A-STABILITY AND COMPOSITE MULTISTEP METHODS

by

WILLIAM BENJAMIN RUBIN

B.S., Clarkson College, 1965

M.S., Syracuse University, 1968

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering in the Graduate School of Syracuse University, May, 1973.

Approved Theodore A Bickart

Date 20 April 1973

ABSTRACT

Many dynamical systems of practical importance in engineering are modelled by stiff ordinary differential equations, that is, equations $\dot{x}(t) = f[x(t),t]$ in which the Jacobian matrix $f'[x(t),t]$ has instantaneous eigenvalues widely distributed in the complex plane. The last decade has seen the emergence and acceptance of A-stability as an appropriate property of numerical methods suitable for solving stiff equations. To achieve high order together with A-stability, it has been proposed to investigate composite multistep methods.

This work contains a careful and comprehensive study of composite multistep methods, as well as a new general characterization of the A-stability property. Composite multistep methods are a generalization of multistep methods in that n equations in m past points are solved simultaneously for n or fewer future points. The present study shows that when a composite multistep method is applied to a linear autonomous differential equation, the generated approximating sequences satisfy a linear matrix difference equation. From this equation, the characteristic polynomial of a composite multistep method is derived in three different formulations, two of them new. The A-stability criterion for polynomials in two variables is defined, and it is shown that a composite multistep method whose poles lie in the closed right half complex plane is A-stable if and only if its characteristic polynomial satisfies the A-stability criterion.

The introduction of the A-stability criterion has the advantage of allowing A-stability to be studied separately from the properties of a particular class of methods. General analytic and algebraic approaches to the determination of A-stability are explored at length in this work. A rigorous derivation of the analytic approach (the "locus technique") for characterizing the A-stability criterion is presented using the concept of the Riemann surface induced by a polynomial in two variables. It is shown that in order for a polynomial in the variables lambda and zeta to satisfy the A-stability criterion, it is necessary and, together with certain side conditions, sufficient that its zeta locus lie inside the closed unit disc. One of the side conditions is that the poles lie in the closed right half plane. A dual result is also given.

The most important result in this work consists in an algebraic characterization of the A-stability criterion for polynomials in two variables. The characterization is given in various primary and dual forms, which are based on the analytic characterizations and the classical theorems of Hurwitz and Sturm. For each form of the algebraic characterization, the necessary and sufficient conditions are reducible to the addition and multiplication of polynomials with integral coefficients, and the examination of the signs of integers. The simplest form of the algebraic characterization has been implemented in APL/360 using the technique of infinite precision integer arithmetic. This implementation is free from roundoff error, and therefore determines with absolute certainty whether a given characteristic polynomial satisfies the A-stability criterion. The algebraic characterization is useful as a direct test of A-stability, not only for composite multistep methods, but also for most other classes of methods for the numerical solution of ordinary differential equations. It can also be used as the basis for a search procedure by which new high-order A-stable methods can be obtained in a systematic way. Other possible applications of the algebraic approach include extensions to deal with stiff stability, and with stability under varying step size.

The last chapter is devoted to a discussion of various fundamental properties of composite multistep methods not directly related to A-stability. Relations for the discretization error and order are derived, and it is shown that the exact order of a composite multistep method, defined in terms of the truncation error, is equal to the minimum of the exact orders of its discretization errors associated with the future points. The concepts of equivalence and canonical form for composite multistep methods are defined, and it is shown that if only such intrinsic properties as A-stability, order, and discretization error constant are of interest, then attention can be restricted to the subclass of canonical composite multistep methods without loss in generality. Particularly noteworthy are the derivations of useful parametric representations for classes of high-order composite multistep methods and their characteristic polynomials. The work concludes with a summary of the present state of knowledge concerning high-order A-stable composite multistep methods, and the construction of some new methods.

ACKNOWLEDGMENTS

I would like to extend my deep appreciation to Theodore A. Bickart for the encouragement, enthusiasm, and guidance which helped make this work possible. He introduced me to stiff differential equations, A-stability, and composite multistep methods, and he provided me and other students with a research environment in which we could all share our thoughts on these subjects. I would like to thank the National Science Foundation and Prof. Bickart, through whom a Research Assistantship was made available to me, under NSF Grant GK-23010. I am deeply indebted to Joel M. Tendler for valuable discussions and numerous specific suggestions, and for unselfishly giving of his time on my behalf. In particular, the discovery of new A-stable composite multistep methods presented in Chapter 4 would not have been possible without his assistance.

A major factor in the development of many of the important results in this work has been the APL/360 interactive programming system at Syracuse University. This tool of research deserves special mention because its influence at many points in the work is not readily apparent. Much of the computing time was supported by the Department of Electrical and Computer Engineering.

Professor Bickart's very careful reading of the manuscript is gratefully acknowledged. Many thanks are also due Louise Capra for a professional typing job, and Zdenek Picel for special help with proofreading.

I am extremely fortunate to have been given encouragement and support in my studies by many friends and relatives. I am especially grateful to Jay M. Land for helping me overcome many personal difficulties. The attainment of the Ph.D. degree represents years of dedicated effort not only by myself, but also by Judith A. Rubin, my devoted wife. For her faith and confidence I am grateful beyond words.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

The analysis or design of engineering systems frequently involves obtaining the transient responses of specific nonlinear system models with specific initial conditions. These computations generally require the application of approximate numerical techniques on a digital computer. Many dynamical systems of practical importance in engineering are modelled by stiff ordinary differential equations, that is, equations $\dot{x}(t) = f[x(t),t]$ in which the Jacobian matrix $f'[x(t),t]$ has instantaneous eigenvalues widely distributed in the complex plane. Equivalently, a stiff system is one whose solutions exhibit both fast transients and relatively slow transients. Stiff equations arise naturally as models for solid-state switching circuits [Brayton, et al], nuclear reactor control systems [Little, et al], photochemical smog [Gelinas], and various other problems of chemical kinetics [Moretti] and applied physics [Gelinas]. A bibliography for stiff systems (containing 134 entries) is given in [Enright].

Most classical methods for the numerical solution of ordinary differential equations are poorly suited to stiff problems. This well-known fact can be explained as follows: The time interval over which the solution is desired is proportional to the slowest time constant of the system, while the maximum step size consistent with numerical stability of classical methods is proportional to the fastest time constant of the system. Therefore, the number of steps, and with it the number of computations, is proportional to the ratio between the slowest and the fastest time constants of the system. For stiff problems this ratio is frequently on the order of 10^4 to 10^6. Thus, classical methods are impractical for stiff problems. Detailed analyses supporting the above statements have appeared numerous times in the literature, for example [Sandberg and Shichman], [Fowler and Warten], [Gelinas], [Moretti].

In 1963 the fundamental paper of [Dahlquist] described an approach to the numerical solution of stiff equations which has since become generally accepted. Dahlquist's suggestion is, roughly speaking, to use solution methods which are numerically stable for any step size. He has termed such methods A-stable*. The maximum step size for A-stable methods is not limited by stability considerations, but only by accuracy considerations. In typical stiff

*The term "A-stable" has been interpreted by some writers to mean "absolutely stable (in the left half plane)" [Gelinas].

problems the initial step size is on the order of the fastest time constants. After a reasonable number of steps the amplitude of the fast transients will have become negligible, and the step size can be increased--perhaps by many orders of magnitude--to the size of the next fastest time constants, etc. This approach is viable only when the solution method has the crucial property of A-stability*.

[Dahlquist] began the investigation of A-stability for multistep methods [Henrici]. One of his most important results is that there do not exist A-stable multistep methods with orders of accuracy exceeding two. This fact is unfortunate, because methods of order as high as 10 or 15 have been shown to allow greater overall computational efficiency for many problems than comparable lower order methods [Ehle, 1972], [Hull], [Hull, et al]. Thus, the first important goal of current research in stiff problems is to obtain high-order A-stable (or at least stiffly stable) solution methods which are practical and efficient to implement in computer algorithms.

At the present writing dozens of new classes of high-order A-stable methods have been presented in the recent literature. This fact shows that there is a high degree of interest in stiff problems, that presently available methods for stiff problems have not proved to be entirely satisfactory, and that better methods are believed likely to be discovered. Some (but by no means all) of the classes of methods which have recently been proposed can be viewed as generalizations of the class of multistep methods. Three of the directions such generalizations have taken are as follows: higher derivative methods [Liniger and Willoughby], averaging multistep methods [Odeh and Liniger, 1972], and composite multistep methods.

The class of composite multistep methods was first introduced and proposed for use in stiff problems by [Sloate and Bickart]. See also [Sloate]. To this writer's knowledge, the only other reference which treats the general class of composite multistep methods is the announcement [Rubin and Bickart]. Certain subclasses, however, have been treated by other researchers. Most prominent are the A-stability analyses for multistep methods of [Dahlquist] and later researchers. The only other known subclasses treated are as follows: The subclass of composite one-step methods was treated in [Watts]. See also [Shampine

---

*Actually, certain properties which are somewhat weaker than A-stability have been found adequate by some researchers. In fact, notable success has been achieved using multistep methods which are only "stiffly stable" [Gear].

and Watts], which contains early references on the subject, [Watts and Shampine], [Hulme], [Handbook of Mathematical Tables], [Bickart, et al], and [Bickart and Picel]. Watts discovered a family of A-stable composite one-step methods whose n-th member has n equations (therefore n future points), and has accuracy of order n, for all n = 1, 2, 3, ... The subclass of cyclic composite multi-step methods was treated in [Donelson and Hanson], but only from the viewpoint of non-stiff problems. The particular methods proposed by [Donelson and Hanson] happen not to be A-stable or stiffly stable*, and therefore are of no use for stiff problems.

It is profitable to compare composite multistep methods with multistep methods, since the latter are among the most successful methods known for stiff problems [Gear]. The multistep methods used by Gear are limited in order of accuracy to six, with the 6-th order formula having only marginally acceptable stability properties. By comparison, there are reasons for believing that a family of A-stable composite multistep methods exists whose n-th member has n equations and accuracy of order 2n [Sloate and Bickart]. Such methods are thought to be considerably less efficient (in terms of number of computations per step) than multistep methods, but considerably more efficient than Watts' methods. However, cyclic composite multistep methods are almost as efficient as multistep methods. Since the former class contains the latter, there is hope that higher order methods with better stability properties can be found in the larger class. If so, then a significant improvement in performance can be expected. Some recent progress in this direction will be announced in the near future by [Tendler].

One goal of the present work is to give a rigorous treatment of some of the fundamental properties of composite multistep methods in general. This is attempted in Chapters 1 and 4. The second goal, motivated by the A-stability criterion of Chapter 1, is to characterize A-stability for polynomials in two variables. The results appear in Chapters 2 and 3. It turns out that the characterizations apply not only to composite multistep methods, but also to most of the other known classes of methods proposed for numerical solution of ordinary differential equations.

---

*Private communication, Joel M. Tendler, Syracuse University Research Corporation, Syracuse, New York (1971).

In Chapter 1 the class of composite multistep methods is defined, as is the concept of A-stability. Characteristic polynomials for composite multistep methods are derived in three different formulations (two of them new), and it is shown that they are intimately related to A-stability. More precisely, a given composite multistep method is A-stable if and only if its characteristic polynomial satisfies the A-stability criterion defined in Chapter 1. Chapter 1 also contains the first rigorous treatment of certain existence and uniqueness questions related to the approximating sequences generated by composite multistep methods. Several new terms are defined for use in Chapter 1, the most important ones relating to poles and to regular composite matrices.

Chapter 2 continues the exploration of polynomials in two variables begun in Chapter 1. A-stability characterizations using primary and dual locus methods are derived. The theory is based on the concept of the Riemann surface induced by a polynomial in two variables.

Chapter 3 presents primary and dual algebraic characterizations of the A-stability criterion, based on the locus methods of Chapter 2 and the classical theorem of Hurwitz. These new results can be used as practical tests to determine whether a characteristic polynomial satisfies the A-stability criterion, and hence whether a given composite multistep method is A-stable. Although motivated by composite multistep methods, the tests are actually applicable to every known class of methods for numerical solution of ordinary differential equations.

Chapter 4 is devoted to a discussion of various fundamental properties of composite multistep methods not directly related to A-stability. First, equivalence and canonical form are explored. Then, relations for the discretization error and order of accuracy are derived. Finally, examples are presented of new high-order composite multistep methods whose A-stability has been verified by the methods of Chapter 3.

In order that this Dissertation be as readable as possible, the formal parts (Chapters 1 through 4 and the Appendices) have been written to form an essentially self-contained unit. In particular, all specialized technical terms are defined when they are first used; also, specialized results from the literature are not depended upon to replace formal proofs, except in a few cases for which this seemed the only practical approach. References to

the literature are cited using square brackets enclosing the surname(s) of the author(s). In ambiguous cases the publication year, and if necessary an identifying letter, are also included. The end of the Dissertation contains two other reference aids to the reader, a List of Frequently Occurring Symbols, and an Index of Defined Terms.

It is intended that the four chapters be read in their order of appearance, but with each appendix read when it is first cited. However, Chapter 4 can be read directly after Chapter 1, since it is essentially independent of Chapters 2 and 3. Also, large portions of Chapters 2 and 3 are independent of Chapter 1.

Chapter 1

COMPOSITE MULTISTEP METHODS: CHARACTERISTIC POLYNOMIALS AND A-STABILITY

§1.1 Introduction to Composite Multistep Methods

Consider the first-order vector ordinary differential equation

$$\dot{X}(t) = f[X(t),t] \,, \tag{1-1}$$

where t is a real parameter (considered to be time), f is a given (vector-valued) function, X is an unknown differentiable vector-valued function of t, and $\dot{X}$ is the derivative of X. If f is sufficiently smooth*, then for every initial condition

$$X(t_0) = X_0 \tag{1-2}$$

(t_0 and X_0 given), there exists a unique differentiable vector-valued function X on $[t_0,\infty)$ satisfying (1-1) and (1-2). Only such smooth problems are of interest with respect to numerical solutions in the present work. Most practical interest is in the situation where f maps into a real Euclidean space; however, for theoretical purposes it is necessary to consider also the case where f is complex valued.

The task in numerical solution of the initial value problem (1-1) with (1-2) can be stated as follows: To compute a "suitable" monotonic sequence $\{t_i\}$ in $[t_0,\infty)$ and a vector-valued sequence $\{x_i\}$ approximating the true solution X in the sense that x_i is near to $X(t_i)$ for each i = 0, 1, 2, ... A simplification important in theory, although not expected in practice, is to require the time sequence $\{t_i\}$ to be uniform; that is, assume $t_i = t_0 + ih$ for i = 0, 1, 2, ... , where h is a positive constant called the step size.

The class of methods to be considered in this work for computing the approximating sequence $\{x_i\}$ is the class of composite multistep methods. A composite multistep method can be defined as an ordered pair (R,k), where R, the composite matrix, is an $n \times 2(m+n)$ matrix of real numbers α_{ij} and β_{ij},

---

*A standard sufficiency condition is as follows: (1) f is continuous, and, (2) f is locally Lipschitz in its first argument [Coddington and Levinson].

$$R = \begin{bmatrix} \alpha_{10} & \alpha_{11} & \cdots & \alpha_{1(m+n-1)} & \beta_{10} & \beta_{11} & \cdots & \beta_{1(m+n-1)} \\ \alpha_{20} & \alpha_{21} & \cdots & \alpha_{2(m+n-1)} & \beta_{20} & \beta_{21} & \cdots & \beta_{2(m+n-1)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{n0} & \alpha_{n1} & \cdots & \alpha_{n(m+n-1)} & \beta_{n0} & \beta_{n1} & \cdots & \beta_{n(m+n-1)} \end{bmatrix} , \qquad (1\text{-}3)$$

and k, the number of retained points, is a positive integer. The composite matrix R uniquely determines the two integers m, the number of past points, and n, the number of future points. It is required that m and n be positive, and that $k \leq n$.

For each smooth function f in (1-1), initial condition t_0, X_0 in (1-2), and positive step size h, the composite multistep method (R,k) specifies* an approximating sequence $\{x_i\}$ as follows: Put $x_0 = X_0$, and compute $x_1, x_2, \ldots, x_{m-1}$ using an external starting procedure. Set $y_i = x_i$ for $i = 0, 1, \ldots, m - 1$. With $\ell = 0$ solve the n algebraic equations

$$\sum_{j=0}^{m+n-1} \{\alpha_{ij} y_j - h\beta_{ij} f[y_j, t_0 + (k\ell+j)h]\} = 0 , \qquad i = 1, 2, \ldots, n , \qquad (1\text{-}4)$$

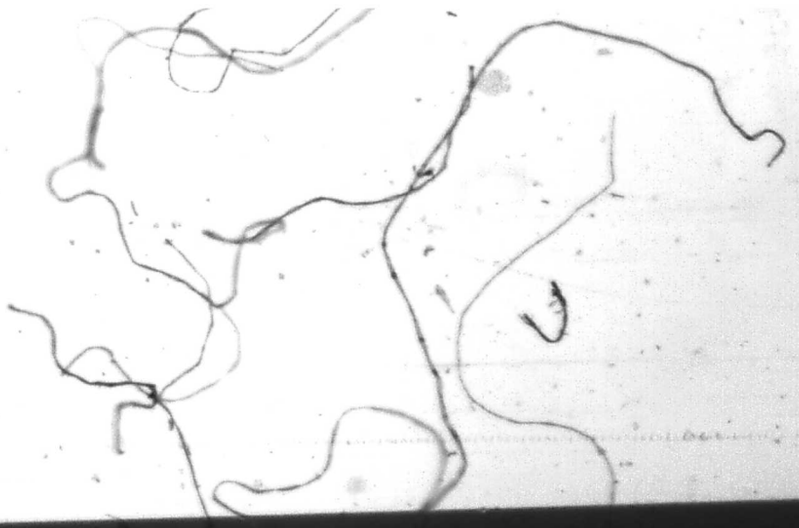
for $y_m, y_{m+1}, \ldots, y_{m+n-1}$; retain the first k elements of this result as the values of $x_m, x_{m+1}, \ldots, x_{m+k-1}$, discarding the remaining n - k. The solution is thus advanced k points at the 0-th iteration. In general, the ℓ-th step of the iteration, $\ell = 0, 1, 2, \ldots$, proceeds as follows:

1) Set $y_j = x_{k\ell+j}$, $\quad j = 0, 1, \ldots, m - 1.$
2) Solve (1-4) for y_j, $\quad j = m, m+1, \ldots, m+n-1.$
3) Set $x_{k\ell+j} = y_j$, $\quad j = m, m+1, \ldots, m+k-1.$

The approximating sequence $\{x_i\}$ is thus advanced a "block" of k points at each iteration by using the n equations (1-4) in m past points to solve for n future points, but retaining only the first k. For n = 1 the above becomes a multistep method [Henrici]; for m = 1 (and k = n) the above becomes a composite one-step method [Shampine and Watts]. Composite multistep methods thus form a natural generalization and unification of these two classes of methods.

---

*Subject to existence, uniqueness, and starting conditions to be discussed later.



If the differential equation (1-1) is linear, that is, if f is linear in its first argument, then the algebraic equation (1-4) which must be solved at each iteration is linear. For this reason the composite multistep methods are considered to be linear methods. Of course, the important case in practice is nonlinear differential equations; in this case it can be seen that nonlinearities are introduced into the algebraic equation (1-4), requiring, in principle, an iterative solution for each $\ell = 0, 1, 2, \ldots$.

§1.2 A-Stability and Linear Differential Equations

For a given function f and a given positive step size h, a composite multistep method (R,k) is said to be asymptotic to the origin if for every starting condition $x_0, x_1, \ldots, x_{m-1}$, the approximating sequence $\{x_i\}$ exists, is unique, and approaches zero as $i \to \infty$.

1.1. Definition: A composite multistep method is said to be A-stable if for every asymptotically stable linear autonomous differential equation and every positive step size, the composite multistep method is asymptotic to the origin.

In this section the significance of this definition is explored. First the class of differential equations of interest is characterized.

The differential equation (1-1) is linear and autonomous if f is of the form

$$f(\zeta,t) = q\zeta \tag{1-5}$$

where q is a square matrix of real (or complex) numbers. It is well known that (1-1) with (1-5) is asymptotically stable if and only if all eigenvalues of q are in the open left half complex plane $\mathcal{L} = \{\lambda \in \mathbb{C}: \operatorname{Re}\lambda < 0\}$. It can be shown that for the purposes of A-stability, attention can be restricted to one-dimensional problems in which q is a 1×1 matrix (or, equivalently, a scalar, as will be assumed here)*. In such case (1-1) with (1-5) is asymptotically stable if and only if $q \in \mathcal{L}$.

*For example, asymptotic properties of (1-1) or the composite multistep method are invariant with respect to a change of basis for $x(t)$. Therefore it is sufficient to consider only matrices q in Jordan canonical form; the eigenvalues of q are just its diagonal elements.

Consider the application of a composite multistep method (R,k) to a one-dimensional linear autonomous differential equation. With $q \in \mathcal{C}$ substitute (1-5) into (1-4) to give

$$\sum_{j=0}^{m+n-1} (\alpha_{ij} - \lambda\beta_{ij}) y_j = 0, \qquad i = 1, 2, \ldots, n, \tag{1-6}$$

where $\lambda = qh$. Note that the algebraic equation no longer depends explicitly on the iteration ℓ. Furthermore, since $qh = \lambda$, the condition "$q \in \mathcal{L}$ and $h > 0$" is equivalent to the condition "$\lambda \in \mathcal{L}$". The above discussion and Definition 1.1 lead to the following:

1.2. Proposition: A composite multistep method is A-stable if and only if for every λ in the open left half plane $\mathcal{L}$ and every starting condition, the approximating sequence $\{x_i\}$ generated through (1-6) exists, is unique, and approaches zero as $i \to \infty$.

In order to write (1-6) in matrix notation, let

$$Y_p = [y_0 \;\; y_1 \;\; \cdots \;\; y_{m-1}]^T \tag{a}$$

$$Y_f = [y_m \;\; y_{m+1} \;\; \cdots \;\; y_{m+n-1}]^T \tag{b}$$

(1-7)

(superscript T denotes transpose), and partition the composite matrix (1-3) along its columns as

$$R = [A_p \;\; A_f \;\; B_p \;\; B_f] \tag{1-8}$$

where A_p and B_p are the $n \times m$ matrices of "past" α's and β's, and A_f and B_f are the $n \times n$ matrices of "future" α's and β's. Then (1-6) with $\lambda = qh$ can be written as

$$(A_f - \lambda B_f) Y_f = -(A_p - \lambda B_p) Y_p \,. \tag{1-9}$$

Define the function D mapping the complex plane $\mathcal{C}$ into the set of $n \times n$ complex matrices as follows:

$$D(\lambda) = A_f - \lambda B_f \,. \tag{1-10}$$

Such a function is called a pencil of matrices [Gantmacher, vol. 2, ch. 12].

Define the function δ by

$$\delta(\lambda) = \det D(\lambda) . \tag{1-11}$$

If δ is the zero function, D is called a singular pencil; otherwise δ is a polynomial with real coefficients of degree n or less*, and D is called a regular pencil. The composite matrix R (and hence the composite multistep method (R,k)) is called singular [regular] whenever D is a singular [regular] pencil. Let Λ be the set of zeros** of δ: $\Lambda = \{\lambda \in C : \delta(\lambda) = 0\}$. The elements of Λ will be called the poles of the composite multistep method. The poles of a method are important for both theoretical and practical reasons, some of which are discussed in this section.

1.3. Proposition: Let a composite multistep method (R,k) be applied with a given step size to a one-dimensional linear autonomous differential equation. If λ is not a pole, then there exists a unique approximating sequence $\{x_i\}$ for each starting condition.

---

*At many points in this work, functions arise which are either polynomials or the zero function. The term "polynomial" will always be used in this work to refer to a function f which can be represented in the form $f(x) = \sum_{i=0}^{j} f_i x^i$, where $j \geq 0$ and $f_j \neq 0$. According to this definition, the zero function is not a polynomial. The distinction between polynomials and the zero function turns out to be of crucial importance at many points, and will always be maintained. The exact degree j of f is also important in some cases. Except in Chapter 3, all polynomials in this work have real coefficients.

**If δ is of degree n, for example, then Λ contains n elements, provided all zeros of δ have multiplicity one. In general Λ contains from one to n elements, depending upon the multiplicities of the zeros of δ.

It will be seen in most cases in this work that only the zeros of polynomials, and not the polynomials themselves, are of interest. In such cases one can always replace a given polynomial by another which differs from it by a trivial factor (a nonzero real constant factor). This fact is of theoretical importance in this work, and it is also of considerable practical importance, as shown in Chapter 4.

Proof: At each iteration, the method produces a set of n linear algebraic equations in n unknowns Y_f. Combining (1-9) and (1-10), these can be written as

$$D(\lambda)Y_f = -(A_p - \lambda B_p)Y_p \quad . \tag{1-12}$$

There exists a unique solution to (1-12), namely

$$Y_f = -[D(\lambda)]^{-1}(A_p - \lambda B_p)Y_p \quad ,$$

if and only if $D(\lambda)$ is nonsingular, that is, if and only if λ is not a pole of the method. The existence of a unique solution to (1-12) for every right-hand side is sufficient to guarantee the existence and uniqueness of $\{x_i\}$, proving the proposition. □

The converse of Proposition 1.3 does not hold in general. For example, consider the composite multistep method

$$\left(\begin{bmatrix} -1 & 1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 1 & -1 & 4 & -1 \end{bmatrix}, 1 \right) \quad . \tag{1-13}$$

For this composite matrix (1-12) becomes

$$\begin{bmatrix} 1-\lambda & 0 \\ -4\lambda & 1+\lambda \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1-\lambda \end{bmatrix} [y_0] \quad .$$

By inspection the poles are $\Lambda = \{1,-1\}$. At the pole $\lambda = -1$ the above becomes

$$\begin{bmatrix} 2 & 0 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} [y_0] \quad ,$$

which has the general solution

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_0/2 \\ c \end{bmatrix} \quad ,$$

where c is an arbitrary constant. Since k = 1, the arbitrary value of y_2 is discarded; only the unique value $y_1 = y_0/2$ (which always exists) is retained as the next element of the approximating sequence $\{x_i\}$.

For the important special case in which all future points are retained, the converse of Proposition 1.3 holds, since if k = n, nonunique elements of Y_f, which might otherwise be discarded, are retained. The situation for k = n can be summarized as follows:

1.4. Proposition: Let a composite multistep method (R,n) with n future points be applied with a given step size to a one-dimensional linear autonomous differential equation. There exists a unique approximating sequence $\{x_i\}$ for given starting conditions if and only if λ is not a pole of (R,n).

It will be noted that (1-13) gives the same approximating sequence in every case as the classical implicit Euler method ([-1 1 0 1],1), which is known to be a useful method. In general, two methods for the numerical solution of ordinary differential equations (not necessarily composite multistep methods) will be called weakly equivalent if they generate for every function f, step size h, and starting condition the same approximating sequence $\{x_i\}$. It is clear that weak equivalence is an equivalence relation. The only differences between two weakly equivalent methods are algorithmic complexity (such as number of computations), and roundoff error properties (caused by finite precision arithmetic). Equivalence concepts are discussed further in Chapter 4.

Suppose λ is a pole, and suppose existence is satisfied in (1-12) for all Y_p. Then an approximating sequence $\{x_i\}$ exists. But since λ is a pole, the solution Y_f to (1-12) is not unique. In general, for $k < n$, $\{x_i\}$ may or may not be unique. Example (1-13) is typical of those cases in which $\{x_i\}$ is unique. In such cases the nonunique future points in Y_f are discarded in such a way that their values make no contribution to the values of the retained points. In other words, the retained points are independent of some of the values of f in (1-4) corresponding to the future times m + k, m + k + 1, ..., m + n - 1. The potential advantage of future information for improving accuracy and stability is thus never realized. This situation is distinct from the case when λ is not a pole; if $k < n$ and λ is not a pole, then all future information usually* contributes to the values of the retained points.

*One situation in which this does not occur is for cyclic composite multistep methods. A composite multistep method is said to be cyclic [Donelson and Hanson] if A_f and B_f are both lower triangular. (Incidentally, cyclic methods with $k \neq n$ are among the most useful of composite multistep methods, since the n simultaneous equations (1-4) are already triangularized, and can be solved sequentially.)

As a practical matter, it is a burden to solve singular algebraic equations such as those arising when λ is a pole. This fact argues against the application of composite multistep methods at poles.

According to Proposition 1.2 the solutions of (1-12) must be examined for all λ in the open left half plane $\mathcal{L}$, if A-stability is to be determined. For the various reasons discussed above it is desirable, and in many cases necessary, to restrict our attention to those composite multistep methods having no poles in $\mathcal{L}$ ($\Lambda \cap \mathcal{L}$ empty). For example, this restriction is necessary in the important case $k = n$ (by Proposition 1.4), since existence and uniqueness are pre-conditions for A-stability.

If (R,k) is a singular composite multistep method, then, by definition $\Lambda = C$. Thus $\Lambda \cap \mathcal{L} = \mathcal{L}$, which is nonempty. By Proposition 1.4 such methods can never be A-stable if $k = n$. Even if $k < n$ the above remarks show that singular methods are most undesirable, since every point is a pole. Note that the condition "$\lambda \notin \Lambda$", which will be used frequently in this chapter, is vacuous for singular composite multistep methods. Putting it another way, the statement "$\Lambda \cap \mathcal{L}$ is empty" can hold only for regular methods.

§1.3 The Difference Equation

In the last section it was shown that when a composite multistep method is applied to a one-dimensional linear autonomous differential equation, the algebraic equation to be solved at each iteration is of the form (1-12). In this section it is shown that the corresponding approximating sequence $\{x_i\}$ satisfies a linear autonomous difference equation. Furthermore, the stability properties of this equation are governed by a polynomial in two variables.

Let Δ be the $n \times n$ matrix-valued function defined by

$$\Delta(\lambda) = \text{adj } D(\lambda) . \tag{1-14}$$

The elements of Δ are thus polynomials in λ of degree $n - 1$ or less, or the zero function. Since

$$\Delta(\lambda)D(\lambda) = \delta(\lambda)I_n \tag{1-15}$$

holds identically, where I_n is the $n \times n$ identity matrix, $\Delta(\lambda)$ is nonsingular whenever $\lambda \notin \Lambda$. In such case (1-12) is equivalent to

$$\Delta(\lambda)(A_p - \lambda B_p)Y_p + \delta(\lambda)Y_f = 0 . \tag{1-16}$$

At each iteration only the first k elements of Y_f are to be retained; thus let

$$Y_r = J_{kn} Y_f \ , \tag{1-17}$$

where J_{kn} denotes the submatrix of I_n consisting of its first k rows. Premultiplying (1-16) by J_{kn}, multiplying (1-17) by $\delta(\lambda)$, and combining the two results gives

$$J_{kn}\Delta(\lambda)(A_p - \lambda B_p)Y_p + \delta(\lambda)Y_r = 0 \ . \tag{1-18}$$

The above relation (together with (1-17)) is equivalent to (1-16) (with respect to Y_r), and hence to the algebraic equation (1-12), whenever $\lambda \notin \Lambda$. However, since (1-18) does not explicitly involve the n - k discarded elements of Y_f, it can be used to write a relation directly in terms of the approximating sequence $\{x_i\}$. To do so, define the number of past blocks M to be the smallest integer not less than m/k, and let N = kM - m. Thus $0 \le N < k$, with equality if and only if k divides m. Also let E be the $k \times (M+1)k$ matrix-valued function formed by catenating three matrices as follows:

$$E(\lambda) = [0_{kN}, \quad J_{kn}\Delta(\lambda)(A_p - \lambda B_p), \quad \delta(\lambda)I_k] \ , \tag{1-19}$$

where 0_{kN} is the $k \times N$ zero matrix. The elements of E are thus polynomials of degree n or less, or the zero function. Partition E along its columns into M + 1 square matrices as follows:

$$E = [E_0 \ \ E_1 \ \ \cdots \ \ E_M] \ . \tag{1-20}$$

By (1-19) the first N columns of E_0 are zero. Furthermore

$$E_M(\lambda) = \delta(\lambda)I_k \ ; \tag{1-21}$$

hence $E_M(\lambda)$ is nonsingular whenever $\lambda \notin \Lambda$. The above notation can be used to write (1-18) as

$$[E_0(\lambda) \ \ E_1(\lambda) \ \ \cdots \ \ E_M(\lambda)] \begin{bmatrix} Y_p \\ \\ Y_r \end{bmatrix} = 0 \ . \tag{1-22}$$

Now define

$$X_i = [x_{ki-N} \ \ x_{ki+1-N} \ \ \cdots \ \ x_{k(i+1)-1-N}]^T \ , \quad i = 0, 1, 2, \cdots , \tag{1-23}$$

where $x_{-N}, x_{1-N}, \cdots, x_{-1}$ are arbitrary. For $\lambda \notin \Lambda$ the algebraic equation used to define the approximating sequence $\{x_i\}$ has been shown to be (1-22). According to (1-23) and the definition of $\{x_i\}$ it follows that for $\lambda \notin \Lambda$

$$[E_0(\lambda) \;\; E_1(\lambda) \cdots E_M(\lambda)] \begin{bmatrix} X_i \\ X_{i+1} \\ \vdots \\ X_{i+M} \end{bmatrix} = 0 , \qquad i = 0,1,2,\ldots \quad (1\text{-}24)$$

The above relation, written in the form

$$\sum_{j=0}^{M} E_j(\lambda) X_{i+j} = 0 , \qquad i = 0,1,2,\ldots \quad (1\text{-}25)$$

is seen to be an M-th order linear autonomous difference equation for the sequence $\{X_i\}$ of k-vectors. It is possible to investigate the stability properties of (1-25) directly in terms of its characteristic polynomial $\tilde{P}$ defined by

$$\tilde{P}(\lambda,\zeta) = \det \sum_{j=0}^{M} E_j(\lambda)\zeta^j . \qquad (1\text{-}26)$$

Such an approach [Dejon], [Stoate and Bickart] requires complicated arguments. An approach believed to be simpler is developed in Sections 1.3 through 1.5.

The basic idea is to transform the M-th order difference equation on k-vectors into a first order difference equation on kM-vectors. Thus, consider M successive iterations of (1-24), written as

$$\left[\begin{array}{ccc|ccc} E_0(\lambda) & E_1(\lambda) \cdots & E_{M-1}(\lambda) & E_M(\lambda) & 0 \cdots & 0 \\ 0 & E_0(\lambda) \cdots & E_{M-2}(\lambda) & E_{M-1}(\lambda) & E_M(\lambda) \cdots & 0 \\ \vdots & \vdots \;\; \ddots & \vdots & \vdots & \vdots \;\; \ddots & \vdots \\ 0 & 0 \cdots & E_0(\lambda) & E_1(\lambda) & E_2(\lambda) \cdots & E_M(\lambda) \end{array}\right] \left[\begin{array}{c} X_i \\ X_{i+1} \\ \vdots \\ X_{i+M-1} \\ \hline X_{i+M} \\ X_{i+M+1} \\ \vdots \\ X_{i+2M-1} \end{array}\right] = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \quad (1\text{-}27)$$

Let

$$G = \begin{bmatrix} E_0 & E_1 & \cdots & E_{M-1} \\ 0 & E_0 & \cdots & E_{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_0 \end{bmatrix} , \qquad (a)$$

(1-28)

$$H = \begin{bmatrix} E_M & 0 & \cdots & 0 \\ E_{M-1} & E_M & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ E_1 & E_2 & \cdots & E_M \end{bmatrix} , \qquad (b)$$

and

$$\hat{X}_i = [X_{Mi} \quad X_{Mi+1} \quad \cdots \quad X_{M(i+1)-1}]^T , \qquad i = 0, 1, 2, \cdots \qquad (1\text{-}29)$$

In this notation every M-th equation of the form (1-27) can be written as

$$[G(\lambda) \quad H(\lambda)] \begin{bmatrix} \hat{X}_i \\ \\ \hat{X}_{i+1} \end{bmatrix} = 0 \quad , \; i = 0, 1, 2, \cdots .$$

This equation is seen in the form

$$H(\lambda)\hat{X}_{i+1} = -G(\lambda)\hat{X}_i \quad , \qquad i = 0, 1, 2, \cdots \qquad (1\text{-}30)$$

as a first order linear difference equation. From (1-28b) and (1-21)

$$\det H(\lambda) = [\det E_M(\lambda)]^M = [\det \delta(\lambda) I_k]^M = [\delta(\lambda)]^{kM} . \qquad (1\text{-}31)$$

Therefore, when $\lambda \notin \Lambda$, $H(\lambda)$ is nonsingular, and (1-30) can be written as

$$\hat{X}_{i+1} = C(\lambda)\hat{X}_i \quad , \qquad i = 0, 1, 2, \cdots \qquad (1\text{-}32)$$

where

$$C(\lambda) = -[H(\lambda)]^{-1} G(\lambda) . \qquad (1\text{-}33)$$

By induction it can easily be shown that (1-32) is equivalent to

$$\hat{X}_i = [C(\lambda)]^i \hat{X}_0 \quad , \qquad i = 0, 1, 2, \ldots \qquad (1\text{-}34)$$

The vector $\hat{X}_0$ is, in essence, the starting condition used by the composite multistep method. That is,

$$\hat{X}_0 = [* \ \ldots \ * \ x_0 \ x_1 \ \cdots \ x_{N-1}]^T \ ,$$

where the first N elements (shown by asterisks) are arbitrary.

Let $\lambda \notin \Lambda$ be fixed. By transforming $C(\lambda)$ in (1-34) into Jordan canonical form it can be shown that

$$\lim_{i \to \infty} \hat{X}_i = 0 \qquad (1\text{-}35)$$

for all starting conditions $\hat{X}_0$ if and only if all eigenvalues of $C(\lambda)$ are in the open unit disc $U = \{\zeta \in C : |\zeta| < 1\}$. The eigenvalues of $C(\lambda)$ are the complex numbers ζ satisfying $\det[\zeta I - C(\lambda)] = 0$. Let

$$\hat{Q}(\lambda,\zeta) = G(\lambda) + \zeta H(\lambda) \qquad (1\text{-}36)$$

and

$$\hat{P}(\lambda,\zeta) = \det \hat{Q}(\lambda,\zeta) \ . \qquad (1\text{-}37)$$

By (1-33) $\hat{Q}(\lambda,\zeta) = H(\lambda)[\zeta I - C(\lambda)]$. Since $H(\lambda)$ is nonsingular when λ is not a pole, the eigenvalues of $C(\lambda)$ are just the roots ζ of $\hat{P}(\lambda,\zeta) = 0$. This is the essential idea behind the following result:

1.5. Proposition: Let $\lambda \notin \Lambda$ be fixed. Then the approximating sequence approaches zero for every starting condition if and only if $\hat{P}(\lambda,\zeta) = 0$ implies $\zeta \in U$, where U is the open unit disc.

Note that existence and uniqueness of the approximating sequence $\{x_i\}$ for all starting conditions is guaranteed by Proposition 1.3. This conclusion is also a consequence of (1-34), since (1-34) has been shown to be equivalent to (1-6) when λ is not a pole. Since $\{\hat{X}_i\}$ is related to the approximating sequence $\{x_i\}$ by (1-23) and (1-29), it is elementary that (1-35) is equivalent to

$$\lim_{i \to \infty} x_i = 0 \ .$$

The function $\hat{P}$ defined by (1-8), (1-10), (1-11), (1-14), (1-19), (1-20), (1-28), (1-36), and (1-37) is the first formulation of the characteristic polynomial of a composite multistep method (R,k). Note that $\hat{P}$ is well defined even when R is singular. In fact, $\hat{P}$ is of degree kM in ζ if and only if R is regular.

In general $\hat{P}$ is a polynomial in two variables with real coefficients (or the zero function). Some important properties of polynomials in two variables are discussed in the following section. These properties will be used in Section 1.5 to help relate $\hat{P}$ to A-stability of composite multistep methods.

§1.4 The A-Stability Criterion

A real polynomial in two complex variables (λ,ζ), of degree $n \geq 0$ in λ and degree $m \geq 0$ in ζ, is a function P of the form

$$P(\lambda,\zeta) = \sum_{i=0}^{n} \sum_{j=0}^{m} p_{ij}\lambda^i\zeta^j , \qquad (1\text{-}38)$$

where the p_{ij} are real numbers, not all p_{nj} are zero, and not all p_{im} are zero*. If the functions ϕ_j and ψ_i are defined by

$$\phi_j(\lambda) = \sum_{i=0}^{n} p_{ij}\lambda^i \qquad j = 0, 1, \ldots m \qquad (a)$$

and

$$\psi_i(\zeta) = \sum_{j=0}^{m} p_{ij}\zeta^j \qquad i = 0, 1, \ldots n , \qquad (b) \qquad (1\text{-}39)$$

then P can be written as

$$P(\lambda,\zeta) = \sum_{j=0}^{m} \phi_j(\lambda)\zeta^j \qquad (a)$$

or

$$P(\lambda,\zeta) = \sum_{i=0}^{n} \psi_i(\zeta)\lambda^i . \qquad (b) \qquad (1\text{-}40)$$

*The remarks of the footnotes on page 10 are to be extended to the present case.

By definition the ϕ_j [ψ_i] are real polynomials in λ [ζ], or the zero function, but ϕ_m [ψ_n] is not the zero function.

Let ϕ be a greatest common divisor of $\phi_0, \phi_1, \ldots, \phi_m$; that is, ϕ is a real polynomial of maximal degree such that $\phi_i = \phi\hat{\phi}_i$ for each $i = 0, 1, \ldots, m$, where $\hat{\phi}_i$ is a real polynomial (or the zero function). From elementary algebra ϕ exists and is unique up to a trivial factor, that is, a non-zero real constant factor. Let ψ be a greatest common divisor of $\psi_0, \psi_1, \ldots, \psi_n$. It is an elementary fact of algebra that P can be written in the form

$$P(\lambda,\zeta) = \phi(\lambda)\psi(\zeta)\bar{P}(\lambda,\zeta) , \tag{1-41}$$

where $\bar{P}$ is a polynomial in two variables, unique up to a trivial factor, and satisfying the following property: For every λ, $\bar{P}(\lambda,\cdot)$ is not the zero function; also, for every ζ, $\bar{P}(\cdot,\zeta)$ is not the zero function. The polynomial $\bar{P}$ of (1-41) will be called the reduced polynomial corresponding to P. Let $\bar{n}$ and $\bar{m}$ be the degrees of $\bar{P}$ in λ and ζ, respectively. Then ϕ and ψ are of degrees $n - \bar{n}$ and $m - \bar{m}$, respectively.

Analytic properties of polynomials in two variables are developed most naturally when the domain of λ and ζ is considered to be the Riemann sphere $\bar{C}$ (the complex plane C together with the point at infinity, extending the topology of C in the natural way [Rudin, p. 252]). Thus $P : \bar{C}^2 \to \bar{C}$. For $\zeta \in C$, the point $P(\infty,\zeta)$ is defined to be equal to $\lim_{\lambda \to 0} \lambda^n P(\frac{1}{\lambda},\zeta)$, and similarly for $P(\lambda,\infty)$ and $P(\infty,\infty)$. The first [second] coordinate of $\bar{C}^2$ will be called the lambda sphere [zeta sphere].

Let $\bar{P}$ be a reduced polynomial. Define the Riemann surface S induced by $\bar{P}$ to be the set of zeros of $\bar{P}$; that is, $S = \{(\lambda,\zeta) \in \bar{C}^2 : \bar{P}(\lambda,\zeta) = 0\}$. Clearly S is invariant with respect to multiplication of $\bar{P}$ by a trivial factor. A Riemann surface can be thought of as (the graph of) a multi-valued function (technically, a relation) whose domain [range] is the lambda* [zeta] sphere. Accordingly, if $(\lambda,\zeta) \in S$ we will say that ζ is an image of λ, and that λ is a pre-image of ζ (both with respect to S). Every point in $\bar{C}$ has

---

*It is easy to see that a reduced polynomial is of degree zero in λ if and only if it is of degree zero in ζ. For a reduced polynomial of degree zero, S is empty; hence the domain and range of S are empty. Except where otherwise noted, the following discussion of reduced polynomials applies, usually trivially, to reduced polynomials of degree zero.

at least one image and pre-image, unless $\bar{P}$ is of degree zero. If $\bar{P}$ is of degree $\bar{m}$ in ζ, then every point has at most $\bar{m}$ images, since $\bar{P}$ is reduced. Similarly, if $\bar{P}$ is of degree $\bar{n}$ in λ, then every point has at most $\bar{n}$ pre-images. If $\mathcal{A} \subset \bar{C}$, the image [pre-image] of $\mathcal{A}$ is defined to be the set of all images [pre-images] of all points of $\mathcal{A}$. Also, $\bar{\mathcal{A}}$ shall denote the closure of $\mathcal{A}$ with respect to $\bar{C}$.

The pre-images of ∞ with respect to the Riemann surface S will be called the poles of S, and hence of $\bar{P}$. By this definition ∞ is a pole of S (the infinite pole) if and only if $\bar{P}(\infty,\infty) = 0$, that is, if and only if the degree of $\bar{\phi}_{\bar{m}}$ is strictly less than $\bar{n}$, where

$$\bar{P}(\lambda,\zeta) = \sum_{j=0}^{\bar{m}} \bar{\phi}_j(\lambda)\zeta^j \quad . \tag{1-42}$$

On the other hand, the finite poles of S (those other than ∞) are precisely the zeros of $\bar{\phi}_{\bar{m}}$.

For a polynomial P, not necessarily reduced, the poles are defined to be the values of $\lambda \in \bar{C}$ for which $P(\lambda,\infty) = 0$. From (1-41) it is clear that $P(\lambda,\infty) = 0$ if and only if $\phi(\lambda) = 0$ or $\bar{P}(\lambda,\infty) = 0$. The zeros of ϕ will be called removable poles of P, while the poles of $\bar{P}$ will be called unremovable poles of P. A pole of a polynomial P is thus a removable pole or an unremovable pole or both. Note, however, that ∞ can never be a removable pole. It is clear from (1-40a), (1-41), and (1-42) that $\phi\bar{\phi}_{\bar{m}}$ and ϕ_m are equal, to within a trivial factor. Therefore, the finite poles of P are the zeros of ϕ_m. Also, ∞ is a pole of P if and only if the degree of ϕ_m is strictly less than n. In any case, the number of poles of a polynomial P in two variables cannot exceed n.

An elementary property of polynomials in one variable is that the zeros are continuous functions of the coefficients [Marden, p. 3]; that is, small changes in the values of the coefficients produce only small changes in the values of the zeros. A reduced polynomial $\bar{P}$ can be viewed in this light, with either λ or ζ as a parameter. As a consequence, the Riemann surface S induced by $\bar{P}$ has a smoothness property which is essentially that of continuity. One way of expressing this property in precise language is as follows: The image [pre-image] of every open set in $\bar{C}$ is open in $\bar{C}$. This property will be referred to as the open mapping property of Riemann surfaces.

1.6. Definition: A polynomial $P : \mathbb{C}^2 \longrightarrow \mathbb{C}$ in (λ,ζ) is said to satisfy the A-stability criterion if for every λ in the open left half plane $\mathcal{L}$, the zeros of $P(\lambda,\cdot)$ lie in the open unit disc U.

The intent of the above definition in excluding points at infinity from the domain of P is to prevent the value $\zeta = \infty$ from being considered, before the fact, a zero of $P(\lambda,\cdot)$. In other words, P does not fail to satisfy the A-stability criterion merely on the condition that $P(\lambda,\infty) = 0$ for some $\lambda \in \mathcal{L}$. Nevertheless, it will be shown (Corollary 1.9) that if this condition holds, then does in fact fail to satisfy the A-stability criterion.

The following preliminary result is a consequence of the open mapping property of Riemann surfaces:

1.7. Proposition: Let $\overline{P} : \overline{\mathbb{C}}^2 \longrightarrow \overline{\mathbb{C}}$ be a reduced polynomial, and let S be the induced Riemann surface. If there exists a point $(\lambda,\zeta) \in S$ with λ in the closed left half plane $\overline{\mathcal{L}}$, and ζ not in the closed unit disc $\overline{U}$, then $\overline{P}$ does not satisfy the A-stability criterion.

Proof: Let W be the complement of $\overline{U}$ in $\overline{\mathbb{C}}$. Thus W is open, and $\zeta \in W$. By the open mapping property of S, the pre-image V of W with respect to S is open in $\overline{\mathbb{C}}$, and by construction $\lambda \in V$. Therefore $\lambda \in \overline{\mathcal{L}} \cap V$. Since $\mathcal{L}$ and V are open, and $\overline{\mathcal{L}} \cap V$ is nonempty, $\mathcal{L} \cap V$ is a nonempty open set. Since $\overline{P}$ has only a finite number of poles, there exists a point $\hat{\lambda} \in \mathcal{L} \cap V$ which is not a pole. Now $\hat{\lambda}$ has an image $\hat{\zeta} \in W$ with $\hat{\zeta} \neq \infty$. That is, $\overline{P}(\hat{\lambda},\hat{\zeta}) = 0$, violating the A-stability criterion. □

1.8. Corollary: A polynomial P (not necessarily reduced) satisfying the A-stability criterion has no removable poles in $\mathcal{L}$ and no unremovable poles in $\overline{\mathcal{L}}$.

