since the corrector equation is of form (1.1). The theory also encompasses predictor equations, since (1.1) with \( b_k = 0 \) is a predictor formula.
CHAPTER II: LINEAR MULTISTEP METHODS

Consider the initial value problem
\[ y' = f(x,y), \quad y(a) = \eta. \]
From the theory of ordinary differential equations, it is well-known that if

i) \( f \) is continuous for \( a \leq x \leq b \), and \( -\infty < y < \infty \), and

ii) \( f \) satisfies a Lipschitz condition with respect to \( y \), then, for any real \( \eta \), problem (2.1) has a unique solution on the interval \([a,b]\). Equation (1.1) defines a numerical method which determines a sequence \( \{y_n\} \) which can be used to approximate the solution of (2.1) at the points \( x_n, \ n = 0,1,2,\ldots \), when \( y_0, y_1, \ldots, y_{k-1} \) are initially specified. Such a method is commonly called a linear multistep method when \( k > 1 \) and a linear one-step method when \( k = 1 \). The term linear is used since the values of \( y \) and \( y' \) enter linearly in (1.1).

Introduction of the concept of convergence of (1.1) to the solution of (2.1) leads to some important theoretical results which are due primarily to Dahlquist [2], [3], and Henrici [5]. The definition of convergence used here is essentially that of Henrici. It has been modified slightly in order that it be applicable for the work done in the following chapters. In order to make a specific definition, convergence of the method based on the use of (1.1) is defined below. However, a similar definition can
be made for any numerical method, e.g. the modified predictor-corrector method, which determines a sequence \( \{y_n\} \) that approximates the solution of (2.1).

Consider the set \( F \) of all functions \( f \) which satisfy conditions i) and ii) above. Let \( G \) denote any subset of \( F \). For any \( g \in G \), let \( y_g(x) \) denote the unique solution of problem (2.1).

Definition 2.1: The numerical method defined by (1.1) is said to be convergent over \( G \) if and only if, for every \( g \in G \) and for every set of starting values \( y_q = y(a + qh) \) satisfying \( \lim_{h \to 0} y_q = y(a) = \eta \), for \( q = 0, 1, \ldots, k-1 \), it determines a sequence \( \{y_n\} \) such that

\[
\lim_{n \to \infty} y_n = \overline{y}_g(x_0) \text{ for all } x_0 \text{ in } [a, b].
\]

An intuitive interpretation of the definition is helpful. Let \( x_0 \) be chosen in \( (a, b] \), and divide the interval \( [a, x_0] \) into \( n \) equal parts, where \( n > k-1 \). Then choose \( h = \frac{x_0 - a}{n} \).

By solving (1.1) a sufficient number of times, actually \( n-k+1 \) times since the first \( k \) values of \( y \) must be initially specified, one obtains an approximation \( y_n \) of the true value \( \overline{y}_g(x_0) \). The definition above insures that, for a convergent method, by choosing \( n \) sufficiently large, the approximate value can be made to be arbitrarily close to the true value. Practical limitations are apparent since
it is impossible to allow \( h \) to become arbitrarily small
when one maintains only a finite number of digits in com-
puting \( y_n \). However, the theoretical results which follow
provide a sound basis for determining practical numerical
methods.

Two conditions must be imposed upon equation (1.1) in
order that it define a numerical method that converges over
the set \( F \) mentioned above. They are the condition of con-
sistency and the condition of zero stability. Their ef-
flect can be understood qualitatively by considering two
types of errors which occur in the use of (1.1). The
first, the local truncation error, has been mentioned
above and occurs since, in general, the value \( y_{n+k} \) obtain-
ed from (1.1) would not be exact even if all the previous
values, \( y_{n+j} \) for \( j = 0, 1, \ldots, k-1 \), were exact. This type
of error is introduced each time (1.1) is used in the
process of solving (2.1). The condition of consistency is
introduced to control this error and insures that (1.1) is
locally accurate when \( h \) is sufficiently small. Once an
error is present at a particular step in the solution of
(2.1), it has an effect on succeeding steps. This second
type of error, the propagated error, can result in a very
poor approximation of the solution of (2.1) even though
the local errors are small. This will be demonstrated by
a numerical example below. The condition of zero stabili-
ty is introduced to control the propagated error.

Before introducing the condition of consistency, it is convenient to define, for equation (1.1), a concept of order. Assume, for the present, that \( y(x) \) possesses a termwise differentiable Taylor series expansion about the point \( x_n \). Expand each term of (1.1) about \( x_n \) by use of the relations

\[
(2.2.1) \quad y_{n+j} = \sum_{m=0}^{\infty} \frac{(jh)^m y^{(m)}(x_n)}{m!}, \text{ and}
\]

\[
(2.2.2) \quad y'_{n+j} = \sum_{m=0}^{\infty} \frac{(jh)^m y^{(m+1)}(x_n)}{m!}.
\]

Then, multiply both sides of the resulting equation by \( m! \) and equate like terms, \( h^m y^{(m)}(x_n) \). The conditions

\[
(2.3.1) \quad \sum_{j=0}^{k-1} a_j = 1 \quad \text{for } m = 0, \text{ and}
\]

\[
(2.3.2) \quad \sum_{j=0}^{k-1} (a_j j^m + b_j j^{m-1}) + \frac{m! b_j k^{m-1}}{m!} = k^m
\]

for \( m = 1, 2, \ldots \) result. In (2.3.2), the definition \( 0^0 = 1 \) is assumed. This definition is also assumed in later formulas without further mention.

Definition 2.2: The linear difference equation (1.1) and the numerical method which it defines are said to be of order \( p \) if and only if conditions (2.3) hold for \( m \neq 0, 1, \ldots, p \), but not for \( m = p+1 \).
Definition 2.3: (Condition of consistency) The numerical method defined by (1.1) is said to be consistent if and only if it is at least of order one. —

It is easy to find numerical methods of the form (1.1) which, although consistent, are very impractical. Consider, for example, the formula

\[ y_{n+2} = -2y_{n+1} + 3y_n + \frac{h}{6} (y'_{n+2} + 16y'_{n+1} + 7y'_n) \]  

This method is consistent as can be seen by verifying conditions (2.3) for \( m = 0 \) and \( m = 1 \). In fact, conditions (2.3) are also satisfied for \( m = 2 \) and \( m = 3 \), but not for \( m = 4 \). Hence, (2.4) is of order three. Consider the result when (2.4) is used to solve the trivial initial value problem

\[ y' = 0, \quad y(0) = 0. \]  

In this case, (2.4) reduces to

\[ y_{n+2} + 2y_{n+1} - 3y_n = 0, \]

with the initial condition \( y(0) = 0 \). When \( y(h) = y_1 \) is specified as a second condition, which is necessary to solve (2.6), the solution

\[ y_n = \frac{y_1}{4} [1 - (-3)^n] \]

is obtained. Thus, if \( y_1 = 0 \), the true solution of (2.5) results. However, if \( y_1 \) is not identically zero, but only approximately so, the solution grows exponentially as \( n \)
increases. This example, although trivial, illustrates the phenomenon of numerical instability. The difficulty lies in the numerical method itself, not in the example which was used to illustrate the problem. In particular, the trouble is caused by the root $\rho = -3$ of the characteristic equation $\rho^2 + 2\rho - 3 = 0$ which is associated with the difference equation (2.6). An error introduced at any step in the solution increases exponentially in succeeding steps because of this root. Contrast this situation with that which occurs when

$$ (2.8) \quad y_{n+2} = \frac{1}{2} y_{n+1} + \frac{1}{2} y_n + \frac{h}{8} \left( 3y'_{n+2} + 8y'_{n+1} + y'_n \right) $$

is used to solve (2.5). Equation (2.8) is also of order three. However, in this case, the solution of (2.5) can be written as

$$ (2.9) \quad y_n = \frac{2}{3} y_1 \left[ 1 - \left( -\frac{1}{2} \right)^n \right]. $$

Here, it is apparent that the exponential error growth is not present. Study of these examples suggests that the propagated error, and hence the stability of the method defined by (1.1), depends upon the roots of the equation

$$ (2.10) \quad \rho^k - \sum_{j=0}^{k-1} a_j \rho^j = 0, $$

which is the characteristic equation associated with (1.1) when $h = 0$. This is indeed the case, and it motivates the following definition.
Definition 2.4: (Condition of zero stability) The numerical method defined by (1.1) is said to be zero stable if and only if every root of equation (2.10) has modulus less than or equal to one, and each root of modulus one is simple.

The term zero is used to distinguish this type of stability from others considered later, and also to direct attention to the fact that the characteristic equation involved here is that which is associated with (1.1) when \( h = 0 \). Motivation for requiring that the roots of modulus one be simple can be obtained by considering the effect of a double root of plus one, for instance. With such a double root, the solution of (1.1) would, in general, contain a term proportional to \( n \). Hence, any error introduced would tend to grow linearly in succeeding steps.

Another trivial example illustrates that zero stability alone is not sufficient to insure convergence. Consider the use of the trivial zero stable formula

\[ y_{n+1} = y_n \]

for the solution of the initial value problem

\[ y' = 1, \quad y(0) = 0. \]

The true solution is \( y(x) = x \), but the solution given by the difference equation is \( y(x) = 0 \). Note that the equation \( y_{n+1} = y_n \) is not consistent.

The importance of the conditions of consistency and
zero stability is evident from the following theorem, the proof of which can be found in Henrici [5]. Recall that the set of functions $F$ was defined earlier in this chapter.

**Theorem 2.1:** The numerical method defined by (1.1) is convergent over $F$ if and only if it is consistent and zero stable.

Although zero stability is an important concept and leads to significant results, it does not explain all numerical stability problems which occur with methods defined by equation (1.1). The following example, which uses a zero stable method, illustrates a second type of numerical instability. Consider the use of Simpson's rule,

\begin{equation}
Y_{n+2} = Y_n + \frac{h}{3} \left( Y_{n+2} ' + 4Y'_{n+1} + Y_n ' \right),
\end{equation}

for the solution of the initial value problem

\begin{equation}
y' = -y, \quad y(0) = 1.
\end{equation}

The true solution of (2.12) is $y(x) = e^{-x}$. Equation (2.11), when applied to problem (2.12), reduces to

\begin{equation}
Y_{n+2} (1 + \frac{h}{3}) + \frac{4h}{3} Y_{n+1} + Y_n(-1 + \frac{h}{3}) = 0,
\end{equation}

with characteristic equation

\begin{equation}
\rho^2 (1 + \frac{h}{3}) + \frac{4h}{3} \rho + (-1 + \frac{h}{3}) = 0.
\end{equation}

When $h = 0.1$, the roots, correct to six significant figures, are found to be $\rho_1 = 0.904837$ and $\rho_2 = -1.033870$. 
Thus, the solution of (2.13) for $h = 0.1$ is

\begin{equation}
(2.15) \quad y_n = A(0.904837)^n + B(-1.033870)^n,
\end{equation}

where $A$ and $B$ are determined from the conditions $y(0) = 1$ and $y(0.1) = y_1$. Hence,

\[
A = \frac{1.033870 + y_1}{1.938707}, \quad \text{and} \quad B = \frac{0.904837 - y_1}{1.938707}.
\]

Since $e^{-0.1} = 0.904837$ when six significant figures are retained, it is apparent that the part of the solution which involves $\rho_1$ approximates the true solution and that the part of the solution due to $\rho_2$ is extraneous. Thus, if $B \neq 0$, the extraneous solution ultimately dominates and the solution of (2.12) obtained by use of (2.11), rather than approaching zero, grows without bound. In practice, it is to be expected that $B \neq 0$.

Hence, equation (2.11), although zero stable, cannot be successfully used to approximate the solution of (2.12) when $h = 0.1$. It is easily seen that a similar phenomenon occurs for any positive value of $h$ near zero. For, when the roots of (2.14) are approximated as functions of $h$, it is found that, as $h \to 0$, $\rho_1 \approx 1 - h + \frac{h^2}{2} \approx e^{-h}$, and $\rho_2 \approx -1 - \frac{h}{3} - \frac{h^2}{18}$. Hence, even for arbitrarily small positive $h$, $|\rho_2| > 1$. The following theory is developed in order to characterize methods with properties like those just illustrated. It also provides additional insight into the general problem of numerical stability.
Again, assume that (1.1) is to be used to approximate the solution of problem (2.1). Let \( \overline{y}(x) \) denote the unique solution of (2.1), and write

\[
(2.16) \quad \overline{y}_{n+k} = y_{n+k} + T_n, \text{ where}
\]

\[
(2.17) \quad y_{n+k} = \sum_{j=0}^{k-1} a_j \overline{y}_{n+j} + h \sum_{j=0}^{k} b_j \overline{y}_{n+j},
\]

\( T_n \) is the local truncation error, and \( \overline{y}'_{n+j} = f(x_{n+j}, \overline{y}_{n+j}) \).

