

ITERATIVE METHODS FOR THE SOLUTION OF EQUATIONS

J. F. TRAUB  
BELL TELEPHONE LABORATORIES, INCORPORATED  
MURRAY HILL, NEW JERSEY

- ii -

TO SUSANNE

## PREFACE

This book presents a general theory of iteration algorithms for the numerical solution of equations and systems of equations. The relationship between the quantity and quality of information used by an algorithm and the efficiency of the algorithm are investigated. Iteration functions are divided into four classes depending on whether they use new information at one or at several points and whether or not they reuse old information. Known iteration functions are systematized and new classes of computationally effective iteration functions are introduced. Our interest in the efficient use of information is influenced by the widespread use of computing machines.

The mathematical foundations of our subject are treated with rigor but rigor in itself is not the main object. Some of the material is of wider application than to our theory. Belonging to this category are Chapter 3, "The Mathematics of Difference Relations"; Appendix A, "Interpolation"; and Appendix D "Acceleration of Convergence". The inclusion of Chapter 12, "A Compilation of Iteration Functions" permits the use of this book as a handbook of iteration functions. Extensive numerical experimentation was performed on a computer; a selection of the results are reported in Appendix E.

The solution of equations is a venerable subject. Among the mathematicians who have made their contribution are Cauchy, Chebyshev, Euler, Fourier, Gauss, Lagrange, Laguerre, and Newton. E. Schröder wrote a classic paper on the subject in 1870. A glance at the bibliography indicates the level of contemporary interest. Perhaps the most important recent contribution is the book by Ostrowski; papers by Bodewig and Zajta are also outstanding.

Most of the material is new and unpublished. Every attempt has been made to keep the subject in proper historical perspective. Some of the material has been orally presented at meetings of the Association for Computing Machinery in 1961, 1962, 1963, the American Mathematical Society in 1962 and 1963, and the International Congress of Mathematicians in 1962.

I wish to acknowledge with sincere appreciation the assistance I have received from my friends and colleagues at Bell Telephone Laboratories, Incorporated. I am particularly indebted to M. D. McIlroy, J. Morrison, and H. O. Pollak for numerous important suggestions. I want to thank A. J. Goldstein, R. W. Hamming, and E. N. Gilbert for stimulating conversations and valuable comments. My thanks to Professor G. E. Forsythe of Stanford University for his encouragement and comments during the preparation of the manuscript and for reading the final manuscript. My appreciation to S. P. Morgan and Professor A. Ralston

of Stevens Institute of Technology who also read the manuscript. I am particularly grateful to J. Riordan for suggesting numerous improvements in style.

I want to thank Miss Nancy Morris for always digging up just one more reference and Mrs. Helen Carlson for editing the final manuscript. I am grateful to Mrs. Elizabeth Jenkins for her splendid supervision of the preparation of the difficult manuscript and to Miss Joy Catanzaro for her speedy and accurate typing.

To my wife, for her never-failing support and encouragement as well as her assistance in editing and proofreading, I owe the principal acknowledgment.

J. F. TRAUB

TABLE OF CONTENTS

PREFACE

TERMINOLOGY

GLOSSARY OF SYMBOLS

1. GENERAL PRELIMINARIES

    1.1 Introduction

    1.2 Basic Concepts and Notations

        1.21 Some concepts and notations

        1.22 Classification of iteration functions

        1.23 Order

        1.24 Concepts related to order

2. GENERAL THEOREMS ON ITERATION FUNCTIONS

    2.1 The Solution of a Fixed Point Problem

    2.2 Linear and Superlinear Convergence

        2.21 Linear convergence

        2.22 Superlinear convergence

        2.23 The advantages of higher order iteration  
            functions

    2.3 The Iteration Calculus

        2.31 Preparation

        2.32 The theorems of the iteration calculus

- 3. THE MATHEMATICS OF DIFFERENCE RELATIONS
  - 3.1 Convergence of Difference Inequalities
  - 3.2 A Theorem on the Solutions of Certain Inhomogeneous Difference Equations
  - 3.3 On the Roots of Certain Indicial Equations
    - 3.31 The properties of the roots
    - 3.32 An important special case
  - 3.4 The Asymptotic Behavior of the Solutions of Certain Equations
    - 3.41 Introduction
    - 3.42 Difference equations of type 1
    - 3.43 Difference equations of type 2
- 4. INTERPOLATORY ITERATION FUNCTIONS
  - 4.1 Interpolation and the Solution of Equations
    - 4.11 Statement and solution of an interpolation problem
    - 4.12 Relation of interpolation to the calculation of roots
  - 4.2 The Order of Interpolatory Iteration Functions
    - 4.21 The order of iteration functions generated by inverse interpolation
    - 4.22 The equal information case
    - 4.23 The order of iteration functions generated by direct interpolation
  - 4.3 Examples