Proof: If P has a removable pole $\lambda \in \mathcal{L}$, then, in the notation of (1-41), $P(\lambda,\zeta) = 0 \cdot \psi(\zeta)\overline{P}(\lambda,\zeta) = 0$ for all $\zeta \in \mathbb{C}$, violating the A-stability criterion.

If P has an unremovable pole $\lambda \in \overline{\mathcal{L}}$, then $(\lambda,\infty) \in S$, where S is the Riemann surface induced by the reduced polynomial $\overline{P}$ associated with P. By Proposition 1.7 $\overline{P}$ does not satisfy the A-stability criterion; hence, neither does P. □

1.9. Corollary: A polynomial $P : \bar{C}^2 \to \bar{C}$ satisfies the A-stability criterion if and only if for every λ in the open left half plane $\mathcal{L}$, the zeros of $P(\lambda,\cdot)$ lie in the open unit disc U.

Proof: Points of the form (λ,∞) are allowed as zeros of P in Corollary 1.9, but not in Definition 1.6. Therefore, the "if" part holds trivially.

Suppose P satisfies the A-stability criterion. By Corollary 1.8, P has no poles in $\mathcal{L}$; that is, if $\lambda \in \mathcal{L}$, then (λ,∞) is not a zero of P. This being the case, the "only if" part of Corollary 1.9 is equivalent to that of Definition 1.6. □

§1.5 The Characteristic Polynomial and A-Stability

In this section it is shown that in all cases of interest a composite multistep method is A-stable if and only if its characteristic polynomial $\hat{P}$ of (1-37) satisfies the A-stability criterion, Definition 1.6.

First consider the characteristic polynomial $\hat{P}$ of (1-37) corresponding to a regular composite multistep method. By (1-31) and (1-36) the highest power of ζ in $\hat{P}$ is kM, and the coefficient of ζ^{kM} is $[\delta(\lambda)]^{kM}$. Therefore, the finite poles of $\hat{P}$ are just the zeros of δ, which are, by definition, the poles of the composite multistep method. In other words, for regular composite multistep methods the set Λ of poles of the method is also the set of finite poles of its characteristic polynomial. This fact explains the use of the term "poles" for both situations. On the other hand, for a singular composite multistep method every complex number is a pole of the method; however, the corresponding characteristic polynomial $\hat{P}$ of (1-37) is only of degree kM - 1 or less in ζ, and the poles of $\hat{P}$ form only a finite set. (If $\hat{P}$ is the zero function, the concept of poles of $\hat{P}$ is undefined.)

1.10. Theorem: Let (R,k) be a composite multistep method, and let $\hat{P}$ of (1-37) be its characteristic polynomial.

a) If R is regular and $\hat{P}$ satisfies the A-stability criterion, then $\Lambda \cap \mathcal{L}$ is empty, and (R,k) is A-stable.

b) If $\Lambda \cap \mathcal{L}$ is empty and $\hat{P}$ does not satisfy the A-stability criterion, then R is regular, and (R,k) is not A-stable.

Proof: a) By Corollary 1.8 $\hat{P}$ has no poles in $\mathcal{L}$. Since R is regular, the finite poles of $\hat{P}$ are just the elements of Λ. Thus $\Lambda \cap \mathcal{L}$ is empty.

Now let $\lambda \in \mathcal{L}$. By the A-stability criterion $\hat{P}(\lambda,\zeta) = 0$ implies $\zeta \in U$. Since $\lambda \notin \Lambda$ Proposition 1.5 applies, showing that the approximating sequence approaches zero for every starting condition. Since λ is arbitrary, (R,k) is A-stable by Proposition 1.2. (As before, existence and uniqueness of the approximating sequence follow from Proposition 1.3.)

b) By Corollary 1.9 there exists a $\lambda \in \mathcal{L}$ and a $\zeta \notin U$ such that $\hat{P}(\lambda,\zeta) = 0$. By hypothesis $\lambda \notin \Lambda$. Therefore, by Proposition 1.5 there exists a starting condition which generates an unstable approximating sequence (one for which (1-35) fails to hold). Hence, by Proposition 1.2 (R,k) is not A-stable. Regularity of R follows from the hypothesis that $\Lambda \cap \mathcal{L}$ be empty, since if R is singular, then $\Lambda = C$. □

It is emphasized that the first hypotheses in parts a) and b) of Theorem 1.10 are not entirely superfluous. First consider part a). If R is singular, then of course $\Lambda \cap \mathcal{L}$ is nonempty. The other conclusion of part a) can also be violated if R is singular, as shown by the following example: The characteristic polynomial $\hat{P}$ of the composite multistep method

$$\left(\begin{bmatrix} -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}, 1 \right) \tag{1-43}$$

is computed by (1-8), (1-10), (1-11), (1-14), (1-19), (1-20), (1-28), (1-36), and (1-37) to be

$$\hat{P}(\lambda,\zeta) = (1 - \lambda)\zeta - 1 . \tag{1-44}$$

If $\lambda \in \mathcal{L}$, then $\hat{P}(\lambda,\zeta) = 0$ implies $\zeta = \frac{1}{1-\lambda}$; therefore $|\zeta| = \frac{1}{|1-\lambda|}$, which can easily be shown to be less than unity. Therefore $\hat{P}$ of (1-44) satisfies the A-stability criterion, Definition 1.6. Nevertheless, (R,k) of (1-43) is not A-stable. To see this, observe that the algebraic equation (1-12) corresponding to (1-43) is

$$\begin{bmatrix} 0 & -1 \\ 0 & 1-\lambda \end{bmatrix} \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ \lambda & 0 \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} ,$$

which has no solution $[y_2 \ y_3]^T$ (except when $y_0 = (1 - \lambda)y_1$). Therefore,

for each $\lambda \in C$, there exist starting conditions for which (1-43) does not generate an approximating sequence*.

The hypothesis of part b), that $\Lambda \cap \mathcal{L}$ be empty, is also not superfluous. If $\Lambda \cap \mathcal{L}$ is nonempty, it is obvious that R need not be regular, regardless of the behavior of $\hat{P}$. For example, consider (1-43), but with $\beta_{23} = -1$ (see (1-3)). The composite matrix is still singular. However, the characteristic polynomial becomes $\hat{P}(\lambda,\zeta) = (\lambda + 1)\zeta - 2\lambda - 1$, for which $\hat{P}(-2,3) = 0$.

The other conclusion of part b) can also be violated if $\Lambda \cap \mathcal{L}$ is nonempty, as shown by example (1-13). The characteristic polynomial associated with (1-13) is

$$\hat{P}(\lambda,\zeta) = (1 + \lambda)[(1 - \lambda)\zeta - 1] \quad . \tag{1-45}$$

Since $\hat{P}(-1,2) = 0$, $\hat{P}$ does not satisfy the A-stability criterion. Nevertheless (R,k) of (1-13) is A-stable. To see this, recall that (1-13) is weakly equivalent to the implicit Euler method ([-1 1 0 1],1). The implicit Euler method is clearly regular; furthermore, its characteristic polynomial can be shown to be (1-44), which satisfies the A-stability criterion. Thus by Theorem 1.10a), the implicit Euler method (and hence (1-13)) is A-stable.

A counterexample to both conclusions of part b) (when $\Lambda \cap \mathcal{L}$ is nonempty) is given in Section 1.7 by (1-54).

1.11. Corollary: A composite multistep method (R,k) with $\Lambda \cap \mathcal{L}$ empty is A-stable if and only if its characteristic polynomial $\hat{P}$ defined by (1-37) satisfies the A-stability criterion.

Proof: If $\Lambda \cap \mathcal{L}$ is empty, then R is regular. Therefore the corollary follows immediately from Theorem 1.10. □

The nature of the above examples might lead one to conjecture a somewhat stronger version of Corollary 1.11 as follows: Replace the hypothesis "$\Lambda \cap \mathcal{L}$ empty" with the somewhat weaker one "R regular", and substitute for $\hat{P}$ the factor $\psi\bar{P}$ of $\hat{P}$, where it is assumed that $\hat{P}$ has been factored according to (1-41). In other words, take away the removable poles of $\hat{P}$ before testing it for the A-stability criterion. The "only if" part of the conjecture is

*This example is characteristic of the situation occurring when R is singular, as shown by Theorem 1.16a).

believed to be true. However, the "if" part definitely is not, as the following counterexample shows:

The characteristic polynomial corresponding to the regular composite multistep method ([1 2 -1 -2],1) is $\hat{P}(\lambda,\zeta) = (\lambda + 1)(2\zeta + 1)$. Therefore $\psi(\zeta)\bar{P}(\lambda,\zeta) = 2\zeta + 1$, which is easily seen to satisfy the A-stability criterion. However, at $\lambda = -1$ the algebraic equation (1-12) holds identically for all Y_p, so that uniqueness is violated. Since $k = n = 1$, Propositions 1.2 and 1.4 show that (R,k) is not A-stable.

§1.6 An Improved Formulation of the Characteristic Polynomial

In its present formulation, the characteristic polynomial $\hat{P}$ of (1-37) is of degree at most nkM in λ and of degree kM in ξ (if R is regular). Such a formulation is, in general, of much higher degree than necessary for the purpose of establishing A-stability. The following result shows that in the statement of Theorem 1.10, $\hat{P}$ of formulation (1-37) can be replaced by $\tilde{P}$ of formulation (1-26), which is only of degree at most nk in λ and of degree kM in ζ. This section concludes with an even simpler formulation, of degree at most n in λ and of degree m in ζ. The following theorem is proved in Appendix A:

1.12. Theorem: Let $\lambda \notin \Lambda$ be fixed in (1-26) and (1-37). The zeros ζ of $\hat{P}(\lambda,\cdot)$ are the M-th power of the zeros of $\tilde{P}(\lambda,\cdot)$. More precisely, if $\tilde{P}(\lambda,\zeta) = 0$ then $\hat{P}(\lambda,\zeta^M) = 0$. Conversely, if $\hat{P}(\lambda,\zeta) = 0$, then there exists a $z \in \bar{C}$ such that $z^M = \zeta$ and $\tilde{P}(\lambda,z) = 0$.

1.13. Corollary: A composite multistep method (R,k) with $\Lambda \cap \mathcal{L}$ empty is A-stable if and only if its characteristic polynomial $\tilde{P}$ defined by (1-26) satisfies the A-stability criterion.

Proof: Clearly $\zeta^M \in U$ if and only if $\zeta \in U$, which in turn holds if and only if $\zeta^{1/M} \in U$. Therefore, by Definition 1.6 and Theorem 1.12, $\tilde{P}$ satisfies the A-stability criterion if and only if $\hat{P}$ does. The corollary now follows directly from Corollary 1.11. □

A third formulation for the characteristic polynomial of a composite multistep method (R,k) is defined as follows: Partition R along its columns into two $n \times (m + n)$ matrices:

$$R = [A \quad B] \quad . \tag{1-46}$$

Define the function $\hat{V}$ by

$$\hat{V}(\lambda) = A - \lambda B \quad ; \tag{1-47}$$

partition $\hat{V}$ along its columns to give

$$\hat{V} = [V \quad K] \quad , \tag{1-48}$$

where V is $n \times (m + k)$ and K is $n \times (n - k)$. Let M be the smallest integer not less than m/k, and let $N = kM - m$. Extend V with N columns of zeros, and partition the result after every k-th column, so that

$$[V \quad 0_{nN}] = [V_0 \; V_1 \; \cdots \; V_M] \quad , \tag{1-49}$$

where each V_i is $n \times k$. Finally define the functions W, Q, and P by

$$W(\lambda,\zeta) = \sum_{i=0}^{M} V_i(\lambda)\zeta^i \quad , \tag{1-50}$$

$$Q(\lambda,\zeta) = [W(\lambda,\zeta) \quad K(\lambda)] \quad , \tag{1-51}$$

and

$$P(\lambda,\zeta) = \det Q(\lambda,\zeta) \quad . \tag{1-52}$$

It is clear that P is defined for all (R,k), and that P is a polynomial in two variables (or the zero function). P is the final formulation for the characteristic polynomial of a composite multistep method.

The following theorem is proved in Appendix B:

1.14. Theorem: The two formulations (1-26) and (1-52) for the characteristic polynomial of a composite multistep method (R,k) are related by the following identity:

$$\tilde{P}(\lambda,\zeta) = (-1)^{(k-1)N}\zeta^N[\delta(\lambda)]^{k-1}P(\lambda,\zeta) \tag{1-53}$$

1.15. Corollary: A composite multistep method (R,k) with $\Lambda \cap \mathcal{L}$ empty is A-stable if and only if its characteristic polynomial P defined by (1-46) through (1-52) satisfies the A-stability criterion.

Proof: If $\lambda \in \mathcal{L}$, then $[\delta(\lambda)]^{k-1} \neq 0$, since $\Lambda \cap \mathcal{L}$ is empty. Therefore, (1-53) shows that for $\lambda \in \mathcal{L}$ the zeros ζ of $\tilde{P}(\lambda,\cdot)$ are just those of $P(\lambda,\cdot)$, and $\zeta = 0$, for the case $N \neq 0$. Since $0 \in U$, P satisfies the A-stability criterion if and only if $\tilde{P}$ does, by Corollary 1.9. The above corollary now follows directly from Corollary 1.13. □

The characteristic polynomial P defined by (1-46) through (1-52) is a new formulation, which has several advantages over the formulations $\tilde{P}$ and $\hat{P}$. Firstly, the degrees of P in both λ and ζ are generally lower than those of $\tilde{P}$ and $\hat{P}$; in fact, the degrees of P are believed to be the lowest possible. This fact, discussed in detail below, is not only esthetically pleasing; it also makes computations more manageable. Secondly, P is a multilinear function of the rows of R. This fact, discussed further in Chapter 4, is of crucial importance there in the exploration of certain important classes of composite multistep methods. The multilinear property of P is not generally shared with $\tilde{P}$ and $\hat{P}$. Thirdly, it is claimed that P reflects the behavior of a larger class of composite multistep methods than either $\tilde{P}$ or $\hat{P}$ does. Some technical results in this direction are given below.

A disadvantage of the new formulation (1-52), at least in the present development, is in the devious route by which P is related to A-stability of (R,k). The relationship of $\hat{P}$ to A-stability of (R,k) is very easily established through its associated difference equation. The relationship of $\tilde{P}$ to A-stability of (R,k) is established much less easily, either through a technically difficult transformation to $\hat{P}$ (Theorem 1.12), or through its own associated difference equation [Sloate and Bickart]. In the present work $\tilde{r}$ is related to A-stability of (R,k) only through $\tilde{P}^*$, by use of complicated matrix manipulations (Theorem 1.14). Apparently P is not related in a direct way with a difference equation for the approximating sequence, in contrast with $\tilde{P}$ and $\hat{P}$.

The degrees of $\hat{P}$, $\tilde{P}$, and P in λ and ζ are summarized in Table 1.1, along with those of a fourth formulation discussed in the previous footnote. The entries for $\tilde{P}$ follow directly from (1-21) and (1-26). These entries and (1-53) can be used to deduce the entries for P, since kM - N = m. Alternatively, the

---

*An alternate course is possible from $\hat{P}$ to P, bypassing $\tilde{P}$ entirely. A fourth formulation can be defined, of degree nM in λ and of degree m in ζ. This fourth formulation can be transformed into $\hat{P}$ by a derivation similar to that in Appendix B, and into P by a derivation similar to that in Appendix A. However, the resulting development is somewhat more complicated than the present one.

Formulation	Notation	Maximum degree in λ	Degree in ζ when R is regular	Coefficient of the highest power of ζ
1	$\hat{P}$	nkM	kM	$[\delta(\lambda)]^{kM}$
2	$\tilde{P}$	nk	kM	$[\delta(\lambda)]^{k}$
3	P	n	m	$\pm\ \delta(\lambda)$
4	see text	nM	m	$\pm\ [\delta(\lambda)]^{M}$

Table 1.1. The Degrees of Various Formulations for the Characteristic Polynomial

entries for P can be deduced directly from its definition. For example, since each element of Q in (1-51) is a polynomial of degree 1 in λ, it follows from (1-52) that P is of degree at most n in λ. Also, the last N columns of V_M in (1-49) are zero. Therefore, columns k - N + 1, ..., k - 1, k of Q are only of degree M - 1 in ζ. Thus P is of degree kM - N = m in ζ. To show that the coefficient of ζ^m is $\pm\ \delta(\lambda)$, take the coefficients of the highest powers of ζ in each column of Q separately. The matrix formed from these coefficient can be seen to be D, to within a permutation of the columns.

In summary, the degrees of P in λ and ζ are just the numbers of future points and past points, respectively. There is no doubt that these degrees are the lowest possible in any general formulation. With n future points the method has n poles; if P reflects the behavior of the method, then P must also have n poles. On the other hand, if there are m past points, then for a given λ, the set of approximating sequences forms an m-dimensional space. (A basis for this space can be generated from the m different starting conditions formed by setting exactly m - 1 of the past points in the starting condition to zero.) In such a system, there are m natural modes; the eigenvalue of each mode is a zero of $P(\lambda,\cdot)$. Thus $P(\lambda,\cdot)$ must have at least m zeros, if it is to reflect the behavior of the method.

The class of composite multistep methods for which the new formulation P gives useful information is larger than that for which $\tilde{P}$ or $\hat{P}$ does. For example, suppose R is a singular composite matrix. Then by definition δ is the zero function. Therefore, according to (1-53), $\tilde{P}$ is the zero function

(unless $k = 1$, in which case $\tilde{P}$ is essentially equal to P). Thus P, which is generally not the zero function, gives information about some of the class of singular composite multistep methods. On the other hand, suppose (R,k) is a regular composite multistep method, and λ is a finite unremovable pole of P. Since $\delta(\lambda) = 0$, it follows again from (1-53) that λ is a removable pole of $\tilde{P}$ (unless $k = 1$). In other words, for $k \neq 1$, all finite poles of $\tilde{P}$ are removable poles of $\tilde{P}$. It is claimed that this fact prevents $\tilde{P}$ from reflecting the behavior of composite multistep methods in as natural a way as P does. It can be shown that $[\delta(\lambda)]^{(k-1)M}$ is a factor of $\hat{P}(\lambda,\cdot)$, and that $\hat{P}$ therefore suffers from the same deficiencies discussed above as $\tilde{P}$ does.

§1.7 The New Formulation and Existence at Poles

One illustration of the more general theorems available under the new formulation is given below. The behavior of a characteristic polynomial near a pole can be inferred from the proof of Proposition 1.7 (see also Corollary 1.8). But what about the behavior of the composite multistep method itself? Some aspects of the answer, for regular composite multistep methods, are given by Theorem 1.10. Another aspect is given by the "only if" part of Proposition 1.4. The following result gives additional insight into the behavior of composite multistep methods at poles.

1.16. Theorem: Let (R,k) be a composite multistep method, let P of (1-52) be its characteristic polynomial, and assume P is not the zero function.

a) If $\lambda \in \Lambda$, but λ is not a removable pole of P, then there is a starting condition such that for this value of λ no approximating sequence exists.

b) If there exists a $\lambda \in \Lambda \cap \mathcal{L}$ such that λ is not a removable pole of P, then (R,k) is not A-stable.

c) If R is singular, then (R,k) is not A-stable.

Proof: First part a) is proved. Since λ is not a removable pole and P is not the zero function, $P(\lambda,\cdot)$ is not the zero function. That is, there exists a $\zeta \in \mathbb{C}$ such that $P(\lambda,\zeta) \neq 0$. Now by (1-52) rank $Q(\lambda,\zeta) = n$. Using (1-50), (1-51), and the notation of Appendix B, relations (B-1) and (B-2), we can write

$$Q(\lambda,\zeta) = [V(\lambda) \quad 0_{nN} \quad K(\lambda)] \begin{bmatrix} Z(\zeta) & 0_{(M+1)k,(n-k)} \\ 0_{(n-k)k} & I_{n-k} \end{bmatrix} .$$

Therefore

$$\begin{aligned} n &= \text{rank } Q(\lambda,\zeta) \\ &\leq \text{rank } [V(\lambda) \quad 0_{nN} \quad K(\lambda)] \\ &= \text{rank } [V(\lambda) \quad K(\lambda)] \\ &= \text{rank } [(A_p - \lambda B_p) \quad D(\lambda)] \quad , \end{aligned}$$

the last equality following from (B-10). Also, rank $D(\lambda) \leq n - 1$ by (1-11), since $\lambda \in \Lambda$. Combining these two relations gives

$$\text{rank } D(\lambda) < \text{rank } [(A_p - \lambda B_p) \quad D(\lambda)] \quad .$$

Therefore, the range of the linear transformation $A_p - \lambda B_p$ is not contained in the range of the linear transformation $D(\lambda)$. Hence, there exists a Y_p such that (1-12) has no solution Y_f. For such Y_p the approximating sequence does not exist either, proving part a).

Part b) follows directly from part a) and Proposition 1.2. To prove part c), note that a polynomial in two variables can have only a finite number of poles. It follows that there exists a $\lambda \in \mathcal{L}$ such that λ is not a (removable) pole of P. Since R is singular $\lambda \in \Lambda$. Now part c) follows directly from part b). □

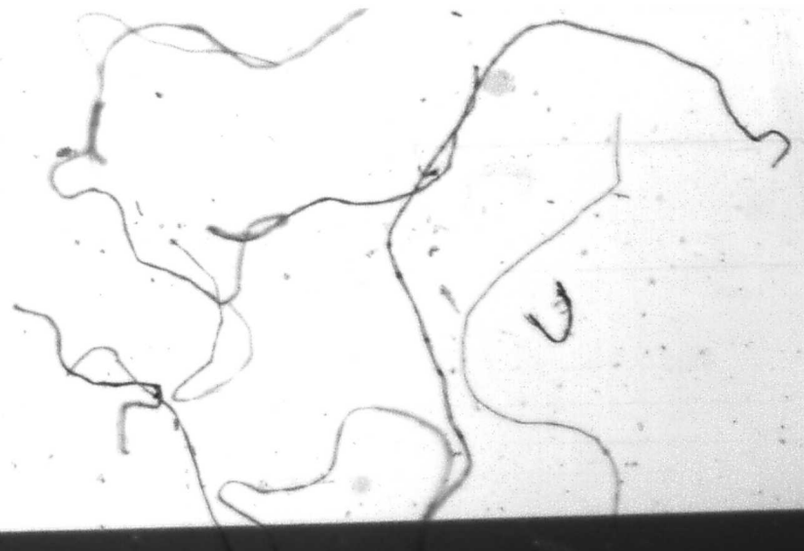
The hypothesis in Theorem 1.16, that P not be the zero function, is quite necessary. To show this, consider the singular composite multistep method

$$\left(\begin{bmatrix} -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, 1 \right) , \quad . \tag{1-54}$$

whose characteristic polynomial P is the zero function. It is clear that (1-54) is weakly equivalent to the implicit Euler method, and is therefore A-stable.

It is to be noted that if P is replaced by $\tilde{P}$ or $\hat{P}$ in Theorem 1.16, the hypotheses for parts a), b), and c) become vacuous (unless k = 1). This is because if $\tilde{P}$ or $\hat{P}$ is not the zero function, then (1) all poles of $\tilde{P}$ or $\hat{P}$ are removable poles (unless k = 1), and (2) R is regular.

It can be shown that P, $\tilde{P}$ and $\hat{P}$ are all equal to the zero function if the rows of the composite matrix R (that is, the multistep methods constituting



the composite multistep method) are linearly dependent. However, the converse of this statement is false, even for the new iormulation P, as shown by the counterexample (1-54). Thus, P can be the zero function even if the rows of R are linearly independent.

Theorem 1.16a) can profitably be compared with the "only if" part of Proposition 1.4. Neither result includes the other. However, in their common domain of applicability, the conclusion of Theorem 1.16a) is more definite: There is a starting condition for which existence is violated.

Theorem 1.16b) provides a complementary result to Theorem 1.10b), in that the former applies only if $\Lambda \cap \mathcal{L}$ is nonempty, while the latter applies only if $\Lambda \cap \mathcal{L}$ is empty. Together, the two results yield a very powerful necessary condition for A-stability of composite multistep methods.

§1.8 Final Remarks and a Review of Previous Work

The following four conditions discussed in this chapter form a sequence of increasingly strong restrictions on composite multistep methods:

1. The rows of the composite matrix R are linearly independent.
2. The characteristic polynomial P of (1-52) is not the zero function.
3. The composite matrix R is regular.
4. $\Lambda \cap \mathcal{L}$ is empty.

The fourth condition, which is the strongest, has been shown to be particularly important in A-stability investigations.

Some of the material in this chapter is the work of [Sloate and Bickart]. In particular, they derived the difference equation (1-25) and the characteristic polynomial* $\tilde{P}$ of (1-26). They also related $\tilde{P}$ to A-stability of (R,k) by methods involving the Smith canonical form. However, [Sloate and Bickart] failed to consider pathological situations which can arise if the composite multistep method has poles in the open left half plane (for example, see (1-13)). Therefore their result, which is Corollary 1.13 without the restriction that $\Lambda \cap \mathcal{L}$ be empty, is not quite correct in general. The study of poles was initiated by [Watts] for composite one-step methods (with k = n).

---

*The actual derivation in [Sloate and Bickart] treats the important case k = n separately. Their result yields a characteristic polynomial which for k = n is not $\tilde{P}$, but instead $\pm \zeta^N P(\lambda,\zeta)$. Therefore, for k = n the Sloate-Bickart formulation has most of the important properties of P, as it does for k = 1.

The difference equation (1-32) and the characteristic polynomial $\hat{P}$ of (1-37) are new formulations. The connection between $\hat{P}$ and $\tilde{P}$ (Theorem 1.12) is a new development, as is the final formulation (1-52) for the characteristic polynomial P. In this regard, the basic idea of Lemma B.1 was suggested to the author by Tendler*, who recognized that a technique of [Donelson and Hanson] could be applied to the present case. The concepts of regular composite matrices, reduced polynomials, removable poles, and unremovable poles are introduced for the first time in this work, although removable poles are hinted at by [Watts]. Regular composite matrices are discussed further in Chapter 4. The present chapter gives the first extensive treatment of the relationship of poles to existence, uniqueness, and stability questions. Chapter 2 provides another viewpoint from which the important role of poles in A-stability questions can be seen.

*Private communication, Joel M. Tendler, Syracuse University Research Corporation, Syracuse, New York (1972).

Chapter 2

AN ANALYTIC CHARACTERIZATION OF A-STABILITY

§2.1 Introduction

The principal result of Chapter 1 is that a composite multistep method (R,k) with no poles in the open left half plane $\mathcal{L}$ is A-stable if and only if its characteristic polynomial* P satisfies the A-stability criterion, that is, if and only if for every $\lambda \in \mathcal{L}$ the zeros of $P(\lambda,\cdot)$ lie in the open unit disc U. To determine A-stability of (R,k) directly from this result requires, in principle, the examination of the zeros of the polynomial $P(\lambda,\cdot)$ (whose coefficients are complex) as a function of the parameter λ in an open subset $\mathcal{L}$ of the complex plane. The next two sections show that it is enough to examine the zeros for those parameter values lying on the boundary of this subset, and to check some simple side conditions.

The A-stability criterion, in the form of Corollary 1.9, can be restated in the following "dual" form:

2.1. Proposition: A polynomial $P : \bar{C}^2 \to \bar{C}$ satisfies the A-stability criterion if and only if for every ζ not in the open unit disc U, the zeros of $P(\cdot,\zeta)$ lie in the closed right half plane $\bar{\mathcal{R}}$.

Proof: Both Proposition 2.1 and Corollary 1.9 are logically equivalent to the following statement: For every $\lambda \in \mathcal{L}$ and $\zeta \notin U$, $P(\lambda,\zeta) \neq 0$. □

§2.2 Mappings on the Riemann Surface

In this section an important topological property of Riemann surfaces is reviewed. Throughout this section let $\bar{P}$ be a reduced polynomial in (λ,ζ), and let S be the Riemann surface induced by $\bar{P}$. Let I be the extended imaginary axis in the lambda sphere $\bar{C}$ (the imaginary axis together with the point at infinity). In $\bar{C}$, I can be viewed as a simple closed analytic curve [Springer, p. 114], thus dividing $\bar{C}$ into two connected open sets: The complement of I in $\bar{C}$ is the union of the open left half plane $\mathcal{L}$ and the open

---

*In this chapter and the next, any of the three formulations $\hat{P}$, $\tilde{P}$, and P can be used as the characteristic polynomial of a composite multistep method, since $\Lambda \cap \mathcal{L}$ is empty in all cases considered. In practice however, the new formulation P of (1-52) is to be preferred, for the various reasons discussed in Chapter 1.

right half plane $\mathcal{R}$. Define the zeta locus T associated with $\bar{P}$ to be the image of I with respect to S. (If $\bar{P}$ is of degree zero, then T is empty.) If $\bar{P}$ is of degree m in ζ, then T consists of at most m continuous closed curves which, in general, are not simple.

It is claimed that the complement of T in $\bar{C}$ is the union of a finite number of nonempty open connected sets, called components (see Fig. 2.1.). A proof of this fact is sketched in the following paragraph:

If $\bar{P}$ is of degree zero, the result holds trivially. Hence assume $\bar{P}$ is a reduced polynomial of positive degree. Such a polynomial can be factored, in an essentially unique way, into a finite number of irreducible polynomials of positive degree: $\bar{P} = P_1 P_2 \cdots P_r$ [Bocher, p. 209], [Bliss, p. 22]; each P_i induces a Riemann surface S_i which is analytic except for a finite number of singularities, called branch points. The image of any analytic curve in $\bar{C}$ with respect to S_i is therefore a finite collection of continuous curves, which fail to be analytic only at the projections of the branch points [Bliss], [Springer]. Since $S = S_1 \cup S_2 \cup \cdots \cup S_r$, and I is a closed analytic curve in $\bar{C}$, the image of I with respect to S is just the union of its images with respect to S_1, S_2, $\cdots$, S_r; that is, the zeta locus T consists of a finite number of closed, piecewise-analytic curves. If two such analytic arcs intersect in more than a finite number of points, they are identical*. Viewing identical curves as a single curve, the zeta locus T can be represented as a finite collection of simple analytic curves, joined to each other only at the endpoints, forming a finite number of simple closed piecewise-analytic curves. Any such curve separates $\bar{C}$; that is, the complement of a simple closed piecewise-analytic curve in $\bar{C}$ is the union of two nonempty disjoint open connected sets [Springer, pp. 91, 115-117]. Therefore T separates $\bar{C}$ into a finite number of components.

A component P will be called an instability component of the complement of T if the pre-image V of P is a subset of the open right half plane $\mathcal{R}$.

2.2. Lemma: Let T be the zeta locus associated with a reduced polynomial $\bar{P} : \bar{C}^2 \to \bar{C}$, let P be a component of the complement of T, and let $\zeta \in P$. If all the pre-images of ζ (with respect to the Riemann surface S induced by $\bar{P}$) are in $\mathcal{R}$, then P is an instability component.

*This fact is proved analytically by [Springer, p. 117] using only the fact that S is compact. An algebraic proof is given by [Bliss, p. 22], using the concept of irreducible polynomials.

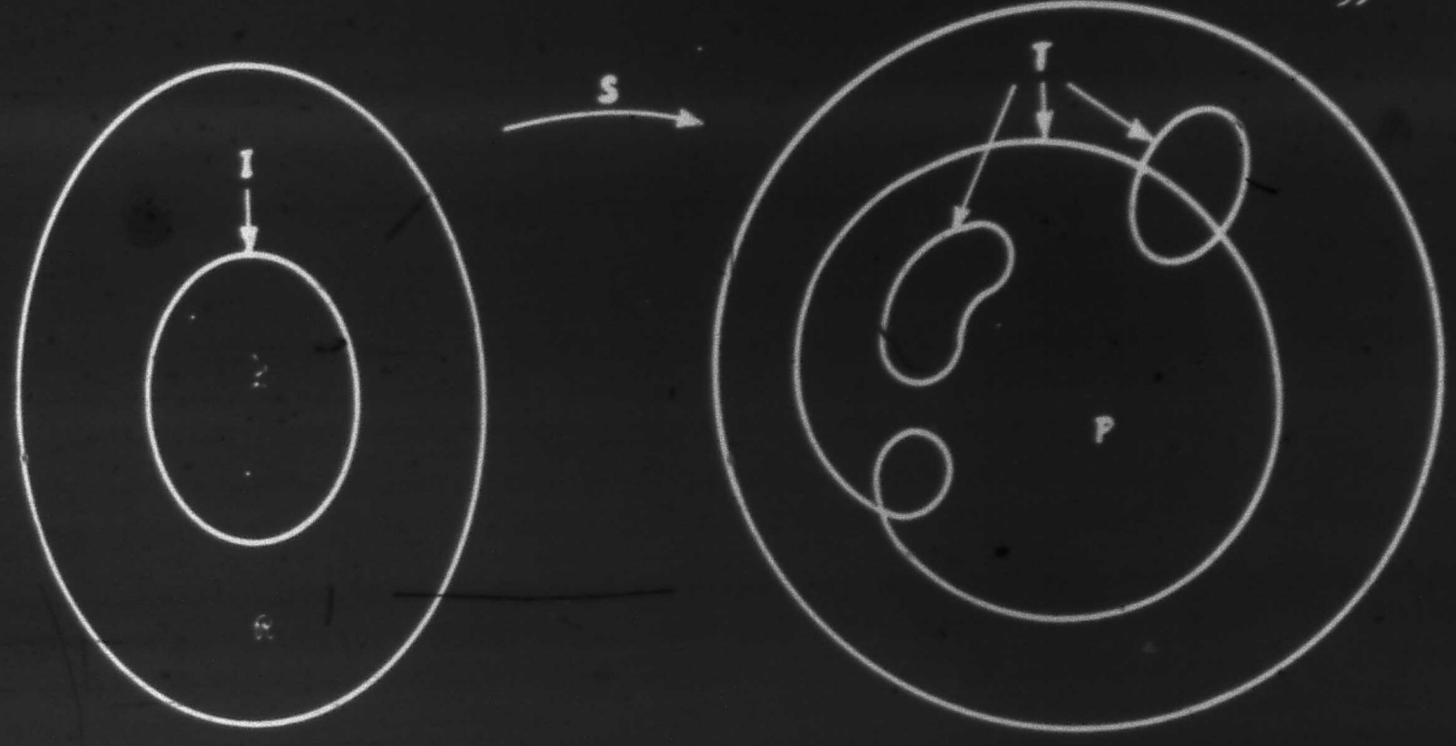


Fig. 2.1. Schematic Sketch for Definition of the Zeta Locus

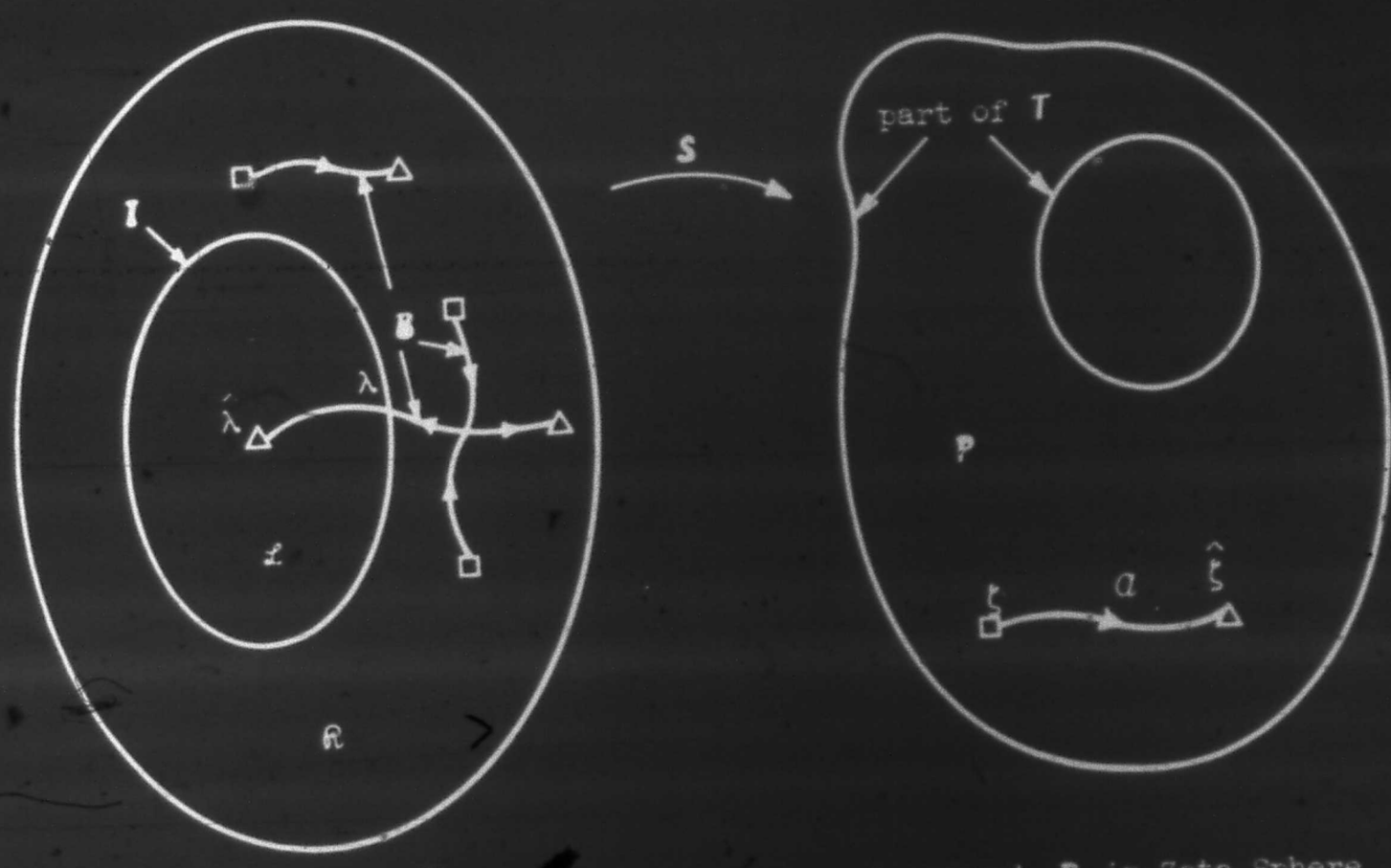


Fig. 2.2. Pre-Images for a Component

Proof: Suppose there exists a point $\hat{\zeta} \in P$ having a pre-image $\hat{\lambda}$ not in $\mathcal{R}$ (see Fig. 2.2.). Since $P \cap T$ is empty, $\hat{\lambda} \notin I$; hence $\hat{\lambda} \in \mathcal{L}$. Since P is a connected subset of $\overline{C}$, it is arcwise connected [Springer, p. 55]; therefore, there exists a (continuous) arc $\mathcal{A} \subset P$ beginning at ζ and ending at $\hat{\zeta}$. The pre-image $\mathcal{B}$ of $\mathcal{A}$ consists of a finite collection of arcs in the lambda sphere beginning at the pre-images of ζ and ending at the pre-images of $\hat{\zeta}$. In particular, $\mathcal{B}$ contains an arc from a pre-image of ζ (which is in $\mathcal{R}$ by hypothesis) to $\hat{\lambda} \in \mathcal{L}$. This arc intersects the boundary I of $\mathcal{L}$ and $\mathcal{R}$ at, say, λ. Now $\lambda \in I$ is a pre-image of a point in $\mathcal{A} \subset P$, which is a contradiction, since all pre-images of components are in $\mathcal{L} \cup \mathcal{R}$. □

§2.3 The Analytic A-Stability Characterization

Lemma 2.2 is the key to the following fundamental result:

2.3. Theorem: Let $\overline{P} : \overline{C}^2 \to \overline{C}$ be a reduced polynomial, let T be the corresponding zeta locus, and let $\zeta \notin \overline{U}$. In order for $\overline{P}$ to satisfy the A-stability criterion, it is necessary and sufficient that

a) all pre-images of ζ are in $\mathcal{R}$, and,

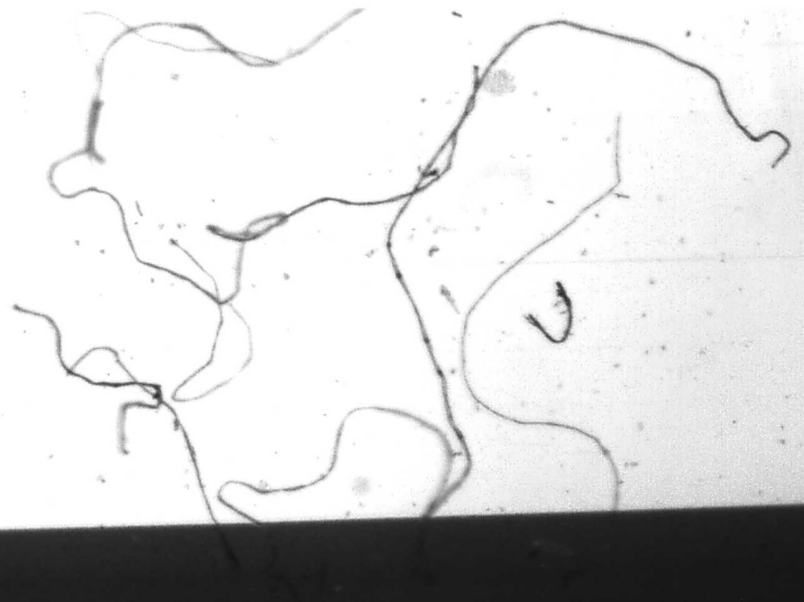
b) $T \subset \overline{U}$.

Proof: The necessity of condition a) follows immediately from Proposition 1.7. To show that condition b) is necessary, suppose there exists a $\hat{\zeta} \in T \cap W$, where W is the complement of $\overline{U}$ in $\overline{C}$. By definition of T, $\hat{\zeta}$ has a pre-image $\lambda \in I \subset \overline{\mathcal{L}}$. Proposition 1.7 now shows the A-stability criterion to be violated.

To prove sufficiency, first note that by condition b) $\infty \notin T$. Let P be the unbounded component (the component containing ∞) of the complement of T. By condition b) $\zeta \in W \subset P$. By condition a) and Lemma 2.2, P is an in-stability component. That is, the pre-image of P is contained in $\mathcal{R}$. Since $W \subset P$, the pre-image V of W is contained in $\mathcal{R}$. It can be shown that the pre-image J of $\overline{W}$ is contained in $\overline{V}$*. But $V \subset \mathcal{R}$ implies $\overline{V} \subset \overline{\mathcal{R}}$; therefore

---

*To show this, let $\hat{\zeta} \in \overline{W}$, and suppose $\hat{\zeta}$ has a pre-image $\lambda \notin \overline{V}$. Let $\mathcal{D}$ be the image of the complement of $\overline{V}$. By the open mapping property of Riemann surfaces $\mathcal{D}$ is open. By construction $\hat{\zeta} \in \mathcal{D} \cap \overline{W}$. Since $\mathcal{D}$ and W are open and $\mathcal{D} \cap \overline{W}$ is nonempty, $\mathcal{D} \cap W$ is nonempty. Choose $\zeta \in \mathcal{D} \cap W$. Now ζ has a pre-image $\tilde{\lambda} \notin \overline{V} \supset V$, which is a contradiction, since all pre-images of elements of W are in V.



$J \subset \overline{\mathcal{R}}$. In other words, the pre-image of $\overline{W}$ (the set of points not in U) is in $\overline{\mathcal{R}}$. By Proposition 2.1, $\overline{P}$ satisfies the A-stability criterion. □

A convenient way of applying Theorem 2.3 to arbitrary polynomials (not necessarily reduced) is as follows:

2.4. Theorem: Let $P : \overline{C}^2 \to \overline{C}$ be a polynomial, factored according to (1-41), so that $P = \phi\psi\overline{P}$. Let T be the zeta locus corresponding to the reduced polynomial $\overline{P}$. In order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) all poles of P lie in the closed right half plane $\overline{\mathcal{R}}$,

b) all zeros of ψ lie in the open unit disc U, and,

c) the zeta locus is contained in the closed unit disc; that is,

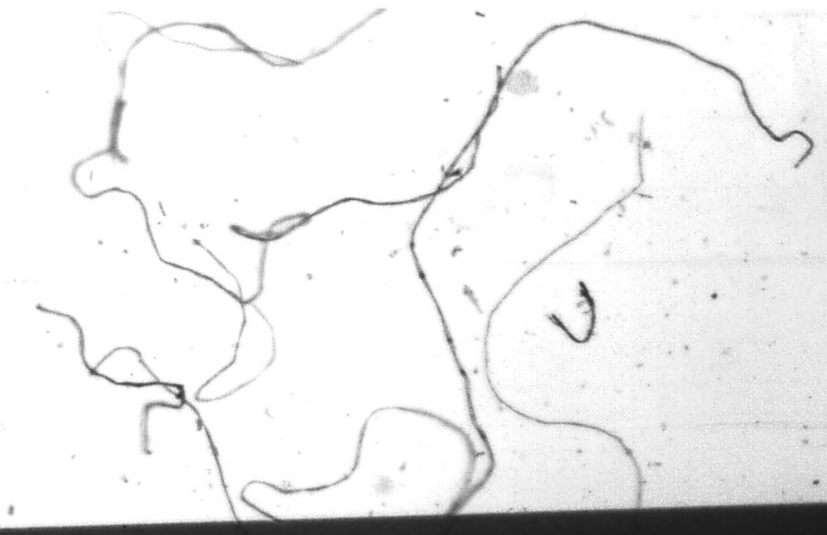
$T \subset \overline{U}$.

Proof: Condition a) is necessary by Corollary 1.8. Suppose $\zeta \notin U$ is a zero of ψ. By (1-41), $P(\lambda,\zeta) = 0$ for all λ, in particular for $\lambda \in \mathcal{L}$. Therefore P violates the A-stability criterion, Definition 1.6. Hence condition b) is necessary. It is clear from (1-41) that if P satisfies the A-stability criterion, so does $\overline{P}$. Therefore, by Theorem 2.3 condition c) is necessary.

To prove sufficiency, first note that for all $\lambda \in \mathcal{L}$ and $\zeta \notin U$, $\phi(\lambda)\psi(\zeta) \neq 0$, by conditions a) and b). Therefore P satisfies the A-stability criterion if and only if $\overline{P}$ does. The pre-images of ∞ with respect to the Riemann surface S induced by $\overline{P}$ are, by definition, the poles of $\overline{P}$. By condition a) all these poles are in $\overline{\mathcal{R}} = \mathcal{R} \cup I$. But by condition c) all images of I are in $\overline{U}$, which does not contain ∞. Therefore, all pre-images of ∞ are in $\mathcal{R}$. Now Theorem 2.3 can be applied to $\overline{P}$ with $\zeta = \infty$, showing that $\overline{P}$, and hence P, satisfies the A-stability criterion. □

The analytic characterization of A-stable composite multistep methods now follows immediately from Theorem 2.4 and the principal result of Chapter 1, in the form of Corollary 1.11, 1.13, or 1.15*.

*Recall that if $\Lambda \cap \mathcal{L}$ is empty, then R is regular. Hence, the finite poles of P are the poles of (R,k), which are just the elements of Λ. In such case, condition a) of Theorem 2.4 holds.



2.5. Corollary: Let (R,k) be a composite multistep with $\Lambda \cap \mathcal{L}$ empty, and let P be its characteristic polynomial, factored according to (1-41), so that $P = \phi\psi\bar{P}$. Let T be the zeta locus corresponding to the reduced polynomial $\bar{P}$. In order for (R,k) to be A-stable, it is necessary and sufficient that

a) all zeros of ψ lie in the open unit disc U, and,

b) the zeta locus is contained in the closed unit disc; that is,

$T \subset \bar{U}$.

§2.4 The Dual Analytic A-Stability Characterization

A natural duality is apparent in the A-stability criterion, as can be seen from Corollary 1.9 and Proposition 2.1, as well as from the symmetrical properties of Riemann surfaces. Thus, instead of taking the image of the boundary of the left half plane (in the lambda sphere), dual results can be developed by taking the pre-image of the boundary of the unit disc (in the zeta sphere). The mathematical development is the same in all essential respects. Therefore, the major results will be presented without proof.

The development begins as before with the Riemann surface S induced by a reduced polynomial $\bar{P}$. Define the lambda locus M to be the pre-image with respect to S of the unit circle E. (If $\bar{P}$ is of degree zero, then M is empty.) If $\bar{P}$ is of degree n in λ, then M consists of at most n closed curves. As before, the complement of M is the union of a finite number of nonempty open connected sets, called components. A component P will be called a stable component if the image V of P is a subset of the open unit disc U. The situation is the same, from the Riemann surface viewpoint, as in Lemma 2.2. That is, a single element of P serves to determine whether P is a stable component.