The term \( \overline{y}'_{n+k} \) in equation (2.17) is correct for the non-iterative uses of (1.1) considered in later chapters. If (1.1) is used iteratively, \( \overline{y}'_{n+k} \) should be replaced by \( y'_{n+k} \). This replacement only affects terms of \( O(h^{p+2}) \).

Hence, the following results are valid for either case.

In view of the work done in the following chapters, it is helpful to determine the form of \( T_n \) when \( \overline{y}(x) \in C^{p+2} \) and equation (1.1) is of order \( p \). With these assumptions, it follows that

\[
(2.18) \quad \overline{y}_{n+k} = \sum_{m=0}^{p+1} \frac{k^m}{m!} h^m \overline{y}^{(m)}(x_n) + O(h^{p+2}), \text{ and}
\]

\[
(2.19) \quad y_{n+k} = \sum_{m=0}^{p+1} \frac{k!}{m!} h^m \overline{y}^{(m)}(x_n)
\]

\[
+ \sum_{j=0}^{k-1} a_j \sum_{m=0}^{p+1} \frac{j^m}{m!} h^m \overline{y}^{(m)}(x_{n+j})
\]

\[
+ \sum_{j=0}^{k} b_j \sum_{m=0}^{p+1} \frac{j^m}{m!} h^m \overline{y}^{(m+1)}(x_{n+j}),
\]

\( + O(h^{p+2}) \).

Thus, from (2.16), (2.18), and (2.19), it follows that
\[ T_n = \overline{y}_{n+k} - y_{n+k} \]

\[
= (1 - \sum_{j=0}^{k-1} a_j) \overline{y}(x_n) + \sum_{m=1}^{p+1} \left\{ \frac{x^m}{m!} - \sum_{j=0}^{k-1} \frac{a_j^m}{m!} \frac{b_j^m (m-1)!}{(m-1)!} \right\}
\]

\[- \frac{b_k^{m-1}}{(m-1)!} \right\} h^m \overline{y}^{(m)}(x_n) + O(h^{p+2}).\]

Since equation (1.1) is of order \( p \), it follows from conditions (2.3) that

\[
(2.20) \quad T_n = \left\{ \frac{k^p+1}{(p+1)!} - \sum_{j=0}^{k-1} \left[ \frac{a_j^{p+1}}{(p+1)!} + \frac{b_j^p}{p!} \right] \right\}
\]

\[- \frac{b_k^p}{p!} \right\} h^{p+1} \overline{y}^{(p+1)}(x_n) + O(h^{p+2}).\]

Let \( C_p \) be defined by the coefficient of \( h^{p+1} \overline{y}^{(p+1)}(x_n) \) above to get

\[
(2.21) \quad T_n = C_p h^{p+1} \overline{y}^{(p+1)}(x_n) + O(h^{p+2}).\]

Hence, from (2.16), (2.17), and (2.21), it follows that

\[
(2.22) \quad \overline{y}_{n+k} = \sum_{j=0}^{k-1} a_j \overline{y}_{n+j} + h \sum_{j=0}^{k} b_j \overline{y}'_{n+j}
\]

\[ + C_p h^{p+1} \overline{y}^{(p+1)}(x_n) + O(h^{p+2}).\]

Subtract equation (2.22) from equation (1.1) and make use of the relations \( y_{n+j}' = f(x_{n+j}, y_{n+j}) \) and \( \overline{y}_{n+j}' = f(x_{n+j}, \overline{y}_{n+j}) \), to obtain
\begin{align}
(2.23) \quad y_{n+k} - \overline{y}_{n+k} &= \sum_{j=0}^{k-1} a_j (y_{n+j} - \overline{y}_{n+j}) \\
&+ h \sum_{j=0}^{k} b_j \left[ f(x_{n+j}, y_{n+j}) - f(x_{n+j}, \overline{y}_{n+j}) \right] \\
&- C_p h^{p+1} \overline{y}^{(p+1)}(x_n) + O(h^{p+2}).
\end{align}

Now, let \( y_{n+j} - \overline{y}_{n+j} = \epsilon_{n+j} \) and define the quantities \( g_{n+j} \) by the relation

\[ g_{n+j} = \begin{cases} 
\frac{f(x_{n+j}, y_{n+j}) - f(x_{n+j}, \overline{y}_{n+j})}{\epsilon_{n+j}} & \text{if } \epsilon_{n+j} \neq 0 \\
0 & \text{if } \epsilon_{n+j} = 0.
\end{cases} \]

With these definitions, equation (2.23) becomes

\begin{align}
(2.24) \quad \epsilon_{n+k} - \sum_{j=0}^{k-1} a_j \epsilon_{n+j} &= h \sum_{j=0}^{k} b_j g_{n+j} \epsilon_{n+j} \\
&- C_p h^{p+1} \overline{y}^{(p+1)}(x_n) + O(h^{p+2}).
\end{align}

Since, by definition, \( \epsilon_{n+j} \) is the difference between the approximate value, \( y_{n+j} \), given by (1.1) and the true value of the solution of (2.1) at the point \( x_{n+j} \), equation (2.24) is a linear difference equation for the error which occurs when (1.1) is used to solve (2.1). Recall that in order to derive equation (2.24), it was assumed that the true solution of (2.1) had continuous derivatives of order \( p+2 \) and that (1.1) was of order \( p \). Before proceed-
ing, a further restriction on the class of problems under consideration will be made.

In the following, it will be assumed that the function \( f \) of (2.1) is such that its partial derivative with respect to \( y \) exists, and further, is equal to a constant, \( A \). The apparent severity of this restriction requires comment. The purpose of making this assumption is to allow the precise development of the mathematical results which follow. It is recognized that particular numerical methods of the general type studied here will be used to solve the more general problem (2.1) in which \( f_y \) is not constant. While the theory developed for the restricted case is not rigorously applicable to the general case, it does provide considerable insight as to the stability problems which arise. For instance, when the function \( f \) of (2.1) is such that \( f_y \) is approximately constant over the region in the \( xy \) plane at which (1.1) is evaluated, one would expect that the local behavior of the solution of (2.1) might be approximated by considering a corresponding problem of the form \( y' = Ay + g(x) \). Further, the assumption that \( f_y \) is essentially constant over a region of the \( xy \) plane is not unreasonable if the region is sufficiently restricted. Since the size of the region can be restricted by choosing \( h \) sufficiently small, the restriction that \( f_y \) be constant is not as severe as it would
first appear.

When \( f_y = A \), it follows from the mean value theorem and from the definition of \( g_{n+j} \) which precedes equation (2.24), that \( g_{n+j} = A \) for \( j = 0,1,\ldots,k \). Hence, (2.24) becomes, with the definition that \( \Delta h = h \),

\[
\varepsilon_{n+k} = \sum_{j=0}^{k-1} a_j \varepsilon_{n+j} + \sum_{j=0}^{k} b_j \varepsilon_{n+j} = h \sum_{j=0}^{k} b_j \varepsilon_{n+j} - C_p h^{p+1} \bar{y}^{(p+1)}(x_n) + O(h^{p+2}).
\]

The characteristic equation associated with (2.25) can be written as

\[
(1 - b_k h) \rho^k - \sum_{j=0}^{k-1} (a_j + b_j h) \rho^j = 0.
\]

It is now possible to generalize the definition of zero stability given earlier.

**Definition 2.5:** The numerical method defined by (1.1) is said to be \( h \)-stable if and only if every root of equation (2.26) has modulus less than or equal to one, and each root of modulus one is simple.

Note that when \( h = 0 \), the definition coincides with Definition 2.4.

In the example cited above, Simpson's rule was applied to an initial value problem for which \( f_y = A = -1 \).
Recall that as $h \to 0$, one root of equation (2.14) is of the form $\rho_1 \approx 1 + Ah + \frac{(Ah)^2}{2}$, where $A = -1$. Hence $\rho_1 \approx e^h$. It is apparent that for arbitrarily small positive $h$, $|\rho_1| > 1$. Thus, there does not exist any interval $[0,h^*]$ of the positive $h$ axis such that Simpson's rule is $h$ stable for $0 < h < h^*$. This is more than a coincidence. In fact, it is known that, for any consistent method, there exists a root $\rho_1 \approx e^h$ as $h \to 0$ when the problem $y' = Ay$ is considered [5, p. 237]. The same characteristic equation, namely (2.26), is obtained for the general problem in which $f_y$ is equal to the constant $A$ as for the particular problem $y' = Ay$. Thus, for the general case in which $f_y$ is constant, a consistent numerical method defined by (1.1) cannot be $h$ stable for arbitrarily small positive $h$. No similar argument applies, in general, to the negative $h$ axis. There exist formulas of the form (1.1) which are $h$ stable for all $h$ contained in some closed interval $[h^*,0]$. As seen from the example above, however, Simpson's rule does not have this property. There is no non-trivial interval containing zero throughout which Simpson's rule is $h$ stable. The following definitions are based on the preceding discussion.

Definition 2.6: The numerical method defined by (1.1) is said to be strongly stable if and only if there exists a
negative $h^*$ such that the method is $h$ stable for every $h$ satisfying $h^* \leq h \leq 0$.

Definition 2.7: The numerical method defined by (1.1) is said to be weakly stable if and only if it is zero stable and not strongly stable.

The terms strongly stable and weakly stable or weakly unstable appear in the literature. See, for instance, [2], [3], and [5]. Definitions given in these works are closely related to those given above. However, one aspect in which they differ is the following. In the definition above, a method is strongly or weakly stable depending upon its behavior over some interval containing zero. Since a definition of stability for non-zero $h$ is not used in the works cited above, the definitions of strong and weak stability given there are based on the properties, at $h = 0$, of the method defined by (1.1). With the definition of $h$ stability introduced above, it is natural to characterize strong and weak stability as interval properties. This lends unity to the presentation here and emphasizes the importance of $h$ stability.

In the following three chapters, the concepts of convergence, order, consistency, and stability are extended to cover the predictor-corrector and modified predictor-corrector methods previously introduced. Stability
is studied in Chapter III and results pertaining to local accuracy, i.e. order and consistency, are presented in Chapter IV. The relation of stability and accuracy to convergence is discussed in Chapter V.
CHAPTER III: STABILITY OF MODIFIED PREDICTOR - CORRECTOR METHODS

The first goal of this chapter is the derivation of the difference equation satisfied by the error that results when the modified predictor-corrector method defined by equations (1.4) is used to approximate the solution of the initial value problem (2.1). This equation is analogous to equation (2.24) which resulted when equation (1.1) was used to obtain an approximate solution of (2.1).

As in the derivation of equation (2.24), let \( \tilde{y}(x) \) denote the unique solution of (2.1). Assume that equation (1.4.1), the predictor, is of order \( p \), and that equation (1.4.3), the corrector, is of order \( q \). For the present, assume that \( \tilde{y}(x) \) has continuous derivatives of order \( r \), where \( r = \max(p+2, q+2) \), and define \( \tilde{f}_{n+j} \equiv f(x_{n+j}, \tilde{y}_{n+j}) \). Then, in correspondence with equations (1.4), write the equations

\[
(3.1.1) \quad \overline{P}_{n+k} = \sum_{j=0}^{k-1} (a_j \tilde{y}_{n+j} + h b_j \tilde{f}_{n+j}) + c_{(P)}^{(p)} h^{p+1} \overline{y} (p+1) (x_n) + o(h^{p+2}),
\]

\[
(3.1.2) \quad \overline{M}_{n+k} = \overline{P}_{n+k} + K (\overline{P}_{n+k-1} - \overline{C}_{n+k-1}),
\]
(3.1.3) \[ \bar{c}^{n+k} = \sum_{j=0}^{k-1} (c_j \bar{y}^{n+j} + h d_j \bar{f}^{n+j}) + h d_k f(x^{n+k}, M^{n+k}) \]
\[ + c_q h^{q+1} \bar{y}(q+1)(x^n) + O(h^{q+2}), \text{ and} \]

(3.1.4) \[ \bar{y}^{n+k} = \bar{c}^{n+k} + L(\bar{p}^{n+k} - \bar{c}^{n+k}). \]

The error terms in (3.1.1) and (3.1.3) result from considerations similar to those preceding equation (2.22).