5. ONE-POINT ITERATION FUNCTIONS

5.1 The Basic Sequence  $E_s$

5.11 The formula for  $E_s$

5.12 An example

5.13 The structure of  $E_s$

5.2 Rational Approximations to  $E_s$

5.21 Iteration functions generated by  
rational approximation to  $E_s$

5.22 The formulas of Halley and Lambert

5.3 A Basic Sequence of Iteration Functions

Generated by Direct Interpolation

5.31 The basic sequence  $\Phi_{o,s}$

5.32 The iteration function  $\Phi_{o,3}$

5.33 Reduction of degree

5.4 The Fundamental Theorem of One-Point Iteration  
Functions

5.5 The Coefficients of the Error Series of  $E_s$

5.51 A recursion formula for the coefficients

5.52 A theorem concerning the coefficients

6. ONE-POINT ITERATION FUNCTIONS WITH MEMORY

6.1 Interpolatory Iteration Functions

6.11 Comments

6.12 Examples

6.2 Derivative Estimated One-Point Iteration

Functions with Memory

6.21 The secant iteration function and its  
generalization

6.22 Estimation of  $f^{(s-1)}$

6.23 Estimation of  $\tilde{g}^{(s-1)}$

6.24 Examples

6.3 Discussion of One-Point Iteration Functions

with Memory

6.31 A conjecture

6.32 Practical considerations

6.33 Iteration functions which do not use  
all available information

6.34 An additional term in the error equation

7. MULTIPLE ROOTS

7.1 Introduction

7.2 The Order of  $E_s$

7.3 The Basic Sequence  $\xi_s$

7.31 Introduction

7.32 The structure of  $\xi_s$

7.33 Formulas for  $\xi_s$

7.4 The Coefficients of the Error Series of  $\xi_s$

7.5 Iteration Functions Generated by Direct  
Interpolation

7.51 The error equation

7.52 On the roots of an indicial equation

7.53 The order

7.54 Discussion and examples

7.6 One-Point Iteration Functions with Memory

7.7 Some General Results

7.8 An Iteration Function of Incommensurate Order

8. MULTIPONT ITERATION FUNCTIONS

8.1 The Advantages of Multipoint Iteration Functions

8.2 A New Interpolation Problem

8.21 A new interpolation formula

8.22 Application to the construction of  
multipoint iteration functions

8.3 Recursively Formed Iteration Functions

8.31 Another theorem of the iteration calculus

8.32 The generalization of the previous theorem

8.33 Examples

8.34 The construction of recursively formed  
iteration functions

8.4 Multipoint Iteration Functions Generated by  
Derivative Estimation

8.5 Multipoint Iteration Functions Generated by  
Composition

8.6 Multipoint Iteration Functions with Memory

9. MULTIPONT ITERATION FUNCTIONS: CONTINUATION
  - 9.1 Introduction
  - 9.2 Multipoint Iteration Functions of Type 1
    - 9.21 The third order case
    - 9.22 The fourth order case
  - 9.3 Multipoint Iteration Functions of Type 2
    - 9.31 The third order case
    - 9.32 The fourth order case
  - 9.4 Discussion of Criteria for the Selection of an Iteration Function
10. ITERATION FUNCTIONS WHICH REQUIRE NO EVALUATION OF DERIVATIVES
  - 10.1 Introduction
  - 10.2 Interpolatory Iteration Functions
    - 10.21 Direct Interpolation
    - 10.22 Inverse Interpolation
  - 10.3 Some Additional Iteration Functions
11. SYSTEMS OF EQUATIONS
  - 11.1 Introduction
  - 11.2 The Generation of Vector-Valued Iteration Functions by Inverse Interpolation
  - 11.3 Error Estimates for Some Vector-Valued Iteration Functions

- 11.31 The generalized Newton iteration function
- 11.32 A third order iteration function
- 11.33 Some other vector-valued iteration functions
- 11.34 A test function
- 11.4 Vector-Valued Iteration Functions which Require No Derivative Evaluations
- 12. A COMPILATION OF ITERATION FUNCTIONS
  - 12.1 Introduction
  - 12.2 One-Point Iteration Functions
  - 12.3 One-Point Iteration Functions with Memory
  - 12.4 Multiple Roots
    - 12.41 Multiplicity known
    - 12.42 Multiplicity unknown
  - 12.5 Multipoint Iteration Functions
  - 12.6 Multipoint Iteration Functions with Memory
  - 12.7 Systems of Equations

## APPENDICES

- A. INTERPOLATION
  - A.1 Introduction
  - A.2 An Interpolation Problem and Its Solution
    - A.21 Statement of the problem
    - A.22 Divided differences
    - A.23 The Newtonian formulation
    - A.24 The Lagrange-Hermite formulation
    - A.25 The interpolation error