The above considerations lead to the following fundamental result, which is dual to Theorem 2.3:

2.6. Theorem: Let $\bar{P} : \bar{C}^2 \to \bar{C}$ be a reduced polynomial, let M be the corresponding lambda locus, and let $\lambda \in \mathcal{L}$. In order for $\bar{P}$ to satisfy the A-stability criterion, it is necessary and sufficient that

a) all images of λ are in U, and,

b) $M \subset \bar{R}$.

The above theorem leads to an A-stability characterization of arbitrary polynomials which is dual to Theorem 2.4, but not quite as elegant:

2.7. Theorem: Let $P : \bar{C}^2 \to \bar{C}$ be a polynomial, factored according to (1-41), so that $P = \phi\psi\bar{P}$. Let M be the lambda locus corresponding to the reduced polynomial $\bar{P}$. In order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) all poles of P lie in the closed right half plane $\bar{\mathcal{R}}$,

b) all the zeros of $P(-1,\cdot)^*$ lie in the open unit disc $\mathcal{U}$, and,

c) the lambda locus is contained in the closed right half plane; that is, $M \subset \bar{\mathcal{R}}$.

The above result, together with Corollary 1.11, 1.13, or 1.15, yields an analytic characterization of A-stable composite multistep methods, dual to Corollary 2.5:

2.8. Corollary: Let (R,k) be a composite multistep method with $\Lambda \cap \mathcal{L}$ empty, and let P be its characteristic polynomial, factored according to (1-41), so that $P = \phi\psi\bar{P}$. Let M be the lambda locus corresponding to the reduced polynomial $\bar{P}$. In order for (R,k) to be A-stable, it is necessary and sufficient that

a) all the zeros of $P(-1,\cdot)^*$ lie in the open unit disc $\mathcal{U}$, and,

b) the lambda locus is contained in the closed right half plane; that is, $M \subset \bar{\mathcal{R}}$.

A slight modification of condition a) above produces a sufficiency condition for A-stability which is useful in many practical situations:

2.9. Corollary: Assume the hypotheses of Corollary 2.8. In order for (R,k) to be A-stable, it is sufficient that

a) all the zeros of $P(\infty,\cdot)$ lie in $\mathcal{U}$, and,

b) $M \subset \bar{\mathcal{R}}$.

Proof: If condition a) holds, then $\infty \notin M$. The unbounded component P of the complement of M is therefore a stability component. If condition b) holds, then $\mathcal{L} \subset P$, proving the corollary. □

*Any point in $\mathcal{L}$ can be substituted in this sentence in place of -1.

The lambda locus M and the zeta locus T associated with a reduced polynomial $\bar{P}$ have complementary properties, as is to be expected. Let $\lambda \in I$ and $\zeta \in E$ be fixed, where I is the imaginary axis, and E is the unit circle. Then $\lambda \in M$ and $\zeta \in T$ if and only if $\bar{P}(\lambda,\zeta) = 0$, by definition of M and T. In other words, for every point of intersection of the lambda locus with the imaginary axis, there corresponds a point of intersection of the zeta locus with the unit circle, and conversely. Suppose $I \subset M$. Since $\bar{P}$ is reduced, T contains an infinite number of points of E. But this is possible only if T contains all of E, since analytic continuations of curves in T are in T. . The roles of M and T are clearly interchangeable in this discussion. Therefore, the following result has been established:

2.10. Proposition: The lambda locus M associated with a reduced polynomial $\bar{P}$ contains the imaginary axis I if and only if the zeta locus T associated with $\bar{P}$ contains the unit circle E.

A plausible converse of Proposition 2.10 is false. Therefore, it is not true that $M = I$ if and only if $T = E$. As a counterexample, consider $\bar{P}(\lambda,\zeta) = (\lambda^2 + \lambda - 2)\zeta - (\lambda^2 - \lambda - 2)$. If $\lambda \in I$ and $\bar{P}(\lambda,\zeta) = 0$, then $|\zeta| = |\lambda^2 - 2 - \lambda| \,/\, |\lambda^2 - 2 + \lambda| = 1$. Therefore $T = E$. On the other hand, $\bar{P}(\sqrt{2},-1) = 0$, showing that $M \neq I$.

§2.5 Final Remarks and a Review of Previous Work

Lambda loci have been used to determine the stability regions of multistep methods at least since 1968 [Gear]. [Sloate] used both lambda and zeta loci in A-stability tests for composite multistep methods. The first attempt at a rigorous justification of locus methods in stability was made in 1971 [Watts]. Watts proved the sufficiency part of Corollary 2.5 (his Lemma 3.1) for the special case of one past point. In this case the Riemann surface reduces to a rational function. Watts invoked the maximum modulus principle for the proof. This approach can be extended to deal with the general case; however, the resulting sufficiency condition includes the hypothesis that the branch points of the Riemann surface (above the left half lambda plane) have all their zeta-sphere projections in the open unit disc. This hypothesis is awkward to test in practice, since it involves deriving and factoring a high-degree polynomial (the discriminant); furthermore, Corollary 2.5 shows the

hypothesis to be superfluous. Finally, Watts' approach apparently cannot provide stability theorems for lambda loci.

The stability component in the lambda sphere was treated rigorously by [Cooke] for multistep methods. The situation here is the dual of Watts', since λ is a rational function of ζ. Cooke presents a sufficiency condition for stiff stability which includes the hypothesis that this rational function be one-to-one on the unit circle. The conclusion of his theorem cannot readily be used to determine A-stability. Also, it is not apparent whether his method can be extended to the general case.

The work of [Odeh and Liniger, 1971], as with [Cooke], is concerned with the determination of a stability component in the lambda sphere for multistep methods. The approach of [Odeh and Liniger, 1971] appeals to abstract degree theory [Olum], and is very much in the spirit of the present work. Although their main result applies only to multistep methods (one future point), they discuss an extension to multistep methods "involving higher derivatives". Such methods lead to characteristic polynomials of the same type as with composite multistep methods. [Odeh and Liniger, 1971] does not specifically investigate A-stability or the primal (zeta locus) formulation, although their approach can be applied directly to both.

Locus methods in both the lambda and zeta spheres are quite useful in practice as a qualitative measure of how close a given method is to A-stability. A software package for producing computer-generated plots of lambda and zeta loci from a given polynomial in two variables has been written by Joel M. Tendler*, and has been available to the writer. The plots generated using this package have been found to be indispensible in the search for composite multistep methods with good stability properties. Lambda and zeta plots for certain new A-stable methods are given in Chapter 4. The proof that these methods are A-stable is based on the work of Chapter 3.

---

*See footnote at end of Chapter 1.

Chapter 3

AN ALGEBRAIC CHARACTERIZATION OF A-STABILITY

§3.1 Introduction

The analytic characterizations of the last chapter are of great practical value in exploring classes of methods to find A-stable ones. However, direct application of these characterizations can never be used to conclude with certainty that a given method is A-stable, because the locus to be computed still requires the exact factorization of an infinite* number of polynomials. The most that can be obtained in practice is the approximate factorization of a finite number of polynomials. Even if it were somehow possible to factor exactly a finite number of polynomials, the partial locus would contain gaps of finite length. Some examples discussed in Chapter 4 indicate that however small these gaps may be, there exists a polynomial P not satisfying the A-stability criterion, whose partial locus does not intersect the forbidden region. On the other hand, if it were somehow possible to approximately factor an infinite number of polynomials, the A-stability question would still not be resolved, for the following reason:

Fundamental accuracy requirements discussed in Chapter 4 lead to the fact that the reduced characteristic polynomial $\overline{P}$ for any composite multistep method of potential usefulness satisfies

$$\overline{P}(0,1) = 0 \ . \tag{3-1}$$

Since $0 \in \mathcal{I}$, it follows that $1 \in \mathcal{T}$, where $\mathcal{T}$ is the zeta locus corresponding to $\overline{P}$. Thus $\mathcal{T}$ contains an analytic curve passing through the point unity. Since $\overline{P}$ is a real polynomial, $\mathcal{T}$ is symmetrical about the real axis. It follows that the zeta locus contains a curve asymptotic to the unit circle, and thus to the forbidden region, at unity. Similarly, the lambda locus is asymptotic to the forbidden region (the left half plane) at the origin. Therefore, even if an entire locus were computed, but with uniform error, the result would not be delicate enough to decide the A-stability question in either

---

*It is not necessary, even in principle, to compute every point in the locus. Since the locus is the union of continuous curves, it is enough to compute the images (locus points) on a dense subset (of the domain), which can be taken to be countable.

direction, since if (3-1) holds, then a portion of either locus lies arbitrarily close to the boundary. It is literally true that every A-stable method of potential usefulness is a "borderline case" with respect to the locus criterion.

The zeta or lambda locus itself provides much more information than is needed for the determination of A-stability. At a given point in the domain one needs only to know in what region the images [pre-images] lie. For polynomials in one variable such questions are the subject of many classical and modern theorems [Marden], [Gantmacher, ch. 15], [Barnett]. This chapter shows how these theorems can be applied to the A-stability question in such a way that the pitfalls of discretization and approximation of the locus are avoided. The main results provide algebraic necessary and sufficient conditions, which are practical to implement, for a polynomial to satisfy the A-stability criterion. Because the computations can be carried out without introducing any approximations, the algebraic characterizations determine with certainty whether a given method is A-stable.

The algorithms developed in this chapter can be applied--in principle--to any polynomial P in two variables with real coefficients. However, in practice it is necessary to restrict consideration to polynomials with integral coefficients to avoid making approximations. As shown in Chapter 4, characteristic polynomials with integral coefficients arise in a natural way from accuracy specifications. In any case, the requirement of integral coefficients is not a serious restriction, for two reasons: (1) The class* of polynomials with integral coefficients is dense in the class of polynomials with real coefficients. Therefore, the behavior of the loci of any real polynomial can be approximated arbitrarily closely using polynomials with integral coefficients. (2) If a given polynomial P cannot be represented with integral coefficients, then there is no practical way of even representing its coefficients (let alone performing a stability analysis) without already introducing approximations.

For the A-stability criterion of Definition 1.6, the important regions in the Riemann sphere are the left half plane and the unit disc. By contrast,

---

*In much of the discussion on polynomials it has tacitly been assumed that two polynomials are equivalent if they differ by a trivial factor. Thus, classes of polynomials are, strictly speaking, sets of equivalence classes of polynomials.

the algebraic theory of the present chapter is most easily formulated in terms of the upper and lower half planes. Therefore, a natural procedure is to transform the A-stability criterion and the polynomials in two variables to this more suitable domain. This is discussed in the next section. In Section 3.3 the important Bezoutian matrix T formed from the transformed polynomial is defined, and its utility in finding the greatest real divisor of the transformed polynomial is demonstrated. The Bezoutian matrix is used again in Section 3.4 to formulate the fundamental characterization theorem for the transformed A-stability criterion, Theorem 3.9. In Section 3.5 the auxiliary polynomials are introduced, and the significance of their positive real zeros is shown.

Sections 3.6 and 3.7, together with the last part of Section 3.8, contain the main results of this chapter. Theorem 3.14 provides an algebraic characterization of the A-stability criterion, applicable to almost all polynomials in two variables. A completely general extension of this result is given by Theorem 3.16, which, however, is considerably more complicated. The direct application of these two theorems to composite multistep methods is indicated by Corollaries 3.15 and 3.17, respectively. It is shown in Theorem 3.19 that particularly simple algebraic conditions arise in the characterization of strong A-stability, a concept introduced in Section 3.7.

The analytic A-stability characterization of Section 2.3 is the foundation for the algebraic theory of Sections 3.2 through 3.7. By starting with the dual analytic A-stability characterization of Section 2.4, a dual algebraic theory can be developed. Section 3.8 outlines the main points in the parallel development, and presents dual theorems for all the principal results. Particularly noteworthy are Theorems 3.25, 3.26, and 3.28, which are the dual versions of Theorems 3.14, 3.16, and 3.18, respectively.

The practical implementation of the algebraic characterizations of this chapter is straightforward, although certain specialized programming techniques such as polynomial multiplication in many variables and infinite precision integer arithmetic are needed. These and other implementation considerations are discussed in Appendix E.

§3.2 The Transformed A-Stability Criterion

Let P be a real polynomial in (λ,ζ). No further restrictions are needed in the following developments; however, the manner in which integral coefficients are preserved will be emphasized.

Transform P into the polynomial P", and transform P" into the polynomial P' according to the identities

$$P''(\lambda,y) = (y-1)^m P(\lambda, \frac{y+1}{y-1}) \tag{3-2}$$

and

$$P'(\omega,z) = P''(\hat{j}\omega,-\hat{j}z) , \tag{3-3}$$

where m is the degree of P in ζ, and $\hat{j}$ is the imaginary unit. The coefficients of P" are sums of products of the coefficients of P. Therefore, if P has integral coefficients, so does P". Note that P" is a real polynomial while P' is a complex* polynomial. However, the transformation (3-3) requires complex arithmetic of only a trivial nature: multiplying real numbers by integral powers of $\hat{j}$. Thus, if P" has integral coefficients, the real and imaginary parts of the coefficients of P' are integers.

Combining (3-2) and (3-3) yields the composite transformation

$$P'(\omega,z) = (z-\hat{j})^m P(\hat{j}\omega, \frac{z+\hat{j}}{z-\hat{j}}) \tag{3-4}$$

after removing the trivial** factor $(-\hat{j})^m$. The polynomial P' defined by (3-2) and (3-3) will be called the transformed polynomial associated with P. The transformation $\omega \longrightarrow \hat{j}\omega$ maps the extended real axis conformally onto the extended imaginary axis I, while the transformation $z \longrightarrow (z+\hat{j})/(z-\hat{j})$ maps the closed lower half plane $\overline{H} = \{z \in \overline{C} : \text{Im } z \leq 0\}$ conformally onto the closed unit disc $\overline{U}$ [Rudin, Ch. 14]. Because the transformations are conformal, the important properties of P are generally preserved by P', as interpreted with respect to the new domains. The qualification "generally" is explained by the following preliminary result:

---

*A polynomial will be called complex [real] if all its coefficients are complex [real].

**With respect to complex polynomials, it is clear that a nonzero complex constant factor is to be regarded as a trivial factor. This additional application of the term "trivial factor" causes ambiguity with respect to polynomials which happen to be real. In such cases the context will make clear whether real or complex factors are intended.

3.1. Lemma: Let P be a real polynomial in (λ,ζ), of degree m in ζ, and let P' be its transformed polynomial, of degree m' in z. Let P be factored according to (1-41), repeated here as

$$P(\lambda,\zeta) = \phi(\lambda)\psi(\zeta)\overline{P}(\lambda,\zeta) \ . \tag{3-5}$$

The multiplicity of unity as a zero of ψ is equal to m - m'. In particular, m' = m if and only if $\psi(1) \neq 0$.

Proof: Formally applying the transformation (3-4) to (3-5) gives

$$P'(\omega,z) = \phi'(\omega)\psi'(z)\overline{P}'(\omega,z) \ , \tag{3-6}$$

where

$$\phi'(\omega) = \phi(\hat{j}\omega) \ , \tag{a}$$

$$\psi'(z) = (z-\hat{j})^{i}\psi\left(\frac{z+\hat{j}}{z-\hat{j}}\right) \ , \tag{b}$$

$$\overline{P}'(\omega,z) = (z-\hat{j})^{m-i}\overline{P}\left(\hat{j}\omega, \frac{z+\hat{j}}{z-\hat{j}}\right) \ , \tag{c}$$

(3-7)

and i is the degree of ψ. The "z-transformation" of $\overline{H}$ onto $\overline{U}$ has infinity as the pre-image of unity. Therefore (3-7b) shows the degree of ψ' to be i - j, where j is the multiplicity of unity as a zero of ψ. By (3-5) $\overline{P}$ is of degree m - i in ζ. Since $\overline{P}$ is reduced, (3-7c) shows that $\overline{P}'$ is of degree m - i in z. By (3-6) the degree of P' in z is m' = (i-j) + (m-i) = m - j, proving the first conclusion of Lemma 3.1. The second conclusion follows immediately from the first. □

3.2. Definition: A complex polynomial P' in (ω,z) is said to satisfy the transformed A-stability criterion if for every ω in the open upper half plane $G = \{\omega \in C : \text{Im } \omega > 0\}$, the zeros of $P'(\omega,\cdot)$ lie in the open lower half plane H.

3.3. Proposition: Let P be a real polynomial in (λ,ζ), of degree m in ζ, and let P' be its transformed polynomial, of degree m' in z. In order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) m' = m, and,

b) P' satisfies the transformed A-stability criterion.

Proof: If $m' \neq m$, then $\psi(1) = 0$, by Lemma 3.1. In such case $P(\lambda,1) = 0$ for all $\lambda \in C$ by (3-5), violating the A-stability criterion, Definition 1.6. Therefore condition a) is necessary.

If condition a) holds, then the transformation (3-4) has an inverse, which can be written as

$$P(\lambda,\zeta) = (\zeta-1)^{m'} P'(-\hat{j}\lambda, \hat{j}\,\frac{\zeta+1}{\zeta-1}) \quad . \tag{3-8}$$

The conformal mappings in (3-4) and the above, together with Definition 3.2, now show that condition b) is equivalent to the A-stability criterion, Corollary 1.9. □

It is clear that all the results of Chapter 2 have counterparts in the domain of the transformed polynomial. The following result is essentially Theorem 2.4 in the transformed domain, but in a form more useful for present purposes. The proof is given in Appendix C.

3.4. Proposition: In order for a complex* polynomial P' in (ω,z) to satisfy the transformed A-stability criterion, it is necessary and sufficient that

a) All zeros of $P'(\cdot,\hat{j})$ lie in the closed lower half plane $\overline{H}$,

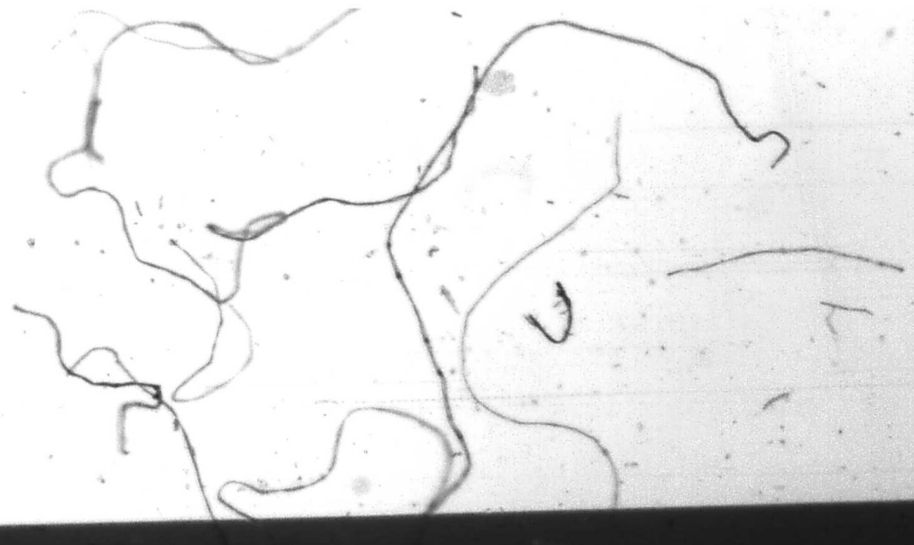
b) $P'(\cdot,z)$ is not the zero function for any real z, and,

c) For every real ω, if $P'(\omega,\cdot)$ is not the zero function. then all zeros of $P'(\omega,\cdot)$ lie in $\overline{H}$.

§3.3 Factorization of the Transformed Polynomial

A completely general treatment of the A-stability question requires the factorization described in this section (Proposition 3.6 and Theorem 3.7). In most cases of interest this factorization is trivial, and the results of Section 3.4 can be applied directly. However, the definitions of the next three paragraphs are fundamental to the rest of this chapter.

*Our only interest in the present work is in complex polynomials P' which are the transformed polynomials of real polynomials P.. Not every complex polynomial P' is of this type. Yet Proposition 3.4 and the given proof are valid in the general case as stated, even though they are based on the results of Chapter 2. The reason is that the entire development of Chapters 1 and 2 extends without modification to the general case of complex polynomials P (or complex composite matrices R).



Any complex polynomial P' in (ω,z) can be represented uniquely in the form

$$P'(\omega,z) = [1 \quad \hat{j}]\tau(\omega)[1 \quad z \quad z^2 \ \ldots \ z^{m'}]^T \tag{3-9}$$

where τ is a $2 \times (m'+1)$ matrix, each of whose elements is a real polynomial in ω or the zero function, and m' is the degree of P' in z. If P' is of degree n in ω, then τ is evidently of degree n. Also, the elements of τ have integral coefficients if the real and imaginary parts of the coefficients of P' are integers.

Let T be the $m' \times m'$ symmetric matrix whose elements t_{ij} are sums of 2×2 minors* of τ as follows:

$$t_{ij} = \sum_{s=\max(0,m'+1-i-j)}^{m'-\max(i,j)} \tau\begin{pmatrix} 1, & 2 \\ 2m'+1-i-j-s, & s \end{pmatrix}, \quad i,j = 1,2,\ldots,m'. \tag{3-10}$$

The matrix defined by (3-10) is simply the classical Bezoutian matrix** associated with the two polynomials in z whose coefficients form the two rows of τ. It can be seen from (3-10) that the elements of T are real polynomials in ω of degree at most 2n, or the zero function. Again note that if the elements of τ have integral coefficients, so do those of T, since only additions and multiplications appear in (3-10).

The nested principal minors of T are defined as follows: For each $i = 0, 1, 2, \ldots, m'$ let

$$\nabla_i' = T\begin{pmatrix} 1, & 2, & \ldots, & i \\ 1, & 2, & \ldots, & i \end{pmatrix}. \tag{3-11}$$

Clearly ∇_i' is a real polynomial in ω of degree at most 2ni, or the zero function. Also, if the elements of T have integral coefficients, so does ∇_i'. Note that $\nabla_0'(\omega) = 1$ for all ω.

One of the motivations for the above definitions is given in the proof of the following preliminary result:

---

*The notation for minors is the same as that in the proof of Lemma B.2.
**Properties of Bezoutian matrices are discussed in [Householder, 1970]. See also some of the references of that paper.

3.5. Lemma: Let P' be a complex polynomial in (ω,z), of degree m' in z. Let τ, T, and $\nabla'_{m'}$, be given by (3-9), (3-10), and (3-11), respectively. Let ω be a fixed real number. Then $\nabla'_{m'}(\omega) = 0$ if and only if there exists a $z \in \overline{C}$ such that

$$P'(\omega,z) = P'(\omega,z^*) = 0 \ , \tag{3-12}$$

where * denotes complex conjugate.

Proof: Let the subscript R [I] denote the real polynomial formed by replacing every coefficient of P' by its real [imaginary] part. Thus

$$P' = P'_R + \hat{j}P'_I \tag{3-13}$$

where P'_R and P'_I are real polynomials in (ω,z) or the zero function, but both are not the zero function. The coefficients of P'_R [P'_I] appear in the first [second] row of τ.

By (3-11) $\nabla'_{m'}(\omega) = \det T(\omega)$; thus $\nabla'_{m'}(\omega) = 0$ if and only if $T(\omega)$ is singular. But by [Householder, 1970] $T(\omega)$ is precisely* the Bezoutian matrix associated with the polynomials $P'_R(\omega,\cdot)$ and $P'_I(\omega,\cdot)$. Therefore, $T(\omega)$ is singular if and only if these two polynomials have a common zero [Householder, 1970, Theorem 1], that is, if and only if there exists a $z \in \overline{C}$ such that

$$P'_R(\omega,z) = P'_I(\omega,z) = 0 \ . \tag{3-14}$$

Relations (3-14) are easily shown to be equivalent to the relations

$$P'_R(\omega,z) + \hat{j}P'_I(\omega,z) = P'_R(\omega,z^*) + \hat{j}P'_I(\omega,z^*) = 0 \ ,$$

since $P'_R(\omega,\cdot)$ and $P'_I(\omega,\cdot)$ are real polynomials. Applying (3-13) to the above yields (3-12), thus proving the lemma. □

The greatest real divisor $\hat{D}$ of a complex polynomial P' in (ω,z) is, by definition, the real polynomial in (ω,z) of maximal degrees in ω and z such that

*Through private correspondence with Alston Householder, this writer has concluded that the sequence of polynomial coefficients in [Householder, 1970] is not given in the correct order, but in reverse order. This reversal presents a minor technical difficulty, which it is desirable to correct. If the given order of Householder's coefficient sequences is reversed (as will be assumed in all that follows), then Householder's definition of the Bezoutian matrix agrees with (3-10).

$$P' = \hat{D}\hat{P}' \tag{3-15}$$

where $\hat{P}'$ is a complex polynomial in (ω,z). It is evident that for every P' the factorization (3-15) exists and is essentially unique. The factorization (3-15) will be called trivial if $\hat{D}$ is of degree zero in z. Only nontrivial factorizations require special treatment, as shown in the next section. It is indicated in [Householder, 1970] that the factorization can be computed from the Bezoutian matrix T, although no explicit formulas are given. The result below can be used to determine whether the factorization (3-15) is trivial, and to actually compute the factorization.

3.6. Proposition: Let P' be a complex polynomial in (ω,z), of degree m' in z. Let τ, T, and ∇'_i, $i = 0, 1, \ldots, m'$, be given by (3-9), (3-10), and (3-11), respectively. Let $\hat{m}$ be the largest integer such that $\nabla'_{\hat{m}}$ is not the zero function. Then the greatest real divisor $\hat{D}$ of P' is of degree $m' - \hat{m}$ in z. Furthermore,

$$\hat{D}(\omega,z) = \begin{cases} P'_R(\omega,z) & \text{when } \hat{m} = 0 \text{ and } P'_R \text{ is not the zero function,} \\ P'_I(\omega,z) & \text{when } \hat{m} = 0 \text{ and } P'_R \text{ is the zero function,} \\ \dfrac{1}{\nabla(\omega)} \displaystyle\sum_{i=0}^{m'-\hat{m}} \nabla_{\hat{m},m'-i}(\omega) z^i & \text{when } \hat{m} > 0, \end{cases} \tag{3-16}$$

where

$$\nabla_{\hat{m},j} = T\begin{pmatrix} 1, 2, \ldots, \hat{m} \\ 1, 2, \ldots, \hat{m}-1, j \end{pmatrix} \qquad \begin{matrix} \hat{m} = 1, 2, \ldots, m' \\ j = \hat{m}, \hat{m}+1, \ldots, m' \end{matrix} \tag{3-17}$$

and ∇ is a polynomial in ω which divides all the coefficients $\nabla_{\hat{m},m'-i}$.

Proof: The first conclusion of Proposition 3.6 follows directly from [Householder, 1970, Theorem 1] and the proof of Lemma 3.5. The second conclusion can be pieced together as described in the next paragraph.

It is evident that $\hat{D}$ is precisely the greatest common divisor of the real polynomials P'_R and P'_I. For polynomials in one variable such divisors are classically computed by means of Euclid's algorithm, a process essentially the same as the computation of Sturm sequences [Jacobson]. The method of [Barnett, 1971c, Theorem 3] for computing Sturm sequences through a companion matrix

formulation can be modified to yield an expression for the greatest common divisor for polynomials in one variable. The transformation relating this result to the Bezoutian matrix can be performed using [Barnett, 1972]. When the resulting procedure is applied to polynomials containing the parameter ω in addition to the variable z, the "divisor" contains, as an extraneous factor, a polynomial ∇ in ω. This factor can be removed by a technique of [Bareiss]. The end result is (3-16) and (3-17). □

Once $\hat{D}$ has been found, by means of Proposition 3.6 or otherwise*, the factorization (3-15) can be completed by computing $\hat{P}' = P'/\hat{D}$. Whenever P'_R and P'_I have integral coefficients it can be arranged, using trivial factors if necessary, that $\hat{D}$, $\hat{P}'_R$ and $\hat{P}'_I$ have integral coefficients.

It is clear from (3-15) and Definition 3.2 that P' satisfies the transformed A-stability criterion if and only if both $\hat{P}'$ and $\hat{D}$ do. Since $\hat{D}$ is a real polynomial, the conditions under which it satisfies the transformed A-stability criterion can be made simpler than in the general case of a complex polynomial. These considerations yield the following:

3.7. Theorem: Let P' be a complex polynomial in (ω,z), factored according to (3-15). In order for P' to satisfy the transformed A-stability criterion, it is necessary and sufficient that

a) All zeros of $P'(\cdot,\hat{j})$ lie in the closed lower half plane $\overline{H}$,

b) $\hat{D}(\cdot,z)$ is not the zero function for any $z \in C$,

c) For every real ω, if $\hat{D}(\omega,\cdot)$ is not the zero function, then all zeros of $\hat{D}(\omega,\cdot)$ are real, and,

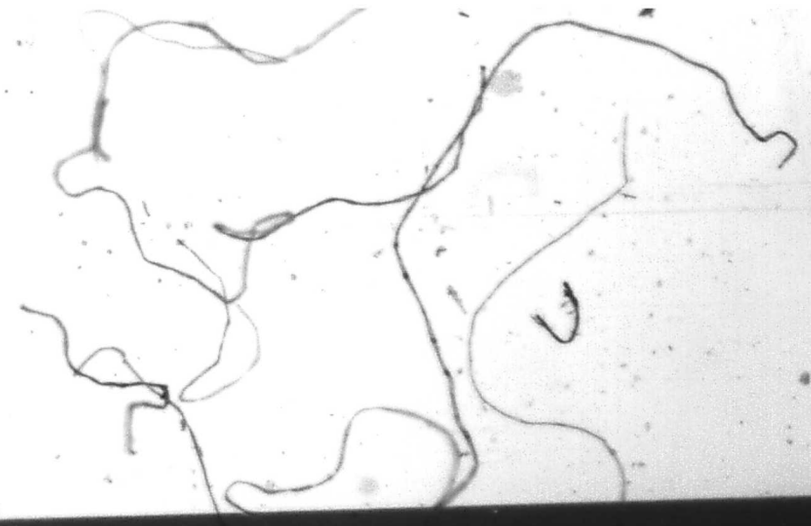
d) $\hat{P}'$ satisfies the transformed A-stability criterion.

Proof: P' satisfies the transformed A-stability criterion if and only if both $\hat{P}'$ and $\hat{D}$ do. Also, condition a) of Theorem 3.7 implies that all zeros of $\hat{D}(\cdot,\hat{j})$ lie in $\overline{H}$. In the following consider Proposition 3.4 as applied to $\hat{D}$. It is enough to show that conditions b) and c) of Theorem 3.7 taken together are equivalent to conditions b) and c) of Proposition 3.4.

The first half of the proof is immediate, since conditions b) and c) of Theorem 3.7 trivially imply conditions b) and c) of Proposition 3.4, respectively.

---

*A classical treatment of the greatest common divisor problem for polynomials in two variables can be found in [Bocher, Ch. 16].



Conversely, suppose conditions b) and c) of Proposition 3.4 hold. Let ω be real, $z \in \mathcal{C}$, and suppose $\hat{D}(\omega,z) = 0$. Since $\hat{\Gamma}$ is a real polynomial it follows that $\hat{D}(\omega,z^*) = 0$, where * denotes complex conjugate. If $\hat{D}(\omega,\cdot)$ is not the zero function, then by condition c) of Proposition 3.4 $z \in \overline{H}$ and $z^* \in \overline{H}$. Therefore z is real. This shows that condition c) of Theorem 3.7 holds.

To complete the proof, suppose condition b) of Theorem 3.7 fails to hold. Thus, there exists a $z \in \mathcal{C}$ such that $\hat{D}(\cdot,z)$ is the zero function. Since $\hat{D}$ is a real polynomial $\hat{D}(\cdot,z^*)$ is also the zero function. Let ω be a real number such that $\hat{D}(\omega,\cdot)$ is not the zero function. Then $\hat{D}(\omega,z) = \hat{D}(\omega,z^*) = 0$. Therefore, as before, z is real. This violates condition b) of Proposition 3.4. □

Conditions a) and b) of Theorem 3.7 can be dealt with by standard methods. If P' is the transformed polynomial corresponding to a real polynomial P in (λ,ζ) (the only case of interest in this work), then condition a) of Theorem 3.7 is equivalent to the condition that all zeros of the real polynomial δ in λ be in the closed right half plane. Alternatively, the zeros of δ_M must lie in the closed left half plane, where δ_M is constructed from δ by changing the sign of every other coefficient. The problem can now be solved by the classical Hurwitz criterion [Gantmacher, vol. 2, Ch. 15]. A particularly convenient form for the Hurwitz criterion is given by [Cutteridge, 1959].

Condition b) of Theorem 3.7 is equivalent to the condition that the coefficients of the powers of ω in $\hat{D}$ (which are polynomials in z) have no nonconstant common factor. This problem can be solved directly by approaches based on Euclid's algorithm. A modern approach (which must be modified to deal with the present case) is given by [Barnett, 1971d].

Conditions a) and b) of Theorem 3.7 have thus been reduced to the computation of sums of products (usually expressed in terms of determinants) of real numbers, and to the examination of the signs (positive, negative, or zero) of the results. If the coefficients of P'_R and P'_I are integers, then it is easy on a digital computer to perform all these computations exactly, since integers are preserved. This fact is one of the crucial principles which make the present stability analysis practical*. The actual computations generally require extended precision or infinite precision integer arithmetic.

*The well-known Routh criterion [Gantmacher, vol. 2, Ch. 15] is unsuitable in this regard, since it requires divisions, thereby destroying the absolute precision of integer representations within the computer. (Of course, absolute precision can be preserved within the Routh scheme by representing all numbers as the ratio of two integers; however, it is questionable whether such an approach has any advantages over methods of the Hurwitz type.)

In order to complete the algebraic characterization of A-stability, it is necessary to reduce conditions c) and d) of Theorem 3.7 to sums of products and sign determinations. Conditions c) and d) are of a less elementary nature than conditions a) and b), because in the former, two variables must be considered simultaneously. The reduction of condition d) to conditions on polynomials in only one variable is the subject of the next section. In Section 3.6 it will be indicated how the same kind of technique can be applied to deal with condition c). In this way the general problem of the transformed A-stability criterion is reduced to simple integer arithmetic.

§3.4 The Fundamental Characterization Theorem

The following fundamental result, which is the key to the algebraic characterization of A-stability, is a form of the Hurwitz criterion as applied to complex polynomials in one variable.

3.8. Theorem: Let P' be a complex polynomial in (ω,z), of degree m' in z. Let τ, T, and ∇'_i, $i = 0, 1, 2, \ldots, m'$ be given by (3-9), (3-10), and (3-11), respectively. Let ω be a fixed real number, and assume $\nabla'_{m'}(\omega) \neq 0$. Then $P'(\omega,\cdot)$ has no real zeros. Furthermore, in order for all zeros of $P'(\omega,\cdot)$ to lie in the open lower half plane H, it is necessary and sufficient that $\nabla'_i(\omega) > 0$ for all $i = 1, 2, \ldots, m'$.

Proof: Theorem 3.8 is the immediate consequence of [Gantmacher, vol. 2, Ch. 15, sect. 18] and [Cutteridge, 1959] together with the following observations. Gantmacher's result gives the number of zeros of a complex polynomial in the open right half plane, in terms of Hurwitz determinants. It is a simple matter to restate his Theorem 23 (page 250) to determine the number of zeros in the open upper half plane G. It has been shown by [Cutteridge, 1959, eq. 38], in a notation suited to his more specialized application, that the $2i$-th order Hurwitz determinant used by Gantmacher (see his eq. (157)) is equal to $\nabla'_i(\omega)$.

According to Gantmacher's result (transformed as indicated in the above paragraph), $P'(\omega,\cdot)$ has no real zeros. Furthermore, the number of zeros of $P'(\omega,\cdot)$ in G is equal to the number of sign changes in the sequence $\nabla'_0(\omega), \nabla'_1(\omega), \ldots, \nabla'_{m'}(\omega)$. It follows that no zeros lie in G (that is, all zeros lie in H) if and only if there are no sign changes.

A special rule is given [Gantmacher, p. 250] to handle the situation in which some elements of the sequence (other than $\nabla'_{m'}(\omega)$) are zero. It turns out (his eq. (160)) that if there are any zero elements in the sequence, then there is at least one equivalent sign change. Of course, if any $\nabla'_i(\omega)$ is negative, there must also be a sign change, since $\nabla'_0(\omega) = 1$, which is positive. In summary, there are no sign changes if and only if $\nabla'_i(\omega) > 0$ for all $i \geq 1$. □

The fundamental characterization theorem for the transformed A-stability criterion is presented below. The central idea of the proof is that Theorem 3.8 can be applied for every real value of ω.

3.9. Theorem: Let P' be a complex polynomial in (ω,z), of degree m' in z. Let τ, T, and ∇'_i, $i = 1, 2, \ldots, m'$, be given by (3-9), (3-10), and (3-11), respectively. Assume $\nabla'_{m'}$ is not the zero function. Then $P'(\cdot,z)$ is not the zero function for any real z. Furthermore, in order for P' to satisfy the transformed A-stability criterion, it is necessary and sufficient that

a) All zeros of $P'(\cdot,\hat{j})$ lie in the closed lower half plane $\overline{H}$,

b) ∇'_i is not the zero function, for all $i = 1, 2, \ldots, m'-1$, and,

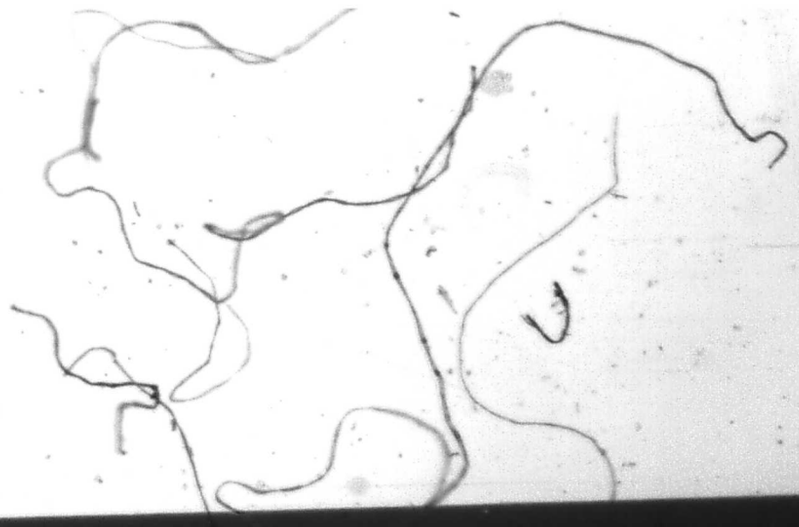
c) $\nabla'_i(\omega) \geq 0$ for all $i = 1, 2, \ldots, m'$, and all real ω.

Proof: If $\nabla'_{m'}$ is not the zero function, it is a polynomial. Therefore, there exists a real ω such that $\nabla'_{m'}(\omega) \neq 0$. By Theorem 3.8 $P'(\omega,z) \neq 0$ for all real z. This proves the first conclusion of Theorem 3.9.

Observe that conditions a) of Proposition 3.4 and Theorem 3.9 are the same, and that condition b) of Proposition 3.4 is just the first conclusion of Theorem 3.9. Therefore, Theorem 3.9 will be proved if it can be shown that conditions b) and c) of Theorem 3.9 are equivalent to condition c) of Proposition 3.4.

Let ω be a real number, and suppose $P'(\omega,\cdot)$ is the zero function. It follows that $\tau(\omega) = 0$, and hence that $T(\omega) = 0$. Therefore $\nabla'_{m'}(\omega) = 0$. This shows that if $\omega \in \mathcal{Y}$, where $\mathcal{Y} = \{\omega : \nabla'_{m'}(\omega) \neq 0, \omega \text{ real}\}$, then $P'(\omega,\cdot)$ is not the zero function. By assumption $\nabla'_{m'}$ is not the zero function, and hence $\mathcal{Y}$ is the real line except possibly for a finite set of points.

First assume condition c) of Proposition 3.4 holds. Let $\omega \in \mathcal{Y}$. By condition c) of Proposition 3.4 all zeros of $P'(\omega,\cdot)$ lie in $\overline{H}$. But by Theorem 3.8 $P'(\omega,\cdot)$ has no real zeros. Therefore all zeros of $P'(\omega,\cdot)$ lie



in H. By Theorem 3.8 it follows that $\nabla_i'(\omega) > 0$ for all $i = 1, 2, \ldots, m'$. Therefore condition b) of Theorem 3.9 holds. Furthermore, each ∇_i' is a continuous function, and the real axis is the closure of Y. This shows that condition c) of Theorem 3.9 holds.

Conversely, assume conditions b) and c) of Theorem 3.9 hold. Let $\hat{Y} = \{\omega : \nabla_i'(\omega) \neq 0$ for all $i = 1, 2, \ldots, m'$, ω real$\}$. By condition b) the real axis is the closure of $\hat{Y}$. From condition c) it follows that if $\omega \in \hat{Y}$, then $\nabla_i'(\omega) > 0$ for all $i = 1, 2, \ldots, m'$. Theorem 3.8 now shows that all zeros of $P'(\omega,\cdot)$ lie in H whenever $\omega \in \hat{Y}$. Extending $\hat{Y}$ to the real axis and using the usual continuity argument shows that condition c) of Proposition 3.4 holds. □

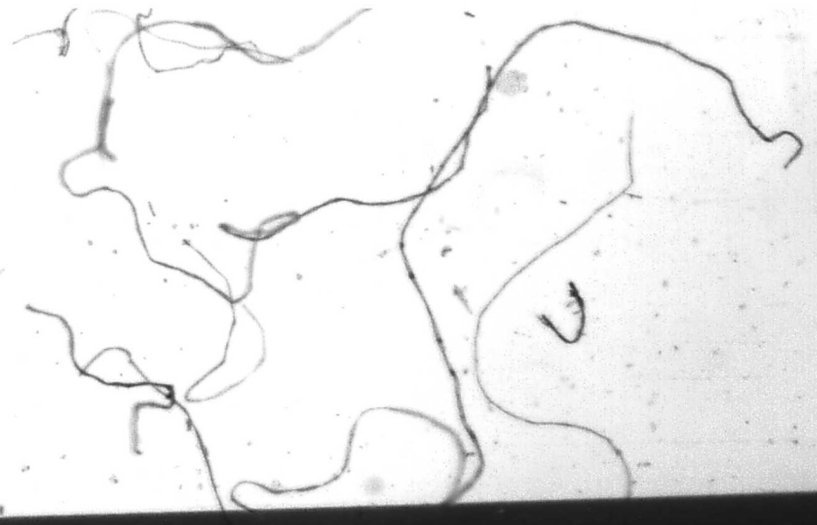
The converse of the first conclusion of Theorem 3.9 is false, in general. That is, if $\nabla_{m'}'$ is the zero function it does not follow that $P'(\cdot,z)$ is the zero function for some real z. The exact situation is that if $\nabla_{m'}'$ is the zero function, then P' has a nontrivial factorization (3-15), as indicated by Proposition 3.6. For real z, $P'(\cdot,z)$ is the zero function if and only if $\hat{D}(\cdot,z)$ is.

§3.5 Real and Transformed Polynomials

The discussion of this chapter, except for Lemma 3.1 and Proposition 3.3, has dealt with arbitrary complex polynomials in (ω,z). For the remainder of the chapter, attention will be restricted to those complex polynomials which are the transformed polynomials of real polynomials in (λ,ζ). This class of polynomials is, of course, exactly the class of interest for dealing with composite multistep methods. For this class, the polynomials ∇_i' of Theorem 3.9 can be replaced by polynomials of much lower degree, as shown in this section and the next.

Recall that a function ∇ (of a real variable) is even, if $\nabla(-\omega) = \nabla(\omega)$ identically on the real axis. A polynomial in one variable is even if and only if all its odd coefficients are zero. Also, the zero function is even. The following algebraic result is proved in Appendix D:

3.10. Proposition: Let P' be the transformed polynomial in (ω,z), of degree m' in z, associated with a real polynomial P in two variables. Let τ, T, and ∇_i', $i = 0, 1, 2, \ldots, m'$ be given by (3-9), (3-10), and (3-11), respectively. Then each ∇_i' is even.



By Proposition 3.10 it is convenient to apply the transformation $\omega \rightarrow \omega^2 = \Omega$ to ∇'_i for i = 0, 1, 2, ..., m', defining the function ∇''_i as follows:

$$\nabla''_i(\Omega) = \nabla'_i(\Omega^{1/2}) \tag{3-18}$$

If ∇'_i is the zero function, so is ∇''_i; otherwise, ∇''_i is a real polynomial of degree at most ni, whose coefficients are just the even coefficients of ∇'_i.

3.11. Proposition: Assume the hypotheses of Proposition 3.10, and let ∇''_i be given by (3-18). If (3-1) holds, then $\nabla''_i(0) = 0$ for all i = 1, 2, ..., m'.

Proof: By (3-1) and (3-7c) $\bar{P}'(0,\infty) = 0$. Hence, by (3-6) $P'(0,\infty) = 0$. Therefore, the last column (column m') of $\tau(0)$ in (3-9) is zero. It follows that for every i = 1, 2, ..., m', $\tau(^{1}_{m'},{}^{2}_{m'-i})(0) = 0$. But (3-10) for j = 1 gives $t_{il} = \tau(^{1}_{m'},{}^{2}_{m'-i})$. Combining these two relations shows that the first column of T(0) is zero. Hence $\nabla'_i(0) = 0$ by (3-11). Setting $\Omega = 0$ in (3-18) now gives $\nabla''_i(0) = 0$, which was to be proved. □

Let k_i be the multiplicity* of 0 as a zero of ∇''_i. (If ∇''_i is the zero function, let $k_i = 0$.) Define the function ∇_i for i = 0, 1, 2, ..., m' by

$$\nabla_i(\Omega) = \nabla''_i(\Omega)/\Omega^{k_i} . \tag{3-19}$$

If ∇''_i is the zero function, so is ∇_i; otherwise, ∇_i is a real polynomial of degree at most ni - k_i, whose coefficients are just those of ∇''_i after dropping all leading zero coefficients. The polynomials ∇_i (any of which may be the zero function) generated by (3-9), (3-10), (3-11), (3-18), and (3-19) will be called the auxiliary polynomials associated with the transformed polynomial P'.

The above discussion has shown that the definition of auxiliary polynomial makes sense for a polynomial P' related to a real polynomial P by (3-2) and (3-3). Actually, the definition of auxiliary polynomial associated with P' also makes sense when $\hat{j}P'$ is related to P by (3-2) and (3-3). To see this,

*Proposition 3.11 shows that $k_i \geq 1$ in all cases of interest (except that $k_0 = 0$). For composite multistep methods of potential usefulness, k_i has been observed empirically to exceed this lower bound substantially. This point is discussed further in Chapter 4.

let τ [$\hat{\tau}$] be the matrix associated with P' [$\hat{j}$P'] in the representation (3-9). Then $\hat{\tau} = U\tau$, where $U = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. It follows that all the corresponding 2 x 2 minors of τ and $\hat{\tau}$ are equal. But the auxiliary polynomials depend upon only these minors (see (3-10)). Therefore, either P' or $\hat{j}$P' can be used to generate the same auxiliary polynomials, when either of them is related to a real polynomial through (3-2) and (3-3).

Auxiliary polynomials cannot be defined for arbitrary complex polynomials in two variables. The following technical result can be used to show that auxiliary polynomials are well defined for the polynomial $\hat{P}'$ of (3-15), whenever P' is the transformed polynomial associated with a real polynomial in two variables.

3.12. Lemma: Let P' be a complex polynomial in two variables, let $\hat{D}$ be its greatest real divisor, and let $\hat{P}'$ be given by the factorization (3-15). If P' is the transformed polynomial associated with a real polynomial in two variables according to (3-2) and (3-3), then so are $\hat{D}$ and $\hat{P}'$ or $\hat{\hat{j}}\hat{D}$ and $\hat{\hat{j}}\hat{P}'$.

Proof: The proof is based on a triple application of Lemma D.1. To begin, observe that the relations (D-1) hold for P' by the necessity part of Lemma D.1 and the hypothesis of Lemma 3.12.

The polynomial $\hat{D}$ is precisely the greatest common divisor of P'_R and P'_I, as noted in the proof of Proposition 3.6. Therefore $\hat{D}_M$ is the greatest common divisor of P'_{RM} and P'_{IM}. But by (D-1) $\hat{D}_M$ is the greatest common divisor of P'_R and P'_I. Since greatest common divisors are essentially unique, it follows that $\hat{D}_M = \pm \hat{D}$. Analogously to (3-13) write

$$\hat{P}' = \hat{P}'_R + \hat{\hat{j}}\hat{P}'_I .$$

By (3-13), (3-15), and the above, it follows that $P'_R = \hat{\hat{D}}\hat{P}'_R$ and $P'_I = \hat{\hat{D}}\hat{P}'_I$. Now $\pm \hat{\hat{D}}\hat{P}'_{RM} = \hat{D}_M\hat{P}'_{RM} = (\hat{\hat{D}}\hat{P}'_R)_M = P'_{RM} = P'_R = \hat{\hat{D}}\hat{P}'_R$. Since $\hat{D}$ is not the zero function this shows that $\hat{P}'_{RM} = \pm \hat{P}'_R$. In the same way it can be shown that $\hat{P}'_{IM} = \pm \hat{P}'_I$.