In equations (1.4), use the definition that \( f_{n+j} = f(x_{n+j}, y_{n+j}) = y'_{n+j} \). Now, subtract from each member of equations (1.4) the corresponding member of equations (3.1) to get

(3.2.1) \[ p^{n+k} - \bar{p}^{n+k} = \sum_{j=0}^{k-1} a_j (y_{n+j} - \bar{y}_{n+j}) + h b_j (f_{n+j} - \bar{f}_{n+j}) \]
\[ - c^{(P)} h^{p+1} \bar{y}(p+1)(x^n) + O(h^{p+2}). \]

(3.2.2) \[ m^{n+k} - \bar{m}^{n+k} = (p^{n+k} - \bar{p}^{n+k}) \]
\[ + k[(p^{n+k-1} - \bar{p}^{n+k-1}) - (c^{n+k} - \bar{c}^{n+k})], \]

(3.2.3) \[ c^{n+k} - \bar{c}^{n+k} = \sum_{j=0}^{k-1} c_j (y_{n+j} - \bar{y}_{n+j}) \]
\[ + h d_j (f_{n+j} - \bar{f}_{n+j}) + h d_k [f(x^{n+k}, M^{n+k}) \]
\[ - f(x^{n+k}, \bar{M}^{n+k})] - c_q h^{q+1} \bar{y}(q+1)(x^n) + O(h^{q+2}), \text{ and} \]
(3.2.4) \[ Y_{n+k} - Y\hat{y}_{n+k} = (C_{n+k} - C\hat{C}_{n+k}) \]

\[ + L[(P_{n+k} - P\hat{P}_{n+k}) - (C_{n+k} - C\hat{C}_{n+k})]. \]

In equations (3.2), make the definitions

\[ \epsilon^{(p)}_{n+k} \equiv P_{n+k} - P\hat{P}_{n+k}, \]
\[ \epsilon^{(c)}_{n+k} \equiv C_{n+k} - C\hat{C}_{n+k}, \]
\[ \epsilon^{(m)}_{n+k} \equiv M_{n+k} - M\hat{M}_{n+k}, \]
\[ \epsilon_{n+j} = Y_{n+j} - Y\hat{y}_{n+j}. \]

\[ g^{(M)}_{n+k} \equiv \begin{cases} \frac{f(x_{n+k}, M_{n+k}) - f(x_{n+k}, M\hat{M}_{n+k})}{\epsilon^{(M)}_{n+k}} & \text{if } \epsilon^{(M)}_{n+k} \neq 0 \\ 0 & \text{if } \epsilon^{(M)}_{n+k} = 0. \end{cases} \]

and

\[ g_{n+j} \equiv \begin{cases} \frac{f_{n+j} - f\hat{f}_{n+j}}{\epsilon_{n+j}} & \text{if } \epsilon_{n+j} \neq 0 \\ 0 & \text{if } \epsilon_{n+j} = 0. \end{cases} \]

With these definitions, equations (3.2) become

(3.3.1) \[ \epsilon^{(p)}_{n+k} = \sum_{j=0}^{k-1} (a_j + h b_j g_{n+j}) \epsilon_{n+j} \]
\[ - C^{(p)}_p h^{p+1} \nabla^{(p+1)} x_n + O(h^{p+2}), \]

(3.3.2) \[ \epsilon^{(M)}_{n+k} = \epsilon^{(p)}_{n+k} + K(\epsilon^{(p)}_{n+k-1} - \epsilon^{(c)}_{n+k-1}), \]
(3.3.3) \[ \epsilon_{n+k} = \sum_{j=0}^{k-1} (c_j + h d_j g_{n+j}) \epsilon_{n+j} + h d_{x} g_{n+k}^{(M)} \epsilon_{n+k}^{(M)} - C_q (c^{(C)} h^{q+1} y^{(q+1)}(x_n) + 0(h^{q+2}), \text{ and} \]

(3.3.4) \[ \epsilon_{n+k} = \epsilon_{n+k}^{(C)} + L(\epsilon_{n+k}^{(P)} - \epsilon_{n+k}^{(C)}). \]

Equations (3.3) are simultaneous difference equations in the four unknowns \( \epsilon_{n+k}, \epsilon_{n+k}^{(P)}, \epsilon_{n+k}^{(M)}, \text{ and } \epsilon_{n+k}^{(C)}. \) It is possible, by straightforward manipulation, to eliminate the unknowns \( \epsilon_{n+k}^{(P)}, \epsilon_{n+k}^{(M)}, \text{ and } \epsilon_{n+k}^{(C)}, \) and thus to obtain a linear difference equation involving only the errors \( \epsilon_{n+j}, \text{ for } j = -1,0,\cdots,k. \) A brief description of this lengthy manipulation follows. It is assumed that \( L \neq 1. \) This is essentially no restriction since, if \( L = 1, \) equation (1.4.4) becomes \( y_{n+k} = p_{n+k}, \) and the effect of the corrector is ignored. This could be accomplished more easily by simply using the linear multistep method defined by (1.4.1).

It is possible to obtain an expression for \( \epsilon_{n+k-1}^{(P)} \) in terms of \( \epsilon_{n+j} \) by replacing \( n \) by \( n - 1 \) in equation (3.3.1). Note that, because of the assumption that \( \bar{y}(x) \) has continuous derivatives of order \( r, \) it follows that \( \bar{y}^{(p+1)}(x_{n-1}) = \bar{y}^{(p+1)}(x_n) + 0(h). \) An expression for \( \epsilon_{n+k-1}^{(C)} \) in terms of \( \epsilon_{n+j} \) results from solving (3.3.4) for
\[ \epsilon_{n+k}^{(C)}, \text{ replacing } n \text{ by } n-1, \text{ and using the expression for } \\
\epsilon_{n+k-1}^{(P)} \text{ mentioned above. Then, } \epsilon_{n+k}^{(M)} \text{ is expressed in terms} \\
of \epsilon_{n+j} \text{ through equation (3.3.2) since } \epsilon_{n+k}, \epsilon_{n+k-1}, \text{ and} \\
\epsilon_{n+k-1}^{(C)} \text{ are all expressible in terms of } \epsilon_{n+j}. \text{ By use of} \\
this expression for } \epsilon_{n+k}^{(M)}, \text{ equation (3.3.3) thus determines} \\
\epsilon_{n+k}^{(C)} \text{ in terms of } \epsilon_{n+j}. \text{ The desired difference equation is} \\
then obtained from (3.3.4) by use of this expression for} \\
\epsilon_{n+k}^{(C)}, \text{ since } \epsilon_{n+k}^{(P)} \text{ is known from equation (3.3.1). After} \\
considerable rearrangement, this difference equation can} \\
be written as} \\

\begin{equation}
(3.4) \quad \epsilon_{n+k}^{(C)} = \sum_{j=0}^{k-1} \left[ L a_j + (1 - L) c_j \right] \epsilon_{n+j} = h \epsilon_{n+k-1}^{(M)} \left[ L b_{k-1}^{(M)} \\
+ (1 - L) a_{k-1}^{(M)} g_{n+k-1} + d_{k}^{(M)} \left[ (1 - L) (a_{k-1}^{(M)} \\
+ h b_{k-1}^{(M)} g_{n+k-1} - K) \right] \right] + h \sum_{j=0}^{k-2} \epsilon_{n+j}^{(M)} \left[ L b_{j}^{(M)} \\
+ (1 - L) d_{j}^{(M)} g_{n+j} + d_{k}^{(M)} \left[ (1 - L) (a_{j}^{(M)} + h b_{j}^{(M)} g_{n+j} \\
+ K (a_{j+1}^{(M)} + h b_{j+1}^{(M)} g_{n+j}) \right) \right] + h \epsilon_{n-1}^{(M)} \left[ d_{k}^{(M)} g_{n+k-1}^{(M)} (a_{0}^{(M)} \\
+ h b_{0}^{(M)} g_{n-1}) \right] + L \left[ -c_{P}^{(M)} h^{p+1} y^{(p+1)} (x_{n}) \right] \\
+ 0 (h^{p+2}) + (1 - L) \left[ -c_{q}^{(C)} h^{q+1} y^{(q+1)} (x_{n}) \right] \\
+ 0 (h^{q+2}) + (1 - L + K) d_{k}^{(M)} g_{n+k}^{(M)} \left[ -c_{P}^{(M)} h^{p+2} y^{(p+1)} (x_{n}) \right] \\
+ 0 (h^{p+3})].
\end{equation}

For convenience, \( y \) has been replaced by \( y \) in (3.4).
Equation (3.4) is, under the assumptions listed above, the difference equation that is satisfied by the error which occurs when the modified predictor-corrector method defined by (1.4) is used to approximate the solution of (2.1). As indicated in Chapter I, variations of the modified method can be obtained by setting either, or both, of the constants $K$ and $L$ equal to zero. The difference equation satisfied by the error in these methods, the predictor-corrector, the $P$-modified predictor-corrector, and the $C$-modified predictor-corrector can be obtained by making the appropriate choice of $K$ and $L$ in (3.4). This result is easily verified by deriving the appropriate difference equation directly in each particular case.

In Chapter V, equation (3.4) is used to study the convergence of the modified predictor-corrector method defined by equations (1.4). For the study of stability, however, it is appropriate to consider the particular form of (3.4) which results when the function $f$ of problem (2.1) is such that $f_y$ is equal to a constant, $A$. As discussed in Chapter II, it follows that $g_{n+j} = A$ for $j = -1, 0, \ldots, k$, and, also, $g^{(M)}_{n+k} = A$. Then, by use of the definition $Ah = h$, equation (3.4) becomes

$$
(3.5) \quad \epsilon_{n+k} \quad \sum_{j=0}^{k-1} [La_j + (1 - L)c_j] \epsilon_{n+j} = h \epsilon_{n+k-1} [Lb_{k-1} + (1 - L) \delta_{k-1} + \delta_k (1 - L) (a_{k-1} + hb_{k-1} - K)]
$$
\[
\sum_{j=0}^{k-2} h \varepsilon_{n+j} \left\{ Lb_j + (1 - L) d_j + d_k \left[ (1 - L) (a_j + h b_j) \right] + K (a_{j+1} + h b_{j+1}) \right\} + \varepsilon_{n-1} [d_k K (a_0 + h b_0)]
\]

\[
- L C_p^{(P)} h^{P+1} y^{(P+1)} (x_n) - (1 - L) C_q^{(C)} h^{q+1} y^{(q+1)} (x_n)
\]

\[
- (1 - L + K) h d_k C_p^{(P)} h^{P+1} y^{(P+1)} (x_n) + O (h^{p+2}).
\]

After some routine manipulation, the characteristic equation associated with the difference equation (3.5) can be written as

\[
(3.6) \quad \rho^{k+1} - \rho^k \left\{ L a_{k-1} + (1 - L) c_{k-1} + h [L b_{k-1} - K d_k] \right. \\
\left. \quad + (1 - L) (d_{k-1} + d_k a_{k-1}) \right\} + h^2 (1 - L) (d_k b_{k-1})
\]

\[
- \sum_{j=0}^{k-2} \rho^{j+1} \left\{ L a_j + (1 - L) c_j + h [L b_j + K d_k a_{j+1} + (1 - L) b_j + K b_{j+1}] \right\}
\]

\[
- h K d_k (a_0 + h b_0) = 0.
\]

This characteristic equation for the modified predictor-corrector method is analogous to equation (2.26) which was developed for the linear multistep method (1.1).

Equation (3.6) is considerably more complicated than equation (2.26) since two linear multistep formulas, (1.4.1) and (1.4.3) are involved, as well as the constants K and L. Also, (3.6) is one degree higher than (2.26).

Note, however, that when a predictor-corrector or a C-
modified predictor-corrector method is considered, one root of (3.6) is identically zero since \( K = 0 \).

The concepts of h stability, strong stability, and weak stability can now be extended to the modified predictor-corrector method.

**Definition 3.1:** The numerical method defined by equations (1.4) is said to be h stable if and only if every root of equation (3.6) has modulus less than or equal to one, and each root of modulus one is simple.

**Definition 3.2:** The numerical method defined by equations (1.4) is said to be strongly stable if and only if there exists a negative \( h^* \) such that the method is h stable for every \( h \) satisfying \( h^* < h < 0 \).

**Definition 3.3:** The numerical method defined by equations (1.4) is said to be weakly stable if and only if it is zero stable and not strongly stable.

In Chapter VI, h stability for non-zero h will be emphasized. For the present, however, the implications of zero stability will be considered. This work is done in preparation for Chapter V in which zero stability is shown to be necessary for convergence of the modified method defined by (1.4).