A.3 The Equal Information Case

    A.31 The Newtonian formulation

    A.32 The Lagrange-Hermite formulation

    A.33 The interpolation error

A.4 The Approximation of Derivatives

    A.41 Statement of the problem

    A.42 Derivative estimation from the  
        Newtonian formulation

    A.43 Derivative estimation from the  
        Lagrange-Hermite formulation

A.5 The Error in the Approximation of Derivatives

    A.51 Discussion

    A.52 First proof of the error formula

    A.53 Second proof of the error formula

B. ON THE JTH DERIVATIVE OF THE INVERSE FUNCTION

C. SIGNIFICANT FIGURES AND COMPUTATIONAL EFFICIENCY

D. ACCELERATION OF CONVERGENCE

    D.1 Introduction

    D.2 Aitken's  $\delta^2$  Transformation

    D.3 The Steffensen-Householder-Ostrowski Iteration  
        Function

E. NUMERICAL EXAMPLES

E.1 Introduction

E.2 Growth of the Number of Significant Figures

E.3 One-Point and One-Point with Memory Iteration  
Functions

E.4 Multiple Roots

E.5 Multipoint Iteration Functions

E.6 Systems of Equations

F. AREAS FOR FUTURE RESEARCH

BIBLIOGRAPHY

## TERMINOLOGY

	<u>Page</u>
asymptotic error constant	1.2-12
basic sequence	1.2-18
derivative estimated iteration function	6.2-3
informational efficiency	1.2-16
informational usage	1.2-16
interpolatory iteration function	4.1-5
iteration function	1.2-4
multipoint iteration function	1.2-11
multipoint iteration function with memory	1.2-11
one-point iteration function	1.2-10
one-point iteration function with memory	1.2-10
optimal iteration function	1.2-18
optimal basic sequence	1.2-18
order	1.2-12
order is multiplicity-dependent	1.2-13
order is multiplicity-independent	1.2-13

#### GLOSSARY OF SYMBOLS

The following list, which is intended only for reference, contains the symbols which occur most frequently in this book.

		<u>Page</u>
$a_j(x)$	=	1.2-7
$A_j(x)$	=	1.2-7
$A_{\ell,j}^{\gamma_j,n}(t)$		A.2-1
$\alpha$	a zero of $f$	1.2-1
$a_j(y)$	=	1.2-7
$B_{j,m}(x)$	=	1.2-7
$B_{\ell,j}^{s,n}$	=	A.4-6
$\beta_{k,a}$	dominant zero of $g_{k,a}(t)$	3.3-2
$c$	asymptotic error constant	1.2-1
$C[x,\ell]$	binomial coefficient	7.3-7
$c_{\ell,j}^{\gamma_j}(t)$	Newton coefficient	A.2-8
$d$	informational usage	1.2-1
$D_{\ell,j}^s$	$\left[ c_{\ell,j}^s(t) \right]_{t=x_1}^{(s)}$	A.4-2
$e$	$x - \alpha$	1.2-6
$e_1$	$x_1 - \alpha$	1.2-6
EFF	informational efficiency	1.2-1

	<u>Page</u>
$E_s$	a certain family of iteration functions
$*E_{n,s}$	a family of iteration functions generated by derivative estimation
$\frac{1}{m}E_{n,s}$	a family of iteration functions generated by derivative estimation
$e_s$	$e_s(x, f, m) = E_s(x, f^{1/m}, 1)$
$f$	function whose zero is sought
$*f_n^{(s)}$	an estimate of $f^{(s)}$
$f[x_1, \gamma_0; x_{1-1}, \gamma_1; \dots; x_{1-n}, \gamma_n]$	confluent divided difference
$f$	a certain kind of interpolatory function
$\tilde{x}$	the inverse function to $f$
$\frac{1}{m}\tilde{x}_n^{(s)}$	an estimate of $\tilde{x}^{(s)}$
$g_{k,a}(t)$	$t^k - a \sum_{j=0}^{k-1} t^j$
$H_{ij}$	the inverse of the Jacobian matrix
I.F.	iteration function
$I_p$	class of iteration function of order $p$

	<u>Page</u>
$d^I_p$	class of iteration functions with informational usage d and order p 1.2-16
$J_{ij}$	Jacobian matrix 11.1-3
$\lambda_{\ell,s}(m)$	$\varepsilon_s(x,f,m) = \sum \lambda_{\ell,s}(m)(x-\alpha)^\ell$ 7.4-3
$m$	the multiplicity of $\alpha$ 1.2-2
$n$	the number of points at which old information is reused 1.2-17
$v_\ell$	$u(x) = \sum v_\ell(x-\alpha)^\ell$ 5.5-2
$\underline{o}$	order (in sense of order of magnitude) 1.2-8
$\omega_\ell(m)$	$mu(x) = \sum \omega_\ell(m)(x-\alpha)^\ell$ 7.4-1
$p$	order 1.2-12
$P_{n,s}(t)$	hyperosculatory polynomial for $f$ 4.1-2
$\varphi, \Phi, \psi, \Psi$	names of iteration functions 1.2-4
$\Phi_{n,s}$	family of iteration functions generated by inverse interpolation 4.2-8
$\Phi_{n,s}$	family of iteration functions generated by direct interpolation 4.2-22
$\psi_{a,b}$	family of iteration functions generated by rational approximation to $E_s$ 5.2-2
$Q_{n,s}(t)$	hyperosculatory polynomial for $\mathfrak{F}$ 4.1-2
$r$	$= s(n+1)$ 1.2-17