In case $\hat{D}_M = \hat{D}$, note that since $\hat{D}$ is real, $\hat{D}_{RM} = \hat{D}_R = \hat{D}$ and $\hat{D}_{IM} = -\hat{D}_I = 0$. Therefore, by the sufficiency part of Lemma D.1, $\hat{D}$ is the transformed polynomial of a real polynomial. When $\hat{D}_M = \hat{D}$ the above paragraph shows that

$\hat{P}'_{RM} = \hat{P}'_R$ and $\hat{P}'_{IM} = -\hat{P}'_I$. Therefore $\hat{P}'$ is also the transformed polynomial of a real polynomial.

In case $\hat{D}_M = -\hat{D}$ one can write $(\hat{j}\hat{D})_{RM} = (\hat{j}\hat{D})_R = 0$ and $(\hat{j}\hat{D})_{IM} = -(\hat{j}\hat{D})_I = -\hat{D}$. Therefore $\hat{j}\hat{D}$ is the transformed polynomial of a real polynomial. Furthermore, $(\hat{j}\hat{P}')_{RM} = -\hat{P}'_{IM} = -\hat{P}'_I = (\hat{j}\hat{P}')_R$ and $(\hat{j}\hat{P}')_{IM} = \hat{P}'_{RM} = -\hat{P}'_R = -(\hat{j}\hat{P}')_I$, showing that $\hat{j}\hat{P}'$ is the transformed polynomial of a real polynomial. □

Lemma 3.12 is not so shallow a result as it might seem. The hypothesis that $\hat{D}$ be "the greatest real divisor" of P' cannot be weakened to "a divisor" or even to "a real divisor". For simplicity, the following pathological counterexample is offered: Let P' be given by (3-15), where $\hat{D}(\omega,z) = z + 1$ and $\hat{P}'(\omega,z) = z - 1$. By Lemma D.1 P' is the transformed polynomial associated with a real polynomial in two variables, but $\hat{D}$, $\hat{j}\hat{D}$, $\hat{P}'$, and $\hat{j}\hat{P}'$ are not.

By Lemma 3.12 and the discussion of this section, auxiliary polynomials can be defined for the polynomial $\hat{P}'$. This fact is needed in order to state the general algebraic A-stability characterization of Section 3.6, Theorem 3.16. Lemma 3.12 is also needed elsewhere in the proof of Theorem 3.16.

The positive real zeros of the auxiliary polynomials, especially those of $\nabla_{m'}$, are of particular significance, as shown by the following result, as well as by the developments of the next two sections.

3.13. Proposition: Let P be a real polynomial in (λ,ζ), and let P' be its transformed polynomial. Assume both are of degree m' in their second arguments. Let $\nabla_{m'}$ be the m'-th auxiliary polynomial associated with P'. Let ω be a fixed nonzero real number. Then $\nabla_{m'}(\omega^2) = 0$ if and only if there exists a $\zeta \in \bar{C}$ such that $P(\hat{j}\omega,\zeta) = P(\hat{j}\omega, \frac{1}{\zeta^*}) = 0$. Therefore, if the lambda locus M associated with the reduced polynomial $\bar{P}$ of P intersects the imaginary axis I at $\pm \hat{j}\omega$, then $\nabla_{m'}(\omega^2) = 0$.

Proof: It is obvious that $\nabla_{m'}(\omega^2) = 0$ if and only if $\nabla'_{m'}(\omega) = 0$. The latter condition is equivalent, by Lemma 3.5, to the existence of a $z \in \bar{C}$ such that (3-12) holds. Since P and P' are of the same degree in their second arguments, the transformation (3-4) is invertible. Therefore, (3-12) holds if and only if $P(\hat{j}\omega,\zeta) = P(\hat{j}\omega, \frac{1}{\zeta^*}) = 0$, where $\zeta = \frac{z+\hat{j}}{z-\hat{j}}$ (and thus $\frac{1}{\zeta^*} = \frac{z^*+\hat{j}}{z^*-\hat{j}}$). This proves the first conclusion of Proposition 3.13.

The second conclusion follows easily from the "if" part of the first. If $\hat{j}\omega \in M$, then $\bar{P}(\hat{j}\omega,\zeta) = 0$ for some ζ on the unit circle E, by definition of M. Therefore $P(\hat{j}\omega,\zeta) = 0$. But $\zeta \in E$ implies that $\frac{1}{\zeta^*} = \zeta$. Therefore $P(\hat{j}\omega, \frac{1}{\zeta^*}) = 0$. □

Proposition 3.13 can be used as a direct method to compute the points at which the lambda locus crosses the imaginary axis. As noted in Section 2.4, each such crossing corresponds to a crossing by the zeta locus of the unit circle.

The developments of this section have shown some of the special properties which simplify the analysis of transformed polynomials associated with real polynomials in two variables. These simplifications are exploited in the remainder of Chapter 3.

§3.6 The Algebraic A-Stability Characterization

Consider the specialization of the Fundamental Characterization Theorem (Theorem 3.9) to transformed polynomials associated with real polynomials in two variables. In such case it is advantageous to employ the auxiliary polynomials of Section 3.5. The resulting theorem, presented below, provides an algebraic characterization of the A-stability criterion for all real polynomials whose transformed polynomials have trivial factorizations. Such real polynomials are precisely the ones for which $\nabla_{m'}$ is not the zero function, according to Proposition 3.6.

3.14. Theorem: Let P be a real polynomial in (λ,ζ), of degree m in ζ, let P' be its transformed polynomial, of degree m' in z, and let ∇_i, i = 1, 2, ..., m', be the auxiliary polynomials associated with P'. Assume $\nabla_{m'}$ is not the zero function. Then the lambda locus M associated with the reduced polynomial $\bar{P}$ of P does not contain the imaginary axis I, nor does the zeta locus T contain the unit circle E. Furthermore, in order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) All poles of P lie in the closed right half plane $\bar{\mathcal{R}}$,

b) $m' = m$,

c) $\nabla_i(0) > 0$ for all i = 1, 2, ..., m', and,

d) ∇_i has no positive real zeros of odd multiplicity, for i = 1, 2, ..., m'.

Proof: By assumption $\nabla_{m'}(\omega^2) \neq 0$ for almost all ω. Therefore $M \cap I$ is a finite set, by Proposition 3.13. The fact that $E \notin T$ is immediate from Proposition 2.10. This proves the first part of Theorem 3.14.

To prove the second part, first note that by Proposition 3.3 P satisfies the A-stability criterion if and only if P' satisfies the transformed A-stability criterion, and condition b) of Theorem 3.14 holds. Now consider Theorem 3.9. Conditions a) of Theorems 3.9 and 3.14 are equivalent, as shown in Appendix C. Also, if condition b) of Theorem 3.9 fails to hold, so does condition c) of Theorem 3.14, since ∇_i is the zero function whenever ∇'_i is. Hence, it is enough to show that if condition b) of Theorem 3.9 holds, then condition c) of Theorem 3.9 is equivalent to conditions c) and d) of Theorem 3.14.

By (3-18) we can write $\nabla''_i(\omega^2) = \nabla'_i(\omega) = \nabla'_i(-\omega)$ for all (real) ω. Therefore condition c) of Theorem 3.9 is equivalent to the following statement:

d) $\nabla''_i(\Omega) \geq 0$ for all $i = 1, 2, \ldots, m'$, and all real $\Omega \geq 0$.

Since $\Omega^{k_i} > 0$ for all real $\Omega > 0$, (3-19) shows, using continuity of ∇_i at zero, that condition d) is equivalent to the statement

e) $\nabla_i(\Omega) \geq 0$ for all $i = 1, 2, \ldots, m'$, and all real $\Omega \geq 0$.

The following result is an elementary exercise in analysis: Let ∇ be a real polynomial, let $\Omega_0 \geq 0$ be fixed, and suppose $\nabla(\Omega_0) \neq 0$. Then $\nabla(\Omega) \geq 0$ for all real $\Omega \geq 0$ if and only if $\nabla(\Omega_0) > 0$ and ∇ has no positive real zeros of odd multiplicity [Siljak, 1971]. For the present application put $\Omega_0 = 0$, and note that $\nabla_i(0) \neq 0$ by construction. (See (3-19), and recall the assumption that ∇_i is not the zero function.) The above considerations show that condition e) is equivalent to conditions c) and d) of Theorem 3.14. This proves the second part of Theorem 3.14. □

In applications of Theorem 3.14, condition a) is easily checked, as discussed at the end of Section 3.3. Checking conditions b) and c) is trivial; $\nabla_i(0)$ is merely the first coefficient of ∇_i, if ∇_i is not the zero function. Condition d) can be dealt with by methods based on the classical results of Sturm. Sturm's theorem [Jacobson, p. 283] implies that the number of positive real zeros of a real polynomial ∇ can be determined from the Sturm sequence of polynomials generated* from ∇, by examining the signs of certain coefficients.

*A modern approach to the computation of both Sturm sequences and the number of zeros is given by [Barnett, 1971c]. See also [Cutteridge, 1960] and [Siljak, 1971].

The multiplicities of positive real zeros of ∇ can be determined by examining the successive derivatives of ∇.

The application of Theorem 3.14 to the A-stability problem for composite multistep methods is straightforward. According to Corollary 1.11, 1.13, or 1.15, a composite multistep method (R,k) with no poles in the open left half plane $\mathcal{L}$ is A-stable if and only if its characteristic polynomial P satisfies the A-stability criterion. Combining this fact with Theorem 3.14 gives an algebraic characterization of such methods. For this application of Theorem 3.14 note that condition a) of Theorem 3.14 always holds, by the hypothesis that $\Lambda \cap \mathcal{L}$ be empty.

3.15. Corollary: Let (R,k) be a composite multistep method with $\Lambda \cap \mathcal{L}$ empty, let P be its characteristic polynomial, of degree m in ζ, let P' be the transformed polynomial associated with P, of degree m' in z, and let ∇_i, $i = 1, 2, \ldots, m'$, be the auxiliary polynomials associated with P'. Assume $\nabla_{m'}$ is not the zero function. In order for (R,k) to be A-stable, it is necessary and sufficient that

a) $m' = m$,

b) $\nabla_i(0) > 0$ for all $i = 1, 2, \ldots, m'$, and,

c) ∇_i has no positive real zeros of odd multiplicity, for $i = 1, 2, \ldots, m'$.

In Section 3.5 it was shown that if the transformed polynomial P' is associated with a real polynomial, then auxiliary polynomials can be defined for the polynomial $\hat{P}'$ of (3-15). Therefore, the techniques of Theorem 3.14 can be applied to $\hat{P}'$. Combining this result with Theorem 3.7 provides a completely general algebraic characterization of the A-stability criterion for real polynomials in two variables:

3.16. Theorem: Let P be a real polynomial in (λ,ζ), of degree m in ζ, and let P' be its transformed polynomial, of degree m' in z. Let $\hat{D}$ and $\hat{P}'$ be given by (3-15), (3-16), and (3-17), where $\hat{m}$ is the degree of $\hat{P}'$ in z. Let ∇_i, $i = 1, 2, \ldots, \hat{m}$, be the auxiliary polynomials associated with $\hat{P}'$.

In order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) All poles of P lie in the closed right half plane $\bar{\mathcal{R}}$,

b) $m' = m$,

c) $\hat{D}(\cdot,z)$ is not the zero function for any $z \in C$,

d) For every positive real ω, if $\hat{D}(\omega,\cdot)$ is not the zero function, then all zeros of $\hat{D}(\omega,\cdot)$ are real,

e) $\nabla_i(0) > 0$ for all $i = 1, 2, \ldots, \hat{m}$, and,

f) ∇_i has no positive real zeros of odd multiplicity, for $i = 1, 2, \ldots, \hat{m}$.

Proof: Conditions b) and a) of Theorem 3.16 arise from Proposition 3.3 and Theorem 3.7, respectively, in the same manner as in Theorem 3.14. Condition c) of Theorem 3.16 is identical with condition b) of Theorem 3.7. By Lemmas 3.12 and D.1 $\hat{D}_M = \pm \hat{D}$, since P' is associated with a real polynomial. (This relation is also derived in the proof of Lemma 3.12.) A little thought shows that if $\hat{D}_M = \pm \hat{D}$, then conditions c) of Theorem 3.7 and d) of Theorem 3.16 are equivalent. The proof of Theorem 3.16 from Theorem 3.7 will be complete if it can be shown that condition d) of Theorem 3.7 is equivalent to conditions e) and f) of Theorem 3.16, whenever condition a) of Theorem 3.7 holds.

By construction $\hat{P}'$ has a trivial factorization. Therefore $\nabla_{\hat{m}}$ is not the zero fuction, by Proposition 3.6. The same arguments used in the proof of Theorem 3.14 can now be applied to $\hat{P}'$ to show the equivalence of conditions e) and f) of Theorem 3.16 with conditions b) and c) of Theorem 3.9. In this way Theorem 3.9 provides the final desired equivalence. □

The application of Theorem 3.16 requires the checking of two kinds of conditions--conditions c) and d)--in addition to the kinds of conditions in Theorem 3.14. The checking of condition c) is discussed in Section 3.3. Condition d), however, still presents a nontrivial problem. Just as condition d) of Theorem 3.7 was reduced to the Fundamental Characterization Theorem (Theorem 3.9) through the theorem of Hurwitz, so also can condition d) of Theorem 3.16 be reduced to a similar kind of result by means of the theorem of Sturm. The same technique of performing only additions and multiplications of polynomials in ω can be maintained. In this way the entire problem of characterizing the A-stability criterion for real polynomials in two variables is reduced to such operations.

The importance of Theorem 3.16 is that it applies to the entire class of real polynomials P in two variables; by contrast, Theorem 3.14 applies only

to those for which P' has a trivial factorization. Yet Theorem 3.14 is not quite a special case of Theorem 3.16, but provides instead a simplification of the computations of Theorem 3.16 for the restricted problem. Thus, suppose P' is a complex polynomial with a trivial factorization (3-15), that is, with $\nabla'_{m'}$ not the zero function. In order to apply Theorem 3.16 it is still necessary to compute the greatest real divisor $\hat{D}$ (which may be a polynomial in ω) by means of Proposition 3.6. Then the polynomial $\hat{P}'$ must be found, $\nabla'_{\hat{m}}$ must be computed to replace $\nabla'_{m'}$, etc. Furthermore, conditions c) and d) of Theorem 3.16 are trivially satisfied. On the other hand, Theorem 3.14 makes all these computations unnecessary in this case. That is, if $\nabla'_{m'}$ is not the zero function, then Theorem 3.14 is more convenient to apply than Theorem 3.16.

The characterization of A-stability for composite multistep methods based on Theorem 3.16 follows directly from the same considerations as led to Corollary 3.15.

3.17. Corollary: Let (R,k) be a composite multistep method with $\Lambda \cap \mathcal{L}$ empty, and let P be its characteristic polynomial. (Therefore, P is real.) Assume the hypotheses of the first paragraph of Theorem 3.16 to hold. In order for (R,k) to be A-stable, it is necessary and sufficient that conditions b) through f) of Theorem 3.16 hold.

Consider the following trivial application of Corollary 3.17: Let (R,k) be a composite one-step method with $\Lambda \cap \mathcal{L}$ empty, and with characteristic polynomial P of the special form $P(\lambda,\zeta) = \delta(\lambda)\zeta - \delta_M(\lambda)$, where δ is a polynomial of degree 1 or greater. In such a case it is easy to show that $\hat{P}'(\omega,z) = 1$ and $\hat{D}$ is of the form $\hat{D}(\omega,z) = \theta_1(\omega)z + \theta_0(\omega)$, where θ_0 [θ_1] is an even [odd] polynomial, neither of which is the zero function. Therefore, it can be shown that all conditions of Theorem 3.16 are satisfied, so that (R,k) is A-stable. This conclusion was derived in [Watts] (for the case k = n).

§3.7 Strong A-Stability

For many important classes of A-stable composite multistep methods*, the transformed polynomials have nontrivial factorizations, and hence the

---

*For example, see [Bickart, et al] and [Watts].

lambda loci contain the imaginary axis. For such methods Corollary 3.15 is not applicable, and only the general result, Corollary 3.17, can be used. On the other hand, among the A-stable composite multistep methods for which Corollary 3.15 is applicable, most satisfy a slightly stronger property than A-stability; this stronger property will be called strong A-stability. Strong A-stability is of particular interest because it can be examined most easily by an algebraic characterization.

3.18. Definition: A composite multistep method (R,k) is said to be strongly A-stable if

a) (R,k) is A-stable,

b) R is regular,

c) Its characteristic polynomial P has no removable poles in the closed left half plane $\bar{\mathcal{L}}$, and,

d) The lambda locus M associated with the reduced polynomial $\bar{P}$ corresponding to P does not intersect the imaginary axis I, except possibly at the origin and infinity.

From the discussion of Section 2.4, it can be seen that condition d) of Definition 3.18 is equivalent to the following condition:

d') The zeta locus T associated with $\bar{P}$ does not intersect the unit circle E, except at zeros of $\bar{P}(0,\cdot)$ and $\bar{P}(\infty,\cdot)$ on E.

The condition that the lambda locus M not include the point at infinity (or alternatively, that the zeta locus T not contain zeros of $\bar{P}(\infty,\cdot)$ on the unit circle) is equivalent to the condition that $\nabla''_{m'}$, be of degree m'n, where $\nabla''_{m'}$ is given by (3-18).

The following algebraic characterization derives its importance from the fundamental simplicity of its necessary and sufficient conditions.

3.19. Theorem: Let (R,k) be a composite multistep method with m past points, let Λ be its set of poles, let P of (1-52) be its characteristic polynomial, let P' be the transformed polynomial associated with P, of degree m' in z, and let ∇_i, i = 1, 2, ..., m', be the auxiliary polynomials associated with P'. In order for (R,k) to be strongly A-stable, it is necessary and sufficient that

a) Λ is contained in the open right half plane $\mathcal{R}$,

b) m' = m,

c) $\nabla_i(0) > 0$ for all i = 1, 2, ..., m', and,

d) $\nabla_{m'}$ has no positive real zeros.

Proof: To prove necessity, suppose (R,k) is strongly A-stable. First it is shown that P satisfies the A-stability criterion. By condition b) of Definition 3.18 P is a polynomial of degree m in ζ; in particular, P is not the zero function. By condition c) of Definition 3.18 P has no removable poles in $\mathcal{L}$. Now Theorem 1.16b) can be applied to show that $\Lambda \cap \mathcal{L}$ is empty. It follows from condition a) of Definition 3.18 and Corollary 1.15 that P satisfies the A-stability criterion.

By condition c) of Definition 3.18, P has no removable poles in $\mathcal{I}$. Combining this fact with Corollary 1.8 verifies condition a) of Theorem 3.19.

Assume $\nabla_{m'}$ is the zero function. Then $\hat{D}$ of (3-15) is of positive degree, by Proposition 3.6. In such case it can be shown from conditions b) and c) of Theorem 3.7 that $\mathcal{I} \subset M$, which is false, by condition d) of Definition 3.18. Hence $\nabla_{m'}$ is not the zero function. Now Theorem 3.14 can be applied, verifying conditions b) and c) of Theorem 3.19.

Suppose condition d) of Theorem 3.19 fails to hold. Let ω be a positive real number such that $\nabla_{m'}(\omega) = 0$. By Proposition 3.13 there exists a $\zeta \in \overline{C}$ such that $P(\hat{j}\omega,\zeta) = P(\hat{j}\omega, \frac{1}{\zeta^*}) = 0$. In the notation of (3-5) this becomes $\phi(\hat{j}\omega)\psi(\zeta)\overline{P}(\hat{j}\omega,\zeta) = \phi(\hat{j}\omega)\psi(\frac{1}{\zeta^*})\overline{P}(\hat{j}\omega, \frac{1}{\zeta^*}) = 0$. But $\phi(\hat{j}\omega) \neq 0$, by condition c) of Definition 3.18. Conditions b) and c) of Theorem 2.4 can be used to show that $\zeta \in \overline{U}$ and $\frac{1}{\zeta^*} \in \overline{U}$; that is, $\zeta \in E$. Therefore $\psi(\zeta) \neq 0$, by condition b) of Theorem 2.4. It follows that $\overline{P}(\hat{j}\omega,\zeta) = 0$, which is false, by condition d) of Definition 3.18. This contradiction shows that condition d) of Theorem 3.19 holds.

To prove sufficiency, suppose the conditions of Theorem 3.19 hold. Conditions b) and c) of Definition 3.18 follow directly from conditions b) and a) of Theorem 3.19, respectively. By condition b) of Theorem 3.19 P is of degree m in ζ. Now condition d) of Definition 3.18 follows from Proposition 3.13 and condition d) of Theorem 3.19.

It remains to verify condition a) of Definition 3.18. This can be done by use of the sufficiency part of Corollary 3.15:

By condition a) of Theorem 3.19 $\Lambda \cap \mathcal{L}$ is empty. Since P is of degree m in ζ, condition a) of Corollary 3.15 holds. By condition c) of Theorem 3.19

$\nabla_{m'}$ is not the zero function, and condition b) of Corollary 3.15 holds. It still must be shown that condition c) of Corollary 3.15 holds. The next paragraph shows that an even stronger condition holds: ∇_i has no positive real zeros, for all i = 1, 2, ..., m'.

Suppose the contrary, and let ω_0 be the smallest positive real number such that $\nabla_i(\omega_0^2) = 0$ for some i = 1, 2, ..., m'. By construction of ω_0, $\nabla_i'(\omega) > 0$ for all i = 1, 2, ..., m' and all positive real $\omega < \omega_0$. For such ω all zeros of $P'(\omega,\cdot)$ lie in the open lower half plane H, by the sufficiency part of Theorem 3.8. By condition c) of Definition 3.18 $\hat{j}\omega_0$ is not a removable pole of P. Hence the zeros of $P'(\omega,\cdot)$ are continuous functions of the first argument of P' at the point ω_0. From this it follows that the zeros of $P'(\omega_0,\cdot)$ lie in $\overline{H}$. But by the necessity part of Theorem 3.8, $P'(\omega_0,\cdot)$ has a zero outside $\overline{H}$. (Note that $\nabla_{m'}'(\omega_0) \neq 0$, by condition d) of Theorem 3.19.) This contradiction proves the supposition of this paragraph to be false.

The above result implies that condition c) of Corollary 3.15 holds. Application of Corollary 3.15 now verifies condition a) of Definition 3.18. This completes the proof of sufficiency. □

Inspection of the above proof shows that condition a) of Theorem 3.19 can be replaced by the stronger condition

a') All poles of P lie in $\mathcal{R}$.

The difference between condition a) and condition a') is that P may have a pole at infinity, which is not in $\mathcal{R}$, while Λ never contains the point at infinity. See Section 1.5. In Theorem 3.19 condition a') provides a stronger result than condition a) for the necessity part, while the reverse is true for the sufficiency part.

Among the theorems for A-stability of composite multistep methods in Chapters 2 and 3, Theorem 3.19 is unique, in that it is the only one for which a condition on poles is part of the characterization rather than part of the hypothesis. This special situation is, of course, due to the incorporation of a condition on removable poles into the definition of strong A-stability (condition c) of Definition 3.18).

As the basis for a practical A-stability test, Theorem 3.19 has three significant advantages over both Corollaries 3.15 and 3.17:

1. The hypothesis "$\Lambda \cap \mathcal{L}$ empty" of Corollaries 3.15 and 3.17 requires that the test for $\Lambda \subset \overline{\mathcal{R}}$ be implemented, while condition a) of Theorem 3.19

only requires that the test for $\Lambda \subset \mathcal{R}$ be implemented. The latter test is much simpler than the former. The reason is that elaborate procedures are needed in the "$\Lambda \subset \bar{\mathcal{R}}$" test to properly account for roots of a polynomial lying on the imaginary axis [Gantmacher, vol. 2].

2. The test for positive real zeros must be performed m' times [$\hat{m}$ times] in Corollary 3.15 [Corollary 3.17], but only once in Theorem 3.19.

3. The multiplicities of positive real zeros must be determined in Corollaries 3.15 and 3.17, but not in Theorem 3.19.

A practical A-stability test has been implemented in a set of computer programs based on the strong A-stability characterization of Theorem 3.19 [Rubin, 1973]. Appendix E discusses some of the special considerations and techniques involved in the practical realization of algebraic A-stability characterizations in general, and describes briefly the main steps of the implemented algorithm. Examples of application of the algorithm to composite multistep methods are given in Chapter 4.

§3.8 The Dual Algebraic A-Stability Characterization

The A-stability characterizations developed in Sections 3.2 through 3.7 are based on the primal analytic A-stability theory, that is, the theory of Section 2.3. By starting with the dual analytic A-stability theory of Section 2.4, dual algebraic characterizations of A-stability can be derived. Much of the primal theory applies without modification to the dual case, as will be seen in this section. Even when modifications are required, the mathematical considerations are essentially the same. Therefore, the major results will be presented without proof.

The development of dual characterizations begins with the same transformed polynomial P' as in Section 3.2. However, instead of using the transformed A-stability criterion directly as in Definition 3.2, it is necessary to use a dual formulation, as follows:

3.20. Proposition: A complex polynomial P' in (ω,z) satisfies the transformed A-stability criterion if and only if for every z in the closed upper half plane $\overline{G}$, the zeros of $P'(\cdot,z)$ lie in the closed lower half plane H.

The proof of Proposition 3.20 is identical in spirit to that of Proposition 2.1.

The main result of Section 3.2, Proposition 3.4, is essentially a statement of the primal analytic characterization (Theorem 2.4) in the transformed domain. The dual result given below is a similar transformation of the dual analytic characterization (Theorem 2.7).

3.21. Proposition: In order for a complex* polynomial P' in (ω,z) to satisfy the transformed A-stability criterion, it is necessary and sufficient that

a) All zeros of $P'(\cdot,\hat{j})$ lie in the closed lower half plane $\overline{H}$,

b) All zeros of $P'(\hat{j},\cdot)$ lie in the open lower half plane H, and,

c) For every real z, all zeros of $P'(\cdot,z)$ lie in $\overline{H}$.

Just as Proposition 3.4 utilizes Definition 3.2 in its proof, Proposition 3.21 utilizes Proposition 3.20.

The factorization (3-15) of the transformed polynomial P' can be computed in the same way as in Section 3.3. However, rather than pursuing this course, it is desirable to perform the factorization in a dual manner, in order that the computations be compatible with the dual characterizations. To begin, it is easily seen from (3-4) that a polynomial P and its transformed polynomial P' are always of the same degree in their first argument. Let this degree be n, as usual. Then P' can be represented uniquely in the form

$$P'(\omega,z) = [1 \quad \hat{j}]\tilde{\tau}(z)[1 \quad \omega \quad \omega^2 \quad \ldots \quad \omega^n]^T \qquad (3\text{-}20)$$

where $\tilde{\tau}$ is a $2 \times (n+1)$ matrix whose elements are real polynomials in z or the zero function. If P' is of degree m' in z, then $\tilde{\tau}$ is evidently of degree m'. The situation is evidently dual to that in (3-9).

Next the $n \times n$ symmetric matrix $\tilde{T}$ is defined from $\tilde{\tau}$ using a relation of the form (3-10), but with m' replaced by n. The elements of $\tilde{T}$ are thus real polynomials in z of degree at most 2m', or the zero function. Finally define the nested principal minors $\tilde{\nabla}'_i$, i = 0, 1, 2, ..., n, using a relation of the form (3-11). Each $\tilde{\nabla}'_i$ is a real polynomial in z of degree at most 2m'i, or the zero function.

It is noted in passing that the above dual definitions can be used to derive a result dual to Lemma 3.5. The resulting dual lemma reads like Lemma 3.5, except for an obvious change of notation.

---

*Refer to the footnote of Proposition 3.4.

The factorization (3-15) is defined as in the primal case. The only difference is that the factorization will be called trivial in the dual sense if $\hat{D}$ is of degree zero in ω. The computation of $\hat{D}$ from the dual Bezoutian matrix $\tilde{T}$ can be performed using the dual of Proposition 3.6. This dual proposition reads like Proposition 3.6, except for an obvious change of notation.

The main result of Section 3.3, Theorem 3.7, has the following dual; the proof rests upon Proposition 3.21.

3.22. Theorem: Let P' be a complex polynomial in (ω,z), factored according to (3-15). In order for P' to satisfy the transformed A-stability criterion, it is necessary and sufficient that

a) All zeros of $P'(\cdot,\hat{j})$ lie in the closed lower half plane $\overline{H}$,

b) All zeros of $P'(\hat{j},\cdot)$ lie in the open lower half plane H,

c) For every real z, all zeros of $\hat{D}(\cdot,z)$ are real, and,

d) $\hat{P}'$ satisfies the transformed A-stability criterion.

Continuing the parallel development into Section 3.4, it is easy to state a result dual to Theorem 3.8 by making the usual notational changes. This dual result can be used to derive the following dual of the Fundamental Characterization Theorem:

3.23. Theorem: Let P' be a complex polynomial in (ω,z), of degree n in ω. Let $\tilde{\tau}$, $\tilde{T}$, and $\tilde{\nabla}'_i$, $i = 1, 2, \ldots, n$, be given by (3-20), (3-10)*, and (3-11)*, respectively. Assume $\tilde{\nabla}'_n$ is not the zero function. Then $P'(\omega,\cdot)$ is not the zero function for any real ω. Furthermore, in order for P' to satisfy the transformed A-stability criterion, it is necessary and sufficient that

a) All zeros of $P'(\cdot,\hat{j})$ lie in the closed lower half plane $\overline{H}$,

b) All zeros of $P'(\hat{j},\cdot)$ lie in the open lower half plane H,

c) $\tilde{\nabla}'_i$ is not the zero function, for all $i = 1, 2, \ldots, n-1$, and,

d) $\tilde{\nabla}'_i(z) \geq 0$ for all $i = 1, 2, \ldots, n$, and all real z.

Consider the situation in which P' is the transformed polynomial associated with a real polynomial in two variables, as in Section 3.5. Then the dual of Proposition 3.10 holds; that is, each $\tilde{\nabla}'_i$ is even. Therefore, polynomials $\tilde{\nabla}''_i$ can be defined by a relation of the form (3-18).

*Substitution of dual quantities in these equations is assumed, as discussed in this section.

The obvious dual of Proposition 3.11 does not hold. Instead, the degree of $\tilde{\nabla}_i''$ is strictly less than m'i whenever (3-1) holds. However, independently of considerations relating to (3-1), it may still happen that $\tilde{\nabla}_i''(0) = 0$. Therefore, define $\tilde{\nabla}_i$, $i = 1, 2, \ldots, n$, by a relation of the form (3-19). The polynomials $\tilde{\nabla}_i$ (any of which may be the zero function) will be called the dual auxiliary polynomials associated with the transformed polynomial P'. It is evident, for the reason given in Section 3.5, that dual auxiliary polynomials are well defined if either P' or $\hat{j}P'$ is related to a real polynomial by (3-2) and (3-3). Hence, dual auxiliary polynomials are well defined for $\hat{P}'$, by Lemma 3.12.

The following result is the dual of Proposition 3.13.

3.24. Proposition: Let P be a real polynomial in (λ,ζ), of degree n in λ, and let P' be its transformed polynomial. Assume both are of the same degree in their second arguments. Let $\tilde{\nabla}_n$ be the n-th dual auxiliary polynomial associated with P'. Let z be a fixed nonzero real number, and let $\zeta = \frac{z+\hat{j}}{z-\hat{j}}$. (Thus ζ is contained in the unit circle E.) Then $\tilde{\nabla}_n(z^2) = 0$ if and only if there exists a $\lambda \in \bar{C}$ such that $P(\lambda,\zeta) = P(-\lambda^*,\zeta) = 0$. Therefore, if the zeta locus T associated with the reduced polynomial $\bar{P}$ of P intersects E at $\pm\zeta$, then $\tilde{\nabla}_n(z^2) = 0$.

When Theorem 3.23 is applied to transformed polynomials associated with real polynomials in two variables, the dual auxiliary polynomials can be brought to bear. The resulting theorem, which is dual to Theorem 3.14, provides an algebraic characterization of the A-stability criterion for all real polynomials whose transformed polynomials have trivial factorizations in the dual sense.

3.25. Theorem: Let P be a real polynomial in (λ,ζ), of degree n in λ, let P' be its transformed polynomial, and let $\tilde{\nabla}_i$, $i = 1, 2, \ldots, n$, be the dual auxiliary polynomials associated with P'. Assume $\tilde{\nabla}_n$ is not the zero function. Then the lambda locus M associated with the reduced polynomial $\bar{P}$ of P does not contain the imaginary axis I, nor does the zeta locus T contain the unit circle E. Furthermore, in order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) All poles of P lie in the closed right half plane $\bar{\mathcal{R}}$,

b) All zeros of $P(-1,\cdot)^*$ lie in the open unit disc $\mathcal{U}$,

c) $\tilde{\nabla}_i(0) > 0$ for all $i = 1, 2, \ldots, n$, and,

d) $\tilde{\nabla}_i$ has no positive real zeros of odd multiplicity, for $i = 1, 2, \ldots, n$.

The conditions of Theorems 3.14 and 3.25 correspond in a natural way, except that condition b) of the former is much simpler than condition b) of the latter. A direct test for condition b) of Theorem 3.25 is the classical Schur-Cohn criterion [Marden].

The application of Theorem 3.25 to composite multistep methods is straightforward. The dual to Corollary 3.15 is obvious, and therefore will not be presented.

The general algebraic characterization of the A-stability criterion for real polynomials in two variables has the following dual form:

3.26. Theorem: Let P be a real polynomial in (λ,ζ), of degree n in λ, and let P' be its transformed polynomial, factored according to (3-15), where $\hat{n}$ is the degree of $\hat{P}'$ in ω. Let $\tilde{\nabla}_i$, $i = 1, 2, \ldots, \hat{n}$, be the dual auxiliary polynomials associated with $\hat{P}'$.

In order for P to satisfy the A-stability criterion, it is necessary and sufficient that

a) All poles of P lie in the closed right half plane $\bar{\mathcal{R}}$,

b) All zeros of $P(-1,\cdot)^*$ lie in the open unit disc $\mathcal{U}$,

c) For every positive real z, all zeros of $\hat{D}(\cdot,z)$ are real,

d) $\tilde{\nabla}_i(0) > 0$ for all $i = 1, 2, \ldots, \hat{n}$, and,

e) $\tilde{\nabla}_i$ has no positive real zeros of odd multiplicity, for $i = 1, 2, \ldots, \hat{n}$.

It is noteworthy that Theorem 3.16 contains six conditions, while Theorem 3.26 contains only five. In fact, condition b) of Theorem 3.26 implies both conditions b) and c) of Theorem 3.16. There may be an advantage in using condition b) of Theorem 3.26 as a sufficiency check for condition c) of Theorem 3.16.

Theorem 3.26 leads to the following characterization of A-stability for composite multistep methods, dual to Corollary 3.17:

*Refer to the footnote of Theorem 2.7.

3.27. Corollary: Let (R,k) be a composite multistep method with $\Lambda \cap \mathcal{L}$ empty, and let P be its characteristic polynomial. (Therefore P is real.) Assume the definitions of the first paragraph of Theorem 3.26. In order for (R,k) to be A-stable, it is necessary and sufficient that conditions b) through e) of Theorem 3.26 hold.

In order to present a natural dual of Theorem 3.19 on strong A-stability, it is necessary to introduce a slightly different kind of strong A-stability. Thus, a composite multistep method (R,k) is said to be strongly A-stable in the dual sense if the four conditions of Definition 3.18 are satisfied, but with condition d) replaced by the following:

d") The zeta locus T associated with the reduced polynomial $\bar{P}$ corresponding to P does not intersect the unit circle E, except possibly at ± 1.

The above condition is equivalent to the statement that the lambda locus M associated with $\bar{P}$ not intersect the imaginary axis I, except at zeros of $\bar{P}(\cdot,1)$ and $\bar{P}(\cdot,-1)$ on I.

The natural dual of Theorem 3.19 is given below:

3.28. Theorem: Let (R,k) be a composite multistep method with n future points, let Λ be its set of poles, let P of (1-52) be its characteristic polynomial, let P' be the transformed polynomial associated with P, and let $\tilde{\nabla}_i$, i = 1, 2, ..., n, be the dual auxiliary polynomials associated with P'. In order for (R,k) to be strongly A-stable in the dual sense, it is necessary and sufficient that

a) Λ is contained in the open right half plane $\mathcal{R}$,

b) All zeros of $P(-1,\cdot)^*$ lie in the open unit disc U,

c) $\tilde{\nabla}_i(0) > 0$ for all i = 1, 2, ..., n, and,

d) $\tilde{\nabla}_n$ has no positive real zeros.

As with the other dual characterizations, Theorem 3.28 may not be quite as convenient as its primal counterpart, due to condition b).

Condition a) of Theorem 3.28 can be replaced by the stronger condition

a') All poles of P lie in $\mathcal{R}$.

The same remarks apply as in Section 3.7.

*Refer to the footnote of Theorem 2.7.

§3.9 Background and Review of Previous Work

The work of Chapter 3 is best regarded as a new application of a very old mathematical subject: the qualitative analysis of polynomials (in one variable) in terms of their coefficients. One aspect of this subject, the computation of greatest common divisors, was first solved by the algorithm of Euclid (330? - 275? B.C.), probably the oldest known algorithm [Knuth]. A related aspect, elimination (reduction of simultaneous polynomial equations to one equation), was solved using Sylvester's determinants in 1841, although Bezout's determinant formulation of 1779 is also applicable [Muir]. A third aspect, localization (determining the number of zeros of a polynomial in a half plane, etc.) was solved using Hurwitz determinants in 1894. Thus, the fundamentals of qualitative analysis of polynomials in one variable using various determinant formulations were well understood by the turn of the century [Muir]. In fact, this classical subject may have been more familiar to mathematicians in 1900 than at present* [Burnside and Panton], [Bocher].

On the other hand, all these problems and the various determinantal formulations for solution (as well as many other problems and formulations) are intimately related, a fact which seems not to have been widely appreciated. Thus [Householder, 1970] was able to suggest the use of Bezout's formulation for the greatest common divisor problem, and for localization. Thus, the essence of Theorem 3.8 appears in [Householder, 1970]. However, it is not certain whether the explicit formula of Proposition 3.6 (for polynomials in one variable) has appeared in the literature.

Modern writers [Gantmacher, vol. 2], [Siljak, 1969], [Householder, 1970] have recognized the value of applying the localization results to polynomials whose coefficients are functions of parameters. An interesting treatment of stability for linear systems with free parameters is given in [Siljak, 1969, pp. 365-367]; included are a set of polynomial inequalities.

A finitary algebraic A-stability test, applicable to only a very special class of problems, is implicit in the work of [Watts]. One property of each method of this class is that the zeta locus is precisely the unit circle E.

---

*However, in the qualitative analysis of polynomials there have been important modern developments, among which are the new matrix formulation of [Barnett], and the generalization of localization results to complex polynomials, as reported in [Gantmacher, vol. 2] and [Householder]. See also [Marden].

Thus, Watts' test uses none of the main ideas of Chapter 3, but only the check for poles in the left half plane. His result is easily derived from Corollary 2.5; it is also a trivial special case of Corollary 3.17, as shown in Section 3.6.

A criterion for A-stability of multistep methods was described in [Liniger]. Liniger reduces the two-dimensional A-stability definition to a one-dimensional property in the spirit of Chapter 2, essentially deriving Corollary 2.9 (for the special case of multistep methods). Condition a) of Corollary 2.9 is then transformed, in the same way as in (3-2), and the Routh criterion is applied. (The term "transformed polynomial" used in Chapter 3 is derived from Liniger's work, although in reference to a different transformation and a different polynomial.) Condition b) of Corollary 2.9 is also transformed, by means of Chebyshev polynomials, into a polynomial inequality, but [Liniger] does not mention the idea of using Sturm's theorem to reduce the problem to finitary inequalities. A result comparable with Liniger's is easily derived as a special case of Corollary 3.27.

Chapter 4

COMPOSITE MULTISTEP METHODS: EQUIVALENCE, ERROR, AND ORDER

§4.1 Intrinsic and Extrinsic Properties

In order for a composite multistep method (R,k) to be useful in the solution of stiff ordinary differential equations, it is not sufficient that (R,k) be A-stable. (A-stability is not necessary either [Gear], although it is generally acknowledged to be desirable.) This chapter explores properties, other than A-stability, considered to be desirable in a useful composite multistep method. While A-stability is the important property for stiff problems, the additional properties discussed in this chapter are important even for non-stiff problems. A-stability is one example of what will be defined in the following paragraph as an intrinsic property of a composite multistep method.

In order to discuss the nature of properties in general, it is useful to adopt momentarily a more abstract point of view. A method for the numerical solution of ordinary differential equations (not necessarily a composite multistep method) can be thought of as a particular realization of an abstract operator which maps each ordered pair--(initial value problem, step size)--into an approximating sequence*. It follows that two methods are weakly equivalent (as defined in Section 1.2) if and only if they are realizations of the same abstract operator. Any given property of an abstract operator can be thought of as being induced into each of its realizations. Such induced properties will be called intrinsic properties of the realization. In other words, a property of a method is intrinsic if and only if it is shared by all methods weakly equivalent to it. Alternatively, weakly equivalent methods have the same intrinsic properties. Properties of realization which are not intrinsic will be called extrinsic. Extrinsic properties depend not only upon the approximating sequences computed, but upon <u>how</u> they are computed.

---

*To be technically correct, it must be stated that each value of the abstract operator is a <u>set</u> of approximating sequences. This set is normally a singleton. However, if existence fails, then the set is empty, while if uniqueness fails, the set contains more than one sequence.

A-stability is an intrinsic property because it is defined in terms of the behavior of approximating sequences for certain initial value problems. Let S be a finite set of complex numbers, and consider the property that a composite multistep method have S as its set of unremovable poles. This property is intrinsic with respect to composite multistep methods, because an unremovable pole can be characterized in terms of the behavior of approximating sequences. On the other hand, the property that a composite multistep method have Λ as its set of poles is an extrinsic property. For example, consider the composite multistep method (1-13) and the implicit Euler method, which is weakly equivalent to it. Both methods have an unremovable pole at unity, but only (1-13) has -1 as a removable pole.

Instead of considering the universe of all possible methods for computing approximating sequences, it will be found useful in this chapter to restrict the discussion to sub-classes of the class of composite multistep methods, and to speak of properties intrinsic with respect to a given subclass. For example, given fixed positive integers m and n, the property that a composite multistep method have a given polynomial P as its characteristic polynomial is an intrinsic property with respect to the class of composite multistep methods having m past points and n future and retained points. This fact follows from the results of Section 4.2.

§4.2 Equivalence, Canonical Form, and Strong Regularity

Two composite multistep methods (R_1,k_1) and (R_2,k_2) will be called equivalent if they have the same number of past points m, future points n, and retained points $k = k_1 = k_2$, and if in addition R_1 and R_2 are row equivalent*. It is evident that equivalence of composite multistep methods is an equivalence relation.

4.1. Proposition: Equivalent composite multistep methods are weakly equivalent.

Proof: It is to be proved that equivalent methods generate for every function f, step size h, and starting condition the same approximating sequence.

---

*A good treatment of row equivalence and many of its characterizations is given in [Finkbeiner, Section 6.3]. Two matrices of the same dimensions are defined to be row equivalent if one can be obtained from the other by a finite sequence of elementary row operations [Finkbeiner, p. 122].

If the corresponding nonlinear algebraic equations (1-4) have the same solutions at the first iteration $\ell = 0$, then this condition holds.

In order to write (1-4) in matrix notation, let F denote the function corresponding to the application of f at m + n consecutive steps; more precisely, if $Y = [y_0 \;\; y_1 \;\; \cdots \;\; y_{m+n-1}]^T$, where each y_i is a column vector having the same number of components as X(t) in (1-1), then let

$$F(Y,h) = [f(y_0,0) \;\; f(y_1,h) \;\; \cdots \;\; f(y_{m+n-1},(m+n-1)h)]^T \qquad (4\text{-}1)$$

for each such matrix Y and each real $h \geq 0$. Using this notation (1-4) can be written for $\ell = 0$ as

$$AY - hBF(Y,h) = 0 \qquad (4\text{-}2)$$

where A and B are given by (1-46). Alternatively, (4-2) can be written as

$$R\begin{bmatrix} Y \\ -hF(Y,h) \end{bmatrix} = 0 \quad . \qquad (4\text{-}3)$$

Let (R_1,k) and (R_2,k) be two equivalent composite multistep methods. Since R_1 and R_2 are row equivalent, their associated linear homogeneous algebraic equations have the same solutions [Finkbeiner, p. 128]. It follows in particular that for every F, h, and Y,

$$R_1\begin{bmatrix} Y \\ -hF(Y,h) \end{bmatrix} = 0 \qquad (a)$$

(4-4)

if and only if

$$R_2\begin{bmatrix} Y \\ -hF(Y,h) \end{bmatrix} = 0 \; . \qquad (b)$$

Therefore, (R_1,k) and (R_2,k) are weakly equivalent. □

Proposition 4.1 shows that the equivalence classes of composite multistep methods are sub-classes of the weak equivalence classes. Therefore,

composite multistep methods which are equivalent have the same intrinsic properties.

The converse of Proposition 4.1 does not hold in general, even for composite multistep methods having the same number of past, future, and retained points. For example, the composite multistep method

$$\left(\begin{bmatrix} -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, 1\right) \tag{4-5}$$

and (1-13) are weakly equivalent. (In fact, each is weakly equivalent to the implicit Euler method.) But the respective composite matrices of (1-13) and (4-5) are not row equivalent, since their ranks differ [Finkbeiner, Theorem 6.4]. Hence (1-13) and (4-5) are not equivalent.

For the important special case in which all future points are retained, the converse of Proposition 4.1 holds (for methods with the same number of past and future points, respectively). The situation for k = n can be summarized as follows:

4.2. Proposition: Let (R_1,n) and (R_2,n) be two composite multistep methods with the same number of past points m and future points n, and with all n future points retained. Then (R_1,n) and (R_2,n) are equivalent if and only if they are weakly equivalent.

Proof: The "only if" part is just a special case of Proposition 4.1.

To prove the "if" part, suppose (R_1,n) and (R_2,n) are not equivalent. Then R_1 and R_2 are not row equivalent. In such case the linear homogeneous algebraic equations associated with R_1 and R_2 do not have the same solutions [Finkbeiner, p. 128]. For definiteness, let X and Y be (m+n)-vectors such that

$$R_1\begin{bmatrix} Y \\ X \end{bmatrix} = 0$$

but

$$R_2\begin{bmatrix} Y \\ X \end{bmatrix} \neq 0 .$$

Let f be a continuous, real-valued function satisfying $f(y,i) = -x_i$ for all

real y and all i = 0, 1, ..., m+n-1, where x_i is the i-th component of X. Then X = -F(Y,1). With h = 1, (4-4a) holds, but (4-4b) does not. That is, Y_f satisfies the algebraic equation associated with R_1, but not that associated with R_2, where

$$Y = \begin{bmatrix} Y_p \\ \\ Y_f \end{bmatrix} . \qquad (4\text{-}6)$$

(The partition is after the m-th row of Y.) Since all elements of Y_f are retained, (R_1,n) generates an approximating sequence not generated by (R_2,n). Therefore, (R_1,n) and (R_2,n) are not weakly equivalent. □

It will be noted that Propositions 4.1 and 4.2 form a kind of parallel development to Propositions 1.3 and 1.4. According to these four results, the theory for k = n is somewhat more elegant than the general theory.

4.3. Proposition: Two composite multistep methods (R_1,k) and (R_2,k) with the same number of past points m, future points n, and retained points k, respectively, are equivalent if and only if there exists an n × n nonsingular matrix S such that $SR_1 = R_2$.

Proof: R_1 and R_2 are row equivalent if and only if there exists a nonsingular matrix S such that $SR_1 = R_2$ [Finkbeiner, Theorem 6.9]. Thus, Proposition 4.3 follows directly from the definition of equivalence of composite multistep methods. □

A composite matrix R, partitioned according to (1-8), will be said to be canonical, or in canonical form, if the matrix $[A_f \;\; B_p \;\; B_f \;\; A_p]$ is in reduced echelon form [Finkbeiner, pp. 123-124]. A composite multistep method will be said to be canonical, or in canonical form, if its composite matrix is canonical.

4.4. Proposition: For every composite multistep method, there exists a unique canonical composite multistep method to which it is equivalent.

Proof: Reduced echelon form is canonical with respect to row equivalence, in the sense that for every matrix there exists a unique matrix in reduced echelon form to which it is row equivalent [Finkbeiner, p. 128]. The proposition is established by applying this fact to $[A_f \;\; B_p \;\; B_f \;\; A_p]$, which differs from R of (1-8) by a column rotation. □

An alternate way of stating Proposition 4.4 is as follows: There is exactly one canonical composite multistep method in every equivalence class of composite multistep methods. The importance of equivalence and canonical form is in the fact that they allow the study of intrinsic properties to be separated from the study of extrinsic properties. Most of the important properties to be discussed in this chapter are intrinsic properties. In order to optimize a composite multistep method with respect to any collection of intrinsic properties, it is enough to consider just the class of canonical methods. Over the set of methods equivalent to a given canonical method (R,k), optimization with respect to a collection of extrinsic properties can then be performed by considering all methods of the form (SR,k), where S is a nonsingular matrix.