Brief comment on the origin of modified predictor-
corrector methods was given in Chapter I. Expansion of this discussion provides motivation for the work done in the remainder of this chapter. Assume that one has chosen a predictor formula, (1.4.1), and a corrector formula, (1.4.3), for use in an iterative predictor-corrector method. It was stated in Chapter II that, for convergence, it is necessary that the numerical method defined by the corrector be zero stable. Assume, for sake of argument, that the predictor also defines a zero stable numerical method. Now, assume that, rather than to use the corrector iteratively, it is desired to correct only once. Assume also that one knows how to choose K and L such that \( M_{n+k} \), as computed from (1.4.2), and \( y_{n+k} \), as computed from (1.4.4), are more accurate than \( P_{n+k} \) and \( C_{n+k} \), respectively. This choice will be discussed in Chapter IV. Then it seems natural to use the algorithm defined by (1.4). This has been recommended and used in practice, but its validity is subject to question. As mentioned above, it is shown in Chapter V that the numerical method defined by (1.4) must be zero stable if it is to converge. However, in the comments above, the zero stability of the modified method was not considered. Instead, zero stability of the method defined by (1.4.1) and zero stability of the method defined by (1.4.3) were considered separately. One must investigate the relation between zero stability of the
methods based on (1.4.1) and (1.4.3) and zero stability of the modified method.

From the definition given in Chapter II, it follows that the method defined by (1.4.1) is zero stable if and only if every root of the equation

\[(3.7) \quad \rho^k - \sum_{j=0}^{k-1} a_j \rho^j = 0\]

has modulus less than or equal to one, and each root of modulus one is simple.

Similarly, the method defined by (1.4.3) is zero stable if and only if the roots of the equation

\[(3.8) \quad \rho^k - \sum_{j=0}^{k-1} c_j \rho^j = 0\]

satisfy the same conditions. From the definition of \(h\) stability given above and the observation that an identically zero root does not affect stability, the modified method defined by equations (1.4) is zero stable if and only if the roots of the equation

\[(3.9) \quad \rho^k - \sum_{j=0}^{k-1} [L a_j + (1 - L)c_j] \rho^j = 0\]

satisfy these conditions.

From (3.9) it is apparent that the choice of \(K\) in equation (1.4.2) does not affect the zero stability of the modified method. This fact, and the observation that, if \(L = 0\), equations (3.8) and (3.9) are identical, give
proof the following results.

**Theorem 3.1:** The predictor-corrector method defined by equations (1.4) with \( K = L = 0 \) is zero stable if and only if the method defined by the corrector equation, (1.4.3), is zero stable.

**Theorem 3.2:** The P-modified predictor-corrector method defined by equations (1.4) is zero stable if and only if the method defined by the corrector equation, (1.4.3), is zero stable.

If \( a_j = c_j \) for \( j = 0,1,\ldots,k-1 \), then equations (3.8) and (3.9) again coincide. Hence, the following result is immediate. Recall that in the derivation of equation (3.4), it was assumed that \( L \neq 1 \).

**Theorem 3.3:** Let \( a_j = c_j \) for \( j = 0,1,\ldots,k-1 \), and assume that \( L \neq 1 \). Then, the modified method defined by equations (1.4) is zero stable if and only if the method defined by the corrector equation, (1.4.3), is zero stable.

Of even more interest than the preceding is the example to be given below. This example illustrates the difficulty that can arise when modified methods are derived in the heuristic manner suggested above. A detailed derivation of the example is given. It is clear from this
construction that many other examples could be used to illustrate the same difficulty.

First, choose \( k = 2 \) in equations (1.4.1) and (1.4.3). Since it is apparent from (3.9) that the choice of \( K \) does not affect the zero stability of the modified method, let \( K = 0 \). Equations (1.4.1) and (1.4.3) become

\[
(3.10.1) \quad p_{n+2} = a_1 y_{n+1} + a_0 y_n + h(b_1 y'_{n+1} + b_0 y'_n), \quad \text{and}
\]
\[
(3.10.2) \quad c_{n+2} = c_1 y_{n+1} + c_0 y_n + h(d_2 y'_{n+2} + d_1 y'_{n+1} + d_0 y'_n),
\]
respectively. In order to reduce the number of free parameters in these equations, it is convenient to require that they each be of order two. This implies the six constraints

\[
\begin{align*}
    a_0 + a_1 &= 1, \\
    a_1 + b_0 + b_1 &= 2, \\
    a_1 + 2b_1 &= 4, \\
    c_0 + c_1 &= 1, \\
    c_1 + d_0 + d_1 + d_2 &= 2,
\end{align*}
\]

and

\[
\begin{align*}
    c_1 + 2d_1 + 4d_1 &= 4.
\end{align*}
\]

Since (3.10.1) and (3.10.2) are each of order two, it follows from results to be shown in Chapter IV, that \( L \) should be chosen so that
in order to increase the local accuracy of the modified method. Hence, another constraint can be obtained by fixing the value of \( L \). This has the advantage, also, of simplifying equation (3.9). For this example, let \( L = -1/2 \). The new constraint is then

\[ -a_1 - 3b_1 + 3c_1 + 9d_1 + 36d_2 = 16. \]

Equation (3.9) then becomes

\[ (\rho - 1)(\rho + \frac{-a_0 + 3c_0}{2}) = 0, \]

when the constraints \( a_0 + a_1 = 1 \) and \( c_0 + c_1 = 1 \) are used. Also, by use of these constraints, equation (3.7) becomes

\[ (\rho - 1)(\rho + a_0) = 0, \]

and equation (3.8) becomes

\[ (\rho - 1)(\rho + c_0) = 0. \]

Hence, any choice of \( a_0 \) and \( c_0 \) such that \(|a_0| < 1, |c_0| < 1, a_0 \neq -1, c_0 \neq -1, \) and \(|-a_0 + 3c_0| > 2\) results in a predictor which defines a zero stable method and a corrector which defines a zero stable method, but, at the same time, a modified method that is not zero stable.

From the many possible choices, let \( a_0 = 0 \) and \( c_0 = 1 \) for ease in solving the constraint equations. With these two parameters selected, there remain seven equations which uniquely determine the seven remaining parameters. The
C-modified method is then determined as

\[(3.15.1) \quad P_{n+2} = Y_{n+1} + \frac{h}{2} (3y'_{n+1} - y'_n), \]

\[(3.15.2) \quad C_{n+2} = Y_n + \frac{h}{36}(7y'_{n+2} + 58y'_{n+1} + 7y'_n), \text{ and} \]

\[(3.15.3) \quad Y_{n+2} = C_{n+2} - \frac{1}{2} (P_{n+2} - C_{n+2}). \]

To illustrate the practical effect of using the method defined by equation (3.15), consider the initial value problem \( y' = 1, y(0) = 0 \). Note that, in this case, \( f_y = 0 \), and, therefore, \( h = 0 \). For purpose of illustration, first assume that equation (3.15.3) is not to be used in the solution. Since \( y' \) is constant, it follows that the solution obtained by use of (3.15.2) is independent of the predicted value even if the corrector is used only once. The solution, in this case, satisfies the difference equation

\[(3.16) \quad Y_{n+2} = Y_n + 2h. \]

If \( y(0) = 0 \) and \( y(h) = Y_1 \), the solution of (3.16) is given by

\[(3.17) \quad Y_n = \frac{Y_1 - h}{2} + \left(\frac{-Y_1 + h}{2}\right)(-1)^n + nh. \]

The term \( nh \) in (3.17) corresponds to the true solution \( \bar{y}(x) = x \). Hence, the error is given by

\[(3.18) \quad \epsilon_n = Y_n - \bar{y}_n = \frac{Y_1 - h}{2} \left[ 1 + (-1)^{n+1} \right]. \]
If \( y_1 = h \), then \( \epsilon_n \) is identically zero. However, if \( y_1 \) is only approximately equal to \( h \), then the error is zero for \( n \) even and equal to \( y_1 - h \) for \( n \) odd. In particular, the magnitude of \( \epsilon_n \) never exceeds \( |y_1 - h| \).

Now, consider the result when equation (3.15.3) is used in conjunction with (3.15.1) and (3.15.2) to solve the same initial value problem. In this case, the solution satisfies

\[
(3.19) \quad y_{n+2} = c_{n+2} - \frac{1}{2} (p_{n+2} - c_{n+2}) = -\frac{1}{2} y_{n+1} + \frac{3}{2} y_n + \frac{5}{2} h.
\]

When \( y(0) = 0 \), and \( y(h) = y_1 \), the solution of (3.19) is given by

\[
(3.20) \quad y_n = \frac{2}{5} (y_1 - h)[1 - \left(-\frac{3}{2}\right)^n] + nh.
\]

Hence, the error is

\[
(3.21) \quad \epsilon_n = y_n - \overline{y}_n = \frac{2}{5} (y_1 - h) \left[1 - \left(-\frac{3}{2}\right)^n\right].
\]

It is apparent that if \( y_1 \) is only approximately correct, there is a component of the error that grows exponentially in magnitude and alternates in sign.

This example illustrates the importance of considering, directly, the stability of a modified predictor-corrector method rather than considering, separately, the stability of the methods defined by the predictor and corrector. It
is of interest to note that there exist zero stable modified methods in which the predictor and corrector each define methods that are not zero stable. Such examples are easily constructed. For instance, by reference to equations (3.12), (3.13), and (3.14), it is apparent that if \( a_0 \) and \( c_0 \) are chosen so that \( |a_0| > 1, |c_0| > 1, |-a_0 + 3c_0| \leq 2, \) and \(-a_0 + 3c_0 \neq -2\), such a method results.
CHAPTER IV: LOCAL ACCURACY OF MODIFIED PREDICTOR - CORRECTOR METHODS

The relation of the accuracy of the predictor and the accuracy of the corrector to the accuracy of the modified method defined by equations (1.4) is investigated in this chapter. In Chapter II, the concept of order was introduced for equation (1.1). Conditions (2.3) resulted. For convenience, analogous conditions for the predictor defined by (1.4.1) and for the corrector defined by

\[ c_{n+k} = y_{n+k} = \sum_{j=0}^{k-1} (c_jy_{n+j} + hd_jy'_n + j) + hd_ky'_{n+k} \]

are listed below.

For equation (1.4.1), the conditions are

\[ \sum_{j=0}^{k-1} a_j = 1 \quad \text{for } m = 0, \text{ and} \]

\[ \sum_{j=0}^{k-1} (a_j m^m + m b_j m^{m-1}) = k^m \quad \text{for } m = 1, 2, \ldots. \]

For equation (4.1), the conditions are

\[ \sum_{j=0}^{k-1} c_j = 1 \quad \text{for } m = 0, \text{ and} \]

\[ \sum_{j=0}^{k-1} (c_j m^m + m d_j m^{m-1}) + m d_k m^{m-1} = k^m \quad \text{for } m = 1, 2, \ldots. \]
Thus, (1.4.1) is of order $p$ if and only if conditions (4.2) hold for $m = 0, 1, \ldots, p$, but not for $m = p+1$, and (4.1) is of order $q$ if and only if conditions (4.3) hold for $m = 0, 1, \ldots, q$, but not for $m = q+1$.

Also, in Chapter II, an expression for the local truncation error of formula (1.1) was developed. From these results, it follows that the true value, $y_{n+k}$, of the solution of problem (2.1) at the point $x_{n+k}$ can be written as

$$y_{n+k} = y_n + c^{(P)}_p h^{p+1} y^{(p+1)}(x_n) + o(h^{p+2}),$$

or as

$$y_{n+k} = c^{(C)}_n + c^{(C)}_q h^{q+1} y^{(q+1)}(x_n) + o(h^{q+2}),$$

where $P_{n+k}$ is defined by (1.4.1) and $C_{n+k}$ is defined by (4.1). By use of equation (2.20), the constants $c^{(P)}_p$ and $c^{(C)}_q$ can be written as

$$c^{(P)}_p = \frac{1}{(p+1)!} \left\{ k^{p+1} - \sum_{j=0}^{k-1} [a_j^{p+1} + (p+1)b_j^{p}] \right\},$$

and

$$c^{(C)}_q = \frac{1}{(q+1)!} \left\{ k^{q+1} - d_k^{(q+1)}k^{q} - \sum_{j=0}^{k-1} [c_j^{q+1} + (q+1)d_j^{q}] \right\}.$$

Definition 4.1: When $p$ and $q$ denote the respective orders of equations (1.4.1) and (4.1), the constants $c^{(P)}_p$ and
\( c_{q}^{(C)} \) are called error coefficients. The superscript denotes whether the constant is related to the predictor or to the corrector.