		<u>Page</u>
R	= S(n+1)	6.2-4
$\rho_{s,j}(m)$	$\epsilon_{s+1}(x,f,m) = x - \sum_{j=1}^s \rho_{s,j}(m) Z_j(x,f,1)$	7.3-4
s	s - 1 derivatives are used in many functions	1.2-17
S	= s - 1	6.2-4
$S_{\ell,j}$	Stirling numbers of the first kind	7.3-7
$\sigma_{\ell,j}(m)$	$w_{\ell}(x,f,m) = \sum_{j=1}^{\ell} \sigma_{\ell,j}(m) Z_j(x,f,1)$	7.3-5
$T_{\ell,j}$	Stirling numbers of the second kind	7.3-7
$\tau_{\ell,s}$	$E_s(x) = \sum \tau_{\ell,s}(x-\alpha)^{\ell}$	5.5-4
u(x)	= $f(x)/f'(x)$	1.2-6
v(x)	= $\frac{\phi(x) - \alpha}{(x-\alpha)^p}$	2.3-3
w(x)	= $\frac{\phi(x) - \alpha}{u^p(x)}$	2.3-5
$w_j$	$w_j(x,f,m) = Z_j(x,f^{1/m},1)$	7.3-3
$x_1$	approximant to $\alpha$	1.2-4
$Y_j(x)$	= $\frac{(-1)^{j-1} \mathfrak{F}(j)(y)}{j! [\mathfrak{F}'(y)]^j} \Big _{y=f(x)}$	1.2-7
$Z_j(x)$	= $Y_j(x) u^j(x)$	5.1-13
$Z_{i_1 j_1 \dots j_r}$		11.3-9

	<u>Page</u>
~	approximate equality between numbers
≈	same order of magnitude
$\{x   p(x)\}$	the set of $x$ for which the proposition $p(x)$ is true

1.0-1

## CHAPTER 1

### GENERAL PRELIMINARIES

The basic concepts and notations to be used throughout this book will be introduced in Section 1.2.

## 1.1-1

### 1.1 Introduction

The general area into which this book falls may be labeled algorithmics. By algorithmics we mean the study of algorithms in general and the study of the convergence and efficiency of numerical algorithms in particular.

More specifically, we shall study algorithms for the solution of equations. Our approach will be to examine the relationship between the quantity and quality of information used by an algorithm and the efficiency of that algorithm. We shall investigate the effect of reusing old information and of gathering new information at certain felicitous points.

Our interest in the efficient use of information is influenced by the widespread use of high-speed computing machines. The introduction of computers has meant that many algorithms which were formerly of only academic interest become feasible for calculation. They are, in fact, used many times in many establishments on a wide variety of problems. The efficiency of these algorithms is therefore most important. Furthermore, there are situations where the acquisition of more than a certain amount of data is prohibitively expensive. It is then imperative that as much information as possible be squeezed from the available data. Here again the question of efficiency is of paramount importance.

Iteration algorithms for the solution of equations will be studied in a systematic fashion. In the course of this study, new families of computationally effective iteration algorithms will be introduced and certain well-known iteration algorithms will be identified as special cases. It is hoped that this comprehensive approach will moderate, if not prevent, the rediscovery of special cases. This uniform approach will lead to certain natural classification schemes and will permit the uniform rigorous error analysis and the uniform establishment of convergence criteria for families of iteration algorithms. The final verdict on the usefulness of the new methods will not be available until the new methods have been tried on a variety of problems arising in practice. At present, however, extensive numerical experimentation on test problems support the theoretical error analysis.

Although we shall confine ourselves to the solution of real equations and systems of real equations, the field of potential application of this work is of much broader scope. Thus, analogous techniques may be applied to such problems as the solution of differential and integral equations and the calculation of eigenvalues. The generalization of our results to abstract spaces is of interest. The reader is referred to Appendix F for some additional discussion of this point.

## 1.2 Basic Concepts and Notations

1.21 Some concepts and notations. The symbols to be introduced below are global; they will preserve their meaning for the remainder of this book. A few of these symbols will be used with other than their global meanings (for example, as dummy indices in summation) but context will always make this clear.

Our investigation concerns the approximation of a zero of  $f(x)$ , or equivalently, the approximation of a root of the equation  $f(x) = 0$ . The zero and root formulations will be used interchangeably. There does not seem to be a good phrase for labeling this problem. If one says that one is solving an equation, one may be concerned with the solution of a differential equation. The adjectives algebraic and transcendental are usually used to distinguish between the cases where  $f$  is or is not a polynomial. Perhaps the best generic term is root-finding.

We will restrict ourselves to  $f(x)$  which are real single-valued functions of a real variable, possessing a certain number of continuous derivatives in the neighborhood of a real zero  $\alpha$ . The number of continuous derivatives assumed will vary upwards from zero. The restriction to real zeros is not essential. With the exception of Chapter 11,  $f(x)$  will be a scalar function of a scalar variable; in Chapter 11,  $f(x)$  will be a vector function of a vector variable.

The "independent variable" will sometimes appear explicitly and will sometimes not appear explicitly. This will depend on whether the function or the function evaluated at a point in its domain is what is meant. It will also depend on the readability of formulas. Thus we will use both  $f$  and  $f(x)$ . See Boas [1.2-1, pp. 67-68].

Derivatives of  $f$  will be denoted either by  $f^{(\ell)}$  with  $f^{(0)} = f$ , or by  $f', f'', \dots$ . If  $f'$  does not vanish in a neighborhood of  $\alpha$  and if  $f^{(\ell)}$  is continuous in this neighborhood, then  $f$  has an inverse  $\tilde{f}$ , and  $\tilde{f}^{(\ell)}$  is continuous in a neighborhood of zero.