A composite matrix R will be said to be strongly regular if its submatrix A_f of (1-8) is nonsingular. Strongly regular composite matrices are regular, according to Proposition 4.6 (see next section). A composite multistep method (R,k) will be said to be strongly regular if R is strongly regular. Strong regularity will be a particularly important property in the remainder of Chapter 4. The following proposition relates strong regularity with canonical form.

4.5. Proposition: Let (R,k) be a strongly regular composite multistep method. Then the canonical composite multistep method equivalent to (R,k) is $(A_f^{-1}R,k)$, where A_f is given by (1-8). Furthermore, (R,k) is in canonical form if and only if A_f is the identity matrix.

Proof: The two composite multistep methods (R,k) and $(A_f^{-1}R,k)$ are equivalent by Proposition 4.3. Furthermore,

$$A_f R = [A_f^{-1}A_p \quad I \quad A_f^{-1}B_p \quad A_f^{-1}B_f]$$

is in canonical form, since

$$[I \quad A_f^{-1}B_p \quad A_f^{-1}B_f \quad A_f^{-1}A_p]$$

is in reduced echelon form [Finkbeiner, Section 6.3]. Hence $(A_f^{-1}R,k)$ is in canonical form. This proves the first conclusion of Proposition 4.5.

The second conclusion follows from the first and Proposition 4.4. □

§4.3 Relations with the Characteristic Polynomial

The characteristic polynomial is related to strong regularity as follows:

4.6. Proposition: Let (R,k) be a composite multistep method, and let P be its characteristic polynomial. Then R is strongly regular if and only if R is regular and the origin is not a pole of P.

Proof: By definition R is strongly regular if and only if A_f of (1-8) is nonsingular, that is, if and only if $\det A_f \neq 0$. By (1-10) and (1-11) $\det A_f = \delta(0)$. Therefore, R is strongly regular if and only if $0 \notin \Lambda$.

If R is singular, then $\Lambda = C$, and hence $0 \in \Lambda$. Therefore regularity of R is necessary for strong regularity of R.

When R is regular, Λ is precisely the set of finite poles of P. In such case R is strongly regular if and only if the origin is not a pole of P. □

It follows from Proposition 4.6 and Theorems 3.19 and 3.28 that only strongly regular composite multistep methods can be strongly A-stable, or strongly A-stable in the dual sense. Also, a composite multistep method with $\Lambda \cap \mathcal{L}$ empty which fails to be strongly regular cannot be A-stable if its pole at the origin is unremovable. This statement follows from Corollaries 1.8 and 1.15. Even if its pole at the origin is removable, such a method is impractical for small step size. (See Section 1.2.) Thus, composite multistep methods useful for stiff problems should not be just regular, but strongly regular*.

It is easy to show that weakly equivalent composite multistep methods can have different characteristic polynomials, even among methods with the same numbers of past, future, and retained points. Consider the composite matrix

$$R = \begin{bmatrix} -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

*In all publications to date on composite multistep and composite one-step methods (see Introduction), strong regularity is viewed so important as to be assumed at the very start. The present work is the first to even consider methods which are not strongly regular.

created as an application of the implicit Euler method twice cyclically. The following five composite multistep methods are evidently weakly equivalent to each other, but their corresponding characteristic polynomials are all different:

1.) (R,1), which has $(1-\lambda)[(1-\lambda)\zeta-1]$ as its characteristic polynomial,
2.) (R,2), which has $(1-\lambda)^2\zeta-1$ as its characteristic polynomial,
3.) The method (1-13), which has (1-45) as its characteristic polynomial,
4.) The method (4-5), which has the zero function as its characteristic polynomial,
5.) The implicit Euler method, which has (1-44) as its characteristic polynomial.

Note that cases 1, 3, and 4 all have $m = k = 1$, and $h = 2$.

On the other hand, equivalent composite multistep methods have the same characteristic polynomial (to within a trivial factor). A simple and direct algebraic proof of this fact can be given from the following useful technical lemma:

4.7. Lemma: Let (R,k) be a composite multistep method, and let Q be defined by (1-46) through (1-51). Then

$$Q(\lambda,\zeta) = RL(\lambda)\hat{Z}(\zeta) \tag{4-7}$$

where

$$L(\lambda) = \begin{bmatrix} I_{m+n} \\ -\lambda I_{m+n} \end{bmatrix} \tag{4-8}$$

and

$$\hat{Z}(\zeta) = \begin{bmatrix} I_k & \zeta I_k & \zeta^2 I_k & \cdots & \zeta^{M-1} I_k & \zeta^M J^T_{(k-N)k} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & I_{n-k} \end{bmatrix}^T . \tag{4-9}$$

Proof: Let

$$\tilde{Z}(\zeta) = [I_k \quad \zeta I_k \quad \zeta^2 I_k \cdots \zeta^{M-1} I_k \quad \zeta^M J^T_{(k-N)k}]^T . \tag{4-10}$$

It follows from the identity

$$\left[\begin{array}{c} J_{(k-N)k} \\ \hline 0_{N(k-N)} \quad I_N \end{array}\right] = I_k$$

that

$$\left[\begin{array}{c|c} \tilde{Z}(\zeta) & \\ \hline O_{N(k-N)} & \zeta^M I_N \end{array}\right] = Z(\zeta) ,$$

where Z is given by (B-1). Therefore,

$$V(\lambda)\tilde{Z}(\zeta) = [V(\lambda) \quad O_{nN}]\left[\begin{array}{c|c} \tilde{Z}(\zeta) & \\ \hline O_{N(k-N)} & \zeta^M I_N \end{array}\right]$$

$$= [V(\lambda) \quad O_{nN}]Z(\zeta)$$

$$= W(\lambda,\zeta) , \qquad (4\text{-}11)$$

the last equality being (B-2). Now

$$RL(\lambda)\hat{Z}(\zeta) = [A \quad B]\left[\begin{array}{c} I_{m+n} \\ \hline -\lambda I_{m+n} \end{array}\right]\hat{Z}(\zeta) \qquad \text{by (1-46) and (4-8)}$$

$$= (A-\lambda B)\hat{Z}(\zeta)$$

$$= \hat{V}(\lambda)\hat{Z}(\zeta) \qquad \text{by (1-47)}$$

$$= [V(\lambda) \quad K(\lambda)]\hat{Z}(\zeta) \qquad \text{by (1-48)}$$

$$= [V(\lambda) \quad K(\lambda)]\left[\begin{array}{cc} \tilde{Z}(\zeta) & 0 \\ 0 & I_{n-k} \end{array}\right] \qquad \text{by (4-9) and (4-10)}$$

$$= [V(\lambda)\tilde{Z}(\zeta) \quad K(\lambda)]$$

$$= [W(\lambda,\zeta) \quad K(\lambda)] \qquad \text{by (4-11)}$$

$$= Q(\lambda,\zeta) ,$$

the last equality being (1-51). □

4.8. Proposition: Let (R_1,k) and (R_2,k) be equivalent composite multi-step methods, and let P_1 and P_2 be their respective characteristic polynomials, given by (1-46) through (1-52). Then P_1 and P_2 are equal to within a trivial factor.

Proof: By Proposition 4.3 there exists an n × n nonsingular matrix S such that $SR_1 = R_2$. Thus det S is a trivial factor, since det S ≠ 0. A double application of (1-52) and Lemma 4.7 shows that

$$\begin{aligned} P_2(\lambda,\zeta) &= \det R_2 L(\lambda)\hat{Z}(\zeta) \\ &= \det SR_1 L(\lambda)\hat{Z}(\zeta) \\ &= (\det S)\det R_1 L(\lambda)\hat{Z}(\zeta) \\ &= P_1(\lambda,\zeta)\det S \end{aligned}$$

for all $(\lambda,\zeta) \in C^2$. □

If P_1 and P_2 are equal, it does not follow in general that (R_1,k) and (R_2,k) are equivalent, or even weakly equivalent. For example, the parameterized class of methods

$$\left(\begin{bmatrix} \alpha_{10} & 1 & 0 & \beta_{10} & -1 & \beta_{12} \\ -1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}, 2 \right) \qquad (4\text{-}12)$$

has (1-45) for its characteristic polynomial independently of the values of α_{10}, β_{10}, and β_{12}. By Proposition 4.6, (4-12) is in canonical form for all α_{10}, β_{10}, and β_{12}. Therefore, no two distinct methods of the form (4-12) are equivalent, by Proposition 4.4. Thus, by Proposition 4.2, no two such methods are weakly equivalent.

§4.4 Truncation Error

No practical method for the numerical solution of ordinary differential equations can, of course, compute the exact values of the true solution at the discretization points. This fact is embodied in the very term "approximating sequence". The inaccuracies (or errors) in the approximating sequences are of two kinds: The first kind consists of roundoff errors, caused by finite precision representation of numbers, finite precision arithmetic, and even finite precision evaluation of functions, including solution of algebraic equations to finite precision. The second kind is called discretization error, and is caused by the fact that, for a practical method, the function f upon which the true solution depends can be evaluated at only a finite number of points in a given interval.

Roundoff error is regarded as an extrinsic property of a method for the numerical solution of ordinary differential equations, since it depends upon the actual implementation of the method, and not just upon the method itself. For example, the roundoff error properties of two equivalent composite multistep methods may differ, even in the "same" implementation. The analysis of roundoff error for composite multistep methods is beyond the scope of this work. Some consideration of this problem is given in [Sloate]. In principle, roundoff error can be decreased simply by increasing the precision of number representations and arithmetic; however, this remedy is usually impractical. In any case, no practical method can avoid the introduction of roundoff error.

Discretization error is present in any practical method for solution of ordinary differential equations (including any composite multistep methods), since the values of f on a finite set do not uniquely determine the true solution. According to the definition given in the next section, discretization error is an intrinsic property. This section and the next present an introduction to the analysis of discretization error for composite multistep methods.

In the following, only local errors will be considered. The term "local error" refers to the error committed in one iteration of a method, assuming error-free data to have been given at the start of the iteration. Therefore, the analysis of local error can be made in terms of the zero-th iteration with exact starting values. For any composite matrix R, function f, starting value Y_p, and step size h, the algebraic equation to be solved for Y_f at the zero-th iteration is (4-3), where Y is given by (4-6). By a straightforward generalization of standard techniques [Henrici], it can be shown that if R is strongly regular and f is smooth, then there exists a unique solution Y_f to (4-3) for any Y_p and any sufficiently small $h > 0$. The essential ideas for composite multistep methods are discussed in [Watts, Section 2.2], and are also touched upon in Appendix F.

An indirect, but conventional, way of measuring the accuracy of a composite matrix is to determine how closely the true solution satisfies the algebraic equation (4-3). The residual obtained when the true solution is substituted into (4-3) is called the truncation error (see formal definition below). This concept, which is a natural generalization of the truncation error for multistep methods [Henrici], turns out to be related to more direct

measures of accuracy, as shown in Sections 4.5 and 4.7. Let Φ denote the matrix of true solution points to which Y is an approximation; that is, let

$$\Phi(h) = [X(t_0) \quad X(t_0+h) \quad \dots \quad X(t_0+(m+n-1)h)]^T \tag{4-13}$$

where X is the true solution to (1-1) with initial condition (1-2). Thus, the ij-th element of $\Phi(h)$ is the j-th component of the true solution X evaluated at $t_0 + ih$.

4.9. Definition: The truncation error ϵ_T associated with a given composite matrix R, differential equation (1-1), and initial condition (1-2) is defined by

$$\epsilon_T(h) = R \begin{bmatrix} \Phi(h) \\ -hF[\Phi(h),h] \end{bmatrix} \tag{4-14}$$

for all $h > 0$, where F is given by (4-1), and Φ is given by (4-13).

It can be seen from (4-14) that truncation error is an extrinsic property of composite multistep methods. Thus, let R_1 and R_2 be row equivalent composite matrices, and let S be a nonsingular matrix such that $R_2 = SR_1$. By (4-14) the truncation error associated with R_2 is precisely that of R_1 premultiplied by S. Therefore, equivalent composite multistep methods do not have the same truncation error in general.

According to (4-14) the truncation error depends explicitly upon both the function f of the differential equation, and the true solution X of that equation. Furthermore, the dependence of ϵ_T upon X is apparently nonlinear in general. However, f and X are intimately related by (1-1). This fact can be used to express ϵ_T as a linear function of X and R alone, as shown below.

Let $\mathcal{D}$ be the formal differential operator, so that $\dot{X} = \mathcal{D}X$. Then (1-1) can be written formally as

$$\mathcal{D}X = f \circ X \tag{4-15}$$

where $\circ$ denotes composition of functions (with respect to the first argument of f). A power series expansion shows that $e^{h\mathcal{D}}$ is the formal anticipation operator, in the sense that

$$(e^{hD}x)(t) = x(t+h) \tag{4-16}$$

for all (smooth) vector-valued functions x and all real h and t. Both D and e^{hD} are linear operators, and can be manipulated formally in matrix relations as if they were scalars. In particular, D and e^{hD} commute with each other and with scalars.

4.10. Proposition: Let ϵ_T be the truncation error associated with a given $n \times 2(m+n)$ composite matrix R, differential equation (1-1), and initial condition (1-2). Let U denote the $(m+n) \times 1$ matrix polynomial

$$U(\zeta) = [1 \quad \zeta \quad \zeta^2 \quad \cdots \quad \zeta^{m+n-1}]^T \tag{4-17}$$

and let J denote the $2(m+n) \times 1$ matrix of functions

$$J(\lambda) = L(\lambda)U(e^{\lambda}) , \tag{4-18}$$

where L is given by (4-8). Then

$$\epsilon_T(h) = R[J(hD)x^T](t_0) \tag{4-19}$$

for all $h > 0$, where x is the solution to (1-1) and (1-2).

Proof: Using (4-16) and (4-17), (4-13) can be written as

$$\Phi(h) = [U(e^{hD})x^T](t_0) .$$

Then (4-1) and (4-13) can be used to write

$$\begin{aligned} F[\Phi(h),h] &= [U(e^{hD})(f \circ x)^T](t_0) \\ &= [U(e^{hD})(Dx)^T](t_0) \\ &= [DU(e^{hD})x^T](t_0) \end{aligned}$$

the second equality following from (4-15). Substituting the above into (4-14) and applying (4-8) gives

$$\epsilon_T(h) = R\begin{bmatrix} [U(e^{hD})x^T](t_0) \\ -h[DU(e^{hD})x^T](t_0) \end{bmatrix}$$

$$= R[L(hD)U(e^{hD})x^T](t_0)$$

$$= R[J(hD)x^T](t_0) ,$$

the last equality following from (4-18). □

§4.5 Discretization Error

The truncation error associated with a composite matrix gives only an indirect measure of the error of interest. A more direct measure, applicable in principle to any discrete method, is the actual absolute error in the future points. This error, which will be called the discretization error (see formal definition below), can be examined by use of the algebraic equation (4-3). Partitioning (4-3) according to (1-8) and (4-6) yields

$$[A_p \;\; A_f \;\; B_p \;\; B_f][Y_p \;\; Y_f \;\; -hF_p(Y_p,h) \;\; -hF_f(Y_f,h)]^T = 0 ,$$

or

$$A_fY_f - hB_fF_f(Y_f,h) = -A_pY_p + hB_pF_p(Y_p,h) , \tag{4-20}$$

where

$$F_p(Y_p,h) = [f(y_0,0) \;\; f(y_1,h) \;\; \dots \;\; f(y_{m-1},(m-1)h)]^T \tag{a}$$

(4-21)

and

$$F_f(Y_f,h) = [f(y_m,mh) \;\; f(y_{m+1},(m+1)h) \;\; \dots \;\; f(y_{m+n-1},(m+n-1)h)]^T . \tag{b}$$

Define the functions C_h and D_h for every $h > 0$ as follows:

$$C_h(Y_p) = -A_pY_p + hB_pF_p(Y_p,h) \tag{a}$$

(4-22)

$$D_h(Y_f) = A_fY_f - hB_fF_f(Y_f,h) . \tag{b}$$

If A_f is nonsingular, then D_h is invertible for sufficiently small $h > 0$. In such case (4-20) and (4-22) imply

$$Y_f = D_h^{-1}[C_h(Y_p)] \quad .$$

On the other hand, the true solution corresponding to Y_f is the n-rowed matrix Φ_f, where

$$\Phi = \begin{bmatrix} \Phi_p \\ \Phi_f \end{bmatrix} \tag{4-23}$$

and Φ is given by (4-13).

4.11. Definition: Let R be a strongly regular composite matrix, and let f, x_0, and t_0 determine a differential equation (1-1) with initial condition (1-2). The discretization error ϵ_D associated with R, f, x_0, and t_0 is defined by

$$\epsilon_D(h) = \Phi_f(h) - D_h^{-1}\{C_h[\Phi_p(h)]\} \tag{4-24}$$

for all sufficiently small $h > 0$, where C_h, D_h, Φ_p, and Φ_f are given by (4-21) through (4-23).

Each row of ϵ_D is evidently the absolute error associated with the corresponding future point. Thus, the i-th row of ϵ_D will be called the discretization error associated with the i-th future point, for $i = 1, 2, \ldots, n$. Discretization error is intrinsic with respect to a given equivalence class of composite multistep methods, since equivalent methods have the same algebraic equation (4-3).

The following preliminary result gives an exact relation between the discretization error ϵ_D and the truncation error ϵ_T.

4.12. Lemma: Let ϵ_D and ϵ_T be the discretization error and truncation error, respectively, associated with a given strongly regular composite matrix R, differential equation (1-1), and initial condition (1-2). Then

$$\epsilon_D(h) = D_h^{-1}\{C_h[\Phi_p(h)] + \epsilon_T(h)\} - D_h^{-1}\{C_h[\Phi_p(h)]\} \quad . \tag{4-25}$$

Proof: Beginning with (4-22)

$$\begin{bmatrix} \Phi_p(h) \\ \Phi_f(h) \\ -hF_p[\Phi_p(h),h] \\ -hF_f[\Phi_f(h),h] \end{bmatrix}$$

$$D_h[\Phi_f(h)] - C_h[\Phi_p(h)] = [A_p \quad A_f \quad B_p \quad B_f]$$

$$= R\begin{bmatrix} \Phi(h) \\ -hF[\Phi(h),h] \end{bmatrix}$$

$$= \epsilon_T(h) \ .$$

The second equality follows from (1-8), (4-21), and (4-23); the last equality is (4-14). The above relation shows that

$$\Phi_f(h) = D_h^{-1}\{C_h[\Phi_p(h)] + \epsilon_T(h)\} \ .$$

Combining this last relation with (4-24) yields (4-25). □

The truncation and discretization errors are of greatest interest for small step size, as will be seen in Section 4.7. A useful approximation to ϵ_D in terms of ϵ_T, valid for small h, can be given for smooth differential equations. The differential equation (1-1) determined by the function f will be said to be smooth if

a) f is continuous*,

b) f is differentiable with respect to its first argument, and,

c) f' is** continuous.

The following approximation theorem is proved in Appendix F.

---

*As usual, continuity of a function of two variables is defined to mean continuity in both arguments jointly.

**In this section (and Appendix F) the prime symbol denotes the partial derivative with respect to the first argument.

4.13. Theorem: Let ϵ_D and ϵ_T be the discretization error and truncation error, respectively, associated with a given strongly regular composite matrix R, smooth differential equation (1-1), and initial condition (1-2). Assume $\epsilon_T(0) = 0$. Then the approximation

$$\epsilon_D(h) \approx [\hat{\Delta}(h)]^{-1}\epsilon_T(h) \tag{4-26}$$

holds in the sense that

$$\lim_{h \to 0} \frac{\|\epsilon_D(h)-[\hat{\Delta}(h)]^{-1}\epsilon_T(h)\|}{\|\epsilon_T(h)\|} = 0 \tag{4-27}$$

where $\hat{\Delta}$ is given by

$$\hat{\Delta}(h) = A_f - hB_fF_f'[D_h^{-1}[C_h[\Phi_p(h)]],h] \ . \tag{4-28}$$

According to (F-13), $[\hat{\Delta}(h)]^{-1}$ is the derivative of D_h^{-1} evaluated at the point $C_h[\Phi_p(h)]$. Thus, the right side of (4-26) is a natural "first-order" approximation to the right side of (4-25).

§4.6 Error in the Linear Autonomous Case

In this section the general relations for truncation and discretization error are specialized to the linear autonomous (one-dimensional) case, an application of particular interest with respect to A-stability. In this case the function f of the differential equation is given by (1-5), with q a complex constant. Since the behavior in the linear autonomous problem is most naturally described in terms of $\lambda = qh$, rather than h itself, the notation for errors is modified accordingly. For example, the truncation error is denoted by $\tilde{\epsilon}_T$ as a function of λ, so that $\epsilon_T(h) = \tilde{\epsilon}_T(qh)$ for linear autonomous problems.

4.14. Proposition: The truncation error $\tilde{\epsilon}_T$ associated with a given composite matrix R, linear autonomous one-dimensional differential equation (1-1) with (1-5), and initial condition (1-2) is given by

$$\tilde{\epsilon}_T(\lambda) = RJ(\lambda)\chi_0 \ . \tag{4-29}$$

Proof: When f is given by (1-5), the solution to (1-1) and (1-2) is, of course,

$$x(t) = e^{q(t-t_0)} x_0 \quad .$$

It follows that for $i = 0, 1, 2, \ldots$

$$[(e^{h\mathcal{D}})^i x](t_0) = (e^{ih\mathcal{D}} x)(t_0) = x(t_0+ih) = e^{q(t_0+ih-t_0)} x_0 = (e^{\lambda})^i x_0 \quad ,$$

and hence that

$$[U(e^{h\mathcal{D}})x](t_0) = U(e^{\lambda})x_0 \quad . \tag{4-30a}$$

Also $(h\mathcal{D}x)(t_0) = (hqx)(t_0) = (\lambda x)(t_0)$, so that

$$[-h\mathcal{D}\, U(e^{h\mathcal{D}})x](t_0) = -\lambda U(e^{\lambda})x_0 \quad . \tag{4-30b}$$

Combining (4-8), (4-18), and (4-19) gives

$$\tilde{\epsilon}_T(\lambda) = \epsilon_T(h) = R\left(\begin{bmatrix} U(e^{h\mathcal{D}}) \\ \\ -h\mathcal{D}U(e^{h\mathcal{D}}) \end{bmatrix} x\right)(t_0)$$

$$= R \begin{bmatrix} U(e^{\lambda})x_0 \\ \\ -\lambda U(e^{\lambda})x_0 \end{bmatrix}$$

$$= RJ(\lambda)x_0 \quad .$$

The third equality follows from (4-30), and the last from (4-8) and (4-18). □

4.15. Proposition: Let $\tilde{\epsilon}_D$ and $\tilde{\epsilon}_T$ be the discretization error and truncation error, respectively, associated with a given strongly regular composite matrix R, linear autonomous one-dimensional differential equation (1-1) with (1-5), and initial condition (1-2). Then

$$\tilde{\epsilon}_D(\lambda) = [D(\lambda)]^{-1}\tilde{\epsilon}_T(\lambda) \tag{4-31}$$

for $\lambda \notin \Lambda$, where D is given by (1-8) and (1-10).

Proof: If (1-5) holds, then (4-21b) becomes

$$F_f(Y_f,h) = qY_f \quad . \tag{4-32}$$

Therefore, (4-22b) becomes

$$D_h(Y_f) = A_fY_f - hB_fqY_f = (A_f-\lambda B_f)Y_f = D(\lambda)Y_f \quad .$$

That is, D_h is linear, and is equal to the linear transformation $D(\lambda)$. Its inverse exists for $\lambda \notin \Lambda$ and is also linear. Thus, (4-31) is immediate from (4-25). □

Proposition 4.15 is valid even when the hypothesis "strongly" is deleted; however, the proof is longer. If the hypothesis "regular" is also deleted, Proposition 4.15 remains valid. However, for singular composite matrices, the conclusion is vacuous.

It is noteworthy that the approximation (4-26) to the local discretization error in the nonlinear case turns out to be exact for the linear autonomous one-dimensional case. To see this, observe that if (4-32) holds, then

$$F_f'(Y_f,h) = qI_n \quad ,$$

so that $\hat{\Delta}(h) = A_f - hB_fqI_n = A_f - \lambda B_f = D(\lambda)$. Substituting this relation into (4-31) shows that (4-26) holds exactly.

§4.7 Order and Error Constant

The degree of accuracy of a method for solving differential equations is conventionally described in terms of the concept of order [Henrici]. As used here, the concept of order refers to the behavior of an analytic function in a neighborhood of the origin. A matrix-valued* function Γ of a complex variable λ, analytic at the origin, is said to be of order p, where p is a nonnegative integer or -1, if

$$\lim_{\lambda \to 0} \frac{\Gamma(\lambda)}{\lambda^p} = 0 \quad . \tag{4-33}$$

*The intent here is to include, essentially as special cases, vector-valued functions and scalar-valued functions.

Evidently, every analytic function is of order -1. Also, if Γ is of order p, then Γ is of order p-1, p-2, ..., -1. The analytic function Γ is of order p for all p if and only if Γ is the zero function. An analytic function is said to be of exact order p if it is of order p but not of order p+1.

A function Γ, analytic at the origin, can be represented in a power series

$$\Gamma(\lambda) = \sum_{i=0}^{\infty} \Gamma_i \lambda^i$$

convergent in some neighborhood of the origin. It is evident that Γ is of order p if and only if $\Gamma_0 = \Gamma_1 = \ldots = \Gamma_p = 0$. That is, Γ is of order p if and only if it can be represented (in a neighborhood of the origin) in the form

$$\Gamma(\lambda) = \sum_{i=p+1}^{\infty} \Gamma_i \lambda^i = \lambda^{p+1} \sum_{i=0}^{\infty} \Gamma_{p+1+i} \lambda^i . \qquad (4\text{-}34)$$

The conventional notation

$$\Gamma(\lambda) = O[\lambda^{p+1}] \qquad (4\text{-}35)$$

is defined to mean that Γ has a representation of the form (4-34), or equivalently, that (4-33) holds. If Γ is of exact order p, then Γ_{p+1} is called the error constant associated with Γ. It is evident that

$$\Gamma_{p+1} = \lim_{\lambda \to 0} \frac{\Gamma(\lambda)}{\lambda^{p+1}} \qquad (4\text{-}36)$$

whenever Γ is of order p.

A technical requirement for the discussion of order of composite multistep methods is that the solutions to the differential equation (1-1) be analytic functions (or at least differentiable sufficiently many times). A typical sufficiency condition for the solutions to be analytic is that f be analytic jointly in both its arguments.

According to the formal definition given below, a composite multistep method is of order p if the truncation error ϵ_T associated with its composite matrix is of order p, as a function of the step size h. In order for such a

definition to make sense, ϵ_T must be an analytic function of h at the origin. Consider the expression (4-19) for the truncation error associated with a composite matrix R and an analytic true solution x. It is evident from (4-8), (4-17), and (4-18) that $J(h\mathcal{D})$ can be written in the form

$$J(h\mathcal{D}) = \sum_{i=0}^{\infty} J_i (h\mathcal{D})^i \quad , \tag{4-37}$$

where each J_i is a $2(m+n) \times 1$ matrix of real numbers. Since (4-37) is uniformly convergent in any neighborhood of the origin, it can be substituted into (4-19) to yield

$$\epsilon_T(h) = R\left[\sum_{i=0}^{\infty} J_i (h\mathcal{D})^i x^T\right](t_0)$$

$$= \sum_{i=0}^{\infty} [RJ_i x^{(i)T}(t_0)]h^i \tag{4-38}$$

where $x^{(i)}$ denotes the i-th derivative of x. Thus, ϵ_T is a power series in h whose i-th coefficient is proportional to the i-th derivative of the true solution x. In particular, ϵ_T is analytic at the origin.

4.16. Definition: A composite matrix R is said to be of order p if its truncation error ϵ_T for every analytic initial value problem is of order p in the step size h. R is said to be of exact order p if (1) it is of order p, and, (2) there exists an analytic initial value problem for which ϵ_T is of exact order p. If R is of exact order p, then RJ_{p+1} is called the truncation error constant c_T associated with R, where J_{p+1} is given by (4-37).

The exact order of ϵ_T is clearly the minimum of the exact orders of its rows, by definition of exact order. It is evident from (4-14) that the exact order of a given row of ϵ_T depends upon only the corresponding row of the composite matrix R, and not upon other rows of R.

It is natural to extend the terminology with respect to order, exact order, and truncation error constant to refer to a composite multistep method whose composite matrix has the corresponding property. The exact order p and truncation error constant c_T of a composite matrix together give a complete

"first order" description of the behavior of the truncation error, in the sense that

$$\epsilon_T(h) - c_T h^{p+1} \chi^{(p+1)}(t_0) = O[h^{p+2}] \tag{4-39}$$

for all analytic initial value problems.

In Section 4.4 it was shown that truncation error is an extrinsic property of composite multistep methods. The same can be shown for truncation error constant. On the other hand, order and exact order are intrinsic properties with respect to an equivalence class of composite multistep methods. The reason is that $RJ_i = 0$ if and only if $SRJ_i = 0$ for any nonsingular matrix S. However, $RJ_i \neq SRJ_i$ in general. Thus, equivalent composite multistep methods have the same exact order, but not the same truncation error constant.

4.17. Proposition: The composite matrix R is of order [exact order] p if and only if RJ is of order [exact order] p, where J is given by (4-8), (4-17), and (4-18).

Proof: By (4-38) the condition that R be of order p is equivalent to the condition that $RJ_i \chi^{(i)}(t_0) = 0$ for all analytic functions χ and all $i = 0, 1, 2, \ldots, p$. The latter condition is clearly equivalent to the condition that $RJ_i = 0$ for all $i = 0, 1, 2, \ldots, p$. This in turn is equivalent to the statement that RJ be of order p, since RJ can be represented in the form $RJ(\lambda) = \sum_{i=0}^{\infty} RJ_i \lambda^i$.

The assertion of Proposition 4.17 concerning exact order follows logically from that concerning order. □

Note that the truncation error $\tilde{\epsilon}_T$ for linear autonomous differential equations is proportional to RJ, by Proposition 4.14. Therefore, Proposition 4.17 can be interpreted as follows: The exact order of a composite matrix is equal to the exact order of its truncation error $\tilde{\epsilon}_T$ for linear autonomous differential equations.

For smooth differential equations the discretization error ϵ_D is analytic in h near the origin, as can be shown using Lemma 4.12. Therefore, one can speak of the order and error constant associated with ϵ_D.

4.18. Proposition: Let R be a strongly regular composite matrix of exact order p, and let c_T be the associated truncation error constant. Then the discretization error ϵ_D is of exact order p for every analytic initial value problem with $x^{(p+1)}(t_0) \neq 0$. Furthermore, the error constant associated with ϵ_D is equal to $A_f^{-1} c_T x^{(p+1)T}(t_0)$.

Proof: Since R is strongly regular, A_f is nonsingular. In such case (4-28) can be used to show that $[\hat{\Delta}(h)]^{-1}$ is representable as a power series in h, with $[\hat{\Delta}(0)]^{-1} = A_f^{-1}$. Therefore, by (4-26) ϵ_D can be represented approximately from $[\hat{\Delta}]^{-1}$ and ϵ_T using the Cauchy product of their power series. In particular, the first nonzero term of $\epsilon_D(h)$ is the first nonzero term of $[\hat{\Delta}(h)]^{-1}$, namely A_f^{-1}, postmultiplied by the first nonzero term of $\epsilon_T(h)$, namely $c_T h^{p+1} x^{(p+1)T}(t_0)$. See (4-39). This proves both conclusions of Proposition 4.18 □

As noted earlier with regard to ϵ_T, the exact order of ϵ_D is the minimum of the exact orders of its rows. Furthermore, each row of ϵ_D gives the discretization error associated with the corresponding future point. Therefore, the first conclusion of Proposition 4.18 has the following interpretation: The exact order of a strongly regular composite matrix R is equal to the minimum of the exact orders of its discretization errors associated with the future points. This result is sufficient justification for Definition 4.16.

Define the discretization error constant c_D by

$$c_D = A_f^{-1} c_T \quad . \tag{4-40}$$

According to Proposition 4.18

$$\epsilon_D(h) - c_D h^{p+1} x^{(p+1)T}(t_0) = O[h^{p+2}] \tag{4-41}$$

for every strongly regular composite matrix of order p and every sufficiently smooth differential equation. Since ϵ_D is intrinsic with respect to an equivalence class of composite multistep methods, it is evident from (4-41) that c_D is also intrinsic. Thus, c_D is indicative of only the inaccuracy of the approximating sequence $\{x_i\}$, while c_T, which was shown to be extrinsic, depends upon the manner in which $\{x_i\}$ is computed, as well as upon its value.

In other words, c_D gives a far more legitimate measure of the error committed by a composite multistep method than does c_T.

It is evident from (4-40) and Proposition 4.6 that if R is in canonical form, then $c_D = c_T$. It follows from (4-39) and (4-41) that if R is in canonical form, then the truncation error and discretization error are equal, to a first approximation, in that

$$\epsilon_D(h) - \epsilon_T(h) = O[h^{p+2}]$$

whenever $\epsilon_T(h) = O[h^{p+1}]$. In this sense canonical composite multistep methods are distinguished members of their equivalence classes.

For a composite matrix of exact order p, the coefficients of h^{p+2}, h^{p+3}, ... in ϵ_D depend in a very complicated way upon the function f and the true solution X. However, in the linear autonomous case there is a great simplification. By Propositions 4.14 and 4.15

$$\tilde{\epsilon}_D(\lambda) = [D(\lambda)]^{-1} RJ(\lambda) x_0 \,. \tag{4-42}$$

By (1-10) and the formula for a geometric series

$$[D(\lambda)]^{-1} = (I - \lambda A_f^{-1} B_f)^{-1} A_f^{-1} = \sum_{i=0}^{\infty} (A_f^{-1} B_f)^i \lambda^i A_f^{-1}$$

whenever $|\lambda| < 1 / \|A_f^{-1} B_f\|$. Therefore, using the Cauchy product,

$$\tilde{\epsilon}_D(\lambda) = \sum_{i=0}^{\infty} \left[\sum_{j=0}^{i} (A_f^{-1} B_f)^j A_f^{-1} RJ_{i-j} \right] \lambda^i x_0 \,. \tag{4-43}$$

If R is canonical, then $A_f = I_n$, and the above is equal to

$$\tilde{\epsilon}_D(\lambda) = \sum_{i=0}^{\infty} \left[\sum_{j=0}^{i} B_f^j RJ_{i-j} \right] \lambda^i x_0 \,. \tag{4-44}$$

The corresponding bracketed terms in (4-43) and (4-44) are equal, and are reasonably simple functions of the composite matrix R.

Unfortunately, no such simple formulation applies to the nonlinear case. In the first place, the approximation (4-26), which is exact in the linear

case, is not sufficiently accurate to determine all the higher order coefficients. (If it were, then it would not be a mere approximation.) Secondly, even (4-26) with (4-28) yields a relation for higher order coefficients which is a complicated nonlinear function of χ and f and their derivatives. The main conclusion to be drawn from this discussion is that in general the exact order of the discretization error associated with a given future point may be as small as the exact order of the associated composite matrix R, even when a linear analysis of the type (4-43) makes it appear to be greater.

It can be seen that if R is canonical, the discretization error for the i-th future point depends to first order (that is, in the sense of (4-41)) upon the i-th row of R, but not upon the other rows of R. However, the higher order terms in the discretization error for the i-th future point depend, in general, upon all the rows of R. This fact can be verified for the linear autonomous case using (4-44). For example, if R is of exact order p, the coefficient of $\lambda^{p+2}\chi_0$ in (4-44) is equal to

$$RJ_{p+2} + B_f C_D .$$

The factor B_f couples various rows of R to the error for a given future point, unless B_f is diagonal.

§4.8 Order and the Characteristic Polynomial

The order of a composite multistep method (R,k) is an index of its behavior for arbitrarily small step size. The characteristic polynomial P of (R,k) can also be used to gain information about the behavior of (R,k) for arbitrarily small step size, by considering the zeros of $P(\lambda,\cdot)$ for λ near the origin. Thus, it is not surprising that the characteristic polynomial and the order of a composite multistep method are related.

4.19. Definition: The characteristic polynomial P of (1-52) associated with a composite multistep method (R,k) is said to be of order [exact order] p if $P(\lambda,e^{k\lambda})$ is of order [exact order] p in λ.

4.20. Proposition: If the composite multistep method (R,k) is of order p, then its characteristic polynomial P of (1-52) is of order p.

Proof: Let Γ denote the $n \times n$ matrix of polynomials

$$\Gamma(\zeta) = I_n + [0 \quad \zeta \quad \zeta^2 \quad \cdots \quad \zeta^{k-1} \quad \zeta^{m+k} \quad \cdots \quad \zeta^{m+n-1}]^T J_{1n}, \quad (4\text{-}45)$$

where J_{1n} is the first row of I_n. By direct substitution it is easy to verify from (4-9), (4-17), and (4-45) that $\hat{Z}(\zeta^k)\Gamma(\zeta)J_{1n}^T = U(\zeta)$ for all $\zeta \in C$. Therefore, by Lemma 4.7

$$Q(\lambda,\zeta^k)\Gamma(\zeta)J_{1n}^T = RL(\lambda)\hat{Z}(\zeta^k)\Gamma(\zeta)J_{1n}^T = RL(\lambda)U(\zeta) .$$

Substituting $\zeta = e^\lambda$ into the above, and applying (4-18) and Proposition 4.17, gives

$$Q(\lambda,e^{k\lambda})\Gamma(e^\lambda)J_{1n}^T = RL(\lambda)U(e^\lambda) = RJ(\lambda) = O[\lambda^{p+1}] . \quad (4\text{-}46)$$

That is, the first column of $Q(\lambda,e^{k\lambda})\Gamma(e^\lambda)$ is of order p in λ.

Since the second term in (4-45) is strictly lower triangular, it follows that $\det \Gamma(e^\lambda) = \det I_n = 1$. Therefore, by (1-52)

$$P(\lambda,e^{k\lambda}) = \det Q(\lambda,e^{k\lambda}) \det \Gamma(e^\lambda) = \det[Q(\lambda,e^{k\lambda})\Gamma(e^\lambda)] .$$

But determinants are linear functions of every column. Hence (4-46) shows that $P(\lambda,e^{k\lambda}) = O[\lambda^{p+1}]$. □

The converse of Proposition 4.20 is false. That is, if P is of order p, it does not follow that (R,k) is of order p. In other words, Proposition 4.20 cannot be strengthened to the point of replacing "order" by "exact order" in the hypothesis and conclusion*. Various classes of counterexamples are given by [Watts] and [Shampine and Watts]. One such class, the class of diagonal Pade composite one-step methods, consists of an infinite sequence of

---

*The stronger result is well known to hold [Dahlquist] in the important special case of multistep methods (n = 1). A simple proof of this fact in the present notation is as follows: By (4-18) and Proposition 4.17, the exact order of R is the exact order of $RL(\lambda)U(e^\lambda)$ in λ. If n = 1, then P of (1-52) is essentially equal to Q. Therefore, by Lemma 4.7, the exact order of P is the exact order of $RL(\lambda)\hat{Z}(e^\lambda)$ in λ. The proof is completed with the observation that if n = 1, then $\hat{Z}$ of (4-9) and U of (4-17) are equal.

methods of the form (R_n,n) for all $n = 1, 2, 3, \ldots$, in which R_n is of exact order n, but the characteristic polynomial of (R_n,n) is of exact order 2n.

A simple but important consequence of Proposition 4.20 is the following:

4.21. Proposition: Let (R,k) be a strongly regular composite multistep method, and let P of (1-52) be its characteristic polynomial, factored according to (1-41). If R is of order zero, and $\psi(1) \neq 0$, then $\overline{P}(0,1) = 0$.

Proof: By Proposition 4.20, $P(\lambda,e^{k\lambda}) = O[\lambda]$. Therefore,

$$P(0,1) = \lim_{\lambda \to 0} P(\lambda,e^{k\lambda}) = 0 \quad ,$$

the second equality following from (4-33). Applying the above to (1-41) gives

$$\phi(0)\psi(1)\overline{P}(0,1) = 0 \quad .$$

But $\phi(0) \neq 0$ by Proposition 4.6, since the zeros of ϕ are poles of P. Now the conclusion of Proposition 4.21 is immediate, since $\psi(1) \neq 0$ by hypothesis. □

The conclusion of Proposition 4.21 is precisely the relation (3-1) stated in Section 3.1 as being satisfied for all composite multistep methods of potential usefulness. Therefore, it is important to demonstrate that the hypotheses of Proposition 4.21 are necessary, by any reasonable standard, for (R,k) to be a practical method for numerical solution of ordinary differential equations. The necessity of strong regularity has been discussed in Section 4.3. The necessity of R being of order zero is clear from Proposition 4.18, which implies that the "approximating" sequence may bear no relation whatever to the true solution, unless R is of order zero. Finally, the condition $\psi(1) \neq 0$ is necessary for A-stability, by Theorem 2.5.

A much more compelling argument in justification of the last two hypotheses of Proposition 4.21 is that they are necessary for convergence. By convergence is meant, roughly speaking, that the approximating sequence generated by the numerical method approach the true solution (in some appropriate sense) as the step size approaches zero. Convergence is generally regarded to be the most important fundamental property for a numerical method; it is widely used to distinguish "possibly useful" methods from "definitely not useful" methods. For many classes of methods it has been shown that a given method is convergent

if and only if it is stable* and of order unity. Convergence, stability, and order are evidently intrinsic properties with respect to the class of all methods for numerical solution of ordinary differential equations. In fact, it has recently been shown in a very general setting (which includes composite multistep methods) that stability and order unity are necessary and sufficient for convergence [Chartres and Stepleman]. Of course, the condition "order zero" of Proposition 4.21 is a weak necessary condition for order unity. Likewise, it can be shown that the condition $\psi(1) \neq 0$ is a weak necessary condition for stability. In conclusion, the last two hypotheses of Proposition 4.21 are weak necessary conditions for the fundamental property of convergence.

The order of a characteristic polynomial is reflected in the behavior of its algebraic function, as shown by the following interesting generalization of Proposition 4.21.

4.22. Proposition: Let (R,k) be a strongly regular composite multistep method, and let P of (1-52) be its characteristic polynomial, factored according to (1-41). If P is of order p, $\psi(1) \neq 0$, and $\overline{P}(0,\cdot)$ has 1 as a zero of unit multiplicity, then there exists a branch η of the algebraic function associated with $\overline{P}$ which is analytic at the origin, and for which

$$\eta(\lambda) - e^{k\lambda} = O[\lambda^{p+1}] \quad . \tag{4-47}$$

Proof: By hypothesis, $\overline{P}(0,1) = 0$. Since the multiplicity of 1 as a zero of $\overline{P}(0,\cdot)$ is unity, there exists a unique analytic function η, defined in a sufficiently small neighborhood of the origin, such that

$$\eta(0) = 1 \tag{4-48}$$

and such that $\overline{P}[\lambda,\eta(\lambda)] = 0$ for all λ in this neighborhood. The latter relation can be used to show that there exists a function $G : C^2 \to C$, analytic in a neighborhood of (0,1), such that

*Stability is a property akin to, but essentially much weaker than, A-stability. The relation between the two properties is roughly as follows: For A-stability the approximating sequences are examined for all $h > 0$; for stability they are examined only for sufficiently small $h > 0$.

$$\bar{P}(\lambda,\zeta) = G(\lambda,\zeta)[\eta(\lambda)-\zeta] \qquad (4\text{-}49)$$

in this neighborhood. Since the multiplicity of 1 as a zero of $\bar{P}(0,\cdot)$ is unity, it is evident that $G(0,1) \neq 0$. Therefore, $G(\lambda,e^{k\lambda})$ is of exact order -1 in λ.

By Definition 4.19, $P(\lambda,e^{k\lambda}) = O[\lambda^{p+1}]$. Expanding this relation with (1-41) gives

$$\phi(\lambda)\psi(e^{k\lambda})\bar{P}(\lambda,e^{k\lambda}) = O[\lambda^{p+1}] .$$

Since $\phi(0) \neq 0$ (see proof of Proposition 4.21), the exact order of ϕ is -1. Similarly, $\psi(e^{k\lambda})$ is of exact order -1 in λ, by the hypothesis $\psi(1) \neq 0$. Therefore,

$$\bar{P}(\lambda,e^{k\lambda}) = O[\lambda^{p+1}] .$$

Combining the above relation with (4-49) gives

$$G(\lambda,e^{k\lambda})[\eta(\lambda)-e^{k\lambda}] = O[\lambda^{p+1}] .$$

Now (4-47) follows from the fact that $G(\lambda,e^{k\lambda})$ is of exact order -1 in λ. □

The hypothesis of unit multiplicity in Proposition 4.22 is presumably necessary; in any case, it excludes only pathological characteristic polynomials.

A geometric interpretation of Proposition 4.22 can be given by making the "change of variable" $\lambda = \hat{j}\omega$, and writing (4-47) in the form

$$\eta(\hat{j}\omega) - e^{\hat{j}k\omega} = O[\omega^{p+1}] . \qquad (4\text{-}50)$$

For ω real, $\eta(\hat{j}\omega)$ is part of the zeta locus associated with $\bar{P}$, while $e^{\hat{j}k\omega}$ is part of the unit circle. Therefore, the conclusion of Proposition 4.22 implies that part of the zeta locus "follows" the unit circle (the boundary of the forbidden region) to order p, in the sense of (4-50). In particular, (4-48) holds when $p \geq 0$. Proposition 4.22 can thus be interpreted as follows: The behavior of the zeta locus near unity reflects the order of the characteristic polynomial. (A dual development can be used to show that the order of the characteristic polynomial is reflected in the behavior* of the lambda locus near the origin.)

*In both cases the "behavior" connotes not merely the shape of the locus, but also its "rate of travel". Strictly speaking, this property is not of the

§4.9 The Order Relations and Free Parameters

In this section it is shown that a composite matrix R is of order p if and only if R satisfies a certain set of linear homogeneous algebraic relations, called the order relations. Next a convenient representation for the class of all strongly regular composite matrices of order p is obtained in terms of "free parameters". Finally, an explicit representation in terms of free parameters is given for the set of characteristic polynomials corresponding to such a class. These representations are employed in the next two sections to find high-order A-stable composite multistep methods.

For each i = 1, 2, 3, ... define the infinite sequence of 2i x 1 matrices of integers as follows:

$$\pi_{i0} = [\underbrace{1\ \ 1\ \dots\ 1}_{i}\ \underbrace{0\ \ 0\ \dots\ 0}_{i}]^T \qquad \text{(a)}$$

(4-51)

$$\pi_{ij} = [0^j\ \ 1^j\ \dots\ (i-1)^j\ \ -j0^{j-1}\ \ -j1^{j-1}\ \dots\ -j(i-1)^{j-1}]^T \qquad \text{(b)}$$

for j = 1, 2, 3, ..., where $0^0 = 1$ by convention. Next define, for each i = 1, 2, 3, ... and each p = 0, 1, 2, ... the 2i x (p+1) matrix of integers

$$\Pi_{ip} = [\pi_{i0}\ \ \pi_{i1}\ \dots\ \pi_{ip}] \ . \qquad (4\text{-}52)$$

4.23. Proposition: Let R be a composite matrix with m past points and n future points. Then R is of order p if and only if

$$R\Pi_{(m+n)p} = 0 \qquad (4\text{-}53)$$

where $\Pi_{(m+n)p}$ is defined by (4-51) and (4-52).

If R is of exact order p, its truncation error constant is given by

$$c_T = R\lambda_{(m+n)(p+1)} \ / \ (p+1)! \ . \qquad (4\text{-}54)$$

If in addition R is strongly regular, its discretization error constant is given by

---

locus per se, but rather of its parametric representation, that is, its associated Riemann surface. From this viewpoint it can be said that the Riemann surface follows the graph of the function $e^{k\lambda}$ to order p at the point $(0,1) \in \bar{C}^2$.

$$c_D = (\text{adj } A_f) R\pi_{(m+n)(p+1)} \;/\; (p+1)!\, \det A_f, \qquad (4\text{-}55)$$

where A_f is given by (1-8).