It is now possible to show how to choose the constants \( K \) and \( L \) of equation (1.4.2) and (1.4.4), respectively, in order to increase local accuracy. Assume that \( p = q \) in (4.4) and (4.5), and that \( C_{p}^{(C)} \neq C_{p}^{(P)} \). Then, when (4.4) and (4.5) are equated and solved for \( h_{P+1} y_{(P+1)}(x_{n}) \), the relation

\[
(4.8) \quad h^{P+1} y_{(P+1)}(x_{n}) = \frac{p_{n+k} - c_{n+k}}{c_{p}^{(C)} - c_{p}^{(P)}} + 0(h^{P+2})
\]

results. By use of (4.8), equations (4.4) and (4.5) then become

\[
(4.9) \quad \bar{y}_{n+k} = p_{n+k} + \frac{c_{p}^{(P)}}{c_{p}^{(C)} - c_{p}^{(P)}} (p_{n+k} - c_{n+k}) + 0(h^{P+2}), \text{ and}
\]

\[
(4.10) \quad \bar{y}_{n+k} = c_{n+k} + \frac{c_{p}^{(C)}}{c_{p}^{(C)} - c_{p}^{(P)}} (p_{n+k} - c_{n+k}) + 0(h^{P+2}).
\]

When (4.10) is compared with (1.4.4), it seems reasonable to choose \( L \) so that

\[
(4.11) \quad L = \frac{c_{p}^{(C)}}{c_{p}^{(C)} - c_{p}^{(P)}},
\]
where $C_p^{(P)}$ and $C_p^{(C)}$ are defined by (4.6) and (4.7). Under the assumption that $P_{n+k} - C_{n+k} = P_{n+k-1} - C_{n+k-1}$, it also seems reasonable to choose $K$ in equation (1.4.2) so that

\begin{equation}
K = \frac{C_p^{(P)}}{C_p^{(C)} - C_p^{(P)}}.
\end{equation}

This choice of $K$ is used in practice under the assumption that $P_{n+k} - C_{n+k}$ is approximately equal to $P_{n+k-1} - C_{n+k-1}$. When $K$ and $L$ are defined in this way, it is evident that

\begin{equation}
1 + K = L.
\end{equation}

This relation is used in Chapter V.

The next step in this study is the derivation, for the modified method defined by equations (1.4), of conditions that are analogous to conditions (2.3) for linear multistep methods. For convenience in the derivation, it is assumed that $y(x)$ possesses a Taylor series expansion about $x_n$ that is twice differentiable. Consider equations (1.4) with the term $f(x_{n+k}, M_{n+k})$ of (1.4.3) replaced by

\begin{equation}
M'_{n+k},
\end{equation}

where

\begin{equation}
M'_{n+k} = P'_{n+k} + K(P'_{n+k-1} - C'_{n+k-1}).
\end{equation}

These equations can be solved for $y_{n+k}$ in terms of $y_{n+j}$, $y'_{n+j}$, and $y''_{n+j}$, where $j = -1, 0, \ldots, k-1$. This can be done by proceeding in the same manner as was suggested for the solution of equations (3.3) in Chapter III. The result is
\[(4.15) \quad y_{n+k} - \sum_{j=0}^{k-1} \left[ L \alpha_j + (1-L) c_j \right] y_{n+j} = h y'_{n+k-1} \left\{ L b_{k-1} \right\} \\
+ (1-L) d_{k-1} + d_k \left[ (1-L) a_{k-1} - K \right] \right\} \\
+ h^2 y''_{n+k-1} \left[ (1-L) d_k b_{k-1} \right] + h \sum_{j=0}^{k-2} y'_{n+j} \left\{ L b_j \right\} \\
+ (1-L) d_j + d_k \left[ (1-L) a_j + K a_{j+1} \right] \right\} \\
+ h^2 \sum_{j=0}^{k-2} y''_{n+j} \left\{ d_k \left[ (1-L) b_j + K b_{j+1} \right] \right\} \\
+ h d_k K \left( a_0 y'_{n-1} + h b_0 y''_{n-1} \right). \]

Note that this result can be obtained formally from equation (3.4) by disregarding the error terms, replacing \( \epsilon_{n+j} \) by \( y_{n+j} \) for \( j = -1,0,\ldots,k \), and treating \( g_{n+j} \) and \( g^{(M)}_{n+k} \) as differentiation operators with respect to \( y \).

To obtain the desired conditions, first expand each term of (4.15) in a Taylor series expansion about \( x_n \), making use of (2.2.1), (2.2.2), and

\[(4.16) \quad y''_{n+j} = \sum_{m=0}^{\infty} \frac{(jh)^m y^{(m+2)}(x_n)}{m!}, \]

as well as the relations obtained by replacing \( j \) by \( j-1 \) in these series. Then, multiply both sides of the resulting equation by \( m! \) and equate like terms, \( h^m y^{(m)}(x_n) \). The
Definition 4.2: The modified predictor-corrector method defined by (1.4) is said to be of order \( r \) if and only if conditions (4.17) hold for \( m = 0, 1, \ldots, r \), but not for \( m=r+1 \).

Definition 4.3: The numerical method defined by (1.4)
is said to be consistent if and only if it is at least of order one.

In analogy with the definitions of the constants \( C_p^{(P)} \) and \( C_q^{(C)} \) for the predictor and corrector, define the constant \( C_r^{(M)} \) for the modified predictor-corrector method by the equation

\[
(4.18) \quad C_r^{(M)} = \frac{1}{(r+1)!} \left[ k^{r+1} - (1-L) \sum_{j=0}^{k-1} [a_j j^r + rb_j (j-1)^{r-1}] \right.
+ \left. \sum_{j=0}^{k-1} [c_j j^{r+1} + (r+1) d_j j^r] \right]
- Kd_{k}(r+1) \left[ \sum_{j=0}^{k-1} [a_j (j-1)^r + rb_j (j-1)^{r-1}] \right]
- \sum_{j=0}^{k-1} [a_j j^{r+1} + (r+1) b_j j^r] \right].
\]

Definition 4.4: When \( r \) denotes the order of the modified predictor-corrector method defined by equations (1.4), the constant \( C_r^{(M)} \) is called the error coefficient.

Equations (4.17) and (4.18) are applicable, by proper choice of \( K \) and \( L \), to the predictor-corrector method and to the various modified predictor-corrector methods of Chapter I.

The following lemma relates one of the terms of equation (4.18) to conditions (4.2). It is useful in establishing succeeding results.
Lemma 4.1: If the predictor, (1.4.1), is of order \( m \), where \( m = 1, 2, \cdots \), then 
\[
\sum_{j=0}^{k-1} [a_j(j-1)^m + mb_j(j-1)^{m-1}] = (k-1)^m.
\]

Proof:
\[
\sum_{j=0}^{k-1} [a_j(j-1)^m + mb_j(j-1)^{m-1}]
= \sum_{j=0}^{k-1} \left( \sum_{r=0}^{m} (-1)^r \binom{m}{r} j^m - \sum_{r=0}^{m-1} (-1)^r \binom{m-1}{r} j^{m-1} \right)
= \sum_{j=0}^{m} (-1)^r \binom{m}{r} \left( \sum_{j=0}^{k-1} a_j(j-1)^{m-r} + (m-r)b_j(j-1)^{m-1-r} \right)
= \sum_{r=0}^{m} (-1)^r \binom{m}{r} k^{m-r} = (k-1)^m.
\]

Theorem 4.2: If conditions (4.2) and (4.3) are satisfied for \( m = 0, 1, \cdots, p \), then the modified predictor-corrector method defined by equations (1.4) is of at least order \( p \), independent of the choice of the constants \( K \) and \( L \).

Proof: Conditions (4.17) must be verified for \( m = 0, 1, \cdots, p \). For \( m = 0 \), condition (4.17.1) is easily verified; since with (4.2.1) and (4.3.1) it becomes \( 1 = 1 - L + L \). For \( m = 1 \), (4.2.1) implies that (4.17.2) can be written as
k = (1-L) \left[ d_k + \sum_{j=0}^{k-1} (jc_j + d_j) \right] + L \sum_{j=0}^{k-1} (ja_j + b_j), \text{ which } \text{by (4.2.2) and (4.3.2), for } m = 1, \text{ reduces to } k = (1-L)k + Lk = k. \text{ For } m = 2, 3, \ldots, p, \text{ it follows that }

\sum_{j=0}^{k-1} \left[ a_j j^{m-1} + (m-1)b_j j^{m-2} \right] = k^{m-1} \text{ from (4.2.2),}

k \sum_{j=0}^{k-1} \left[ c_j j^m + md_j j^{m-1} \right] = k^m \text{ from (4.3.2),}

\sum_{j=0}^{k-1} \left[ a_j (j-1)^{m-1} + (m-1)b_j (j-1)^{m-2} \right] - (k-1)^{m-1} = 0 \text{ from Lemma 4.1, and }

\sum_{j=0}^{k-1} \left[ a_j j^m + mb_j j^{m-1} \right] = k^m \text{ from (4.2.2). Hence, (4.17.3) becomes } k^m = (1-L)k^m + Lk^m = k^m, \text{ and the theorem is proved.}

Earlier in the chapter, it was suggested that the constants K and L of equations (1.4.2) and (1.4.4) be chosen as indicated in equations (4.11) and (4.12). These suggestions were motivated by noting in equations (4.9) and (4.10) that such choices resulted in errors of \(0(h^{p+2})\) rather than \(0(h^{p+1})\). Note, however, that these errors were those of the predictor and corrector formulas, individually, and not of the modified method. The following theorem establishes the exact way in which the choice of L affects the accuracy of the modified method.

**Theorem 4.3:** If the predictor, (1.4.1), and the corrector,
(4.1), are both of order \(p\), then the modified predictor-corrector method defined by equations (1.4) is of at least order \(p+1\) if and only if the constant \(L\) of equation (1.4.4) is chosen so that

\[
L = \frac{c_p^{(C)}}{c_p^{(C)} - c_p^{(P)}},
\]

where \(c_p^{(P)}\) and \(c_p^{(C)}\) are defined by (4.6) and (4.7), respectively.

Proof: By the definition of order and by use of Theorem 4.2, it follows that the modified predictor-corrector method is at least of order \(p\). Hence, the order is at least \(p+1\) if and only if \(c_p^{(M)} = 0\). It follows from (4.2.2) that

\[
\sum_{j=0}^{k-1} \left[ a_j j^p + pb_j j^{p-1} \right] = k^p,
\]

and from Lemma 4.1 that

\[
\sum_{j=0}^{k-1} \left[ a_j (j-1)^p + pb_j (j-1)^{p-1} \right] - (k-1)^p = 0.
\]

Hence, from equation (4.18), with \(r\) replaced by \(p\),

\[
c_p^{(M)} = \frac{1}{(p+1)!} \left[ k^{p+1} - (1-L) \left\{ d_k (p+1) k^p + \sum_{j=0}^{k-1} c_j j^{p+1} + (p+1) d_j j^p \right\} - L \sum_{j=0}^{k-1} [a_j j^{p+1} + (p+1) b_j j^p] \right].
\]
\[
\frac{1}{(p+1)!} \left[ k^{p+1} - d_{k} (p+1) k^{p} - \sum_{j=0}^{k-1} \left[ c_{j} j^{p+1} + (p+1) d_{j} j^{p} \right] \right] \\
+ L \left\{ d_{k} (p+1) k^{p} + \sum_{j=0}^{k-1} \left[ c_{j} j^{p+1} + (p+1) d_{j} j^{p} \right] \\
- \sum_{j=0}^{k-1} \left[ a_{j} j^{p+1} + (p+1) b_{j} j^{p} \right] \right\},
\]

which, by use of (4.6) and (4.7),

\[
= c_{p}^{(C)} + L \left[ \frac{k^{p+1}}{(p+1)!} - c_{p}^{(C)} + c_{p}^{(P)} - \frac{k^{p+1}}{(p+1)!} \right]
\]

\[
= c_{p}^{(C)} - L (c_{p}^{(C)} - c_{p}^{(P)}).
\]

Therefore, \( c_{p}^{(M)} = 0 \) if and only if \( L = \frac{c_{p}^{(C)}}{c_{p}^{(C)} - c_{p}^{(P)}} \), and

the theorem is proved.

Additional results can be obtained by specifically considering predictor-corrector methods. These methods are often used in practice and have been studied by Henrici [5]. The following theorems are related to the results obtained by Henrici [5, pp. 261-262]. However, they are not identical and the methods of proof are considerably different. Recall that a predictor-corrector method can be studied by setting \( K = L = 0 \) in equations (1.4). For clarity in the results below, however, equations (1.3) are considered as the defining equations.
This merely emphasizes that \( K = L = 0 \) in (1.4). With \( K = L = 0 \), equation (4.18) becomes

\[
(4.19) \quad C_{p}^{(M)} = \frac{1}{(r+1)!} \left\{ k^{r+1} - d_{k}(r+1) \sum_{j=0}^{k-1} a_{j}j^{r} + rbj^{r-1} \sum_{j=0}^{k-1} [c_{j}j^{r+1} + (r+1)d_{j}j^{r}] \right\}.
\]

**Theorem 4.4:** If the corrector, (4.1), is of order \( p \) and the predictor, (1.3.1), is of at least order \( p \), then the predictor-corrector method defined by equations (1.3) is of order \( p \), and the error coefficient, \( C_{p}^{(M)} \), is equal to \( C_{p}^{(C)} \), the error coefficient that results when the corrector is used iteratively.