A zero  $\alpha$  is of multiplicity  $m$  if

$$f(x) = (x-\alpha)^m g(x),$$

where  $g(x)$  is bounded at  $\alpha$  and  $g(\alpha)$  is nonzero. We shall always take  $m$  as a positive integer. Ostrowski [1.2-2, Chap. 5] deals with the case that  $m$  is not a positive integer. If  $m = 1$ ,  $\alpha$  is said to be simple; if  $m > 1$ ,  $\alpha$  is said to be nonsimple. If  $\alpha$  is nonsimple, it is called a multiple zero.

Perhaps the most primitive procedure for approximating a real zero is the following bisection algorithm. Let  $a$  and  $b$  be two points such that  $f(a)f(b) < 0$ . Then  $f$  has at least one real zero of odd multiplicity on  $(a,b)$ . For the purpose of explaining the algorithm it is sufficient to take

### 1.2-3

the interval as  $(0,1)$  and to assume that  $f(0) < 0$ ,  $f(1) > 0$ . Calculate  $f(\frac{1}{2})$ . If  $f(\frac{1}{2}) = 0$ , a zero has been found. If  $f(\frac{1}{2}) < 0$ , the zero lies on  $(\frac{1}{2},1)$  and one next calculates  $f(\frac{3}{4})$ . If  $f(\frac{1}{2}) > 0$ , the zero lies on  $(0,\frac{1}{2})$  and one next calculates  $f(\frac{1}{4})$ . The zero will either be found by this procedure or the zero will be known to lie on an interval of length  $2^{-q}$  after the bisection operation has been applied  $q$  times. By then estimating the value of the zero to be the midpoint of this last interval, the value of the zero will be known to within a maximum error of  $2^{-q-1}$ . Once the zero has been bracketed it takes just  $q$  evaluations of  $f$  to achieve this accuracy. Observe that the accuracy to which the zero may be found is limited only by the accuracy with which  $f$  may be evaluated. Indeed, it is only necessary to be able to decide on the sign of  $f$ .

The bisection algorithm just described is an example of a one-point iteration function with memory, as defined below. Since no use is made of the structure of  $f$ , such as the values of its derivatives, the rate of convergence is not very high. On the other hand, the method is guaranteed to converge. Gross and Johnson [1.2-3] use a property of  $f$  to achieve faster convergence. See also Hooke and Jeeves [1.2-4] and Lehmer [1.2-5]. Kiefer [1.2-6] gives an optimal search strategy for the case of a maximum of a unimodal function.

By using additional information we can do far better than the bisection algorithm. The natural information available are the values of  $f$  and the values of its derivatives. We shall use the words information, samples, and data quite interchangeably.

Let  $x_i, x_{i-1}, \dots, x_{i-n}$  be  $n + 1$  approximants to  $a$ . Let  $x_{i+1}$  be uniquely determined by information obtained at  $x_i, x_{i-1}, \dots, x_{i-n}$ . Let the function that maps  $x_i, x_{i-1}, \dots, x_{i-n}$  into  $x_{i+1}$  be called  $\varphi$ . Thus

$$x_{i+1} = \varphi(x_i, x_{i-1}, \dots, x_{i-n}). \quad (1-1)$$

We call  $\varphi$  an iteration function. The abbreviation I.F. will henceforth be used for iteration function and its plural. Let

$$f^{(\ell)}(x_{i-j}) \equiv f_{i-j}^{(\ell)}.$$

We shall also write  $y_{i-j}$  for  $f(x_{i-j})$  and  $\tilde{y}_{i-j}^{(\ell)}$  for  $\tilde{f}^{(\ell)}(y_{i-j})$ . Since the information used at  $x_{i-j}$  are the values of  $f$  and its derivatives, we may write

$$\varphi = \varphi \left[ x_i, f_i, \dots, f_i^{(\ell_0)}, x_{i-1}, f_{i-1}, \dots, f_{i-1}^{(\ell_1)}, \dots, x_{i-n}, f_{i-n}, \dots, f_{i-n}^{(\ell_n)} \right]. \quad (1-2)$$

We shall not permit I.F. as general as this; the types of I.F. to be studied will be specified in Section 1.22.

Rather than writing  $\phi(x_1, x_{i-1}, \dots, x_{i-n})$ , it is more convenient to write  $\phi$  or  $\phi(x)$ . Observe further that  $\phi$  is a functional depending on  $f$  and should perhaps be written  $\phi(x, f)$ . This notation is not necessary and will therefore not be used except in Chapter 7.

Even the simplest iteration algorithm must consist of initial approximation(s), an I.F., and numerical criteria for deciding when "convergence" is attained. We shall only be concerned with the I.F. Thus, for us, iteration algorithm will be the same as I.F.

Two I.F. are almost universally known. They are the Newton-Raphson I.F. and the secant I.F. For the former,

$$\phi(x_i) = x_i - \frac{f_i}{f'_i}, \quad (1-3)$$

while for the latter,

$$\phi(x_i, x_{i-1}) = x_i - f_i \left[ \frac{x_i - x_{i-1}}{f_i - f_{i-1}} \right], \quad f_i \neq f_{i-1}. \quad (1-4)$$

We shall henceforth call the former Newton's I.F. The secant I.F. is closely related to regula falsi. This last method, as it is usually defined, keeps two approximants which bracket the root; the secant I.F. always uses the latest two approximants.