Proof: By Proposition 4.17, R is of order p if and only if $RJ(\lambda) = O[\lambda^{p+1}]$, where

$$J(\lambda) = L(\lambda)U(e^{\lambda}) = \begin{bmatrix} U(e^{\lambda}) \\ \\ -\lambda U(e^{\lambda}) \end{bmatrix} . \qquad (4\text{-}56)$$

By (4-17) the i-th component of $U(e^{\lambda})$ can be written for i = 0, 1, ..., m+n-1 as

$$(e^{\lambda})^i = \sum_{j=0}^{\infty} \frac{1}{j!}\, i^j \lambda^j .$$

Therefore, the i-th component of $-\lambda U(e^{\lambda})$ is

$$-\lambda(e^{\lambda})^i = -\lambda \sum_{j=0}^{\infty} \frac{1}{j!}\, i^j \lambda^j = \sum_{j=1}^{\infty} \frac{1}{j!}\, (-ji^{j-1})\lambda^j .$$

Substituting the above two relations into (4-56) and comparing with (4-51) shows that J can be written in the form

$$J(\lambda) = \sum_{j=0}^{\infty} J_j \lambda^j$$

where

$$J_j = \frac{1}{j!}\, \pi_{(m+n)j} , \qquad j = 0, 1, 2, \ldots . \qquad (4\text{-}57)$$

The proof of Proposition 4.17 shows that R is of order p if and only if $RJ_j = 0$ for all j = 0, 1, ..., p. By (4-57) the latter is equivalent to the condition that $R\pi_{(m+n)j} = 0$ for all j = 0, 1, ..., p. This condition is in turn equivalent, by (4-52), to (4-53).

The relation (4-54) follows directly from (4-57) and Definition 4.16. Finally, (4-55) follows from (4-40), (4-54), and the identity $A_f^{-1} = (\text{adj } A_f) / \det A_f$. □

The relation (4-53) consists of a set of p+1 linear homogeneous algebraic equations in the 2(m+n) columns of the composite matrix R. This set of equations will be referred to as the set of order relations.

The remainder of this chapter is devoted to investigation of the class $\mathcal{R}(p,n,m)$ of strongly regular composite matrices having m past points, n future points, and order p. Thus, $R \in \mathcal{R}(p,n,m)$ if and only if R is an $n \times 2(m+n)$ matrix satisfying the order relations (4-53), and with A_f of (1-8) nonsingular. Since only such intrinsic properties as order, discretization error constant, and A-stability are of interest, it is enough to investigate the subclass $\mathcal{R}_c(p,n,m)$ of canonical composite matrices. By Proposition 4.6 these composite matrices are the ones in $\mathcal{R}(p,n,m)$ for which $A_f = I_n$. A convenient representation for the elements of $\mathcal{R}_c(p,n,m)$ is based on the following theorem, which is a statement of the nature of the solutions R to the order relations, subject to $A_f = I_n$. The proof is given in Appendix G.

4.24. Theorem: Let m and n be positive integers, let p be a nonnegative integer or -1, and let μ be given by

$$\mu = 2m + n - p - 1 \quad . \tag{4-58}$$

Then $\mathcal{R}(p,n,m)$ is nonempty if and only if $\mu \geq 0$.

Suppose $\mu \geq 0$. Then there exists a $\mu \times 2(m+n)$ matrix S of integers, partitioned after the m-th and (m+n)-th columns as follows,

$$S = [S_p \quad S_f \quad S_B] \quad , \tag{4-59}$$

such that

$$S_f = 0 \tag{a}$$

$$\text{rank } S = \mu \tag{b} \tag{4-60}$$

and

$$S\Pi_{(m+n)p} = 0 \quad . \tag{c}$$

If $R \in \mathcal{R}_c(p,n,m)$ and X is an $n \times \mu$ matrix, then $(R+XS) \in \mathcal{R}_c(p,n,m)$. Conversely, if R and $\hat{R}$ are in $\mathcal{R}_c(p,n,m)$, then there exists a unique $n \times \mu$ matrix X such that

$$\hat{R} = R + XS \quad . \tag{4-61}$$

According to (4-58) and the first conclusion of Theorem 4.24

$$p \leq 2m + n - 1 \tag{4-62}$$

for every $n \times 2(m+n)$ strongly regular composite matrix R of order p. This relation gives an upper bound on the order of a composite matrix, independent of any stability considerations. By contrast, [Sloate, Theorem 3.7] proves (for $n \geq 2$) that

$$p \leq m + 2n - 2 \tag{4-63}$$

under the additional hypothesis that (R,1) be "stable at infinity". (That is, the zeros of $P(\infty,\cdot)$ lie in the open unit disc, where P is the characteristic polynomial of (R,1).) When $m = n - 1$ the two upper bounds for p are the same; when $m < n - 1$ (4-62) is a tighter bound on p than (4-63). In other words, Sloate's result for $m \leq n - 1$ is not so sharp as possible.

The essential idea of the second paragraph of Theorem 4.24 can be stated as follows: For each $R \in \mathcal{R}_c(p,n,m)$ and each S satisfying (4-60) there exists a one-to-one mapping of $\mathcal{R}_c(p,n,m)$ onto the set of all $n \times \mu$ matrices X. Furthermore, the correspondence between X and an element $\hat{R} \in \mathcal{R}_c(p,n,m)$ is given by (4-61). In this sense (4-61) provides a representation for the class $\mathcal{R}_c(p,n,m)$ in terms of $n \times \mu$ matrices X. Alternatively, $\hat{R}$ of (4-61) can be viewed as a composite matrix parameterized by the matrix X of "free parameters". Each element of X is a free parameter; thus there are $n\mu$ free parameters in all, with μ given by (4-58).

Although the parameterization of (4-61) is elegant, it is not so useful in practice as a modification to be described below. The reason is that it is highly desirable to deal with integers, rather than rational or real numbers, when automatic computation is to be employed. In order to formulate an integer representation, the concept of integer form is introduced. A composite matrix R is said to be in integer form if (1) all its elements are integers, and, (2) its submatrix A_f is diagonal. Not every composite matrix is row equivalent to a composite matrix in integer form. A simple counterexample is $[0 \quad 1 \quad 0 \sqrt{2}]$. However, within each nonempty class $\mathcal{R}(p,n,m)$, composite matrices in integer form are sufficiently plentiful for present purposes.

Let $\mathcal{R}_i(p,n,m)$ denote the subclass of $\mathcal{R}(p,n,m)$ consisting of all composite matrices in integer form. If $\mathcal{R}(p,n,m)$ is nonempty, so is $\mathcal{R}_i(p,n,m)$. This fact can be demonstrated by the same technique as was used with respect to S in the proof of Theorem 4.24. (See the footnote in Appendix G.) Thus, a canonical

composite matrix R in $\mathcal{R}(p,n,m)$ can be constructed with rational elements, since $\Pi_{(m+n)p}$ has integral elements. Multiplying R by the least common multiple of its denominators gives a composite matrix in integer form in $\mathcal{R}(p,n,m)$.

In the following theorem, the representation of Theorem 4.24 is modified by the introduction of an additional free parameter allowing elements of $\mathcal{R}_i(p,n,m)$ to be represented entirely in terms of integers.

4.25. Theorem: Let m and n be positive integers, let p be a non-negative integer or -1, and let μ be given by (4-58). Assume $\mu \geq 0$. Let $R \in \mathcal{R}_i(p,n,m)$, and let S be a $\mu \times 2(m+n)$ matrix of integers satisfying (4-60).

a) If x is a nonzero integer and X is an $n \times \mu$ matrix of integers, then $(xR+XS) \in \mathcal{R}_i(p,n,m)$.

b) For every $\hat{R} \in \mathcal{R}_i(p,n,m)$ there exist a nonzero integer x and an $n \times \mu$ matrix X of integers such that $xR + XS$ is row equivalent to $\hat{R}$.

Proof: Since x, R, X, and S are all integral, $xR + XS$ is obviously integral. By assumption, the submatrix A_f of R is diagonal and nonsingular; therefore, so is xA_f, since $x \neq 0$. Therefore $xR + XS$ is in integer form. In addition, $xR + XS$ is strongly regular, by (4-60a). This proves part a) of Theorem 4.25.

To prove part b), let A_f [$\hat{A}_f$] denote the submatrix of R [$\hat{R}$] given by (1-8). Then $A_f^{-1}R$ and $\hat{A}_f^{-1}\hat{R}$ are in $\mathcal{R}_c(p,n,m)$, by Proposition 4.5. Now by Theorem 4.24 there exists a unique $n \times \mu$ matrix $\tilde{X}$ such that

$$\hat{A}_f^{-1}\hat{R} = A_f^{-1}R + \tilde{X}S \quad . \tag{4-64}$$

Premultiplying (4-64) by A_f and postmultiplying by S^T give

$$(A_f\hat{A}_f^{-1}\hat{R}-R)S^T = A_f\tilde{X}SS^T \quad .$$

By (4-60b) SS^T is nonsingular. Hence

$$A_f\tilde{X} = (A_f\hat{A}_f^{-1}\hat{R}-R)S^T(SS^T)^{-1} \quad . \tag{4-65}$$

Since R, $\hat{R}$, and S are integral matrices, it follows that the right-hand-side of (4-65) is a matrix of rational numbers. That is, the left-hand-side can be written in the form

$$A_f\tilde{X} = X/x \tag{4-66}$$

for some matrix of integers X and some nonzero integer x. From (4-64) and (4-66) it is straightforward to show that

$$(\hat{A}_f A_f^{-1}/x)(xR+XS) = \hat{R} \quad .$$

Therefore, xR + XS is row equivalent to $\hat{R}$ (see Proposition 4.3). □

Theorem 4.25 provides a representation for the subclass $\mathcal{R}_i(p,n,m)$ in terms of nonzero integers x and $n \times \mu$ matrices X of integers. Of course, the entire class $\mathcal{R}(p,n,m)$ can be represented by setting x = 1 and allowing X to have non-integral elements. In either case a member of each row equivalence class, but not necessarily every member, is represented explicitly.

The importance of Theorem 4.25 lies in the fact that $\mathcal{R}_i(p,n,m)$ is a dense subclass of $\mathcal{R}(p,n,m)$ if row equivalent composite matrices are regarded as equal. That is, for every $R \in \mathcal{R}(p,n,m)$ and every $\epsilon > 0$, there exists an $\hat{R} \in \mathcal{R}(p,n,m)$ such that (1) $\|R-\hat{R}\| < \epsilon$, and, (2) $\hat{R}$ is row equivalent to a member of $\mathcal{R}_i(p,n,m)$. To see this, consider the fact that the subclass of $\mathcal{R}_c(p,n,m)$ consisting of composite matrices with rational elements is dense in $\mathcal{R}_c(p,n,m)$.

The main goal of the present development is to find A-stable composite multistep methods of high order. That is, for a given class $\mathcal{R}(p,n,m)$ it is desired to determine whether there exist an $R \in \mathcal{R}(p,n,m)$ and a $k = 1,2,\ldots,n$ such that (R,k) is A-stable. For this purpose it is useful to generate a representation for the characteristic polynomial associated with $\mathcal{R}(p,n,m)$, as a function of the free parameters x and X. Such a representation is given by the following general theorem.

4.26. Theorem: Let (R,k) be a composite multistep method with m past points and n future points, let μ be a nonnegative integer, let S be a $\mu \times 2(m+n)$ matrix, let x be a nonzero scalar, and let X be an $n \times \mu$ matrix. Let $\bar{\mu} = \max(0,n-\mu)$. Then the characteristic polynomial P of (xR+XS,k) in the formulation (1-52) can be represented as

$$P = \sum_i x^{j_i - \bar{\mu}} X \begin{pmatrix} i'_1 & , & i'_2 & , \ldots, & i'_{n-j_i} \\ i_{1+j_i} - n, & & i_{2+j_i} - n, & \ldots, & i_n - n \end{pmatrix} P_i \tag{4-67}$$

where the summation is over all integral sequences $i = i_1, i_2, \ldots, i_n$ satisfying $1 \le i_1 < i_2 < \ldots < i_n \le n + \mu$, j_i denotes the largest integer such that $i_{j_i} \le n$ (if $i_1 > n$, put $j_i = 0$), $i'_1, i'_2, \ldots, i'_\mu$ is the complementary sequence* of i, and P_i is the characteristic polynomial associated with (R_i,k), where R_i is the composite matrix consisting in rows $i_1, i_2, \ldots, i_n$ of $\begin{bmatrix} R \\ S \end{bmatrix}$.

Proof: Applying Lemma 4.7 to xR + XS, and using (1-52) give

$$P = \det(xR+XS)\hat{LZ}$$

$$= \det[xI_n \quad X]\begin{bmatrix} R \\ S \end{bmatrix}\hat{LZ}$$

$$= \sum_i [xI_n \quad X]\begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix}\left(\begin{bmatrix} R \\ S \end{bmatrix}\hat{LZ}\right)\begin{pmatrix} i_1, & i_2, & \ldots, & i_n \\ 1, & 2, & \ldots, & n \end{pmatrix}, \qquad (4\text{-}68)$$

the last equality being an application of the Cauchy-Binet theorem. The second factor in (4-68) is equal to P_i, since rows i of $\begin{bmatrix} R \\ S \end{bmatrix}\hat{LZ}$ are precisely rows i of $\begin{bmatrix} R \\ S \end{bmatrix}$ postmultiplied by $\hat{LZ}$. On the other hand, the first factor in (4-68) can evidently be written as

$$x^{j_i} X\begin{pmatrix} i'_1 & , & i'_2 & , \ldots, & i'_{n-j_i} \\ i_{1+j_i} - n, & & i_{2+j_i} - n, & \ldots, & i_n - n \end{pmatrix}.$$

Observe that $\min_i j_i = \max(0,n-\mu) = \bar{\mu}$. Since $x \neq 0$, $x^{\bar{\mu}}$ is a trivial factor of (4-68), and can therefore be removed. The result is (4-67). □

For given positive integers m and n, integer $p \ge -1$, and positive integer $k \le n$, let $\mathcal{P}(p,n,k,m)$ denote the class of characteristic polynomials associated with (R,k) for all $R \in \mathcal{R}_i(p,n,m)$. Just as a representation for $\mathcal{R}_i(p,n,m)$ can be computed using Theorem 4.25, so also can a representation for $\mathcal{P}(p,n,k,m)$ be computed using Theorem 4.26 together with the representation for $\mathcal{R}_i(p,n,m)$. Evidently, the coefficients in both representations are integers. According

*For definition of complementary sequence see Appendix B.

to Theorem 4.25b), every element of $\mathcal{R}_i(p,n,m)$ is represented to within a diagonal integral row equivalence transformation. Likewise, every element of $P(p,n,k,m)$ is represented to within an integral trivial factor.

If $\mu = 0$ in Theorem 4.26, then X has no free parameters, and (4-67) becomes trivial. The next section discusses, in detail, classes $P(p,n,k,m)$ for which $\mu = 0$.

Consider the next simplest case: $\mu = 1$. In this case (4-67) can be seen to reduce to

$$P = xP_0 + \sum_{i=1}^{n} x_i P_i \qquad (4\text{-}69)$$

where P_0 is the characteristic polynomial associated with (R,k), x_i is the i-th element of X, P_i is the characteristic polynomial associated with (R_i,k), and R_i is $\begin{bmatrix} R \\ S \end{bmatrix}$ with row i removed. Since P is a linear function of the free parameters x and x_i, the case $\mu = 1$ is reasonably simple to analyze, as shown in Section 4.11. It is important to note that the linearity of (4-69) is a property of the new formulation (1-52) for the characteristic polynomial, and is not shared in general with the other formulations of Chapter 1. In other words, it is only the new formulation (1-52) which allows the linear analysis of Section 4.11 to be performed.

If $\mu \geq 2$, the coefficients of P in (4-67) depend upon sums of products of the free parameters, so that P is a nonlinear function of the elements of X. This situation is considerably more difficult to analyze than the linear case $\mu = 1$. For this reason, the writer has made no attempt to explore classes $P(p,n,k,m)$ for which $\mu \geq 2$. Certain subclasses have been explored for "stiff stability" in [Tendler].

§4.10 High-Order A-Stable Methods for $\mu = 0$

According to the preceding development, it is reasonable to limit the search for practical composite multistep methods to those whose composite matrices are strongly regular and in integer form. That is, for given values of p, n, and m, only composite matrices in $\mathcal{R}_i(p,n,m)$ are to be considered. For a given positive integer $k \leq n$, a composite multistep method (K,K) will be

said to be a (p,n,k,m) method, or a method of type (p,n,k,m), if $R \in \mathcal{R}_i(p,n,m)$. By the notation of Section 4.9, the characteristic polynomials associated with all the (p,n,k,m) methods are precisely the elements of P(p,n,k,m). Thus, to find an A-stable (p,n,k,m) method, it is enough to select values of the free parameters x and X (in the representation of Theorem 4.26) to give a polynomial P satisfying the A-stability criterion.

The results of a preliminary investigation along these lines are reported in this section and the next. All other things being equal, it is obvious that a practical composite multistep method should have the fewest possible numbers of past points and future points, in order to minimize storage space and computational time. Of course, such requirements are in conflict with the desire for high order, according to Theorem 4.24 and (4-62). In general, it is to be expected that additional relations must be placed among p, n, k, and m if the property of A-stability is desired. These matters are discussed in more detail below.

Let the number m of past points and the number n of future points be fixed. According to (4-62) the maximum possible order p of a strongly regular $n \times 2(m+n)$ composite matrix is given by

$$p = 2m + n - 1 \quad . \tag{4-70}$$

This value of p corresponds to the case $\mu = 0$ in Theorem 4.24. It is evident from Theorem 4.24 that if $\mu = 0$, the class $\mathcal{R}_c(p,n,m)$ contains exactly one composite matrix. In other words, the (p,n,k,m) methods for which (4-70) holds are unique, and can be investigated for A-stability by direct computation using the algorithm described in Appendix E.

In Table 4.1 the values of p in (4-70) are tabulated for small values of m and n. Each entry in the table is readily associated with a class $\mathcal{R}_i(p,n,m)$, and therefore with all methods of type (p,n,k,m) for $1 \leq k \leq n$. The letters accompanying the entries refer to notes below the table, outlining the present state of knowledge concerning each class.

The note "C" refers to the following conjecture, which has been proposed and motivated in [Sloate and Bickart]:

4.27. Conjecture: Let (R,k) be a composite multistep method of order p, with n future points. If $p > 2n$, then (R,k) is not A-stable.

n \ m	1	2	3	4
1	2BT	4CD	6CD	8CD
2	3B	5CN	7C	9C
3	4B	6N	8C	10C
4	5B	7R	9C	11C
5	6B	8U	10U	12C

Notes:

B: [Bickart, et al] has shown that the (n+1,n,n,1) "Newton-Cotes" methods are A-stable, with $T = E$ and $M = I$. In fact, [Watts] has shown that A-stability holds for all $n \leq 8$, but not for n = 9 or 10.

C: According to Conjecture 4.27, these methods are not A-stable.

D: [Dahlquist] has proved that these methods are not A-stable.

N: These methods are not A-stable. See text.

R: The (7,4,k,2) methods are A-stable for k = 1 (see Fig. 4.1), but not for k = 2, 3, 4. See text.

T: This class consists of the trapezoidal rule, which [Dahlquist] has shown to be A-stable.

U: It is presently unknown whether these methods are A-stable.

Table 4.1. Orders of Composite Matrices in Unique Case ($\mu = 0$).

For the case n = 1, Conjecture 4.27 has been proved in [Dahlquist]. The classes in Table 4.1 to the right of the irregular line (which is effectively a line of slope -2) are those corresponding to the hypothesis of Conjecture 4.27.

It is believed that for a given order, a composite multistep method will be more efficient if its number n of future points is kept small at the expense of the number m of past points. The reason is that the number of simultaneous algebraic equations to be solved at each iteration is proportional to n. In Table 4.1 all "lines with slope 2" are lines of constant order, with the more efficient methods to the right. Assuming Corollary 4.27 to be true, it follows that the most important classes to investigate for A-stability are those classes directly to the left of the irregular line in Table 4.1.

According to note "N", the simplest such nontrivial case, the $\mathcal{R}_i(6,3,2)$ class, contains no A-stable methods. In fact, for both this class and the $\mathcal{R}_i(5,2,2)$ class, all methods are unstable at infinity. That is, for each characteristic polynomial P, $P(\infty,\cdot)$ has zeros outside the closed unit disc. Such characteristic polynomials can never satisfy the A-stability criterion*. The same phenomenon occurs in the (7,4,k,2) methods for $k \geq 2$. However, the (7,4,1,2) method has been found by the algorithm of Appendix E to be strongly A-stable.

The composite matrix R and characteristic polynomial P for the (7,4,1,2) method are displayed in Fig. 4.1a) and b), respectively**. More precisely, the matrix Ψ of coefficients of P is displayed: With P represented in the form (1-38), Ψ is simply the $(n+1) \times (m+1)$ matrix of coefficients p_{ij}. Thus, for any $(\lambda,\zeta) \in \mathbb{C}^2$,

$$P(\lambda,\zeta) = [1 \quad \lambda \quad \lambda^2 \quad \dots \quad \lambda^n]\Psi[1 \quad \zeta \quad \zeta^2 \quad \dots \quad \zeta^m]^T \quad . \qquad (4\text{-}71)$$

In the following, a given coefficient matrix Ψ will be identified with the polynomial P according to (4-71).

The particular polynomial of Fig. 4.1b) is an example of a characteristic polynomial in integer canonical form. A characteristic polynomial will be said to be in integer canonical form if all the elements of its coefficient matrix are integers, and

---

*This fact follows from Proposition 1.7 by arguments similar to those used in the proof of Corollary 1.8.

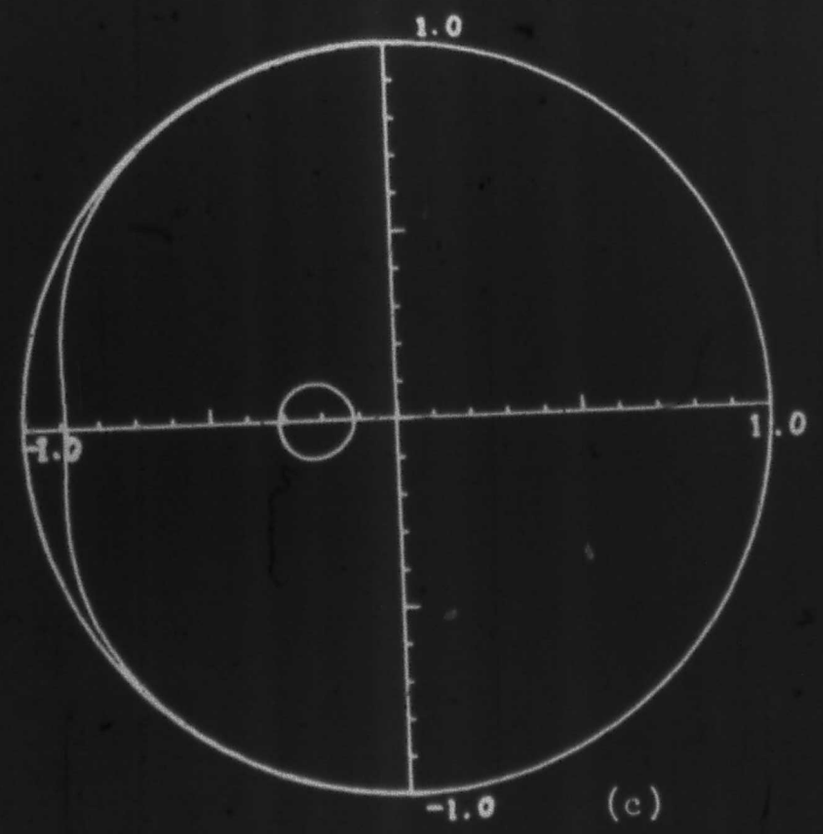
**In all figures of this chapter, the matrices are photographically reproduced from APL terminal print-outs, in order to prevent typographical errors.

$$\begin{bmatrix} -24390 & -53280 & 77670 & 0 & 0 & 0 & 6589 & 58528 & 41608 & -5752 & 1223 & -136 \\ -800 & -7830 & 0 & 8630 & 0 & 0 & 168 & 4053 & 10488 & 3528 & -192 & 15 \\ -1755 & -2560 & 0 & 0 & 4315 & 0 & 498 & 3600 & 3720 & 5280 & 1650 & -48 \\ 4608 & -12375 & 0 & 0 & 0 & 7767 & -1520 & -2150 & 13600 & 2600 & 11600 & 2330 \end{bmatrix}$$

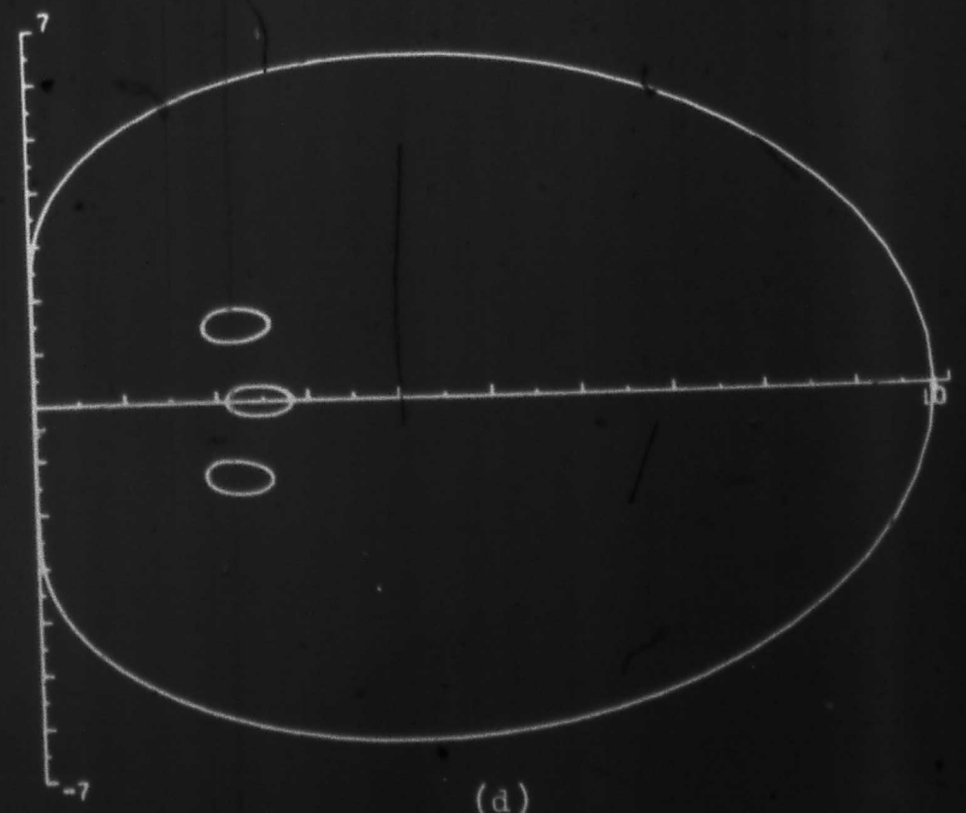
(a)

$$\begin{bmatrix} -4065 & -8880 & 12945 \\ 3330 & 720 & -21060 \\ -598 & 6296 & 14252 \\ -204 & -3768 & -4848 \\ 72 & 720 & 720 \end{bmatrix}$$

(b)



(c)



(d)

Fig. 4.1. The A-Stable (7,4,1,2) Method.

(a) The unique integer canonical $R \in \mathcal{R}_j(7,4,2)$.
(b) The characteristic polynomial P associated with (R,1).
(c) The zeta locus associated with P.
(d) The lambda locus associated with P.

a) the elements have no common (integral) factor,

b) the number of positive elements is not less than the number of negative elements, and,

c) if the number of positive elements is equal to the number of negative elements, then the first nonzero element is positive.

In condition c) "first" refers to the first nonzero element of the first nonzero row of the coefficient matrix. It is obvious that for every characteristic polynomial with integral coefficients, there exists a unique characteristic polynomial in integer canonical form equal to it, to within an integral trivial factor. The use of integer canonical form has the following practical advantages: (1) Since the magnitudes of the coefficients are as small as possible, subsequent computations will be most economical of computation time and storage space when infinite precision representation is used. (See Appendix E.) (2) Because integer canonical form is unique, machine comparison of two given polynomials is simplified. (3) Since there are as few digits and negative signs as possible, the display has maximum readibility.

A composite matrix R will be said to be in integer canonical form if R is in integer form, and in addition, each row of R satisfies conditions a), b), and c) above. For every composite matrix in integer form, there exists a unique integer canonical composite matrix to which it is row equivalent. In fact, the row equivalence transformation is a diagonal matrix whose diagonal elements are reciprocals of nonzero integers. The use of integer canonical composite matrices has all the advantages given above for characteristic polynomials. An additional advantage is the close relation between integer canonical composite matrices and canonical composite matrices. The composite matrix of Fig. 4.1a) is an example of an integer canonical composite matrix. In the following, all elements of $\mathcal{R}_i(p,n,m)$ and $\mathcal{P}(p,n,k,m)$ will be displayed in integer canonical form.

Figure 4.1c) presents a plot of the zeta locus T associated with the characteristic polynomial of Fig. 4.1b). The unit circle E is also shown on the plot. Since the (7,4,1,2) method is strongly A-stable, the zeta locus lies within the open unit disc U, except at unity. This fact can be observed in Fig. 4.1c), although with some difficulty, since the locus lies very close to the unit circle in the first and fourth quadrants. Such close following of the unit circle is characteristic of high-order methods, as explained in Section 4.8.

The lambda locus for the (7,4,1,2) method is given in Fig. 4.1d). Since the (7,4,1,2) method is strongly A-stable in the dual sense, the lambda locus lies in the open right half plane, except at the origin. The dual asymptotic property is easily observable: The lambda locus closely follows the imaginary axis near the origin.

§4.11 High-Order A-Stable Methods for $\mu = 1$

The previous section dealt with classes of methods in the unique case--that for which $\mu = 0$ in Theorem 4.24. This section discusses the next simplest case--the linear case--for which $\mu = 1$. In the latter case, (4-58) can be written as

$$p = 2m + n - 2 \quad . \tag{4-72}$$

The values of p in (4-72) for small values of m and n are tabulated in Table 4.2. As in the previous section, the irregular line of slope -2 is the left boundary for those classes satisfying the hypothesis of Conjecture 4.27. Thus, the most important classes to investigate are those directly to the left of this line.

It is evident from Tables 4.1 and 4.2 that the A-stability question for methods with m = 1 or n = 1 has been relatively well explored. Of greater interest in the present work are the cases with m and n both greater than unity, for which the full power of Chapters 2 and 3 must be employed. The simplest such case in Table 4.2 is the $\mathcal{R}_i(4,2,2)$ class, which has been investigated by [Sloate]. Sloate's approach to the problem of selecting a particular method of type (4,2,1,2) represents an important contribution. An improved and generalized version of his idea is explained below.

Consider a polynomial P in two variables, represented as in (1-40b). In order for P to satisfy the A-stability criterion it is necessary that all zeros of $P(0,\cdot)$ and $P(\infty,\cdot)$, that is, of ψ_0 and ψ_n, lie in the closed unit disc* $\overline{U}$. If P is the characteristic polynomial of a method in Table 4.2, $p \geq 1$. Therefore, Proposition 4.21 can be used to show that $P(0,1) = 0$; in other words $\psi_0(1) = 0$. Let $\hat{\psi}_0$ denote the (m-1)-th degree polynomial such that $\psi_0(\zeta) = (\zeta-1)\hat{\psi}_0(\zeta)$. Then the above necessary condition can be stated as follows: All zeros of $\hat{\psi}_0$ and ψ_n lie in $\overline{U}$. Note that $\hat{\psi}_0$ has m-1 zeros, while

*The reason is that same one referred to in the first footnote of Section 4.10.

n \ m	1	2	3	4
1	1WE	3CD	5CD	7CD
2	2W	4S	6C	8C
3	3W	5R	7C	9C
4	4W	6R	8Y	10C
5	5W	7R	9Y	11C

Notes:

C: According to Conjecture 4.27, these methods are not A-stable.

D: [Dahlquist] has proved that these methods are not A-stable.

E: This class contains the implicit Euler method, which is strongly A-stable.

R: [Rubin and Bickart] has shown that these classes contain strongly A-stable methods. See text and Figs. 4.3, 4.4, and 4.5.

S: [Sloate] has shown that this class contains a strongly A-stable (4,2,1,2) method. See Fig. 4.2.

W: [Watts] has derived two classes of (n,n,n,1) methods which are A-stable for all n. For one class, the diagonal Pade class, $T = E$ and $M = I$. The other class, the subdiagonal Pade class, is strongly A-stable.

Y: These classes contain stiffly stable methods. See text.

Table 4.2. Orders of Composite Matrices in Linear Case ($\mu = 1$).

ψ_n has m zeros (counting multiplicities). Therefore, the above represents 2m-1 "conditions" to be satisfied.

For the classes of Table 4.2, P can be represented in the form (4-69). Therefore, so can $\hat{\psi}_0$ and ψ_n. That is,

$$\hat{\psi}_0 = x\hat{\psi}_{00} + \sum_{i=1}^{n} x_i \hat{\psi}_{0i} \qquad \text{(a)}$$

(4-73)

and

$$\psi_n = x\psi_{n0} + \sum_{i=1}^{n} x_i \psi_{ni} \qquad \text{(b)}$$

where each $\hat{\psi}_{0i}$ and ψ_{ni} is associated with P_i in the natural way, and $x \neq 0$. In the above relations x is essentially a normalizing parameter, in the sense that a given sequence of n ratios x_i/x, i = 1, 2, ..., n determines the polynomials $\hat{\psi}_0$ and ψ_n to within a trivial factor. Recall from Section 4.9 that the free parameters correspond to a strongly regular composite matrix only when $x \neq 0$. The n indexed free parameters give n degrees of freedom in determining the zeros of $\hat{\psi}_0$ and ψ_n. Sloate's procedure is to try to choose these n free parameters so that all zeros of $\hat{\psi}_0$ and ψ_n are zero. Such an approach can be justified on esthetic grounds as follows: Firstly, this approach can provide a simple and natural way to satisfy the necessary conditions of the above paragraph for A-stability. Secondly, the smaller the zeros of $\hat{\psi}_0$ and ψ_n in magnitude, the better are the numerical properties of the composite multistep method for very small and very large step sizes, respectively.

In summary, it is to be assumed that $\hat{\psi}_0(\zeta) = \zeta^{m-1}$ and $\psi_n(\zeta) = \zeta^m$ (to within trivial factors). In such case (4-73) corresponds to a set of 2m-1 linear homogeneous algebraic equations in the n unknowns x_i (and the nonzero normalizing factor x). Generally speaking, solutions to these equations with $x \neq 0$ exist only when

$$n \geq 2m - 1 . \qquad (4\text{-}74)$$

Since all coefficients in (4-73) are integers, all solutions are representible in terms of integers, as usual. However, formal solutions are not always satisfactory. An example is the $P(5,3,1,2)$ class, for which $\psi_3(\zeta) = p_{30} + p_{31}\zeta + p_{32}\zeta^2$ by (1-39b). If x, x_1, x_2, and x_3 are chosen so that $p_{30} = p_{31} = 0$, then it happens that $p_{32} = 0$, in which case ψ_3 is the zero function.

For the P(4,2,1,2) class considered by Sloate, (4-74) is violated, so that a modification is required. Sloate used one degree of freedom to set $p_{00} = 0$, satisfying the condition on $\hat{\psi}_0$. The other degree of freedom was used to set $p_{21} = 0$, guaranteeing that the two zeros of ψ_2 are of equal magnitude. The fact that the conditions $p_{00} = p_{21} = 0$ happen to yield a strongly A-stable composite multistep method in this case should properly be viewed as a matter of good fortune. The (4,2,1,2) method derived by [Sloate] is displayed in Fig. 4.2 (with the composite matrix transformed into integer canonical form), together with the characteristic polynomial and associated loci.

The next class of interest in Table 4.2 is the $\mathcal{R}_i(5,3,2)$ class. In this case (4-74) holds with equality, so that a unique method may exist for each k satisfying the condition $p_{00} = p_{30} = p_{31} = 0$. The method for k = 2 shown in Fig. 4.3 is strongly A-stable. Figures 4.4 and 4.5 display strongly A-stable methods generated from $\mathcal{R}_i(6,4,2)$ and $\mathcal{R}_i(7,5,2)$ by the same procedure. In these last two cases the inequality (4-74) is strict. The "extra" degrees of freedom are explored arbitrarily to arrive at A-stable methods. Note that in part (b) of Figs. 4.2 through 4.5 the pattern $p_{00} = p_{n0} = p_{n1} = 0$ is observable, showing that all zeros of $\hat{\psi}_0$ and ψ_n are zero. It is fair to say that all the A-stable composite multistep methods presented in this section have been generated by an "educated guess" approach.

Many composite multistep methods generated by the above approach and by various other unsystematic parameter searches are not A-stable, but instead possess an essentially weaker property called stiff stability. A method will be said to be stiffly stable* if its stability region in the lambda sphere contains the extended negative real axis, that is, the negative real axis together with the point at infinity. Stiffly stable methods are useful in the numerical solution of stiff ordinary differential equations, although their performance may be inferior in certain cases. Many stiffly stable methods of types (4,2,2,2) and (5,3,3,2) have been found in the course of extensive parameter searches. However, no A-stable methods of these types have been found, and it is believed that none exist.

---

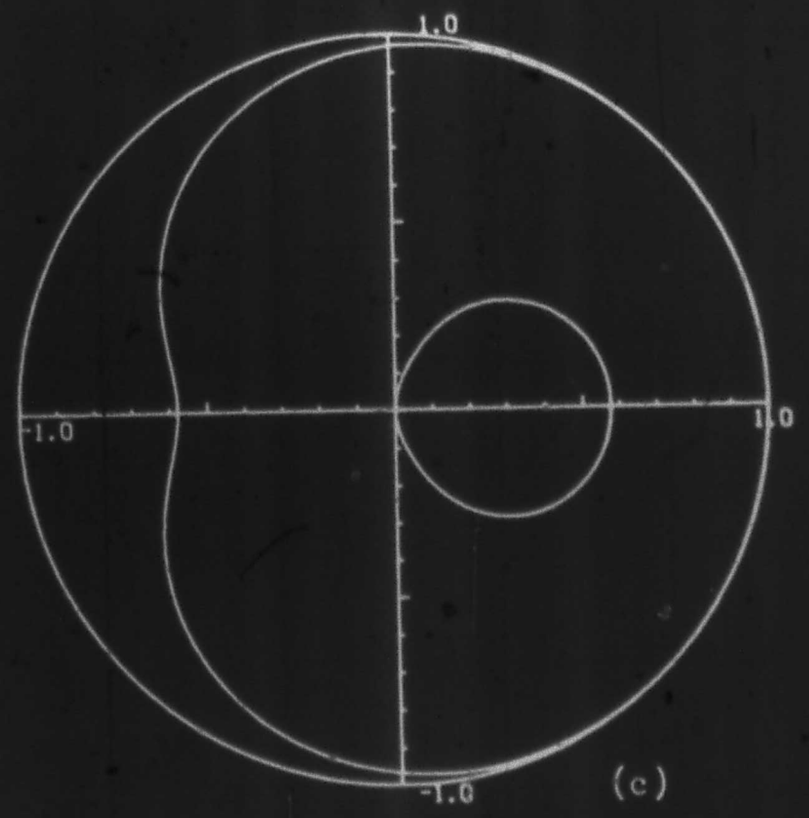
*Various conflicting definitions of stiff stability appear in the literature. The present definition is used only for the purpose of distinguishing methods which may be useful for stiff problems; it is not intended to be compared with alternate definitions. Generally, the particular "stiffly stable" methods mentioned in this section are stiffly stable by any reasonable definition of the term.

$$\begin{bmatrix} 0 & 24 & -24 & 0 & 1 & -13 & -13 & 1 \\ 56 & -72 & 0 & 16 & -21 & -39 & 33 & 3 \end{bmatrix}$$

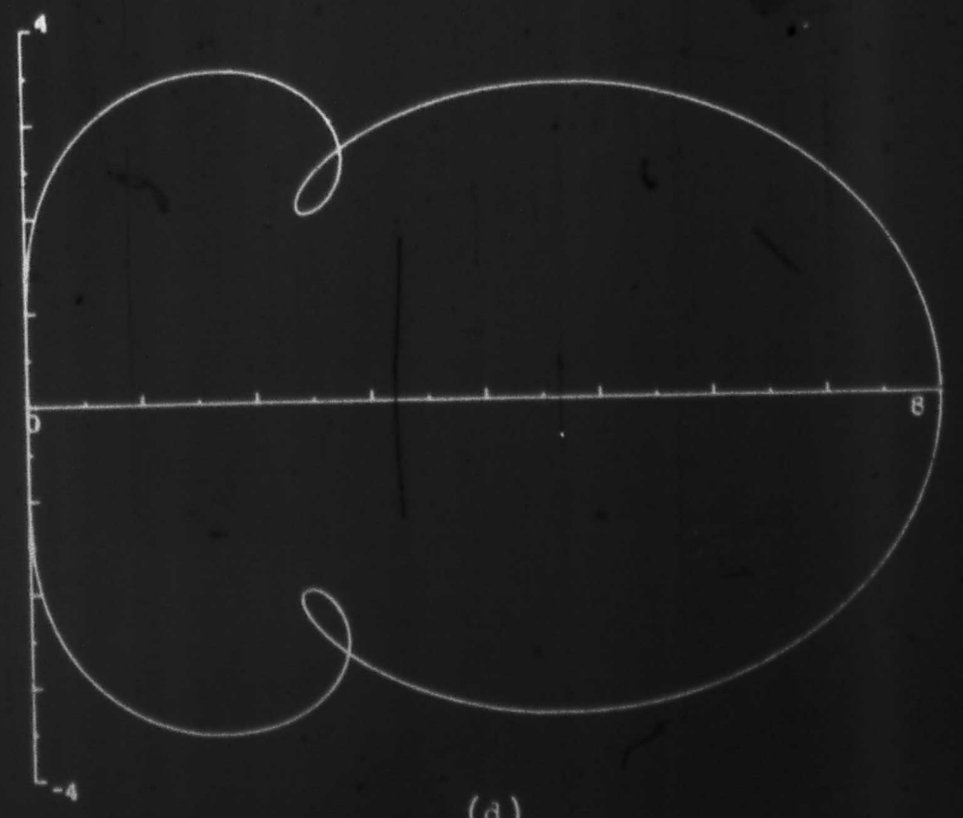
(a)

$$\begin{bmatrix} 0 & 48 & -48 \\ 5 & 8 & 35 \\ 3 & 0 & -9 \end{bmatrix}$$

(b)



(c)



(d)

Fig. 4.2. An A-Stable (4,2,1,2) Method.

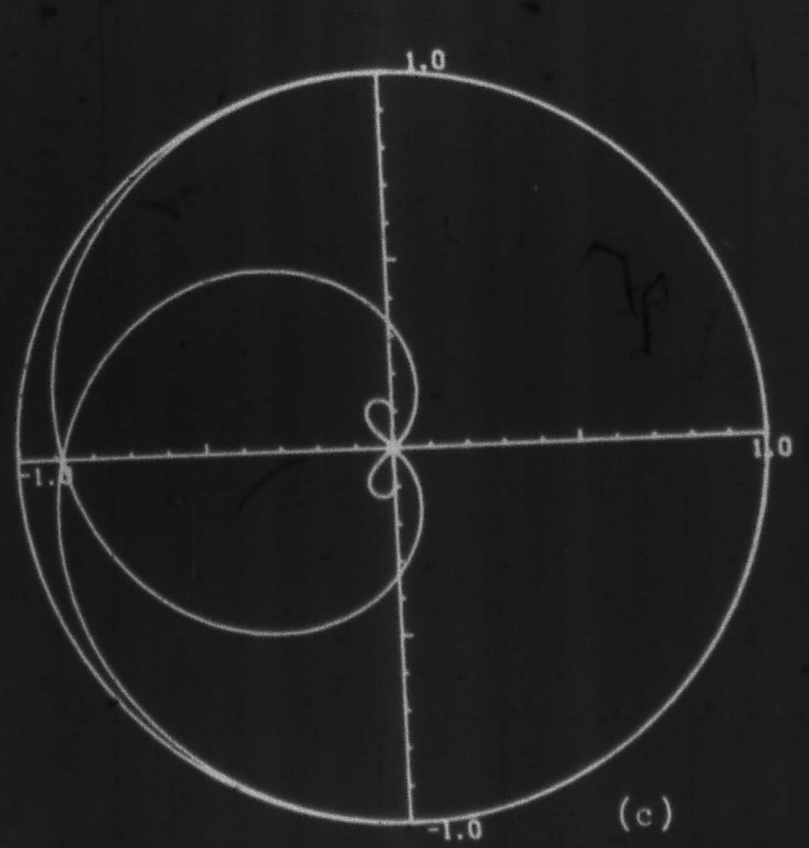
(a) A selected $R \in \mathcal{R}_i(4,2,2)$.
(b) The characteristic polynomial P associated with (R,1).
(c) The zeta locus associated with P.
(d) The lambda locus associated with P.

$$\begin{bmatrix} 53280 & -149040 & 95760 & 0 & 0 & -21101 & -1786 & 80184 & -17686 & 2869 \\ 2960 & -8280 & 0 & 5320 & 0 & -1091 & -646 & 7824 & 1574 & 19 \\ -3060 & 400 & 0 & 0 & 2660 & 967 & 4142 & 1272 & 3842 & 817 \end{bmatrix}$$

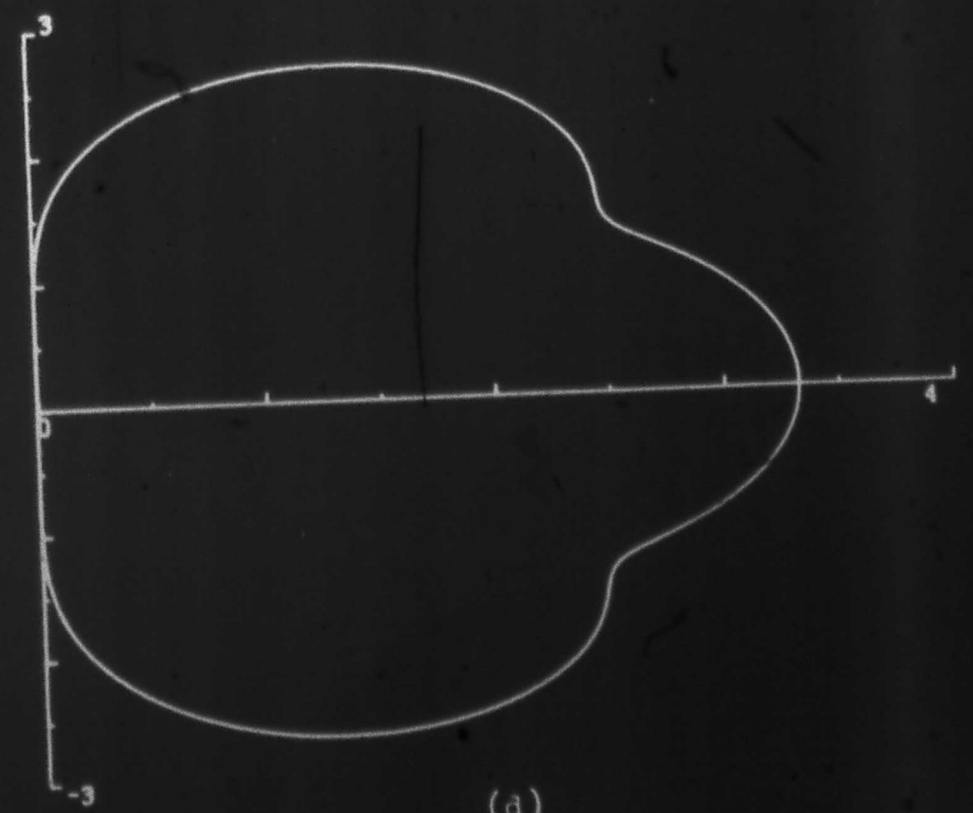
(a)

$$\begin{bmatrix} 0 & -15960 & 15960 \\ 1252 & -10184 & -22988 \\ 727 & 2300 & 13533 \\ 0 & 0 & -3420 \end{bmatrix}$$

(b)



(c)



(d)

Fig. 4.3. An A-Stable (5,3,2,2) Method.