**Proof:** By the definition of order and by use of Theorem 4.2, it follows that the predictor-corrector method is of at least order \( p \). In order to show that the order is exactly \( p \), it is sufficient to show that \( C_{p}^{(M)} = C_{p}^{(C)} \); since \( C_{p}^{(C)} \neq 0 \). By use of (4.2.2), equation (4.19), with \( r = p \), becomes

\[
C_{p}^{(M)} = \frac{1}{(p+1)!} \left\{ k^{p+1} - d_{k}(p+1)kp - \sum_{j=0}^{k-1} [c_{j}j^{p+1} + (p+1)d_{j}j^{p}] \right\},
\]

which coincides exactly with (4.7) when \( q = p \).

**Theorem 4.5:** If the corrector, (4.1), is of order \( p \), where \( p \geq 1 \), and the predictor, (1.3.1), is of order
p-q, where q = 1, 2, ..., p, then the predictor-corrector method defined by equations (1.3) is of at least order p-q+1.

Proof: By the definition of order and by use of Theorem 4.2, it follows that the predictor-corrector method is of at least order p-q. To show that the order is at least p-q+1, it is sufficient to show that $C_{p-q}^{(M)} = 0$. From equation (4.19),

$$C_{p-q}^{(M)} = \frac{1}{(p-q+1)!} \left\{ k^{p-q+1} - d_k(p-q+1) \sum_{j=0}^{k-1} [a_{jj}^{p-q} + (p-q)b_{jj}^{p-q-1} - \sum_{j=0}^{k-1} [c_{jj}^{p-q} + (p-q+1)d_{jj}^{p-q+1} + (p-q)j^{p-q} - \sum_{j=0}^{k-1} [a_{jj}^{p-q} + (p-q)b_{jj}^{p-q-1} - \sum_{j=0}^{k-1} [c_{jj}^{p-q} + (p-q+1)d_{jj}^{p-q+1} = 0. \right.$$}

But,

$$\sum_{j=0}^{k-1} [a_{jj}^{p-q} + (p-q)b_{jj}^{p-q-1}] = k^{p-q} \quad \text{from (4.2),}$$

since the predictor is of order p-q. Also, since the corrector is of order p, it follows that (4.3.2) is satisfied for m = p-q+1. Thus,

$$k^{p-q+1} - d_k(p-q+1)k^{p-q} - \sum_{j=0}^{k-1} [c_{jj}^{p-q+1} + (p-q)j^{p-q}] = 0.$$

Therefore, $C_{p-q}^{(M)} = 0$, and the theorem is proved.
CHAPTER V: CONVERGENCE OF MODIFIED PREDICTOR - CORRECTOR METHODS

In Chapter II, a concept of convergence was introduced for the numerical method defined by equation (1.1). Here, this idea is extended to the modified predictor-corrector method. Again, consider the initial value problem (2.1), and let \( F \) be the set of all functions satisfying conditions i) and ii) as stated following (2.1). For each \( f \in F \), problem (2.1) has a unique solution, \( \overline{y}_f(x) \). Consider the set \( G \subseteq F \) of all functions \( g \) for which \( \overline{y}_g(x) \in C^2 \).

Definition 5.1: The modified predictor-corrector method defined by (1.4) is said to be convergent over \( G \) if and only if for every \( g \in G \) and for every set of starting values \( y_q = y(a + qh) \) satisfying \( \lim_{h \to 0} y_q = y(a) = \eta \) for \( q = 0,1,\ldots, k-1 \), it determines a sequence \( \{y_n\} \) such that

\[
\lim_{n \to \infty} y_n = \overline{y}_g(x_0) \quad \text{for all } x_0 \in [a,b].
\]

\( nh = x_0 \)

Several types of modified methods were discussed in Chapter I. For the first theorem of this chapter, only PC-modified methods are considered. More particularly, the only methods considered are those in which \( K \) and \( L \) are chosen as indicated in equations (4.12) and (4.11),
respectively. Recall that the subscript $p$ in these equations denotes the order of the predictor and of the corrector.

Before attempting to establish necessary and sufficient conditions for the convergence over $G$ of the modified method described above, three lemmas are introduced. The first concerns the truncation error associated with equation (1.1). This concept was discussed in Chapter II, where it was shown that $T_n = C_p h^{p+1} \frac{y^{(p+1)}(x_n)}{y} + O(h^{p+2})$ when the solution of (2.1) satisfies $y(x) \in C^{(p+2)}$ and (1.1) is of order $p$. More information about the term $O(h^{p+2})$ is needed.

Lemma 5.1: Assume that $y(x) \in C^{(p+2)}$ and that equation (1.1) is of order $p$. Then, there exists a positive constant $D$ such that $\left| T_n \right| \leq \left| C_p h^{p+1} \frac{y^{(p+1)}(x_n)}{y} \right| + Dh^{p+2}$.

Proof: With the hypotheses above, it is possible, by use of Taylor's formula, to rewrite equations (2.18) and (2.19) as

$$\begin{align*}
(5.1) \quad \overline{y}_{n+k} &= \sum_{m=0}^{p+1} \frac{k^{m}h^{m}}{m!} \overline{y}(m)(x_n) + \frac{k^{p+2}h^{p+2}}{(p+2)!} \frac{y^{(p+2)}(\eta_k)}{y} , \\
and \quad y_{n+k} &= \sum_{j=0}^{k-1} \left\{ a_j \left[ \sum_{m=0}^{p+1} \frac{j^{m}h^{m}}{m!} \overline{y}(m)(x_n) \\
+ \frac{j^{p+2}h^{p+2}}{(p+2)!} \frac{y^{(p+2)}(\eta_j)}{y} \right] \right\}.
\end{align*}$$
respectively, where \( x_n \leq \eta_j, \xi_j \leq x_n + jh, \) for \( j = 0,1, \ldots, k. \) Proceeding as in Chapter II, it follows that

\[
T_n = y_{n+k} - y_{n+k} = c_p h^{p+1} \frac{1}{y^{(p+1)}(x_n)} \left[ \sum_{m=0}^{\frac{k}{h}} \frac{j!}{m!} \frac{j^{p+1} h^{p+2}}{(p+2)!} (\xi_j) \right],
\]

Since \( y^{(p+2)}(x) \) is continuous for \( a \leq x \leq b, \) it follows that there exists a positive constant \( Y \) such that \( |y^{(p+2)}(x)| \leq Y, \) for all \( a \leq x \leq b. \) Thus,

\[
|T_n| \leq |c_p h^{p+1} y^{(p+1)}(x_n)| + EY h^{p+2}, \quad \text{where}
\]

\[
E = \frac{k^{p+2}}{(p+2)!} + \frac{k-1}{(p+2)!} \sum_{j=0}^{k} \left| \frac{a_j p+2}{h} \right| + \frac{k}{(p+2)!} \sum_{j=0}^{k} \left| \frac{b_j p+1}{h} \right|.
\]

The conclusion of the lemma follows from the definition \( D \equiv EY. \)

The two remaining lemmas are concerned with the growth of solutions of the difference equation

\[
(5.4) \sum_{j=0}^{k+1} a_j z_{n+j} = h \sum_{j=0}^{k} \beta_j, n z_{n+j} + \lambda_n.
\]

The lemmas involve only trivial modifications of results found in Henrici [5, pp. 242-243]; hence, the proofs are
not repeated here. In particular, the modifications consist of replacing \( k \) by \( k+1 \) and setting \( \beta_{k+1,n} = 0 \). Lemma 5.2 is used in the statement and proof of Lemma 5.3.

**Lemma 5.2:** If every root of the equation \( \sum_{j=0}^{k+1} a_j \rho^j = 0 \) has modulus less than or equal to one and each root of modulus one is simple, and if the coefficients \( \gamma_m \) for \( m = 0,1,\ldots \) are defined by

\[
\sum_{j=0}^{k+1} a_{k-j+1} \rho^j / \sum_{m=0}^\infty \gamma_m \rho^m = \gamma_m \rho^m, \text{ then}
\]

\[
\Gamma = \sup_{m = 0,1,\ldots} |\gamma_m| < \infty.
\]

**Lemma 5.3:** If the roots of the equation \( \sum_{j=0}^{k+1} a_j \rho^j = 0 \) satisfy the hypotheses of Lemma 5.2, and there exist constants \( B \) and \( \Lambda \) such that \( \sum_{j=0}^k |\beta_{j,n}| \leq B \) and \( |\lambda_n| \leq \Lambda \), for \( n = 0,1,\ldots,N \), and if \( h > 0 \), then every solution of (5.4) for which \( |z_q| \leq Z \) for \( q = 0,1,\ldots,k \) satisfies

\[
|z_n| \leq \Gamma [N \Lambda + A Z(k+1)] e^{nh} B,
\]

for \( n = 0,1,\ldots,N \), where \( \Gamma \) is defined in Lemma 5.2 and

\[
A = \sum_{j=0}^{k+1} |a_j|.
\]

In addition to these lemmas, the following observation is of use in establishing necessary and sufficient conditions for convergence. When \( K \) and \( L \) are chosen as
suggested above, it is easy to verify that $K = -(1-L)$. Hence, conditions (4.17.1) and (4.17.2), the conditions that the modified method be of at least order one, can be written as

$$\begin{align*}
(5.5.1) \quad 1 &= (1-L) \sum_{j=0}^{k-1} c_j + L \sum_{j=0}^{k-1} a_j, \\
(5.5.2) \quad k &= (1-L) \sum_{j=0}^{k-1} d_j + L \sum_{j=0}^{k-1} b_j \\
&\quad + \sum_{j=0}^{k-1} j[La_j + (1-L)c_j],
\end{align*}$$

respectively.

**Theorem 5.4:** If the constants $K$ and $L$ of equations (1.4.2) and (1.4.4) are defined by (4.12) and (4.11), respectively, and if $G$ is the class of functions described above, then the PC-modified predictor-corrector method defined by equations (1.4) converges over $G$ if and only if it is zero stable and consistent.

**Proof:** From the definition of $h$ stability given in Chapter III, it follows that the method under consideration is zero stable if and only if every root of the equation

$$\rho^{k+1} - \sum_{j=0}^{k-1} [La_j + (1-L)c_j]\rho^{j+1} = 0$$

has modulus less than or equal to one and each root of
modulus one is simple. This condition is satisfied if and only if the roots of the equation

\[ \rho^k - \sum_{j=0}^{k-1} [L a_j + (1-L) c_j] \rho^j = 0 \]

satisfy the same conditions. The proof that zero stability is necessary for convergence now follows from consideration of the initial value problem \( y' = 0, y(0) = 0 \). This proof is not given here since it may be found in Henrici [5, p. 218]. In order to identify that proof with the present problem, let \( \alpha_k = 1 \) and \( \alpha_j = -L a_j - (1-L) c_j \) for \( j = 0,1, \ldots, k-1 \).

The proof that consistency is necessary for convergence involves showing that conditions (5.5.1) and (5.5.2) must be satisfied. This is done by considering two particular initial value problems. Again, the approach is like that found in Henrici [5]. However, in this case, the details vary somewhat. Hence, for clarity, the proof follows.

To show that (5.5.1) must hold, consider the problem \( y' = 0, y(0) = 1 \), with true solution \( y(\cdot) = 1 \). If the method defined by (1.4) is to be convergent over \( G \), it must certainly converge for this problem. The solution of equations (1.4), in this case, can be written as

\[ y_{n+k} = L p_{n+k} + (1-L) c_{n+k} \]
\[ y_{n+k} = L \left( \sum_{j=0}^{k-1} a_j y_{n+j} + (1-L) \sum_{j=0}^{k-1} c_j y_{n+j} \right), \text{ or} \]

\[ y_{n+k} - \sum_{j=0}^{k-1} \left[ L a_j + (1-L) c_j \right] y_{n+j} = 0. \]

(5.8)

Since equation (5.8) does not depend on \( h \), it follows that \( y_n \to 1 \) as \( n \to \infty \). By use of this fact, equation (5.8) yields condition (5.5.1).