## 1.2-6

In much of our work we shall deal not with  $f$ , but with a normalized  $f$  defined by

$$u = \frac{f}{f'}, \quad f' \neq 0. \quad (1-5)$$

If  $f' = 0$ ,  $f \neq 0$ , the  $u$  is undefined. If  $f' = 0$ ,  $f = 0$ , which is the case at a multiple zero, we define  $u = 0$ . The importance of  $u$  in our future work cannot be overestimated. One reason for its importance is that

$$\lim_{x \rightarrow \alpha} \left[ \frac{u(x)}{x-\alpha} \right] = \frac{1}{m}.$$

For simple zeros,

$$\lim_{x \rightarrow \alpha} \left[ \frac{u(x)}{x-\alpha} \right] = 1.$$

Now,  $x - \alpha$  is the error in  $x$  but is not known till  $\alpha$  is known. On the other hand,  $u$  is known at each step of the iteration. In the new notation, Newton's I.F. becomes

$$\varphi = x - u. \quad (1-6)$$

For later convenience we introduce some additional notation. Let

$$e_1 = x_1 - \alpha, \quad e = x - \alpha. \quad (1-7)$$

Thus  $e_i$  is the error of the  $i$ th approximant. Let

$$a_j(x) = \frac{f^{(j)}(x)}{j!},$$

$$A_j(x) = \frac{f^{(j)}(x)}{j! f'(x)},$$

$$B_{j,m}(x) = \frac{a_{j+m-1}(x)}{m a_m(x)}, \quad (1-8)$$

$$a_j(y) = \frac{\mathfrak{f}^{(j)}(y)}{j! \mathfrak{f}'(y)},$$

$$Y_j(x) = \left. \frac{(-1)^{j-1} \mathfrak{f}^{(j)}(y)}{j! [\mathfrak{f}'(y)]^j} \right|_{y=f(x)}.$$

The first of these is the Taylor series coefficient for  $f$ , while the second is a "normalized Taylor series coefficient" for  $f$ . The third is a "generalized normalized Taylor series coefficient" which will be used when the multiplicity  $m$  is greater than unity. Note that  $B_{j,1}(x) \equiv A_j(x)$ . The fourth is a "normalized Taylor series coefficient" for  $\mathfrak{f}$ . The last is a Taylor series coefficient for  $\mathfrak{f}$  with a different normalization and with  $f(x)$  substituted after differentiation.

We shall use the symbols  $\rightarrow$ ,  $\underline{0}$ ,  $\approx$  and  $\sim$  according to the following conventions:

1.2-8

If

$$\lim_{i \rightarrow \infty} g(x_i) = c,$$

we shall write

$$g(x_i) \rightarrow c \quad \text{or} \quad g \rightarrow c.$$

If

$$\lim_{x \rightarrow a} g(x) = c,$$

we shall write

$$g(x) \rightarrow c \quad \text{or} \quad g \rightarrow c.$$

How the limit is taken will be clear from context. If

$$\frac{f}{g} \rightarrow c,$$

where C is a nonzero constant, we shall write

$$f = \underline{o}(g)$$

or

$$f \approx Cg.$$

For approximate equality between numbers, the symbol  $\sim$  will be used. Thus

$$\frac{1}{2}(1+\sqrt{5}) \sim 1.618 \sim 1.62.$$

The use of these symbols is illustrated in

EXAMPLE 1-1. Let

$$e_{i+1} = M_i e_i^2 + L_i e_i^3,$$

where

$$e_i \rightarrow 0, \quad M_i \rightarrow K \neq 0, \quad N_i \rightarrow L \neq 0,$$

and where  $e_i$  is nonzero for all finite  $i$ . Then

$$e_{i+1} = M_i e_i^2 + O(e_i^3),$$

since

$$\frac{L_i e_i^3}{e_i^3} \rightarrow L.$$

Also,

$$e_{i+1} \approx M e_i^2,$$

since

$$\frac{e_{i+1}}{e_i^2} = M_i + L_i e_i \rightarrow M.$$

We shall use the notation

$$J = \{x \mid p(x)\}$$

to denote the set of  $x$  for which the proposition  $p(x)$  is true.

Thus

$$J = \{x \mid |x| < 1\}$$

denotes the interior of the unit interval.

1.22 Classification of iteration functions. We shall classify I.F. by the information which they require. Let  $x_{i+1}$  be determined only by new information at  $x_i$ . No old information is reused. Thus

$$x_{i+1} = \phi(x_i). \quad (1-9)$$

Then  $\phi$  will be called a one-point I.F. Most I.F. which have been used for root-finding are one-point I.F. The most commonly known example is Newton's I.F.

Next let  $x_{i+1}$  be determined by new information at  $x_i$  and reused information at  $x_{i-1}, \dots, x_{i-n}$ . Thus

$$x_{i+1} = \phi(x_i; x_{i-1}, \dots, x_{i-n}). \quad (1-10)$$

Then  $\phi$  will be called a one-point I.F. with memory. The semicolon in (1-10) separates the point at which new data are used from the points at which old data are reused. This type of I.F. is now of special interest since the old information is easily saved in the memory (or store or storage) of a computer. The case of practical interest is when the same information, the values of  $f$  and  $f'$  for example, is used at all points. The best-known example of a one-point I.F. with memory is the secant I.F.