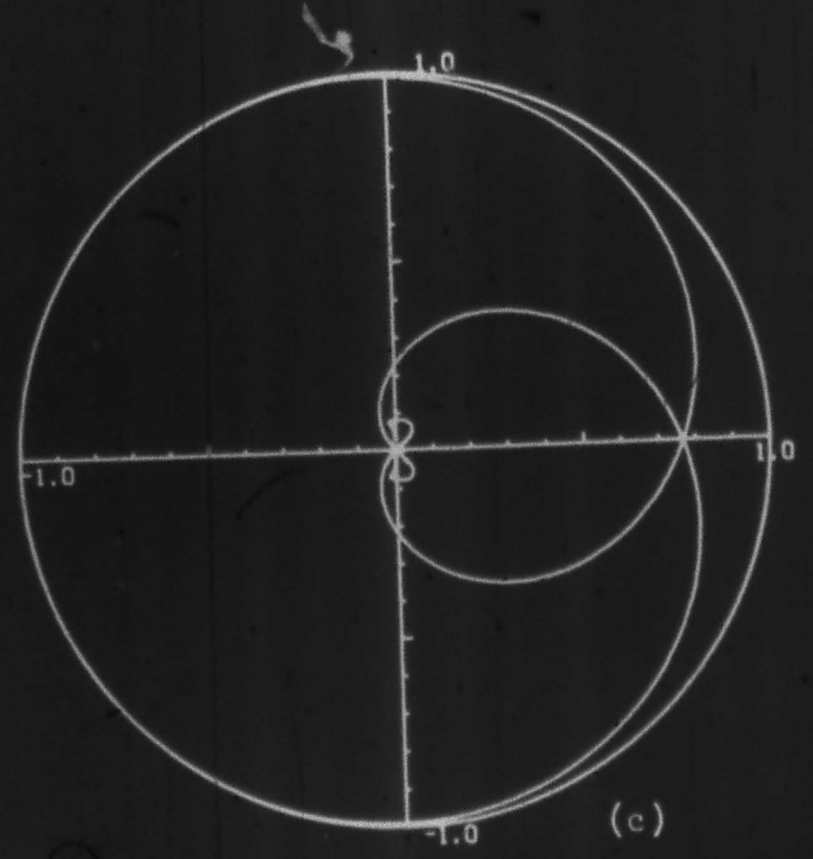
(a) A selected $R \in \mathcal{R}_q(5,3,2)$.
(b) The characteristic polynomial P associated with (R,2).
(c) The zeta locus associated with P.
(d) The lambda locus associated with P.

$$\begin{bmatrix} 720 & -7200 & 6480 & 0 & 0 & 0 & -359 & 2153 & 4998 & -1402 & 433 & -63 \\ -67680 & -714240 & 0 & 781920 & 0 & 0 & 13637 & 362461 & 952926 & 318046 & -16819 & 1269 \\ -4230 & -44640 & 0 & 0 & 48870 & 0 & 479 & 25267 & 50802 & 54562 & 20567 & -837 \\ 186912 & -304200 & 0 & 0 & 0 & 117288 & -61655 & -148735 & 270390 & -10 & 189265 & 32985 \end{bmatrix}$$

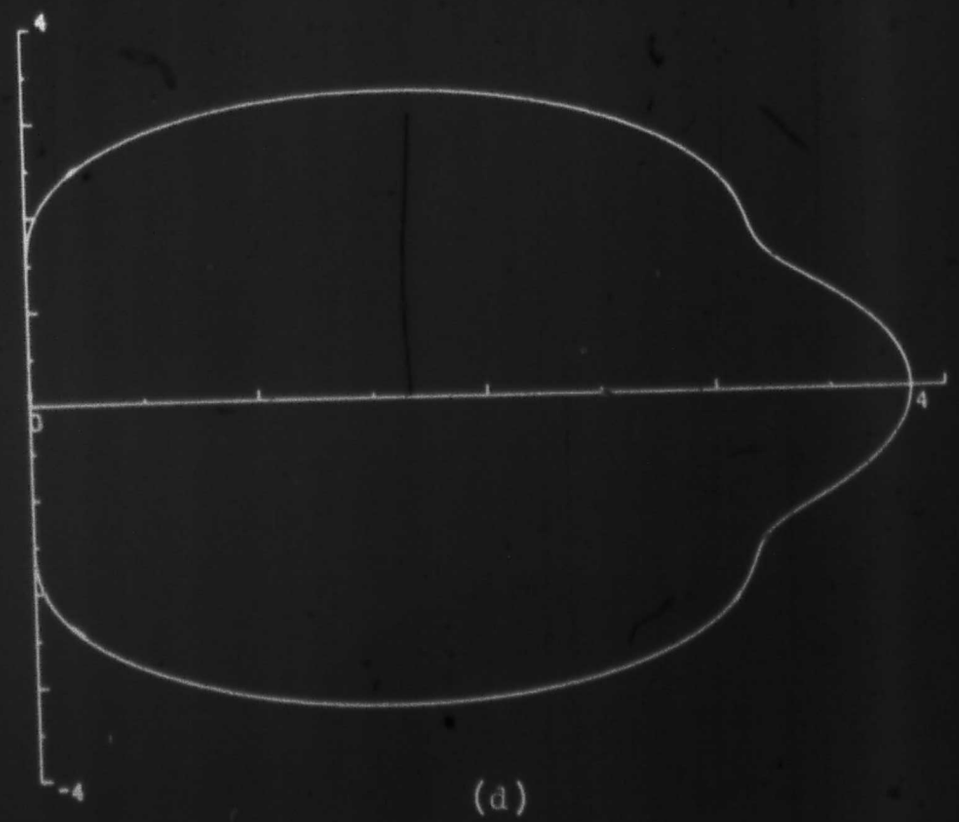
(a)

$$\begin{bmatrix} 0 & 390960 & -390960 \\ 21625 & 416200 & 735055 \\ 22720 & 193480 & -597170 \\ 6876 & -47982 & 255036 \\ 0 & 0 & -50400 \end{bmatrix}$$

(b)



(c)



(d)

Fig. 4.4. An A-Stable (6,4,3,2) Method.

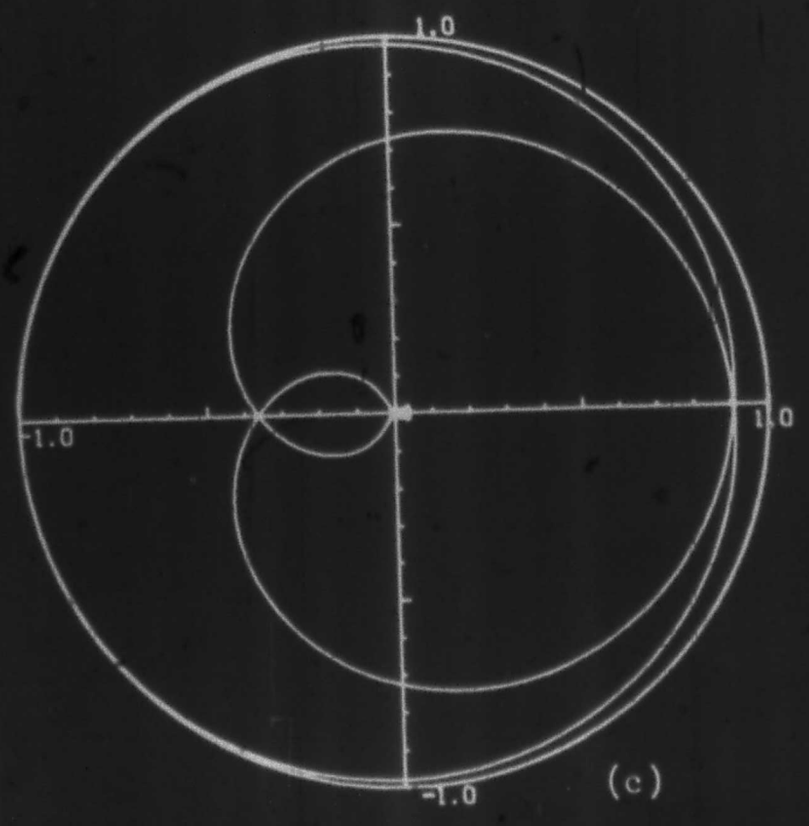
(a) A selected $R \in \mathcal{R}_s(6,4,2)$.
(b) The characteristic polynomial P associated with (R,3).
(c) The zeta locus associated with P.
(d) The lambda locus associated with P.

$$\begin{bmatrix}
60480 & -181440 & 120960 & 0 & 0 & 0 & 0 & -20813 & -14856 & 140439 & -70016 & 34809 & -10488 & 1405 \\
0 & -3780 & 0 & 3780 & 0 & 0 & 0 & -37 & 1398 & 4863 & 1328 & 33 & -30 & 5 \\
-83160 & -280000 & 0 & 0 & 363160 & 0 & 0 & 21547 & 233496 & 362667 & 404352 & 160437 & -10776 & 921 \\
-3592512 & -12096000 & 0 & 0 & 0 & 15688512 & 0 & 1000953 & 9545400 & 17560575 & 13251200 & 19119825 & 6052680 & -184075 \\
31752000 & -63129024 & 0 & 0 & 0 & 0 & 31377024 & -10734025 & -19553640 & 54547275 & 21814400 & 27147525 & 41821800 & 10089785
\end{bmatrix}$$

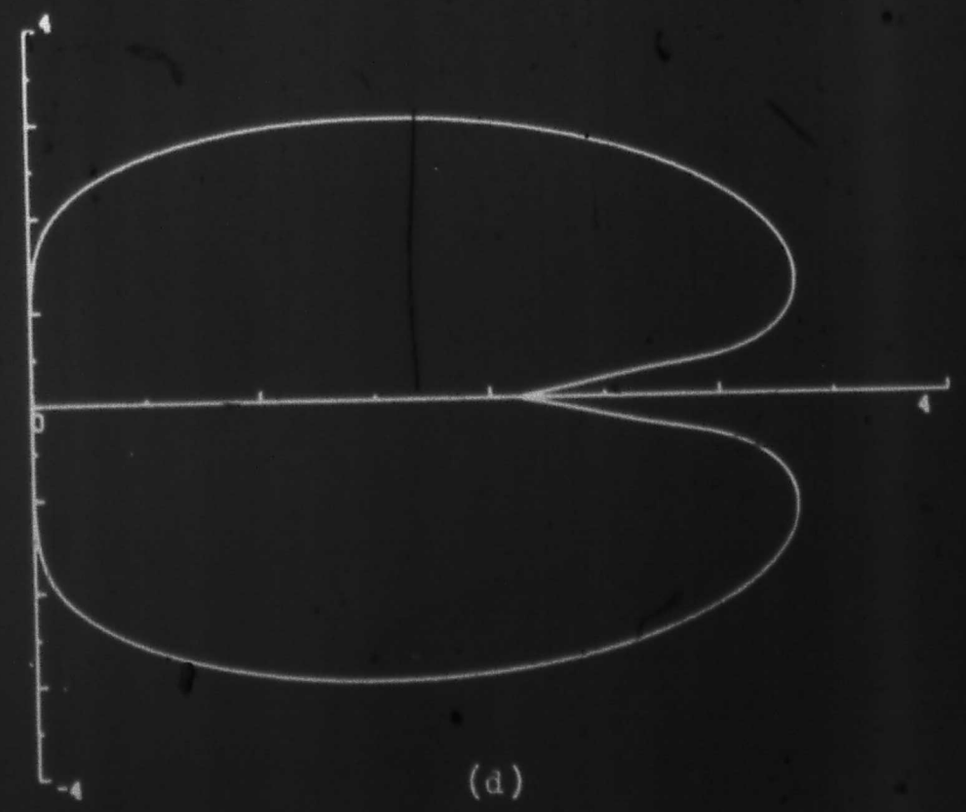
(a)

$$\begin{bmatrix}
0 & 117663840 & -117663840 \\
310443 & 157181514 & 313163403 \\
2298357 & 59678910 & -372078387 \\
2802700 & -10279952 & 252274420 \\
1094820 & -22209240 & -100549500 \\
0 & 0 & 19656000
\end{bmatrix}$$

(b)



(c)



(d)

Fig. 4.5. An A-Stable (7,5,5,2) Method.

(a) A selected $R \in \mathcal{R}_1(7,5,2)$.
(b) The characteristic polynomial P associated with (R,5).
(c) The zeta locus associated with P.
(d) The lambda locus associated with P.

In the remaining important classes of Table 4.2, the $R_i(8,4,3)$ and $R_i(9,5,3)$ classes, preliminary parameter searches produced some stiffly stable methods, but no A-stable methods. The $R_i(8,4,3)$ class is similar to the $R_i(4,2,2)$ class in that (4-74) is violated. Various (8,4,k,3) methods analogous to that of Fig. 4.2 turn out not to be A-stable. However, certain (8,4,3,3) methods are stiffly stable. The $R_i(9,5,3)$ class is similar to the $R_i(5,3,2)$ class in that (4-74) holds with equality. The unique (9,5,4,3) and (9,5,3,3) methods satisfying (4-73) under Sloate's procedure turn out to be stiffly stable but not A-stable.

§4.12 High-Order Methods and the Algebraic Characterization

In this section two applications of the theory of Chapter 3 to methods from Table 2.2 will briefly be examined. In the first, the algorithm of Appendix E is used to show that the (5,3,2,2) method of Fig. 4.3 is strongly A-stable. The elements of the last column in Fig. 4.3b) are the coefficients of the polynomial δ of poles of the characteristic polynomial. All poles are in the open right half plane for these coefficients, according to the Hurwitz test. The transformed polynomial P' turns out to be of degree 2 in z. The auxiliary polynomials are as follows (after removing a positive integral factor):

$$\nabla_2(\omega) = 38580640 + 23002112\omega + 25571970\omega^2 + 6579225\omega^3 \quad \text{(a)}$$

$$\nabla_1(\omega) = 37240 + 1462\omega + 855\omega^2 \quad \text{(b)} \qquad (4\text{-}75)$$

Evidently $\nabla_2(0)$ and $\nabla_1(0)$ are both positive. Also, ∇_2 can be seen by inspection to have no positive real zeros (see Appendix E). Hence, the method of Fig. 4.3 is strongly A-stable by Theorem 3.19.

According to the theory of Section 3.5, the polynomial ∇_2 of (4-75a) could have been expected to be of degree 5, rather than only of degree 3. Numerous computations of auxiliary polynomials for high-order composite multistep methods have consistently shown the degree of $\nabla_{m'}$ to be strictly lower than the upper bound of nm-1. These observations have led to the following conjecture: For a composite multistep method of order p, the integer $k_{m'}$ of (3-19) is not less than (p+1)/2. Note that Proposition 3.11 proves the conjecture for the case p = 0. In general, the conjecture can be made plausible from the following observations: The value of $k_{m'}$ describes the "flatness"

of ∇''_m, in a neighborhood of the origin. The behavior of ∇''_m, near the origin reflects the behavior of the characteristic polynomial near (0,1); the latter behavior in turn reflects the order p of the composite multistep method, as shown in Section 4.8. The conjecture can be given an interesting qualitative interpretation, namely, that for given m and n, the degree of the largest auxiliary polynomial ∇_m, varies inversely with the order of the method; thus, high order methods are simpler to analyze than lower order methods with respect to the Sturm test.

The second application of the theory of Chapter 3 arises from an analysis of the class P(6,4,4,2) of Table 4.2. Applying Sloate's procedure to this case as usual, there results one "extra" degree of freedom. That is, one free parameter can be specified arbitrarily after setting all zeros of $\hat{\psi}_0$ and ψ_4 to zero. Experimental variation of this parameter has not produced an A-stable method. However, a careful analysis of the auxiliary polynomial ∇_2 over many such variations has shown it to have two positive real zeros, one of which is 20/9, with the other variously smaller or larger than 20/9, depending upon the value of the free parameter. The free parameter has been adjusted to make the second zero as close as possible to 20/9, so as to make the method "almost" A-stable*.

The closest that has been achieved is the method given in Fig. 4.6. The zeta locus lies slightly outside the closed unit disc in a neighborhood of the arrows in Fig. 4.6c). These parts of the zeta locus correspond to small unstable intervals on the imaginary axis in the lambda sphere indicated by the arrows in Fig. 4.6d). The approximate size and location of these intervals has been computed using Proposition 3.13, with the following result: The unstable interval on the positive imaginary axis is of the form $(\omega-\epsilon,\omega)$, where $\omega = (20/9)^{1/2} \approx 1.49$, and $0 < \epsilon < 6.33 \times 10^{-11}$. This interval of instability is so small that the (6,4,4,2) method of Fig. 4.6 deserves to be called almost A-stable, by any reasonable definition of the term.

§4.13 Review of Previous Work

About half of the material in this chapter is new, the other half being

*Presumably, there exists a unique value of the free parameter for which the second zero is exactly 20/9. For this value ∇_2 has a zero of multiplicity 2 at 20/9. The resulting method can be shown to be A-stable by Corollary 3.15, but it is not strongly A-stable.

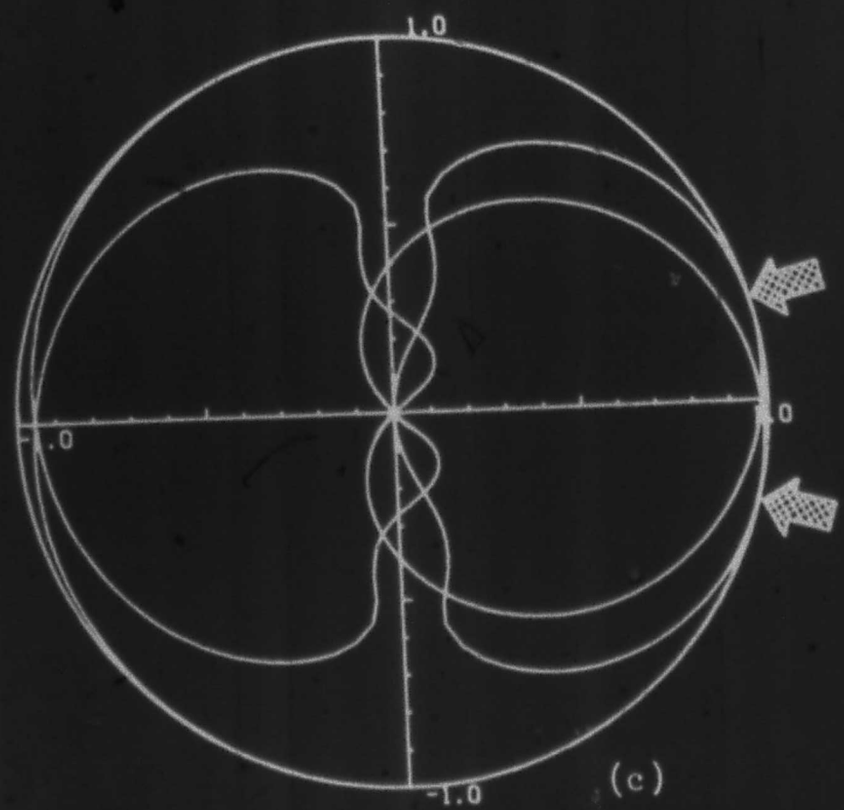
The problem is that even if such a free parameter value could be found, it cannot be expected to be a rational number. If it is irrational, then the associated composite matrix is not in $\mathcal{R}_i(6,4,2)$.

$$\begin{bmatrix}
-3529798560 & 3964033440 & -434234880 & 0 & 0 & 0 & 1172485179 & 3305843699 & -2264282846 & 1259302434 & -447285581 & 69500795 \\
-17279332000 & 18389043360 & 0 & -1109711360 & 0 & 0 & 5712109779 & 16704113739 & -10981264206 & 5364552114 & -2063589621 & 323987475 \\
-22318068240 & 23601172000 & 0 & 0 & -1283103760 & 0 & 7385920983 & 21563247183 & -13763304822 & 6074978058 & -2334606417 & 442521975 \\
-80345045664 & 84964219200 & 0 & 0 & 0 & -4619173536 & 26502706035 & 78182632235 & -51094037390 & 24429713010 & -16222055765 & 69393395
\end{bmatrix}$$

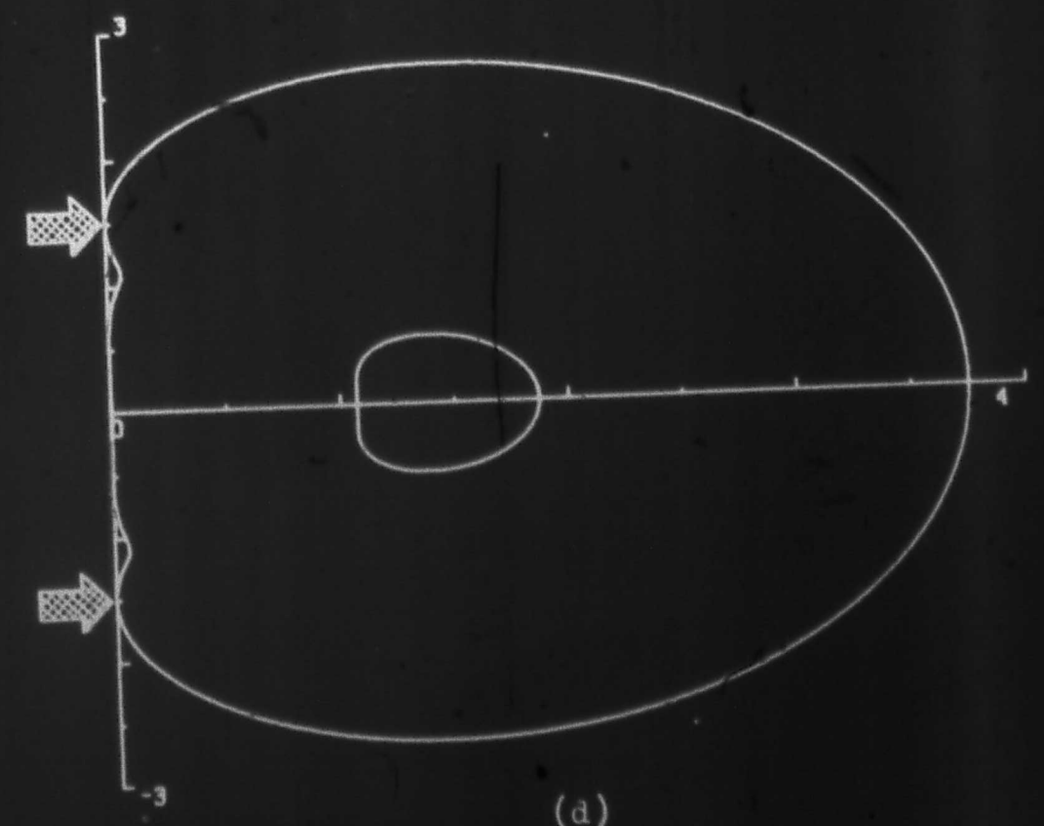
(a)

$$\begin{bmatrix}
0 & 3849311280 & -3849311280 \\
1332976735 & 2954643130 & 11109625255 \\
2249912310 & 3288163940 & -13850180090 \\
1184738481 & -4626099852 & 9360235371 \\
0 & 0 & -3124775880
\end{bmatrix}$$

(b)



(c)



(d)

Fig. 4.6. An Almost A-Stable (6,4,4,2) Method.

(a) A selected $R \in \mathcal{R}_4(6,4,2)$.
(b) The characteristic polynomial P associated with (R,4).
(c) The zeta locus associated with P.
(d) The lambda locus associated with P.

principally a restructuring of material appearing in [Sloate and Bickart] and elsewhere. The importance of strong regularity has been recognized in all previous work on composite one-step and multistep methods, as noted in Section 4.3. Definition 4.9 and Proposition 4.10 on truncation error appear (in a somewhat different form) in [Sloate and Bickart]. Definition 4.16 on the order of a composite matrix is technically different from, but equivalent to, that of [Sloate and Bickart]. The general discussion of order in the beginning of Section 4.7, and of the order relations in Proposition 4.23, follows standard treatments [Henrici].

The distinction between intrinsic and extrinsic properties of composite multistep methods is discussed for the first time. The important concepts of equivalence and canonical form for composite matrices are new. Other new developments include the significance of discretization error, its order, and the discretization error constant, and their relationship to truncation error; the relation of error and order in the linear autonomous case to the general case, including the expansion (4-43) of $\tilde{\epsilon}_D$; the implications of equivalence, weak equivalence, and order with respect to the characteristic polynomial. However, a result in the spirit of Proposition 4.20 was derived by indirect means in [Sloate and Bickart]. The concise notation used in the exposition of truncation error, the order relations, and elsewhere is considered to be an improvement over past attempts.

The results concerning free parameter representations of classes of composite matrices and characteristic polynomials are new. Their rigorous formulation is due to the writer, who was the first to examine the significance of the inequality (4-62). The first representation for the characteristic polynomial as a function of free parameters in the linear case (4-69) was formulated by Burgess*, in unpublished research. The general case, Theorem 4.26, was developed jointly by Tendler** and the writer. The results of preliminary explorations of composite multistep methods in Sections 4.10 - 4.12 are all new, except where otherwise noted.

---

*D. A. Burgess, GTE Sylvania, Bedford, Mass.
**See footnote at end of Chapter 1.

SUMMARY AND CONCLUDING REMARKS

The work which has been presented is a study of composite multistep methods and the A-stability property. A unique feature is the introduction of the A-stability criterion for polynomials in two variables (Definition 1.6), by means of which the general problem of determining whether a given method is A-stable can be separated into two parts. Because of this feature, the general A-stability characterizations of Chapters 2 and 3 can be applied not only to composite multistep methods, but also to most other methods for the numerical solution of ordinary differential equations*. With respect to a given class of such methods, the first part of the A-stability problem is to formulate the characteristic polynomial, and to show that a given method is A-stable if and only if its characteristic polynomial satisfies the A-stability criterion. This part of the A-stability problem is solved for composite multistep methods in Chapter 1. The second part of the A-stability problem is to characterize the A-stability criterion for polynomials in two variables. This part of the problem is solved in Chapters 2 and 3. The result of these developments is a practical and completely general test for determining whether a given composite multistep method is A-stable. This test extends directly to any class of methods whose stability properties can be determined through a characteristic polynomial; it is only necessary to derive the form of the characteristic polynomial for the class of interest.

The attributes of a useful numerical method include not only good stability properties, but also high order of accuracy. In Chapter 4, the fundamentals of error and order for composite multistep methods are developed, and simple parametric representations are derived for classes of high-order composite multistep methods and their characteristic polynomials. These representations are useful in the determination of high-order A-stable composite multistep methods. Chapter 4 concludes with a summary of the present state of knowledge concerning high-order A-stable composite multistep methods, including the presentation of certain new methods.

---

*This wide applicability of the A-stability characterizations was first appreciated by Hans Stetter, Tech. U. of Vienna, A-1040 Vienna, Austria, in an oral communication following the presentation of [Rubin and Bickart].

Chapters 1 and 4 of this work contain the first careful and comprehensive study of composite multistep methods. In Chapter 1 it is shown that when a composite multistep method is applied to a linear autonomous differential equation, the generated approximating sequences satisfy the linear matrix difference equation (1-32). Three successive formulations for the characteristic polynomial of a composite multistep method are derived from this difference equation. For each formulation it is shown that a composite multistep method whose poles lie in the closed right half complex plane is A-stable if and only if its characteristic polynomial satisfies the A-stability criterion. The third formulation, the subject of Corollary 1.15, is indicated to be the most suitable for practical purposes, as well as for some theoretical purposes. The poles of a composite multistep method and of its characteristic polynomial are shown to be related to A-stability and to existence and uniqueness of the approximating sequences in the linear autonomous case.

Chapter 2 consists of a rigorous derivation of locus techniques for characterizing the A-stability criterion for polynomials in two variables. The main result, Theorem 2.4, states that in order for a polynomial in two variables to satisfy the A-stability criterion, it is necessary and, together with certain side conditions, sufficient that its zeta locus lie inside the closed unit disc. One of the side conditions is that the poles lie in the closed right half plane. The dual result, Theorem 2.7, gives a similar characterization in terms of the lambda locus. Not only are these results of interest in themselves, but they are also needed as the starting point for the work of Chapter 3.

In Chapter 3 algebraic necessary and sufficient conditions are derived for a polynomial in two variables to satisfy the A-stability criterion. The conditions are all reducible to the addition and multiplication of polynomials with integral coefficients, and the examination of the signs of integers. The classical tests of Hurwitz and Sturm, and other operations formulated with determinants, play important roles in the computations. The main results of Chapter 3 differ significantly from each other in the complexity of their conditions, as well as in the generality of their application. All results appear in both primal and dual forms. Theorem 3.16 is a completely general algebraic characterization of the A-stability criterion for polynomials in two variables. Its dual is Theorem 3.26. The relatively complicated conditions

in these theorems are needed to account for the many kinds of pathological cases which can arise. A considerable simplification of the algebraic conditions, with some loss in generality, is obtained with Theorem 3.14 and its dual, Theorem 3.25. These two results still apply, however, to almost all cases of importance. The simplest algebraic characterization of A-stability is presented in Theorem 3.19 and its dual, Theorem 3.28. These results characterize strong A-stability and strong A-stability in the dual sense, respectively. Their main virtue is the simplicity of their characterizations, and the fact that they can be applied to most cases of interest.

The simplest form of the algebraic characterization, Theorem 3.19, has been implemented as a practical A-stability test in APL/360 [Rubin, 1973], as discussed in Appendix E. The test employs the technique of infinite precision integer arithmetic, and is therefore free from roundoff error. This technique allows the program to determine with absolute certainty whether a given characteristic polynomial satisfies the A-stability criterion. The implementation of the other algebraic A-stability characterizations of Chapter 3 is a straightforward application of the considerations of Appendix E.

Characteristic polynomials are available to describe the stability properties not only for composite multistep methods, but also for second derivative methods [Liniger and Willoughby], [Enright], higher derivative methods [Odeh and Liniger, 1971], Runge-Kutta methods [Ehle, 1968], off-step methods [Beaudet], and others. In other words, for each of these classes, a result analogous to Corollary 1.15 holds, so that the algebraic A-stability characterizations are immediately applicable. For some other classes, such as the averaging multistep methods of [Odeh and Liniger, 1972], no formulation for the characteristic polynomial is presently known. However, it seems likely that one can be derived, thereby allowing the A-stability theory of Chapter 3 to be applied. Of course, the present work also applies directly to all special cases of composite multistep methods, including multistep methods [Dahlquist], [Henrici], composite one-step methods [Bickart, et al], and cyclic composite multistep methods [Tendler].

In Chapter 4 the concepts of equivalence and canonical form for composite multistep methods are defined, and it is shown that if only such intrinsic properties as A-stability, order, and discretization error constant are of interest, then attention can be restricted to the subclass of canonical composite multistep methods without loss in generality. With regard to the

development of error and order, a noteworthy result is the fact that the exact order p of a composite multistep method, defined in terms of the truncation error ϵ_T, is equal to the minimum of the exact orders of its discretization errors associated with the future points. Furthermore, the exact order of the discretization error in the linear case $\tilde{\epsilon}_D$ does not exceed p. The most important results of Chapter 4 are the free parameter representations for high-order composite matrices, Theorem 4.25, and for their corresponding characteristic polynomials, Theorem 4.26. The representations are practical to obtain and to use, since only integer arithmetic is needed. The significance of these representations rests upon the concepts of equivalence, canonical form, error, order, and integer form developed in Chapter 4. The examples of new high-order A-stable composite multistep methods illustrate the use of characteristic polynomials, zeta and lambda loci, free parameter representations, and the algebraic A-stability characterization. They also show that by considering methods with many past and future points, the goal of efficient high-order A-stable methods is brought closer.

In spite of the length of Chapter 4, certain important topics have been omitted, principally convergence, stability (in the classical sense [Henrici]), and global error. However, the basic results for composite multistep methods in these areas can presumably be obtained by application of the recently-published general theories of [Chartres and Stepleman] and [Butcher].

The Dissertation is now concluded with a brief discussion of some paths to further knowledge indicated by the present research.

During the course of the present research, the lambda loci corresponding to many composite multistep methods have been examined. For a given composite matrix R, the stability region in the lambda sphere for the composite multistep method (R,k) has seemed empirically to vary in size inversely with k. This observation has led to the conjecture that if $k > 1$ and (R,k) is A-stable, then so is (R,k-1). An important consequence of this conjecture is the following: Cyclic composite multistep methods of order greater than 2 are not A-stable*. Since cyclic composite multistep methods are almost as efficient

*This consequence is justified as follows: Let (R,k) be an A-stable cyclic composite multistep method of order p. By induction with respect to the conjecture, (R,1) is A-stable. Since (R,k) is cyclic, (R,1) is weakly equivalent to a multistep method whose coefficients are contained in the first row of R. This multistep method is A-stable and of order p, since A-stability and order are intrinsic properties. It follows from the well-known result of [Dahlquist] that $p \leq 2$.

in implementation as multistep methods, it is important to determine whether this conjecture is correct.

Another direction for future research is the extension of the algebraic theory to characterize stiff stability. For example, a simple transformation allows the stability of a polynomial in two variables to be tested by the algebraic characterization with respect to any given translation of the left half (lambda) plane. Such translations are closely related to the definition of stiff stability proposed by [Gear]. For other definitions of stiff stability, the extension of the present results may be more difficult.

The stability of A-stable or stiffly stable methods under changing step size is a topic of current interest. This problem involves the study of polynomials in two variables [Brayton and Conley], and therefore may be approached from the point of view of the algebraic theory of Chapter 3.

Most past efforts (including that of Sections 4.10 through 4.12) to find "good" methods for the solution of stiff ordinary differential equations seem to have relied upon "educated guess" approaches to achieving high-order A-stable or stiffly stable methods. The algebraic characterization of Chapter 3, together with parameterizations such as those of Section 4.9, makes possible a systematic search for high-order A-stable methods. That is, the algebraic characterization can be applied to a characteristic polynomial given in terms of free parameters. The resulting A-stability conditions are essentially sets of inequalities involving polynomials in the free parameters. Solutions to these inequalities correspond to high-order A-stable methods. The computation of the coefficients of the polynomial inequalities is a completely straightforward, although admittedly nontrivial, task. This systematic approach is clearly not limited to composite multistep methods, but can be applied to any class of methods, once a characteristic polynomial in terms of free parameters has been derived.

Appendix A

PROOF OF THEOREM 1.12

This appendix presents a proof of Theorem 1.12 by use of two identities stated below as lemmas. First define

$$S = \begin{bmatrix} I_{(M-1)k} & 0 \\ 0 & E_M \end{bmatrix} \tag{A-1}$$

and

$$\hat{T} = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I \\ -E_0 & -E_1 & -E_2 & \cdots & -E_{M-1} \end{bmatrix} . \tag{A-2}$$

A.1. Lemma: The following identity holds for all $\zeta \in C$:

$$\det(\zeta S - \hat{T}) = \det \sum_{i=0}^{M} E_i \zeta^i \tag{A-3}$$

Proof: The lemma holds trivially when M = 1. Therefore assume that M = 2, 3, 4, ..., and that the lemma holds for M - 1. If $\zeta \neq 0$ we can write

$$\det(\zeta S - \hat{T}) = \det\left\{ \begin{bmatrix} I & 0 & \cdots & 0 \\ 0 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\zeta^{-1}E_0 & 0 & \cdots & I \end{bmatrix} (\zeta S - \hat{T}) \right\}$$

$$= \det \begin{bmatrix} \zeta I & -I & 0 & \cdots & 0 \\ 0 & \zeta I & -I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -I \\ 0 & \zeta^{-1}E_0 + E_1 & E_2 & \cdots & E_{M-1} + \zeta E_M \end{bmatrix} = \det(\zeta I) \det \sum_{i=-1}^{M-1} E_i \zeta^i = \det \sum_{i=0}^{M} E_i \zeta^i ,$$

where the third equality uses the lemma for M - 1. Therefore, by induction, (A-3) is proved for $\zeta \neq 0$. Since both sides of (A-3) are continuous in ζ (they are both polynomials in ζ or the zero function), (A-3) holds for $\zeta = 0$ as well. □

Note that if E_M is nonsingular, so is S. In such case define

$$T = S^{-1}\hat{T} \quad . \tag{A-4}$$

A.2. Lemma: For each i = 0, 1, ..., M let

$$G_i = \begin{bmatrix} E_0 & E_1 & \cdots & E_{i-1} \\ 0 & E_0 & \cdots & E_{i-2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_0 \end{bmatrix} , \tag{a}$$

$$H_i = \begin{bmatrix} E_M & 0 & \cdots & 0 \\ E_{M-1} & E_M & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ E_{M-i+1} & E_{M-i+2} & \cdots & E_M \end{bmatrix} , \tag{b}$$

and

$$A_i = \begin{bmatrix} 0 & -H_{M-i} \\ G_i & 0 \end{bmatrix} . \tag{A-5, A-6}$$

If E_M is nonsingular, then

$$A_i = -HT^i \quad , \qquad i = 0,1,\ldots,M . \tag{A-7}$$

In particular

$$G = -HT^M \quad . \tag{A-8}$$

Proof: From (A-5), (A-6), and (1-28b)

$$A_0 = \begin{bmatrix} 0 & -H_M \\ G_0 & 0 \end{bmatrix} = -H_M = -H = -HT^0 \quad ,$$

proving (A-7) for the case i = 0. Now suppose (A-7) holds for some i = 0, 1, ..., M-1. Let

$$D_i = [E_1 \quad E_2 \cdots E_i]$$

and

$$F_i = [E_{i+1} \quad E_{i+2} \cdots F_{M-1}] \ .$$

This notation can be used to write

$$G_{i+1} = \begin{bmatrix} E_0 & D_i \\ 0 & G_i \end{bmatrix}, \qquad H_{M-i} = \begin{bmatrix} H_{M-i-1} & 0 \\ F_i & E_M \end{bmatrix},$$

and

$$T = \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ -E_M^{-1}E_0 & -E_M^{-1}D_i & -E_M^{-1}F_i \end{bmatrix}.$$

These relations show that

$$-HT^{i+1} = -HT^iT = A_iT = \begin{bmatrix} 0 & -H_{M-i} \\ G_i & 0 \end{bmatrix} T$$

$$= \begin{bmatrix} 0 & -H_{M-i-1} & 0 \\ 0 & -F_i & -E_M \\ G_i & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & I & 0 \\ 0 & 0 & I \\ -E_M^{-1}E_0 & -E_M^{-1}D_i & -E_M^{-1}F_i \end{bmatrix} = \begin{bmatrix} 0 & 0 & -H_{M-i-1} \\ E_0 & D_i & 0 \\ 0 & G_i & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -H_{M-i-1} \\ G_{i+1} & 0 \end{bmatrix} = A_{i+1}$$

which proves (A-7) by induction. Now (A-8) is simply (A-7) for the case i = M, since $A_M = G_M = G$ by (1-28a). □

The proof of Theorem 1.12 can now be completed. By (1-26) and Lemma A.1

$$\tilde{P}(\lambda,\zeta) = \det(\zeta S - \hat{T}) = (\det S)\det(\zeta I - T) \ , \qquad \text{(A-9)}$$

where the second equality uses (A-4). On the other hand, from (1-36) and (1-37)

$$\hat{P}(\lambda,\zeta) = \det(G + \zeta H) = (\det H)\det(\zeta I - T^M) \ , \qquad \text{(A-10)}$$

where the second equality uses (A-8) of Lemma A.2. Both the above relations make sense for $\lambda \notin \Lambda$, since this is the condition under which E_M is non-singular, by (1-21). In such case (A-9) and (A-10) express the fact that the zeros of $\tilde{P}(\lambda,\cdot)$ are the eigenvalues of T, while those of $\hat{P}(\lambda,\cdot)$ are the eigenvalues of T^M. It is an elementary fact of matrix theory that the eigenvalues of T^M are the M-th power of the eigenvalues of T, and conversely, in the same sense as stated in Theorem 1.12. Thus the theorem is proved.

Appendix B

PROOF OF THEOREM 1.14

This appendix presents a proof of Theorem 1.14 by use of two identities stated below as lemmas. Let Z be the matrix valued function defined by

$$Z(\zeta) = [I_k \quad \zeta I_k \quad \cdots \quad \zeta^M I_k]^T \; . \tag{B-1}$$

From (1-49) and (1-50) we can write

$$W(\lambda,\zeta) = [V(\lambda) \quad 0_{nN}]Z(\zeta) \; . \tag{B-2}$$

Let

$$\tilde{W}(\lambda,\zeta) = [0_{nN} \quad V(\lambda)]Z(\zeta) \; . \tag{B-3}$$

B.1. Lemma: For every $(\lambda,\zeta) \in \mathcal{C}^2$ and every $k \times n$ matrix S

$$\det S\tilde{W}(\lambda,\zeta) = (-1)^{(k-1)N}\zeta^N \det SW(\lambda,\zeta) \; . \tag{B-4}$$

Proof: By (B-1) Z can be partitioned as follows:

$$Z(\zeta) = \begin{bmatrix} Z_1(\zeta) & Z_2(\zeta) \\ 0 & \zeta^M I_N \end{bmatrix} \tag{B-5}$$

Using (B-1) and the definition of Z_1 and Z_2 in (B-5), it can be shown by a tedious but straightforward induction argument on M that

$$Z(\zeta) = \begin{bmatrix} I_N & 0 \\ \zeta Z_2(\zeta) & Z_1(\zeta) \end{bmatrix} . \tag{B-6}$$

Now

$$\begin{aligned}
&\det S\tilde{W}(\lambda,\zeta) \\
&\quad = \det S[0_{nN} \quad V(\lambda)]Z(\zeta) && \text{by (B-3)} \\
&\quad = \det[0_{kN} \quad SV(\lambda)]\begin{bmatrix} I_N & 0 \\ \zeta Z_2(\zeta) & Z_1(\zeta) \end{bmatrix} && \text{by (B-6)} \\
&\quad = \det[\zeta SV(\lambda)Z_2(\zeta) \quad SV(\lambda)Z_1(\zeta)] \\
&\quad = (-1)^{(k-1)N}\zeta^N \det[SV(\lambda)Z_1(\zeta) \quad SV(\lambda)Z_2(\zeta)] \\
&\quad = (-1)^{(k-1)N}\zeta^N \det S[V(\lambda) \quad 0_{nN}]\begin{bmatrix} Z_1(\zeta) & Z_2(\zeta) \\ 0 & \zeta^M I_N \end{bmatrix} \\
&\quad = (-1)^{(k-1)N}\zeta^N \det SW(\lambda,\zeta) && \text{by (B-5) and (B-2),}
\end{aligned}$$

thus proving the lemma. The fourth equality above results from extracting a factor of ζ from each of the first N columns of the matrix in brackets, and permuting its columns as shown. This permutation can be shown to be even [odd] whenever (k-1)N is even [odd]. □

B.2. Lemma: Suppose D is an n × n matrix, Δ = adj D, and δ = det D. If S is any n × k matrix, where $1 \le k \le n$, then

$$\det J_{kn}\Delta S = \delta^{k-1}\det[S \quad D\tilde{J}^T_{(n-k)n}] , \tag{B-7}$$

where J_{kn} [$\tilde{J}_{kn}$] denotes the submatrix of I_n consisting of its first [last] k rows.

Proof: The following notation for the minors of a matrix F will be used: Let $F\begin{pmatrix} i_1,i_2,\ldots,i_k \\ j_1,j_2,\ldots,j_k \end{pmatrix}$ denote the determinant of the k × k submatrix of F formed from rows $i_1,i_2,\ldots,i_k$ and columns $j_1,j_2,\ldots,j_k$.

Applying the Cauchy-Binet theorem [Gantmacher, vol. I, p. 9] to the left side of (B-7) gives

$$\det J_{kn}\Delta S = \sum\left\{(J_{kn}\Delta)\begin{pmatrix} 1, 2,\ldots, k \\ i_1,i_2,\ldots,i_k \end{pmatrix}\right\} S\begin{pmatrix} i_1,i_2,\ldots,i_k \\ 1, 2,\ldots, k \end{pmatrix} \tag{B-8}$$

$$= \sum S\begin{pmatrix} i_1,i_2,\ldots,i_k \\ 1, 2,\ldots, k \end{pmatrix}\Delta\begin{pmatrix} 1, 2,\ldots, k \\ i_1,i_2,\ldots,i_k \end{pmatrix} .$$

The unindexed summations are assumed to apply over all integer sequences $i_1,i_2,\ldots,i_k$ satisfying $1 \le i_1 < i_2 < \ldots < i_k \le n$. Recall Jacobi's theorem on the minors of the adjoint: Any minor of order k in adj D is equal to the complementary signed minor of D^T multiplied by $(\det D)^{k-1}$ [Bocher, p. 31], [Aitken, p. 126]. In the present notation this statement can be written for the first k rows of Δ as

$$\Delta\begin{pmatrix} 1, 2,\ldots, k \\ i_1,i_2,\ldots,i_k \end{pmatrix} = (-1)^{\sum_{j=1}^{k} j+i_j} D\begin{pmatrix} i'_1,i'_2,\ldots,i'_{n-k} \\ k+1,k+2,\ldots,n \end{pmatrix}\delta^{k-1} , \tag{B-9}$$

where $i'_1,i'_2,\ldots,i'_{n-k}$ is the monotonically increasing "complementary" sequence

consisting of all integers from 1 to n not contained in $i_1,i_2,\ldots,i_k$. Substituting (B-9) into (B-8) gives

$$\det J_{kn}\Delta S = \delta^{k-1}\sum (-1)^{\sum_{j=1}^{k} j+i_j} S\begin{pmatrix} i_1,i_2,\ldots,i_k \\ 1,\ 2,\ldots,\ k \end{pmatrix} D\begin{pmatrix} i'_1,i'_2,\ldots,i'_{n-k} \\ k+1,k+2,\ldots,n \end{pmatrix}.$$

The summation above is simply the Laplace expansion for the determinant of $[S \quad D\tilde{J}^T_{(n-k)n}]$ about the first k columns [Browne]. Thus the lemma is proved. □

Now the proof of Theorem 1.14 can be completed. By (1-46) and (1-8), $A = [A_p \quad A_f]$ and $B = [B_p \quad B_f]$. Applying successively (1-48), (1-47), the above, and (1-10), it follows that

$$\begin{aligned} [V(\lambda) \quad K(\lambda)] &= \hat{V}(\lambda) \\ &= A - \lambda B \\ &= [A_p \quad A_f] - \lambda[B_p \quad B_f] \\ &= [(A_p - \lambda B_p) \quad D(\lambda)] \quad . \end{aligned} \tag{B-10}$$

This relation shows that

$$V(\lambda) = [(A_p - \lambda B_p) \quad D(\lambda)J^T_{kn}] \tag{3-11}$$

and

$$K(\lambda) = D(\lambda)\tilde{J}^T_{(n-k)n} \quad . \tag{B-12}$$

Now

$\tilde{P}(\lambda,\zeta)$

$= \det E(\lambda)Z(\zeta)$ by (1-26), (1-20), and (B-1)

$= \det[O_{kN}, \quad J_{kn}\Delta(\lambda)(A_p - \lambda B_p), \quad \delta(\lambda)I_k]Z(\zeta)$ by (1-19)

$= \det[O_{kN}, \quad J_{kn}\Delta(\lambda)(A_p - \lambda B_p), \quad J_{kn}\Delta(\lambda)D(\lambda)J^T_{kn}]Z(\zeta)$ by (1-15)

$= \det J_{kn}\Delta(\lambda)[O_{nN}, \quad (A_p - \lambda B_p), \quad D(\lambda)J^T_{kn}]Z(\zeta)$

$= \det J_{kn}\Delta(\lambda)[O_{nN} \quad V(\lambda)]Z(\zeta)$ by (B-11)

$= \det J_{kn}\Delta(\lambda)\tilde{W}(\lambda,\zeta)$ by (B-3)

$= (-1)^{(k-1)N_\zeta N} \det J_{kn}\Delta(\lambda)W(\lambda,\zeta)$ by Lemma B.1

$= (-1)^{(k-1)N_\zeta N}[\delta(\lambda)]^{k-1} \det [W(\lambda,\zeta) \quad D(\lambda)\tilde{J}^T_{(n-k)n}]$ by Lemma B.2

$= (-1)^{(k-1)N_\zeta N}[\delta(\lambda)]^{k-1}P(\lambda,\zeta)$

by (B-12), (1-51), and (1-52), thus verifying (1-53) of Theorem 1.14.

Appendix C

PROOF OF PROPOSITION 3.4

In this appendix it is shown how Theorem 2.4 and the concept of the transformed polynomial lead to Proposition 3.4. First it is shown that conditions b) and c) of Theorem 2.4 can be replaced jointly by the following two conditions:

d) $P(\cdot,\zeta)$ is not the zero function for any ζ on the unit circle E, and,

e) for every $\lambda \in I$, if $P(\lambda,\cdot)$ is not the zero function, then all zeros of $P(\lambda,\cdot)$ lie in U.

First suppose b) and c) (conditions of Theorem 2.4) hold. Condition d) is equivalent to the statement that ψ has no zeros on E, by (3-5). Therefore, b) implies d). Relation (3-5) also shows that $P(\lambda,\cdot)$ is the zero function only if $\phi(\lambda) = 0$, which occurs at just a finite number of values of λ. Except at these values, the zeros of $P(\lambda,\cdot)$ are the zeros of ψ together with the zeros of $\bar{P}(\lambda,\cdot)$, by (3-5). Thus e) follows from b) and c).

Conversely, suppose d) and e) hold. By e) and the above discussion the zeros of ψ and $\bar{P}(\lambda,\cdot)$ lie in $\bar{U}$ for all $\lambda \in I$, except possibly for λ in a finite set. Now b) follows immediately from d). However, since $\bar{U}$ is a closed set and the zeros of $\bar{P}(\lambda,\cdot)$ are continuous in λ, it follows that the zeros of $\bar{P}(\lambda,\cdot)$ lie in $\bar{U}$ for all $\lambda \in I$, verifying c).

Now Proposition 3.4 can be proved. First consider the case in which $P'(\cdot,\hat{j})$ is the zero function. Then in particular $P'(\hat{j},\hat{j}) = 0$, violating condition a) of Proposition 3.4 and the transformed A-stability criterion, Definition 3.2. Thus the proposition holds in this case.