To show that (5.5.2) must hold, consider the problem \( y' = 1, y(0) = 0 \), with true solution \( y(x) = x \). In this case, the solution given by (1.4) can be written as

\[ y_{n+k} = L p_{n+k} + (1-L) c_{n+k} \]

\[ = L \sum_{j=0}^{k-1} \left( a_j y_{n+j} + h b_j \right) \]

\[ + (1-L) \left[ \sum_{j=0}^{k-1} c_j y_{n+j} + h \sum_{j=0}^{k-1} d_j \right], \text{ or} \]

(5.9)

\[ y_{n+k} - \sum_{j=0}^{k-1} \left[ L a_j + (1-L) c_j \right] y_{n+j} = h \left[ L \sum_{j=0}^{k-1} b_j + (1-L) \sum_{j=0}^{k-1} d_j \right]. \]

Since (5.5.1) must hold for any convergent method, it follows that \( \rho = 1 \) is a root of (5.7). Also, since any convergent method must be zero stable, it follows that \( \rho = 1 \) is not a multiple root of (5.7). Hence, the derivative of (5.7) evaluated at \( \rho = 1 \),
\( (5.10) \quad k - \sum_{j=0}^{k-1} \left[ L a_j + (1-L) c_j \right], \)

is not equal to zero. Define a sequence \( \{y_n\} \) by letting

\[
y_n = nhF, \quad F = \frac{k-1}{k} \sum_{j=0}^{k} \left( \sum_{j=0}^{k} \left[ L b_j + (1-L) d_j \right] \right).
\]

By direct substitution, it can be verified that the sequence \( \{y_n\} \) satisfies the difference equation (5.9).

Also, the starting values satisfy the hypotheses of the theorem since

\[
\lim_{h \to 0} y_q = \lim_{h \to 0} qhF = 0, \quad \text{for } q = 0, 1, \ldots, k-1.
\]

Since the method defined by (1.4) must converge to the solution \( y(x) = x \), for this particular problem, it follows that, for every \( x \) in \([a, b]\),

\[
\lim_{h \to 0} y_n = \lim_{h \to 0} nhF = xF = x.
\]

Hence, \( F = 1 \), which implies that condition (5.5.2) is satisfied.

The fact that zero stability and consistency are sufficient for convergence is shown by relating equation (3.4) to equation (5.4), and then by using Lemma 5.3. Recall that (3.4) is the equation satisfied by the error that occurs when the modified predictor-corrector method
is used to approximate the solution of (2.1).

Since the method is assumed to be consistent, conditions (5.5.1) and (5.5.2) hold. Equation (5.5.2), when solved for $L$ yields

\[
L = \frac{k - \sum_{j=0}^{k-1} d_j - \sum_{j=0}^{k-1} j c_j}{k - \sum_{j=0}^{k-1} d_j - \sum_{j=0}^{k-1} j c_j - d_k - \sum_{j=0}^{k-1} (j c_j + d_j) + \sum_{j=0}^{k-1} (j a_j + b_j)}
\]

(5.12)

Condition (5.5.1) must be satisfied, independent of the value of $L$ determined by (5.12). By writing (5.5.1) as

\[
1 = 2 c_j + L(\sum_{j=0}^{k-1} a_j - \sum_{j=0}^{k-1} c_j),
\]

it is evident that

\[
\sum_{j=0}^{k-1} a_j = \sum_{j=0}^{k-1} c_j = 1.
\]

(5.13)

Conditions (5.12) and (5.13) are satisfied by the PC-modified predictor-corrector method in which the predictor and the corrector are each of order zero. For, it follows from (4.6) and (4.7) that

\[
c_0^{(P)} = k - \sum_{j=0}^{k-1} (j a_j + b_j), \text{ and}
\]

\[
c_0^{(C)} = k - d_k - \sum_{j=0}^{k-1} (j c_j + d_j).
\]

Hence, when $L$ is defined by (4.11), equation (5.12) results.
When the predictor and corrector are each of order zero, the error terms in (3.4) become

\[
(5.14) \quad \lambda_n \equiv L[-c_0^{(P)} h^x(x_n) + 0(h^2)] + (1-L)[-c_0^{(C)} h^x(x_n) \\
+ 0(h^2)] + (1-L+K)d_k g_{n+k}^{(M)} h[-c_0^{(P)} h^x(x_n) \\
+ 0(h^2)].
\]

Since the predictor and corrector equations are both of form (1.1) and \( y \in C^2 \), it follows from Lemma 5.1 that there exists a positive constant \( R \) such that the terms of \( 0(h^2) \) have magnitude less than \( Rh^2 \). By use of (4.13), it is then possible to write

\[
|\lambda_n| \leq |h^x(x_n)| \cdot |LC_0^{(P)} + (1-L)c_0^{(C)}| \\
+ |LR| + |(1-L)R|h^2.
\]

But, \( LC_0^{(P)} + (1-L)c_0^{(C)} = 0 \), by use of (4.11). Therefore,

\[
(5.15) \quad |\lambda_n| \leq Th^2,
\]

where \( T = |LR| + |(1-L)R| \).

To complete the identification of equations (3.4) and (5.4), let \( z_{n+j} = e_{n+j-1} \), for \( j = 0, 1, \ldots, k+1 \), and let \( \alpha_{k+1} = 1, \alpha_0 = 0, \alpha_j = -La_{j-1} - (1-L)c_{j-1} \), for \( j = 1, 2, \ldots, k \). Let

\[
\beta_{k,n} = [Lb_{k-1} + (1-L)d_{k-1}] g_{n+k-1} + d_k g_{n+k}^{(M)} [(1-L)(a_{k-1} \\
+ h^x b_{k-1} g_{n+k-1}) - K],
\]

where \( L = L_1, C = C_1 \), and \( L_1, C_1 \) are positive constants.
\[ \beta_{j,n} = [L b_{j-1} + (1-L) d_{j-1}] g_{n+j-1} + d_k g_{n+k}^{(M)} [(1-L) (a_{j-1} + h b_j g_{n+j-1}) + K (a_j + h b_j g_{n+j-1})], \]

for \( j = 1, 2, \cdots, k-1 \), and

\[ \beta_{0,n} = d_k g_{n+k}^{(M)} K (a_0 + h b_0 g_{n-1}). \]

Recall that \( \lambda_n \) is defined by (5.14).

Now, make the following observations and definitions in preparation for the application of Lemma 5.3. Let

\[ N = \frac{x_n - a}{h}, \]

where \( x_n \in (a,b] \). Since the function \( f \) of (2.1) satisfies a Lipschitz condition with respect to \( y \), \( g_{n+k} \) and \( g_{n+j} \), \( j = -1, 0, \cdots, k-1 \), are bounded. Hence, there exists a constant \( B \) such that \( \sum_{j=0}^{k} |\beta_{j,n}| \leq B \). Let

\[ \Lambda = T h^2. \]

Define \( \epsilon_{-1} = 0 \), and let \( \delta(h) = \max_{q=0,1,\cdots,k-1} |\epsilon_q| \), for \( q = 0, 1, \cdots, k-1 \). Let \( Z = \delta(h) \) and recall that, by hypothesis, \( \delta(h) \to 0 \) as \( h \to 0 \). Let \( \Gamma \) be defined as in Lemma 5.2, and \( A \) be defined as in Lemma 5.3. Then, from the conclusion of Lemma 5.3, it follows that

\[ |\epsilon_n| \leq \Gamma \left[ \left( \frac{x_n - a}{h} \right) T h^2 + A \delta(h) (k+1) \right] e^{x_n \Gamma B}. \]

Hence,

\[ \lim_{h \to 0} \left| \epsilon_n \right| \leq \Gamma e^{x_n \Gamma B} \lim_{h \to 0} \left[ (x_n - a) T h + A \delta(h) (k+1) \right] = 0, \]

and the theorem is proved.
The following theorem can also be established. The proof varies only in minor detail from that just given; thus, it is not included.

Theorem 5.5: If \( G \) is the class of functions defined above and if the predictor, (1.4.1), is at least of order zero, then the predictor-corrector method defined by equations (1.4) with \( K = L = 0 \) is convergent over \( G \) if and only if it is zero stable and consistent.
CHAPTER VI: STRONGLY STABLE PREDICTOR-CORRECTOR METHODS

In Chapter III, a definition of $h$ stability was given for the modified predictor-corrector method defined by equations (1.4). Following this definition, the concept of strong stability was introduced. For the modified method to be strongly stable, it was necessary for it to be $h$ stable for every $h$ in some interval $[h^*,0]$. In this chapter, emphasis is placed on the negative value, $h^*$.

It is desirable to find a numerical method that is $h$ stable throughout some interval of maximum length. In other words, it is desirable to minimize the negative value, $h^*$. A method with such stability properties is useful in solving equations which arise from a particular type of problem often encountered in scientific applications. Dahlquist [2, p. 36] qualitatively describes such a problem as one in which there occur transient effects which die out quickly relative to the time scale of the phenomenon under study. The simple initial value problem $y' = -y$, $y(0) = 1$, which involves exponential decay, illustrates such an effect. In such problems, the step-length, $h$, is often limited by the range of $h$ stability.

The results below are developed by considering predictor-corrector methods involving five points. Hence, $K = L = 0$, and $k=4$ in equations (1.4). The numerical methods to be studied are thus defined by
\begin{equation}
(6.1.1) \quad P_{n+4} = \sum_{j=0}^{3} (a_j y_{n+j} + h b_i y_{j+n+j}), \text{ and}
\end{equation}

\begin{equation}
(6.1.2) \quad y_{n+4} = \sum_{j=0}^{3} (c_j y_{n+j} + h d_j y_{j+n+j} + h d_{j+4}(x_{n+4}, P_{n+4})).
\end{equation}

Whether or not the method defined by equations (6.1) is \( h \) stable depends upon the roots of equation (3.6) with \( K = L = 0 \), and \( k = 4 \). Note that with \( K = 0 \), (3.6) has one root identically zero. This root does not affect the stability; thus, it is sufficient to consider the roots of the equation

\begin{equation}
(6.2) \quad \rho^4 - \sum_{j=0}^{3} \left[ c_j + h d_j + h^2 (d_{j+4} a_j) + h^2 (d_{j+4} b_j) \right] \rho^j = 0.
\end{equation}

These roots are functions of 8 predictor coefficients, 9 corrector coefficients, and \( h \).

To further specialize the problem under study, and to obtain a method with truncation error of \( O(h^5) \), assume that (6.1.1) and (6.1.2) are each of order 4. Thus, conditions (4.2) and (4.3) are satisfied for \( m = 0, 1, \ldots, 4 \). Under these conditions, the roots of (6.2) are functions of 3 predictor coefficients, 4 corrector coefficients, and \( h \).

The condition that each predictor-corrector method under consideration be zero stable depends only upon the coefficients \( c_0, c_1, c_2, \) and \( c_3 \) of the corrector formula.
This is evident from equation (6.2). The conditions that three of these roots be zero, that the term involving $y'_n$ not appear, and that (6.1.2) be of order 4 determines a well-known corrector formula usually derived by other means. This corrector is sometimes called an Adams formula of the closed type and is also called an Adams-Moulton formula. It can be written as

$$y_{n+4} = y_{n+3} + \frac{h}{24} (9y'_{n+4} + 19y'_{n+3} - 5y'_{n+2} + y'_{n+1}).$$

The associated truncation error term is given by

$$T_n = -\frac{19}{720} h^5 y^{(5)}(\eta),$$

where $x_{n+1} < \eta < x_{n+4}$.

On the basis of the following considerations, the class of methods to be studied is specialized even further. In Chapter IV, it was shown that, when the predictor is of at least the order of the corrector, the error coefficient in the predictor-corrector method depends only upon the corrector. Hence, if the corrector is specified by (6.3), $C_4^{(M)} = -\frac{19}{720}$. In deriving (6.3), it was specified that, at $h = 0$, all roots be zero except for one root $p = 1$. Recall that, for any consistent method, there must be a root equal to one at $h = 0$. The fact that the remaining roots are zero is advantageous, especially for cases in which $f_y = 0$ in problem (2.1).
Such problems are common since integrals are often evaluated numerically by solving differential equations of the form \( y' = f(x) \). Thus, in order to retain these properties, it is assumed that the corrector coefficients are specified as in (6.3). The roots of (6.2) are, therefore, functions of 3 predictor coefficients and \( h \). The goal is to choose the 3 predictor coefficients, which are referred to later as parameters, so that a maximum interval of stability is attained.

Due to the nature of the roots of (6.2) at \( h = 0 \), it is apparent that only strongly stable methods are being considered, regardless of the choice of the three predictor coefficients. This is true since the roots of (6.2) are continuous functions of the coefficients and the root \( p = 1 \) at \( h = 0 \) is approximately equal to \( e^h \) as \( h \to 0 \).