## 1.2-11

Now let  $x_{i+1}$  be determined by new information at  $x_1, x_{i-1}, \dots, x_{i-k}$ ,  $k \geq 1$ . No old information is reused. Thus

$$x_{i+1} = \varphi(x_1, x_{i-1}, \dots, x_{i-k}). \quad (1-11)$$

Then  $\varphi$  will be called a multipoint I.F. There are no well-known examples of multipoint I.F. They are being introduced because of certain characteristic limitations on one-point I.F. and one-point I.F. with memory.

Finally let  $x_{i+1}$  be determined by new information at  $x_1, x_{i-1}, \dots, x_{i-k}$  and reused information  $x_{i-k-1}, \dots, x_{i-n}$ . Thus

$$x_{i+1} = \varphi(x_1, x_{i-1}, \dots, x_{i-k}; x_{i-k-1}, \dots, x_{i-n}), \quad n > k. \quad (1-12)$$

Then  $\varphi$  will be called a multipoint I.F. with memory. The semicolon in (1-12) separates the points at which new data are used from the points at which old data are reused. There are no well-known examples of multipoint I.F. with memory.

1.23 Order. We turn to the important concept of the order of an I.F. Let  $x_0, x_1, \dots, x_i, \dots$  be a sequence converging to  $\alpha$ . Let  $e_i = x_i - \alpha$ . If there exists a real number  $p$  and a nonzero constant  $C$  such that

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow C, \quad (1-13)$$

then  $p$  is called the order of the sequence and  $C$  is called the asymptotic error estimate.

The remainder of this section will consist of commentary on this definition. We wish to associate the concept of order with the I.F. which generates the  $x_i$ . To emphasize this point, we can write (1-13) as

$$\frac{|\varphi(x) - \alpha|}{|x - \alpha|^p} \rightarrow C. \quad (1-14)$$

We will associate an order to an I.F. whether or not the sequence generated by  $\varphi$  converges. The order assigned to an I.F. is the order of the sequence it generates when the sequence converges.

Recall that  $\varphi$  is a functional which depends on  $f$ . Hence the order of  $\varphi$  may be different for different classes of  $f$ . For the type of  $f$  and  $\varphi$  that we shall study, we shall insist that  $\varphi$  be of a certain order for all  $f$  whose zeros are

of a certain multiplicity. We shall always assume that we are in the neighborhood of an isolated zero of  $f$  and that the order possibly depends on the multiplicity of the zero but is otherwise independent of  $f$ . (With the exception of Chapter 7, the zero will always be simple unless the contrary is explicitly stated.) To indicate that  $\varphi$  belongs to the class of I.F. of order  $p$ , we shall write

$$\varphi \in I_p. \quad (1-15)$$

The following definitions relate order to multiplicity. If the order is the same for zeros of all multiplicities, we say that the order is multiplicity-independent. If the order depends on the multiplicity we say that the order is multiplicity-dependent. If in particular the order is greater than one for simple zeros and one for all nonsimple zeros, we say that the order is linear for all nonsimple zeros. (The adjectives linear and quadratic will sometimes be used instead of first and second.)

Observe that if the order exists then it is unique. Assume that a convergent sequence has two orders,  $p_1$  and  $p_2$ . Let  $p_2 = p_1 + \delta$ ,  $\delta > 0$ . Then

$$\lim_{i \rightarrow \infty} \left[ \frac{|x_{i+1} - \alpha|}{|x_i - \alpha|^{p_2}} \right] = \lim_{x \rightarrow \infty} \left[ \frac{|x_{i+1} - \alpha|}{|x_i - \alpha|^{p_1+\delta}} \right] = c_2,$$

and

$$\lim_{i \rightarrow \infty} \left[ \frac{|x_{i+1} - \alpha|}{|x_i - \alpha|^p} \right] = 0,$$

which contradicts the assumption that the last limit is nonzero.

If the order is integral, then the absolute values may be dropped in the definition of order. We shall find that I.F. with memory never have integral order. If  $\phi^{(p+1)}$  is continuous and if

$$\phi(x) - \alpha = C(x-\alpha)^p [1 + \underline{o}(x-\alpha)], \quad (1-16)$$

then  $\phi$  is of order  $p$  and  $C = \phi^{(p)}(\alpha)/p!$ . Equation (1-16) may be written as

$$\phi(x) - \alpha \approx C(x-\alpha)^p.$$

In his classic paper of 1870, E. Schröder [1.2-7] defined order as follows:  $\phi(x)$  is of order  $p$  if

$$\phi(\alpha) = \alpha; \quad \phi^{(j)}(\alpha) = 0, \quad j = 1, 2, \dots, p-1; \quad \phi^{(p)}(\alpha) \neq 0. \quad (1-17)$$

This definition is only valid for I.F. of one variable with  $p$  continuous derivatives. Although many authors have followed Schröder's lead, we prefer to state (1-17) in the conclusion of Theorem 2-2. However, a generalization of (1-17) will serve as the definition of order for the case of systems of equations in Chapter 11.

1.2-15

The advantage of high-order I.F. will be discussed  
in Section 2.23.