It can thus be assumed that $P'(\cdot,\hat{j})$ is not the zero function. If P' is represented as in (3-6), with $\bar{P}'$ a reduced polynomial, then $\psi'(\hat{j}) \neq 0$. Define P by (3-8), and factor it according to (3-5). It is claimed that (3-7) holds. The only possible difficulty is with (3-7b), whose formal inverse is

$$\psi(\zeta) = (\zeta - 1)^{i'} \psi'(\hat{j}\,\frac{\zeta+1}{\zeta-1}) , \qquad \text{(C-1)}$$

where i' is the degree of ψ'. The situation is analogous to that in Lemma 3.1. Thus ψ is of degree i' if $\psi'(\hat{j}) \neq 0$. Since the latter relation holds, (3-7b) is established.

Now conditions a) of Theorem 2.4 and Proposition 3.4 are equivalent to each other, since $\hat{j}$ in the z-sphere maps into ∞ in the zeta sphere, and the transformation $\lambda \longrightarrow -\hat{j}\lambda$ of (3-8) maps $\overline{R}$ onto $\overline{H}$. Similarly, conditions b) and c) of Proposition 3.4 are the transformed counterparts of conditions d) and e), respectively. Therefore Proposition 3.4 is proved.

Appendix D

PROOF OF PROPOSITION 3.10

In this appendix the detailed structure of P', τ, T, and ∇_i' is explored in order to prove Proposition 3.10.

Recall the subscript notation defined in the proof of Lemma 3.5. In addition, if P' is a polynomial* in (ω,z), let the subscript M denote the corresponding reflected polynomial defined by $P_M'(\omega,z) = P'(-\omega,-z)$ for all (ω,z). The following lemma, which is of interest in its own right, characterizes real polynomials in terms of their transformed polynomials.

D.1. Lemma: In order for a complex polynomial P' in two variables to be the transformed polynomial associated with a real polynomial in two variables according to (3-2) and (3-3), it is necessary and sufficient that

$$P'_{RM} = P'_R \qquad \text{(a)} \qquad \text{(D-1)}$$

and

$$P'_{IM} = -P'_I \quad . \qquad \text{(b)}$$

Proof: Let P be a polynomial in (λ,ζ) for which P' is the transformed polynomial. It is obvious from (3-2) that P is real if and only if P'' is. Furthermore, P'' is real if and only if

$$[P''(\hat{j}\omega,-\hat{j}z)]^* = P''(-\hat{j}\omega,\hat{j}z)$$

for all real ω and z. By (3-3)

$$[P''(\hat{j}\omega,-\hat{j}z)]^* = [P'(\omega,z)]^* = [P'_R(\omega,z) + \hat{j}P'_I(\omega,z)]^* = P'_R(\omega,z) - \hat{j}P'_I(\omega,z)$$

and

$$P''(-\hat{j}\omega,\hat{j}z) = P'(-\omega,-z) = P'_M(\omega,z) = P'_{RM}(\omega,z) + \hat{j}P'_{IM}(\omega,z) \quad .$$

It follows that P is real if and only if

$$P'_{RM}(\omega,z) + \hat{j}P'_{IM}(\omega,z) = P'_R(\omega,z) - \hat{j}P'_I(\omega,z) \qquad \text{(D-2)}$$

for all real ω and z. Since all four polynomials in (D-2) are real, this last statement is equivalent to (D-1). □

*All the functions discussed in this appendix are either polynomials or the zero function. For brevity, the term "polynomial" will be intended to include the zero function--in this Appendix only.

To prove Proposition 3.10, consider the following interpretation of

Lemma D.1. If $P'_R(\omega,z) = \sum_{j=0}^{m} \theta_j(\omega)z^j$ and (D-1a) holds, then $\sum_{j=0}^{m} (-1)^j\theta_j(-\omega)z^j = \sum_{j=0}^{m} \theta_j(\omega)z^j$. Therefore, for j even [odd] the coefficient of z^j in P'_R is an even [odd*] polynomial in ω. Similarly, (D-1b) yields: For j even [odd] the coefficient of z^j in P'_I is an odd [even] polynomial in ω. But P'_R [P'_I] is the polynomial corresponding to the first [second] row of τ. Thus, the elements of τ in (3-9) form a "checkerboard pattern" of even and odd polynomials as follows: Each row and column of τ is a sequence of polynomials which are alternately even and odd; the first element of the first row is even.

It is claimed that the elements of T form a similar checkerboard pattern of even and odd polynomials, beginning with t_{11} even. For example, consider (3-10) for the case i + j even. If s is even then the 2 x 2 matrix whose determinant is $\tau\begin{pmatrix} 1, & 2 \\ 2m'+1-i-j-s, & s \end{pmatrix}$ has polynomial elements of the form

$$\begin{bmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{bmatrix},$$

since 2m'+1-i-j-s represents an odd column of τ. If s is odd, then the 2 x 2 matrix is of the form

$$\begin{bmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{bmatrix}.$$

In either case the determinant is even**, and hence t_{ij} of (3-10), being the sum of even polynomials, is even. A similar argument*** shows that if i + j is odd, then t_{ij} is odd. In summary, t_{ij} is even [odd] whenever i + j is even [odd].

---

*A real-valued odd function f is one satisfying $f(-\omega) = -f(\omega)$ identically on the real axis. A polynomial in one variable is odd if and only if all its even coefficients are zero. Note that the zero function is odd.

**Clearly the product of two even [odd] polynomials is even, and the sum of even polynomials is even.

***Here use the fact that the product of an even and an odd polynomial is odd, and the sum of odd polynomials is odd.

The determinant (3-11) for ∇_i' can be expressed, using the Laplace expansion recursively, as the algebraic sum (with appropriate signs) of i! terms, each of the form

$$\prod_{s=1}^{i} t_{s,r_s} \qquad \text{(D-1)}$$

where $r_1, r_2, \ldots, r_i$ is a permutation of the sequence $1,2,\ldots,i$. Since t_{s,r_s} is odd if and only if $s + r_s$ is odd, a little thought shows that the product (D-1) is even if and only if $\sum_{s=1}^{i} s + r_s$ is even. But

$$\sum_{s=1}^{i} s + r_s = \sum_{s=1}^{i} s + \sum_{s=1}^{i} r_s = 2 \sum_{s=1}^{i} s \ ,$$

which is evidently even. Hence (D-1) is even for every permutation. Since ∇_i' is the sum of such even terms, ∇_i' is even.

Appendix E

IMPLEMENTATION OF THE ALGEBRAIC CHARACTERIZATION

The simplest form of the algebraic characterization--Theorem 3.19 on strong A-stability--has been implemented in a series of computer programs written in APL/360 [Rubin, 1973]. In this Appendix two important implementation considerations are discussed, and a brief review is given of each main step of the algorithm.

For any A-stability test based on the material of Chapter 3, most of the computations involve addition and multiplication of polynomials in one or two variables (or three, as explained below). The data structure of the APL language is ideally suited to such tasks. A polynomial in r variables is represented by an array with r dimensions of its coefficients. The sum or product of two such polynomials is again an array of r dimensions, which can be computed from the operand arrays in an efficient and elegant way in APL [Rubin, 1972]. When complicated sequences of such sums and products must be executed, as in an A-stability test, APL offers many practical advantages over conventional languages such as FORTRAN: The actual size and shape of arrays is a function of the data, and need not be prespecified. The dynamic storage allocation, which is intrinsic to APL, allows storage to be utilized efficiently. Finally, it is far simpler to program such array-oriented problems in APL, since the minor programming details of array manipulation are, in effect, handled by the system.

Practical A-stability tests based on the algebraic theory of Chapter 3 belong to the class of algorithms known as seminumerical algorithms [Knuth]. Such algorithms are characterized by arithmetic which is free from error. The basic technique used to achieve absolute accuracy in numerical computations for A-stability tests is infinite precision integer arithmetic. In this technique, all input coefficients--which are assumed to be integers--are converted to a vector representation in which the i-th element is the i-th "digit" in the base B representation* of the given coefficient. For addition and multiplication it turns out that each "digital" vector can be treated exactly**

---

*In the present implementation $B \approx 10^4$, and each "digit" is a signed integer stored in a 32 bit word. It can be shown that this provides a good trade-off between computer space and time in APL.

**Actually, a special normalization process (a "propagation of carries") must be performed after each sequence of multiplications and additions, in order to prevent digit overflow. In general, this normalization affects the lengths of operand vectors.

as if it were a polynomial. For example, a polynomial in two variables with infinite precision integral coefficients is represented and manipulated as if it were a polynomial in three variables.

The computations for the present implementation of the strong A-stability characterization (Theorem 3.19) can be divided naturally into five main steps:

1. Compute the characteristic polynomial.
2. Check that all poles lie in the open right half plane.
3. Compute the transformed polynomial.
4. Compute the auxiliary polynomials.
5. Apply the Sturm test to ∇_m.

Associated with each main step are various tests to be performed, as described below. If a particular composite multistep method (R,k) fails any test, then it is not strongly A-stable. For some tests, failure also means that (R,k) is not A-stable, as discussed below. The technique of infinite precision integer arithmetic is used throughout all five steps.

Let a composite multistep method (R,k) be given with m past points, where R has integral elements. To execute Step 1, Q is computed from (1-46) through (1-51). (No actual computations are involved here, only a restructuring of the composite matrix R into an array representing Q.) Then P is computed from (1-52) using a special sub-program which computes the determinant of a matrix with polynomial elements. It turns out that even for small problems ($m \geq n = 3$) the infinite precision technique is generally necessary at this stage to prevent overflow of intermediate results. In such case the elements of Q are effectively polynomials in three variables, as discussed above, so that Q itself is represented by an array of dimensionality five. To complete Step 1 the degree of P in ζ is tested to see if it is equal to m. (If not, then R is singular, and the algorithm terminates.)

To execute Step 2, the polynomial δ is selected from P as the coefficient of ζ^m. The sign of every other coefficient of δ is then changed to form δ_M. A preliminary check is made to see if all coefficients of δ_M have the same sign. (If not, then R has a pole in the closed left half plane, and the algorithm terminates.) Finally, δ_M is tested by the method of [Cutteridge, 1959] to see if it is Hurwitz. (If not, then the algorithm terminates.)

To execute Step 3, P" is computed from (3-2). This computation amounts to the post-multiplication of the array representing P by a constant matrix

whose value depends only on m. Then P' is computed from (3-3). (This stage amounts to only a restructuring of the array representing P", and to changing the signs of some of its elements. Thus, no complex arithmetic is involved.) Finally, the degree of P' in z is tested to see if it is equal to m. (If not, then (R,k) is not A-stable.)

To execute Step 4, note from (3-9) that the array representing P' is essentially the array representing τ. Since the infinite precision representation is used, the elements of τ are effectively polynomials in two variables. First all possible 2 x 2 minors of τ of the form $\tau\begin{pmatrix} 1, & 2 \\ i, & j \end{pmatrix}$, $0 \le j < i \le m$, are computed. Then sums of these minors are computed by (3-10) to form T. Next (3-11) is used to compute ∇_i', i = 1,2,...,m. (The same determinant subprogram is used here as in Step 1; however, note that the elements of T are polynomials in only two variables.) The odd coefficients of ∇_i' are then dropped as in (3-18) to form ∇_i'', and the leading zero coefficients are dropped as in (3-19) to form the auxiliary polynomials ∇_i. Finally, the leading coefficient of each ∇_i is tested to see if it is positive. (If ∇_m has no coefficients, then it is the zero function, and the algorithm terminates. On the other hand, if $\nabla_m(0) \neq 0$, then any failure of the test $\nabla_i(0) > 0$, i = 1,2,...,m, indicates that (R,k) is not A-stable.)

To execute Step 5, two simple preliminary checks are made: If all coefficients of ∇_m are positive, then ∇_m evidently has no positive real zeros, and (R,k) is strongly A-stable. On the other hand, if the highest coefficient of ∇_m is negative, then ∇_m has a positive real zero of odd multiplicity, and (R,k) is not A-stable. Only if the highest coefficient is positive and intermediate coefficients are non-positive is the test based on Sturm's theorem invoked. In such case let $\nabla_m(\omega) = \sum_{i=0}^{j} c_i \omega^i$, where $c_j > 0$. First ∇_m is transformed into a nomic polynomial ∇ by multiplying all its zeros by the positive real factor c_j as follows: Each c_i, i = 0,1,...,j-1, is replaced by the value $c_i c_j^{j-i-1}$, and then c_j is replaced by unity [Turnbull, p. 89]. Now the j x j matrix L of [Barnett, 1971c] is computed for ∇. (Briefly, L is related to the companion matrix of ∇, and can be computed by a simple algorithm [Barnett, 1970, eq. 7 and 8]. The coefficients of the Sturm sequence of ∇ are minors of L.) Two sequences of minors of L are then computed according to [Barnett,

1971c, p. 248]. (Here, the determinant sub-program is again invoked. Since infinite precision integer representation is used, the elements of L are effectively polynomials in one variable.) Both sequences are tested to see if they have the same number of sign variations*. If so, then ∇ has no positive real zeros, and (R,k) is strongly A-stable. If not, then ∇ has positive real zeros, and (R,k) is not strongly A-stable. In fact, if det $L \neq 0$, these zeros are of multiplicity one; which is odd; therefore (R,k) is not A-stable.

It is evident that most of the computations of the algorithm take place in the determinant sub-program. Therefore, it is desirable to make this part of the code as efficient as possible. The present implementation [Rubin, 1973] computes determinants for matrices up to 6×6 whose elements are polynomials in up to three variables, one of which is "normalized" according to a fixed base. Therefore, only problems with $m \leq 6$ and $n \leq 6$ can be handled, by Steps 4 and 1, respectively. If the Sturm test of Step 5 is invoked, then it is also required that $mn - k_m \leq 6$. This capability has proved adequate for present research purposes. For moderate-sized problems such as those of Section 4.11, the application of Step 5 in the present implementation has been observed to give rise to integers as large as 10^{300}.

*One of the sequences contains det L, which is zero exactly when ∇ has multiple zeros. If det L = 0, a modification of the above procedure must be employed [Barnett, 1971c, p. 248].

Appendix F

PROOF OF THEOREM 4.13

This appendix presents a proof of Theorem 4.13 by use of the following lemma:

F.1. Lemma: Let A, U, and V be square matrices of the same dimensions. If A is nonsingular, and

$$\|U\| < 1/2\|A^{-1}\| \qquad \text{(a)} \tag{F-1}$$

$$\|V\| < 1/2\|A^{-1}\| \quad , \qquad \text{(b)}$$

then A - U and A - V are nonsingular, and

$$\|(A - U)^{-1} - (A - V)^{-1}\| \leq 4\|A^{-1}\|^2(\|U\| + \|V\|) \quad . \tag{F-2}$$

Proof: By (F-1a) it follows that

$$\|A^{-1}U\| < 1/2 \quad , \tag{F-3}$$

and hence that the infinite series $\sum_{k=0}^{\infty} (A^{-1}U)^k$ converges to $(I - A^{-1}U)^{-1}$.

Thus $A(I - A^{-1}U) = A - U$ is nonsingular, and

$$\begin{aligned}
\|(A - U)^{-1}\| &= \|(I - A^{-1}U)^{-1}A^{-1}\| \\
&\leq \|A^{-1}\|\|(I - A^{-1}U)^{-1}\| \\
&= \|A^{-1}\|\|\sum_{k=0}^{\infty} (A^{-1}U)^k\| \\
&\leq \|A^{-1}\| \sum_{k=0}^{\infty} \|A^{-1}U\|^k \\
&= \|A^{-1}\|/(1 - \|A^{-1}U\|) \\
&< 2\|A^{-1}\| \quad , \qquad \text{(F-4a)}
\end{aligned}$$

the last inequality following from (F-3). Similarly, A - V is nonsingular, and

$$\|(A - V)^{-1}\| < 2\|A^{-1}\| \quad . \tag{F-4b}$$

Now

$$\begin{aligned}\|(A - U)^{-1} - (A - V)^{-1}\| &= \|(A - U)^{-1}[(A - V) - (A - U)](A - V)^{-1}\| \\ &\leq \|(A - U)^{-1}\|\|U - V\|\|(A - V)^{-1}\| \\ &\leq 4\|A^{-1}\|^2\|U - V\| \\ &\leq 4\|A^{-1}\|^2(\|U\| + \|V\|) \quad .\end{aligned}$$

The last two inequalities follow from (F-4) and the triangle inequality, respectively. □

To prove Theorem 4.13, let an $\epsilon > 0$ be given. Let $W = -A_p\Phi_p(0) = -A_p[1 \;\; 1 \;\; \ldots \;\; 1]^T x_0^T$, let N be the closed ϵ neighborhood of $A_f^{-1}W$, and let*

$$\Gamma = \sup_{\substack{Y\epsilon N \\ 0\leq h\leq\epsilon}} \|F(Y,h)\| + \sup_{\substack{Y\epsilon N \\ 0\leq h\leq\epsilon}} \|F'(Y,h)\| \quad . \tag{F-5}$$

It will be shown in the next paragraph that there exists a positive number δ satisfying $\delta < \epsilon$ and

$$\delta \leq \epsilon/2\Gamma\|A_f^{-1}\|\|B_f\|(\epsilon + 1 + 4\|A_f^{-1}\|) \quad , \tag{F-6}$$

such that

$$\|\epsilon_T(h)\| < \epsilon/4\|A_f^{-1}\| \tag{a}$$

(F-7)

and

$$\|C_h[\Phi_p(h)] - W\| < \epsilon/4\|A_f^{-1}\| \tag{b}$$

whenever $0 < h < \delta$. Following this it will be shown that

$$\|\epsilon_D(h) - [\hat{\Delta}(h)]^{-1}\epsilon_T(h)\| < \epsilon\|\epsilon_T(h)\| \tag{F-8}$$

whenever $0 < h < \delta$. Since ϵ is arbitrary, this result establishes (4-27), thus proving Theorem 4.13.

Since f and X are continuous, so are F and Φ. Therefore ϵ_T of (4-14)

---

*By hypothesis f and f' are continuous. Therefore Γ is finite, since it depends upon the suprema of the continuous functions $\|F\|$ and $\|F'\|$ on a (nonempty) compact set.

is continuous. By hypothesis $\epsilon_T(0) = 0$. Therefore there exists a δ small enough so that (F-7a) holds. Similarly, (4-22a) shows that $C_h[\Phi_p(h)]$ depends continuously on h, and that $C_0[\Phi_p(0)] = W$. Thus, δ can be chosen small enough so that (F-7b) holds.

It remains to verify (F-8). In all that follows, let h be a fixed, positive number, with

$$h < \delta \quad . \tag{F-9}$$

Observe from (4-22b) that

$$D_h'(Y) = A_f - hB_f F_f'(Y,h) \tag{F-10}$$

for all $Y \in N$. Since

$$\begin{aligned} \|hB_f F_f'(Y,h)\| &\leq h\|B_f\|\|F_f'(Y,h)\| \\ &\leq h\|B_f\|\Gamma && \text{by (F-5)} \\ &< \frac{\epsilon\|B_f\|\Gamma}{2\Gamma\|A_f^{-1}\|\|B_f\|(\epsilon + 1 + 4\|A_f^{-1}\|)} && \text{by (F-6) and (F-9)} \\ &< 1/2\|A_f^{-1}\| \quad , && \text{(F-11)} \end{aligned}$$

the first conclusion of Lemma F.1 shows that $D_h'(Y)$ is nonsingular for all $Y \in N$. Now the inverse function theorem [Dieudonne, p. 273] can be used to show that D_h is invertible, and that D_h^{-1} is continuously differentiable, at least in an $\epsilon/2\|A_f^{-1}\|$ neighborhood* M of W. Combine (F-10) with the elementary identity

$$(D_h^{-1})'(Y) = \{D_h'[D_h^{-1}(Y)]\}^{-1}$$

for the derivative of the inverse [Dieudonne, p. 153] of the function D_h, to yield the important relation

$$(D_h^{-1})'(Y) = \{A_f - hB_f F_f'[D_h^{-1}(Y), h]\}^{-1} \quad , \tag{F-12}$$

valid for $Y \in M$.

---

*The truth of this statement depends upon the fact that $M \subset D_h(N)$. But $M \subset D_h(N)$ if and only if the image (under D_h) of the boundary of N does not intersect M; that is, if and only if $\|D_h(Y)^h - W\| \geq \epsilon/2\|A_f^{-1}\|$ whenever $\|Y - A_f^{-1}W\| = \epsilon$. This inequality can be verified directly from (4-22b) and the relation $\|hB_f F_f(Y,h)\| < \epsilon/2\|A_f^{-1}\|$, which is derivable from (F-5), (F-6), and (F-9).

Relation (F-7b) shows that $C_h[\Phi_p(h)] \in M$. For this value of Y in (F-12), the expression in braces is equal to $\hat{\Delta}(h)$, by (4-28). That is,

$$(D_h^{-1})'[C_h[\Phi_p(h)]] = [\hat{\Delta}(h)]^{-1} \quad . \tag{F-13}$$

Let S_h denote the straight line segment defined by the points $C_h[\Phi_p(h)]$ and $C_h[\Phi_p(h)] + \epsilon_T(h)$. The latter point is in M, as can be shown using (F-7) and the triangle inequality. Therefore, $S_h \subset M$. Applying a form of the mean value theorem [Dieudonne, p. 162] to the differentiable function D_h^{-1} at the endpoints of S_h yields the inequality

$$\|D_h^{-1}[C_h[\Phi_p(h)] + \epsilon_T(h)] - D_h^{-1}[C_h[\Phi_p(h)]] - \{(D_h^{-1})'[C_h[\Phi_p(h)]]\}\,\epsilon_T(h)\|$$

$$\leq \|\epsilon_T(h)\| \sup_{Y \epsilon S_h} \|(D_h^{-1})'(Y) - (D_h^{-1})'[C_h[\Phi_p(h)]]\| \quad .$$

By Lemma 4.12 the first two terms above sum to $\epsilon_D(h)$. Also, the third term can be simplified using (F-13). After performing these two substitutions the above inequality becomes

$$\|\epsilon_D(h) - [\hat{\Delta}(h)]^{-1}\epsilon_T(h)\| \leq \gamma(h)\|\epsilon_T(h)\|$$

where

$$\gamma(h) = \sup_{Y \epsilon S_h} \|(D_h^{-1})'(Y) - (D_h^{-1})'[C_h[\Phi_p(h)]]\| \quad . \tag{F-14}$$

The desired relation (F-8) will be established if it can be shown that $\gamma(h) < \epsilon$.

To show this, the identity (F-12) can be applied to each of the two terms inside the norm of (F-14), with the result

$$\gamma(h) = \sup_{Y \epsilon S_h} \|\{A_f - hB_f F_f'[D_h^{-1}(Y), h]\}^{-1}$$

$$- \{A_f - hB_f F_f'[D_h^{-1}[C_h[\Phi_p(h)]], h]\}^{-1}\| \quad .$$

The above norm is in the form of the left side of (F-2). Furthermore, (F-11) verifies the hypothesis (F-1). Applying Lemma F.1 now gives

$$\gamma(h) \le \sup_{Y \epsilon S_h} 4\|A_f^{-1}\|^2(\|hB_f F_f'[D_h^{-1}(Y), h]\| + \|hB_f F_f'[D_h^{-1}[C_h[\Phi_p(h)]], h]\|)$$

$$\le 4\|A_f^{-1}\|^2 h\|B_f\|(\sup_{Y \epsilon S_h} \|F_f'[D_h^{-1}(Y), h]\| + \sup_{Y \epsilon S_h} \|F_f'[D_h^{-1}[C_h[\Phi_p(h)]], h]\|)$$

$$\le 4\|A_f^{-1}\|^2 h\|B_f\|(\Gamma+\Gamma) \qquad \text{by (F-5)}$$

$$< \frac{8\Gamma\|A_f^{-1}\|^2\|B_f\|\epsilon}{2\Gamma\|A_f^{-1}\|\|B_f\|(\epsilon + 1 + 4\|A_f^{-1}\|)} \qquad \text{by (F-6) and (F-7)}$$

$$< \epsilon ,$$

completing the proof.

Appendix G

PROOF OF THEOREM 4.24

This appendix presents a proof of Theorem 4.24 in terms of a corollary to the following lemma:

G.1. Lemma: Let i and p be integers with $p \geq 0$ and $1 \leq i \leq p + 1$. Then the $(p+1) \times (p+1)$ submatrix $\hat{\Pi}_{ip}$ of Π_{pp} consisting of rows $1,2,\ldots,i,p+1,p+2,\ldots,2p-i$ is nonsingular.

Proof: The proof is by induction on p. For p = 0 Lemma G.1 is trivial, since $\hat{\Pi}_{10} = [1]$, which is nonsingular. Let p = 0,1,2,..., and assume that $\hat{\Pi}_{ip}$ is nonsingular for all i = 1,2,...,p+1. Fixing i, the induction proof will be complete if it can be shown that $\hat{\Pi}_{i(p+1)}$ and $\hat{\Pi}_{(i+1)(p+1)}$ are nonsingular.

It can be shown that there exists a (p+1)-th degree polynomial f, unique to within a trivial factor, satisfying the p+1 conditions

$$f(k) = 0 \qquad \text{for } k = 0,1,\ldots,i-1 \qquad \text{(a)} \qquad \text{(G-1)}$$

and

$$f'(k) = 0 \qquad \text{for } k = 0,1,\ldots,p-i \qquad \text{(b)}$$

where f' denotes the derivative of f. It follows from elementary considerations* that

---

*A simple proof of relation (G-2a) can be given by cases. First suppose $p-i \leq i-1$. From (G-1) f has p-i+1 zeros of multiplicity 2, and (i-1) - (p-i) zeros of multiplicity 1, for a total of p+1 zeros, none greater than i-1. Since f is of degree p+1, it can have no other zeros. Therefore (G-2a) holds for the case $p-i \leq i-1$.

Next suppose $p-i > i-1$. Also assume (G-2a) fails to hold. Then f has i+1 integral zeros of multiplicity 2. But a zero of f' lies strictly between any two distinct real zeros of f. Therefore f' has at least i nonintegral zeros. By (G-1b) f' also has p-i+1 integral zeros. In summary, f' has at least p+1 zeros, which is a contradiction, since f' is only of degree p. This proves (G-2a) for the case $p-i > i-1$.

The same principles can be used to prove (G-2b). First suppose $p-i \geq i-1$. From (G-1) f has i integral zeros of multiplicity 2, which implies that f' has at least i-1 nonintegral zeros. By (G-1b) f' also has p-i+1 integral zeros, none greater than p-i. In summary, f' has at least p zeros, all of which are nonintegral or less than p-i+1. Since f' is of degree p, it can have no other zeros. Therefore (G-2b) holds for the case $p-i \geq i-1$.

Next suppose $p-i < i-1$. Also assume (G-2b) fails to hold. Then f has p-i+2 zeros of multiplicity 2, and (i-1) - (p-i+1) zeros of multiplicity 1, for a total of p+2 zeros. This is a contradiction, since f is only of degree p+1. This proves (G-2b) for the case $p-i < i-1$.

$$f(i) \neq 0 \qquad \text{(a)} \qquad \text{(G-2)}$$

and

$$f'(p-i+1) \neq 0 \ . \qquad \text{(b)}$$

Let $\hat{I}$ denote I_{p+2}, but with the last column of I_{p+2} replaced by the sequence of coefficients of f (in order of increasing powers, as usual).

Consider $\hat{I}$ as a transformation postmultiplying $\hat{\Pi}_{i(p+1)}$ $[\hat{\Pi}_{(i+1)(p+1)}]$. By construction, each element of the last column of $\hat{\Pi}_{i(p+1)}\hat{I}$ $[\hat{\Pi}_{(i+1)(p+1)}\hat{I}]$ can be seen to be a value of f or f' at some nonnegative integer. Applying (G-1) shows immediately that

$$\det \hat{\Pi}_{i(p+1)}\hat{I} = f'(p-i+1)\det \hat{\Pi}_{ip}$$

and

$$\det \hat{\Pi}_{(i+1)(p+1)}\hat{I} = f(i)\det \hat{\Pi}_{ip} \ .$$

By (G-2) the right hand side of each relation above is nonzero. It follows that $\hat{\Pi}_{i(p+1)}$ and $\hat{\Pi}_{(i+1)(p+1)}$ are nonsingular, completing the proof. □

G.2. Corollary: Let m and n be positive integers, and let p be a nonnegative integer.

a) If $p + 1 > 2m + n$, then rank $\Pi_{(m+n)p} > 2m + n$. (a) (G-3)

b) If $p + 1 \leq 2m + n$, then rank $\Pi_{mnp} = p + 1$, (b)
where Π_{mnp} is the $(2m+n) \times (p+1)$ submatrix of $\Pi_{(m+n)p}$ consisting of rows $1,2,\ldots,m,m+n+1,m+n+2,\ldots,2(m+n)$.

Proof: Evidently $\hat{\Pi}_{(m+n)(2m+n)}$ is a submatrix of $\Pi_{(m+n)p}$ whenever $p + 1 > 2m + n$. In such case rank $\Pi_{(m+n)p} \geq$ rank $\hat{\Pi}_{(m+n)(2m+n)} = 2m + n + 1$, the equality following from Lemma G.1. This proves (G-3a).

Since Π_{mnp} has $p + 1$ columns, rank $\Pi_{mnp} \leq p + 1$. On the other hand, $\hat{\Pi}_{mp}$ is a $(p+1) \times (p+1)$ submatrix of Π_{mnp}. By Lemma G.1 $\hat{\Pi}_{mp}$ is nonsingular. Therefore (G-3b) is proved. □

To prove Theorem 4.24, consider the order relations (4-53). Partition R using (1-8) and (1-46), and conformably partition $\Pi_{(m+n)p}$ to give

$$[A_p \quad A_f \quad B]\begin{bmatrix}\Pi_p \\ \Pi_f \\ \Pi_B\end{bmatrix} = 0 \ ,$$

or

$$[A_p \quad B]\begin{bmatrix} \Pi_p \\ \Pi_B \end{bmatrix} = - A_f \Pi_f \quad .$$

Note that $\begin{bmatrix} \Pi_p \\ \Pi_B \end{bmatrix}$ is precisely Π_{mnp}, as defined in Corollary G.2. Also, if R is strongly regular, A_f is nonsingular, and the above can be written as

$$- A_f^{-1}[A_p \quad B]\Pi_{mnp} = \Pi_f \quad . \tag{G-4}$$

It is evident that for each $n \times (2m+n)$ matrix Y satisfying

$$Y\Pi_{mnp} = \Pi_f \tag{G-5}$$

and each $n \times n$ nonsingular matrix A_f, the value of $[A_p \quad B]$ given by

$$[A_p \quad B] = - A_f Y$$

satisfies (G-4). Furthermore, all solutions $[A_p \quad B]$ to (G-4) are of this form.

Relation (G-5) is a set of linear inhomogeneous algebraic equations. It is well known [Finkbeiner, Section 5.1] that solutions exist to such a set of relations if and only if

$$\text{rank}\begin{bmatrix} \Pi_{mnp} \\ \Pi_f \end{bmatrix} = \text{rank } \Pi_{mnp} \quad .$$

But $\begin{bmatrix} \Pi_{mnp} \\ \Pi_f \end{bmatrix}$ is equal to $\Pi_{(m+n)p}$, to within a permutation of the rows. To summarize the above discussion, $\mathcal{R}(p,n,m)$ is nonempty if and only if

$$\text{rank } \Pi_{(m+n)p} = \text{rank } \Pi_{mnp} \quad . \tag{G-6}$$

If $\mu < 0$, then $p + 1 > 2m + n$ by (4-58). Therefore, (G-3a) holds. On the other hand, rank $\Pi_{mnp} \leq 2m + n$, since Π_{mnp} has only $2m + n$ rows. Thus, (G-6) fails to hold, and $\mathcal{R}(p,n,m)$ is empty.

Conversely, if $\mu \geq 0$, then $p + 1 \leq 2m + n$. Therefore, (G-3b) holds. But Π_{mnp} is a submatrix of the matrix $\Pi_{(m+n)p}$, which has only $p + 1$ columns. Thus, (G-6) holds, and $\mathcal{R}(p,n,m)$ is nonempty. This proves the first conclusion of Theorem 4.24.

Suppose $\mu \geq 0$. Since rank $\Pi_{mnp} = p + 1$ and Π_{mnp} has $2m + n$ rows, the

solutions to the homogeneous system associated with (G-5) form a vector space of dimension $(2m+n) - (p+1) = \mu$ [Finkbeiner, Theorem 5.1]. That is, there exists a $\mu \times (2m+n)$ matrix $\hat{S}$ with rank $\hat{S} = \mu$ such that

$$\hat{S}\Pi_{mnp} = 0 \quad . \tag{G-7}$$

The rows of $\hat{S}$ form a basis for the solution space. Since the elements of Π_{mnp} are integers and (G-7) is homogeneous, $\hat{S}$ can be taken to have integral elements*. Partition $\hat{S}$ after the m-th column to give $\hat{S} = [S_p \quad S_B]$, and let S be given by (4-59) and (4-60a). Then (4-60b) and (4-60c) hold, the latter since it is equivalent to (G-7). This proves the second conclusion of Theorem 4.24.

The third conclusion is easy. If $R \in \mathfrak{R}_c(p,n,m)$ and X is an $n \times \mu$ matrix, then $R + XS$ is evidently canonical, by (4-60a). Also, (4-53) and (4-60c) show that $R + XS$ is of order p.

To prove the last conclusion of Theorem 4.24, let Y denote columns $1,2,\ldots,m,m+n+1,m+n+2,\ldots,2(m+n)$ of $\hat{R} - R$. Since $\hat{R} - R$ satisfies the order relations, and columns $m+1,m+2,\ldots,m+n$ of $\hat{R} - R$ are zero, it follows that $Y\Pi_{mnp} = 0$. But since the rows of $\hat{S}$ in (G-7) form a basis for the solution space, there exists a unique $n \times \mu$ matrix X such that $Y = X\hat{S}$. By (4-60a), this relation is equivalent to (4-61).

---

*For example, a solution $\tilde{S}$ to (G-7) can be constructed from the elements of Π_{mnp} using only rational arithmetic [Finkbeiner, Section 5.5]. That is, the elements of $\tilde{S}$ are all rational numbers. If ℓ is the least common multiple of the denominators of the elements of $\tilde{S}$, then $\ell\tilde{S}$ is an integral matrix satisfying (G-7).

REFERENCES

Aitken, A.C.
1956 Determinants and Matrices, 9-th Ed., Oliver and Boyd, Ltd., Edinburgh.

Bareiss, E.H.
1968 "Sylvester's Identity and Multistep Integer-Preserving Gaussian Elimination," Math. Comp., 22, pp. 565-578.

Barnett, S.
1970 "Qualitative Analysis of Polynomials Using Matrices," IEEE Trans. Automatic Control, AC-15, pp. 380-382.

1971a "A New Formulation of the Lienard-Chipart Stability Criterion," Proc. Cambridge Philos. Soc., 70, pp. 269-274.

1971b "Location of the Zeros of a Complex Polynomial," Linear Algebra and Appl., 4, pp. 71-76.

1971c "A New Formulation of the Theorems of Hurwitz, Routh, and Sturm," J. Inst. Math. Appl., 8, pp. 240-250.

1971d "Greatest Common Divisor of Several Polynomials," Proc. Cambridge Philos. Soc., 70, pp. 263-268.

1972 "A Note on the Bezoutian Matrix," SIAM J. Appl. Math., 22, pp. 84-86.

Beaudet, P.R.
1972 "Development of Multi-Off-Grid (MOG) Multistep Integration Techniques for Orbital Applications," Vol. 1, Contract Report No. 5035-19100-01TR, Computer Sciences Corp., Falls Church, Va.

Bickart, T.A., Burgess, D.A., and Sloate, H.M.
1971 "High Order A-Stable Composite Multistep Methods for Numerical Integration of Stiff Differential Equations," Proc. Ninth Annual Allerton Conference on Circuit and System Theory, U. of Illinois, pp. 465-473.

Bickart, T. A., and Picel, Z.
1972 "High Order Stiffly Stable Composite Multistep Methods for Numerical Integration of Stiff Differential Equations," SIAM 1972 National Meeting, June, Philadelphia, Pa. Also to appear in BIT, 13 (1973).

Bliss, G.A.
1933 Algebraic Functions, American Math. Soc., New York.

Bocher, M.
1907 Introduction to Higher Algebra, MacMillan Co.

Brayton, R. K., and Conley, C.C.
1972 "Some Results on the Stability and Instability of the Backward Differentiation Methods with Non-Uniform Time Steps," Research Report No. RC 3964, IBM Corporation, Yorktown Heights, N.Y.

Brayton, R.K., Gustavson, F.G., and Liniger, W.
1966 "A Numerical Analysis of the Transient Behavior of a Transistor Circuit," IBM J. Res. Develop., 10, pp. 292-299.

Browne, E.T.
1958 Introduction to the Theory of Determinants and Matrices, U. of North Carolina Press.

Burnside, W.S., and Panton, A.W.
1892 The Theory of Equations, 3-rd Ed., Dublin U. Press.

Butcher, J.C.
1972 "An Algebraic Theory of Integration Methods," Math. Comp., 26, pp. 79-106.

Chartres, B., and Stepleman, R.
1972 "A General Theory of Convergence for Numerical Methods," SIAM J. Numer. Anal., 9, pp. 476-492.

Coddington, E.A., and Levinson, N.
1955 Theory of Ordinary Differential Equations, McGraw-Hill.

Cooke, C.H.
1972 "On Stiffly Stable Implicit Linear Multistep Methods," SIAM J. Numer. Anal., 9, pp. 29-34.

Cutteridge, O.P.D.
1959 "The Stability Criteria for Linear Systems," Proc. IEE (London), 106, part C, pp. 125-132.

1960 "Some Tests for the Number of Positive Zeros and for the Number of Real and Complex Zeros of a Real Polynomial," Proc. IEE (London), 107, part C, pp. 105-110.

Dahlquist, G.
1963 "A Special Stability Problem for Linear Multistep Methods," BIT, 8, pp. 27-43.

Dejon, B.
1967 "Numerical Stability of Difference Equations with Matrix Coefficients," SIAM J. Numer. Anal., 4, pp. 119-128.

Dieudonne, J.
1960 Foundations of Modern Analysis, Academic Press, New York.

Donelson, J., and Hanson, E.
1971 "Cyclic Composite Multistep Predictor-Corrector Methods," SIAM J. Numer. Anal., 8, pp. 137-157.

Ehle, B.L.
1968 "High Order A-Stable Methods for the Numerical Solution of Systems of D.E.'s," BIT, 8, pp. 276-278.

1972 "A Comparison of Methods for Solving Stiff Systems," SIAM-SIGNUM 1972 Fall Meeting, Austin, Texas, Oct. 16-18.

Enright, W.
1972 "Studies in the Numerical Solution of Stiff Ordinary Differential Equations," Tech. Report No. 46, Dept. of Computer Science, U. of Toronto.

Finkbeiner, D.T.
1966 Introduction to Matrices and Linear Transformations, 2-nd Ed., W.H. Freeman and Co., San Francisco.

Fowler, M.E., and Warten, R.M.
1967 "A Numerical Integration Technique for Ordinary Differential Equations with Widely Separated Eigenvalues," IBM J. Res. Develop., 11, pp. 537-543.

Gantmacher, F.R.
1959 Theory of Matrices, Vols. 1 and 2, Chelsea Pub. Co., New York.

Gear, C.W.
1968 "The Automatic Integration of Stiff Ordinary Differential Equations," IFIP Congress, August, pp. A81-A85.

Gelinas, R.J.
1972 "Stiff Systems of Kinetic Equations -- A Practitioner's View," J. Computational Phys., 9, pp. 222-236.

Handbook of Mathematical Tables
1964 2-nd Ed., Chemical Rubber Company, pp. 356-357.

Henrici, P.
1962 Descrete Variable Methods in Ordinary Differential Equations, Wiley.

Householder, A.
1968 "Bigradients and the Problem of Routh and Hurwitz," SIAM Rev., 10, pp. 56-66.

1970 "Bezoutiants, Elimination and Localization," SIAM Rev., 12, pp. 73-78.

Hull, T.E.
1972 "Comparing Numerical Methods for Ordinary Differential Equations," Numerical Solution of Ordinary Differential Equations Conference, October 19-20, Austin, Texas.

Hull, T.E., Enright, W.H., Fellen, B.M., and Sedgewick, A.E.
1971 "Comparing Numerical Methods for Ordinary Differential Equations," SIAM 1971 National Meeting, Seattle, Wash., June 28-30.
Also appearing in SIAM J. Numer. Anal., 9 (1972), pp. 603-637.

Hulme, B.L.
1972 "One-Step Piecewise Polynomial Galerkin Methods for Initial Value Problems," Math. Comp., 26, pp. 415-426.

Jacobson, N.
1964 Lectures in Abstract Algebra, Vol. 3, Van Nostrand Co., Princeton, New Jersey.

Knuth, D.E.
1969 The Art of Computer Programming, Vol. 2, Seminumerical Algorithms, Addison-Wesley.

Liniger, W.
1968 "A Criterion for A-Stability," Computing, 3, pp. 280-285.

Liniger, W., and Willoughby, R.A.
1970 "Efficient Integration Methods for Stiff Systems of Ordinary Differential Equations," SIAM J. Numer. Anal., 7, pp. 47-66.

Little, W.W., Hansen, K.F., Mason, E.A., and Koen, B.V.
1964 "Stable Numerical Solution of Reactor Kinetics Equations," Trans. American Nuclear Society (Philadelphia Meeting), 7, No. 1, p. 3.

Marden, M.
1949 The Geometry of the Zeros of a Polynomial in a Complex Variable, American Math. Soc. See also the 2-nd Ed., Zeros of Polynomials, (1966).

Moretti, G.
1963 "The Chemical Kinetics Problem in the Numerical Analysis of Nonequilibrium Flows," Proc. IBM Scientific Computing Symposium on Large-Scale Problems in Physics, December 9-11, IBM Corp., White Plains, New York, pp. 167-182.

Muir, T.
1906-1923 The Theory of Determinants in the Historical Order of Development, in four volumes, Macmillan, London. Republished by Dover Publications, Inc., New York, (1960).

Odeh, F., and Liniger, W.
1971 "A Note on Unconditional Fixed-h Stability of Linear Multistep Formulae," IFIP Congress, August, Ijubljana, Yugoslavia, Booklet TA-1, pp. 38-41. Also appearing in Computing, 7 (1971), pp. 240-253.

1972 "A-Stable, Accurate Averaging of Multistep Methods for Stiff Differential Equations," IBM J. Res. Develop., 16, pp. 335-348.

Olum, P.
1953 "Mappings of Manifolds and the Notion of Degree," Ann. of Math., 58, pp. 458-480.

Rubin, W.B.
1972 "Addition and Multiplication of Polynomials in n Variables," Algorithm No. 99, APL Quote-Quad, 3, No. 5, June, pp. 73-75.

1973 "APL Algorithms for an A-Stability Test of Composite Multistep Methods," Technical Memorandum No. TM-73-1, Dept. of Electrical and Computer Engineering, Syracuse University, Syracuse, N.Y.

Rubin, W.B., and Bickart, T.A.
1972 "A-Stability of Composite Multistep Methods," SIAM-SIGNUM 1972 Fall Meeting, October 16-18, Austin, Texas.

Rudin, W.
1966 Real and Complex Analysis, McGraw-Hill.

Sandberg, I.W., and Shichman, H.
1968 "Numerical Integration of Systems of Stiff Nonlinear Differential Equations," Bell System Tech. J., 47, pp. 511-527.

Shampine, L.F., and Watts, H.A.
1969 "Block Implicit One-Step Methods," Math. Comp., 23, pp. 731-740.

Siljak, D.D.
1969 Nonlinear Systems, Wiley.

1971 "New Algebraic Criteria for Positive Realness," J. Franklin Inst., 291, pp. 109-120.

Sloate, H.M.
1971 "Simultaneous Implicit Formulas for the Solution of Stiff Systems of Differential Equations," Doctoral Dissertation, Syracuse University.

Sloate, H.M., and Bickart, T.A.
1971 "A-Stable Composite Multistep Methods," SIAM 1971 National Meeting, Seattle, Wash., June 28-30. Also appearing in J. Assoc. Comput. Mach., 20 (1973), pp. 7-26.

Springer, G.
1957 Introduction to Riemann Surfaces, Addison-Wesley, Reading, Mass.

Tendler, J.M.
1973 "A Stiffly Stable Integration Process Using Cyclic Composite Methods," Doctoral Dissertation, Syracuse University.

Turnbull, H.W.
1939 Theory of Equations, Oliver and Boyd, Ltd., Edinburgh.

Watts, H.A.
1971 "A-Stable Block Implicit One-Step Methods," Doctoral Dissertation, U. of New Mexico.

Watts, H.A., and Shampine, L.F.
1972 "A-Stable Block Implicit One-Step Methods," BIT, 12, pp. 252-266.

LIST OF FREQUENTLY OCCURRING SYMBOLS

Symbol	Meaning	Page Where Defined or Introduced
A	matrix of α's	26
A_f	matrix of future α's.	9
A_p	matrix of past α's.	9
α_{ij}	elements of the composite matrix.	6
adj	the adjoint function.	13
B	matrix of β's	26
B_f	matrix of future β's.	9
B_p	matrix of past β's.	9
β_{ij}	elements of the composite matrix.	6
c_D	discretization error constant	97
c_T	truncation error constant	95
C	matrix of rational functions in λ	16
C_h	function of past points	88
C	the complex plane	8
$\bar{C}$	the Riemann sphere.	19
det	the determinant function.	10
D	the pencil of the composite multistep method.	9
$\hat{D}$	the greatest real divisor of P'	49
D_h	function of future points	88
$\mathcal{D}$	the differential operator	86
δ	the polynomial of poles	10
Δ	the adjoint of D.	13
$\hat{\Delta}$	matrix of functions of h.	91
∇_i	auxiliary polynomial.	56
∇'_i	polynomial in ω	48
∇''_i	polynomial in Ω	56
$\tilde{\nabla}_i$	dual auxiliary polynomial	70
e	base of the natural logarithms.	92
E	matrix of polynomials in λ	14
E_i	submatrix of E.	14
$\mathcal{E}$	the unit circle	38

ϵ_D	discretization error (function of h)	89
$\tilde{\epsilon}_D$	discretization error (function of λ)	92
ϵ_T	truncation error (function of h)	86
$\tilde{\epsilon}_T$	truncation error (function of λ)	91
f	the function defining the differential equation	6
F	function induced by f	77
F_f	matrix of future rows of F	88
F_p	matrix of past rows of F	88
G	matrix of polynomials in λ	16
G	the open upper half plane	46
h	the step size	6
H	matrix of polynomials in λ	16
H	the open lower half plane	45
i	index	6
I	the identity matrix	13
I_I	(subscript) the imaginary part of a polynomial	49
I	the extended imaginary axis	33
Im	the imaginary part of	45
j	index	6
$\hat{j}$	the imaginary unit	45
J	matrix of functions of λ	87
J_{kn}	a submatrix of I_n	14
$\tilde{J}_{kn}$	a submatrix of I_n	139
k	the number of retained points	6
k_i	a multiplicity associated with ∇_i''	56
K	a pencil in λ	26
L	a pencil in λ	82
ℓ	index	7
$\mathcal{L}$	the open left half plane	8
λ	the product of q and h	9
Λ	the set of poles	10
m	the number of past points	6
m	the degree of P in ζ	18
m'	the degree of P' in z	46
$\hat{m}$	the degree of $\hat{P}'$ in z	50

INDEX OF DEFINED TERMS

BIOGRAPHICAL DATA

NAME: William Benjamin Rubin

DATE AND PLACE OF BIRTH: February 3, 1943
Manhattan, New York

COLLEGE: Clarkson College of Technology
Potsdam, New York
B.S., 1965

GRADUATE WORK: Syracuse University
Syracuse, New York
M.S., 1968

NSF Traineeship, 1967-1970
Research Assistantship, 1971-1972

PROFESSIONAL EXPERIENCE: IBM Corporation, Owego, New York
Junior Engineer, 1965-1966

Syracuse University
Lecturer, Fall Term 1970

AFFILIATIONS: Institute of Electrical and Electronics Engineers
IEEE Control Systems Society
Sigma Xi

PUBLICATIONS:

"APL Algorithms for an A-Stability Test of Composite Multistep Methods," Technical Memorandum No. TM-73-1, Department of Electrical and Computer Engineering, Syracuse University, April 1973.

"A-Stability of Composite Multistep Methods," (with T.A. Bickart), SIAM-SIGNUM 1972 Fall Meeting, Austin, Texas, October 17, 1972.

"Greatest Common Divisor Pairwise on Vectors," Algorithm No. 97, APL Quote-Quad, 3, No. 5 (1972), p. 72.

"Least Common Multiple," Algorithm No. 98, APL Quote-Quad, 3, No. 5 (1972), pp. 72-73.

"Addition and Multiplication of Polynomials in n Variables," Algorithm No. 99, APL Quote-Quad, 3, No. 5 (1972), pp. 73-75.

"A Simple Method for Finding the Jordan Form of a Matrix," IEEE Trans. Automatic Control, AC-17 (1972), pp. 145-146.

"A General Method for Transforming a Matrix into Jordan Canonical Form," (with A.M. Revington), Technical Report No. TR-68-5, Department of Electrical and Computer Engineering, Syracuse University, May 1968.

XUM