The problem of minimizing \( h^* \) can be approached in more than one way. Wilf [8], for instance, presents one method, and carries it out for corrector equations in some particular cases. This approach, when applied to the problem considered here, leads to an involved set of non-linear inequalities which must be solved. If there exists a unique solution to this set, and if it can be found, this approach has the advantage of assuring one that a minimum is actually attained. An alternate approach for the minimization of \( h^* \) is presented below.
Since every predictor-corrector method considered here is necessarily strongly stable, it is of interest to find a procedure for determining the value $h^*$ for any particular method. A predictor-corrector method is completely determined by specifying the three remaining predictor coefficients. To illustrate this, consider the set of five equations which result when $m = 0, 1, \cdots, 4$ in equations (4.2). This set can be solved for $a_1, a_2, a_3, b_1, \text{ and } b_2$ in terms of $a_0, b_0, \text{ and } b_3$ with the following results.

(6.5.1) \[ a_1 = -17 + 9a_0 - 27b_0 + 3b_3 \]
(6.5.2) \[ a_2 = 9 - 9a_0 + 24b_0 \]
(6.5.3) \[ a_3 = 9 - a_0 + 3b_0 - 3b_3 \]
(6.5.4) \[ b_1 = -6 + 6a_0 - 14b_0 + b_3 \]
(6.5.5) \[ b_2 = -18 + 6a_0 - 17b_0 + 4b_3 \]

It is of interest to consider whether there exist methods for which $h^*$ is infinite. This is not possible as can be seen by replacing $p$ by $\xi^{-1}$ in (6.2) and multiplying the resulting equation by $\xi^4$. The resulting equation has a root which approaches zero as $h \rightarrow -\infty$ unless the coefficients of $h$ and $h^2$ vanish in (6.2) for $j = 0, 1, 2, 3$. This is not possible if equations (6.5) are to be satisfied. Hence, one root of (6.2) is unbounded as $h \rightarrow -\infty$. Therefore, for any method under consideration, $h^*$ is finite. It follows that there exists some negative
h of minimum magnitude at which there occurs a root \( \rho = 1 \), \( \rho = -1 \), or \( \rho = e^{i\theta} \), where \( \theta \neq n\pi \).

Thus, consider these three cases. Recall that the corrector coefficients are determined by (6.3) and the remaining predictor coefficients are specified through (6.5) when \( a_0, b_0, \) and \( b_3 \) are known.

When \( \rho = 1 \), equation (6.2) becomes

\[
F_1(h, a_0, b_0, b_3) = 1 - 3 \sum_{j=0}^{3} [c_j + h(d_j + d_4a_j) + h^2(d_4b_j)] = 0.
\]

This is a quadratic equation in \( h \). Actually, it can be shown that \( h = 0 \) is a root of (6.6), under the conditions already assumed.

When \( \rho = -1 \), equation (6.2) becomes

\[
F_2(h, a_0, b_0, b_3) = 1 - 3 \sum_{j=0}^{3} [c_j + h(d_j + d_4a_j) + h^2(d_4b_j)] (-1)^j = 0.
\]

This equation is also a quadratic in \( h \).

When \( \rho = e^{i\theta} \), \( \theta \neq n\pi \), two equations are obtained from equation (6.2). These result from use of the relation

\[
e^{im\theta} = \cos m\theta + i\sin m\theta,
\]

and from setting the real and imaginary parts of (6.2) equal to zero. The two equations are
\[ G_1(h, a_0, b_0, b_3, \theta) \equiv \cos 4\theta - \sum_{j=0}^{3} [c_j + h(d_j + d_4 a_j)] + h^2(d_4 b_j)] \cos j\theta = 0, \text{ and} \]

\[ G_2(h, a_0, b_0, b_3, \theta) \equiv \sin 4\theta - \sum_{j=0}^{3} [c_j + h(d_j + d_4 a_j)] + h^2(d_4 b_j)] \sin j\theta = 0. \]

When the parameter \( \theta \) is eliminated from \( G_1 \) and \( G_2 \), the equation

\[ F_3(h, a_0, b_0, b_3) \equiv (q_3 - q_1)^2 + q_3(q_3 - q_1)(q_0 - 1) \]

\[ (q_2 - q_0 - 1)(q_0 - 1)^2 = 0 \]

results, where

\[ q_j = -c_j - h(d_j + d_4 a_j) - h^2(d_4 b_j), \]

for \( j = 0, 1, 2, 3 \). Equation (6.8) is a sixth degree polynomial equation in \( h \).

A procedure for the determination of \( h^* \) is now available when any particular method is considered. Namely, solve equations (6.6), (6.7), and (6.8), and from the set of values of \( h \) obtained, select the negative value of least magnitude. A computer program for the solution of these three equations was written.

Consider the predictor formula

\[ P_{n+4} = y_{n+3} + \frac{h}{24} (55y'_{n+3} - 59y'_{n+2} + 37y'_{n+1} - 9y'_{n}). \]
This predictor is sometimes called an Adams formula of the open type and is also called an Adams-Bashforth formula. Hildebrand [6] recommends the use of (6.9) and (6.3) in an iterative predictor-corrector method. Note that (6.9) corresponds to a choice of the parameters \( a_0, b_0 \), and \( c_3 \) equal to 0, -3/8, and 55/24, respectively. By use of the computer program mentioned above, \( h^* \) for the predictor-corrector method defined by (6.9) and (6.3) was found to be equal to -1.2848. At this value of \( h \), a complex pair of roots of magnitude one occurs.

With the preceding techniques available, one possible approach to the problem of minimizing \( h^* \) is to select a set of parameter values and to determine the resulting value of \( h^* \) for each combination of the parameters. The success of such a procedure depends upon having a sufficient number of sample points in the parameter space so that the behavior of \( h^* \) as a function of the parameters can be deduced. The difficulty involved is apparent since, even with only five values of each parameter, equations (6.6), (6.7), and (6.8) would each have to be solved 125 times.

Another approach is based upon the idea of improving a given method, for instance, the method defined by equations (6.9) and (6.3). The following gradient technique was developed to effect such an improvement.
It was mentioned above that at $h = -1.2848$, there occurs a complex pair of roots of magnitude one when $(a_0, b_0, b_3)$ are chosen as $(0, -3/8, 55/24)$. It is now desired to find a way of varying the parameters $(a_0, b_0, b_3)$ from $(0, -3/8, 55/24)$ so that the complex pair of roots of unit magnitude will occur at a more negative value of $h$. Equation (6.8) provides an implicit relation involving $h$, $a_0$, $b_0$, and $b_3$ that is valid when a complex pair of roots is considered. From (6.8), it is possible to compute a gradient vector

\[
V_3(h, a_0, b_0, b_3) \equiv \left( \frac{\partial h}{\partial a_0}, \frac{\partial h}{\partial b_0}, \frac{\partial h}{\partial b_3} \right),
\]

where

\[
\frac{\partial h(j)}{\partial a_0} = -\frac{\partial F_j}{\partial a_0}, \quad \frac{\partial h(j)}{\partial b_0} = -\frac{\partial F_j}{\partial b_0}, \quad \frac{\partial h(j)}{\partial b_3} = -\frac{\partial F_j}{\partial b_3},
\]

and $j = 3$. Thus, $V_3(-1.2848, 0, -3/8, 55/24)$ is the gradient corresponding to the predictor-corrector method defined by (6.9) and (6.3) and to the complex root that limits the interval of stability. The limiting value of $h^*$ can be made to occur at a more negative value by choosing a new set of parameters from the relation
where \( C \) is some positive constant determined experimentally as described below. It is apparent that, when \( \rho = 1 \) is the root which limits the range of stability, a corresponding gradient,

\[
(6.13) \quad \nabla_1(h, a_0, b_0, b_3) \equiv \left( \frac{\partial h^{(1)}}{\partial a_0}, \frac{\partial h^{(1)}}{\partial b_0}, \frac{\partial h^{(1)}}{\partial b_3} \right),
\]

can be computed by letting \( j = 1 \) in equations (6.11). Similarly, when \( \rho = -1 \), the gradient is given by

\[
(6.14) \quad \nabla_2(h, a_0, b_0, b_3) \equiv \left( \frac{\partial h^{(2)}}{\partial a_0}, \frac{\partial h^{(2)}}{\partial b_0}, \frac{\partial h^{(2)}}{\partial b_3} \right),
\]

the latter quantities being computed from (6.11) with \( j = 2 \). A computer program was written to evaluate the gradients (6.10), (6.13), and (6.14).

The procedure for minimizing \( h^* \) can now be described. First, choose some predictor, for instance (6.9). Determine the limiting value, \( h^* \) by solving (6.6), (6.7), and (6.8). In so doing, the limiting root is found to be \( \rho = 1, \rho = -1, \) or \( \rho = e^{i\theta}, \theta \neq n\pi, \) depending upon whether \( h^* \) was obtained from the solution of (6.6), (6.7), or (6.8). Compute the appropriate gradient. Choose some positive constant \( C \), and vary the parameters according to the equation
(6.15) \((a_0^{(k+1)}, b_0^{(k+1)}, b_3^{(k+1)}) = (a_0^{(k)}, b_0^{(k)}, b_3^{(k)})\)

\[-cV_j (h^{(k)}, a_0^{(k)}, b_0^{(k)}, b_3^{(k)}),\]

where \(j = 1,2,3\) depending upon whether (6.13), (6.14), or (6.10) was used, and \(k = 0\) corresponds to the selection of the first predictor formula. For the new parameters, solve (6.6), (6.7), (6.8), and determine the new value of \(h^*\). If this value is satisfactory, compute the appropriate gradient and use (6.15) with \(k\) replaced by \(k + 1\). If not, choose another value of \(C\) and repeat (6.15). After a satisfactory value of \(C\) has been found, compute the appropriate gradient and then use (6.15) with \(k\) replaced by \(k + 1\).

Some complications can arise. For instance, it is possible that, at \(h^*\), both a root \(\rho = 1\) and a root \(\rho = -1\), or some other combination of roots, occurs. In the case of two roots, two gradients are applicable. Unless these gradients are exactly opposite, it is possible to determine a direction in which to move in the parameter space so that the range of stability will be increased. Also, even with only one gradient being considered, it is possible that an attempt to move appreciably in the negative gradient direction results in \(h^*\) occurring at some less negative value. This can happen when the new limiting condition does not correspond to that for which the pre-
vious gradient was computed. In these more complicated cases, one can often devise some technique for continuing the process of minimizing $h^*$ by analyzing a plot of the roots as functions of $h$. Data for such a plot can be obtained by solving (6.2) directly for various values of $h$.

The minimization procedure based on the use of the gradients described above was carried out with the aid of the computer programs already mentioned, starting with the predictor, (6.9). The final result was a predictor formula with coefficients

$$
\begin{align*}
(6.16) \quad a_0 &= -0.69735280, \quad b_0 = -0.71432005, \\
&\quad a_1 = 2.0172069, \quad b_1 = 1.8186108, \\
&\quad a_2 = -1.8675052, \quad b_2 = -2.0316877, \\
&\quad a_3 = 1.5476511, \quad b_3 = 2.0022473.
\end{align*}
$$

The predictor-corrector method defined by (6.16) and (6.3) is $h$ stable for $-2.4809 \leq h \leq 0$. Recall that the method defined by (6.9) and (6.3) is $h$ stable for $-1.2848 \leq h \leq 0$. Hence, the interval of stability was nearly doubled by use of the gradient technique.

The value of $h^* = -2.4809$ has not been shown to be a minimum, either relative or absolute. It is possible that by starting with a different predictor, or by carrying out the gradient procedure with different choices of the constant $C$ in (6.15), for instance, that a different value of $h^*$ might be obtained. This, however, does not
negate the fact that a predictor-corrector method with a considerably increased range of $h$ stability has been derived.

The application described above is one of many that could be devised, based upon the theory developed in the preceding chapters. The following problems indicate some of the possibilities. In the class of predictor-corrector methods defined by equations (6.1), it would be of interest to minimize $h^*$ with the restriction that the predictor and the corrector be of some order other than that considered above. Also, the effect of allowing the corrector to vary, as well as the predictor, might lead to worth-while results. Such analyses could also be made for C-modified and PC-modified predictor-corrector methods. It is possible that such investigations, in addition to providing particular numerical methods of practical significance, might lead to the development of some theory regarding the maximum range of $h$ stability.
BIBLIOGRAPHY


ACKNOWLEDGMENTS

I wish to thank Dr. Robert J. Lambert for his interest and valuable guidance during my years of graduate study. His encouragement and assistance in the preparation of this dissertation are also greatly appreciated.

I would like to thank Dr. Ralph W. Klopfenstein for his many helpful suggestions given at RCA Laboratories during the summer of 1961 when the work described briefly in Chapter VI was done. The assistance of RCA Laboratories in printing this dissertation is sincerely appreciated.