EXAMPLE 1-2. For Newton's method,

$$\frac{e_{i+1}}{e_i^2} \rightarrow A_2(\alpha), \quad A_2 = \frac{f''}{2f'}.$$

For the secant method,

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow A_2(\alpha), \quad p = \frac{1}{2}(1+\sqrt{5}) \sim 1.62.$$

In Section 7-8 we will give an example of an I.F.  
whose order is incommensurate with the scale defined by (1-13).

1.24 Concepts related to order. A measure of the information used by an I.F. and a measure of the efficiency of the I.F. are required. A natural measure of the former is the informational usage,  $d$ , of an I.F. which we define as the number of new pieces of information required per iteration. Since the information to be used are the values of  $f$  and its derivatives, the informational usage is the total number of new derivative evaluations per iteration. We use here the convention that the function is the zeroth derivative. Ostrowski [1.2-8, p. 19] has suggested the "Horner" as the unit of informational usage. To indicate that  $\varphi$  belongs to the class of I.F. of order  $p$  and informational usage  $d$  we write

$$\varphi \in {}_d I_p. \quad (1-18)$$

To obtain a measure of the efficiency, we make the following definition: The informational efficiency,  $EFF$ , is the order divided by the informational usage. Thus

$$EFF = \frac{p}{d}. \quad (1-19)$$

Another measure of efficiency called the computational efficiency, which takes into account the "cost" of calculating different derivatives, will be discussed in Appendix C. An alternative definition of informational efficiency is

$$*EFF = p^{1/d}. \quad (1-20)$$

Ostrowski [1.2-9, p. 20] calls \*EFF an efficiency index; see Appendix C for a discussion of \*EFF. Our definition of informational efficiency was chosen because it permits the simple statement of certain important results.

EXAMPLE 1-3. For Newton's I.F.,

$$p = 2, \quad d = 2, \quad \text{EFF} = 1, \quad *EFF = \sqrt{2}.$$

For the secant I.F.,

$$p = \frac{1}{2}(1 + \sqrt{5}) \sim 1.62, \quad d = 1, \quad \text{EFF} = *EFF = \frac{1}{2}(1 + \sqrt{5}).$$

For one-point I.F. and one-point I.F. with memory there can be only one new evaluation of each derivative. If the first  $s - 1$  derivatives are used, the informational usage will be  $s$  and

$$\text{EFF} = \frac{p}{s}. \quad (1-21)$$

Let  $n$  be the number of points at which old information is reused in a one-point I.F. with memory. Let

$$r = s(n+1). \quad (1-22)$$

Hence  $r$  is the product of the number of pieces of new information with the total number of points at which information is used. The quantities  $s$ ,  $n$ , and  $r$  will characterize certain families of I.F. These symbols will occasionally be used with other meanings as long as there is no danger of confusion.

We shall prove in Section 5-4 that for one-point I.F.,  $\text{EFF} \leq 1$ . We anticipate this result to define a one-point I.F. as optimal if  $\text{EFF} = 1$ . For optimal one-point I.F.,  $p = d = s$ . A basic sequence of I.F. is an infinite sequence of I.F. such that the  $p$ th member of the sequence is of order  $p$ . This concept is defined only for I.F. of integral order. An optimal basic sequence is a basic sequence all of whose members are optimal. In Section 5-1 the properties of a particular basic sequence, which will be used extensively throughout this book, will be developed.

## 2.0-1

### CHAPTER 2

#### GENERAL THEOREMS ON ITERATION FUNCTIONS

In this chapter I.F. will be discussed without regard to their structure. In Section 2.1 the existence and uniqueness of the iterative solution of a fixed point problem will be proven under the assumption that the I.F. satisfies a Lipschitz condition. In Section 2.2 the difference between linear and superlinear convergence will be studied in some detail while in Section 2.3 an "iteration calculus" will be developed for I.F. which possess a certain number of continuous derivatives.

## 2.1-1

2.1 The Solution of a Fixed Point Problem

We propose to study the solution of

$$\varphi(x) = x \quad (2-1)$$

by the iteration

$$x_{i+1} = \varphi(x_i). \quad (2-2)$$

If  $\alpha$  satisfies (2-1), then  $\alpha$  is called a fixed point of  $\varphi$ .

The problem of finding the fixed points of a function occurs in many branches of mathematics. To see its relation to the solution of  $f(x) = 0$  we proceed as follows. Let  $g(x)$  be any function such that  $g(\alpha)$  is finite and nonzero. Let

$$\varphi(x) = x - f(x)g(x). \quad (2-3)$$

Then  $\alpha$  is a solution of  $f(x) = 0$  if and only if  $\alpha$  is a solution of (2-1).

We shall show that under certain hypotheses the fixed point problem has a unique solution and that the iteration defined by (2-2) converges to this solution. Because of the restrictive nature of our hypotheses, the proof will be very simple. Many other proofs have been given both for real functions and in abstract spaces and under a variety of hypotheses. See, for example Antosiewicz and Rheinboldt [2.1-1], Collatz [2.1-2, Chap. II], Coppel [2.1-3], Ford [2.1-4], Householder [2.1-5, Sect. 3.3], John [2.1-6, Chap. 4, Sect. 6], and Ostrowski [2.1-7, Chap. 4].

## 2.1-2

We assume that  $\varphi$  is defined on some closed and bounded interval  $J = [a, b]$  and that its values are in the same interval. Then if  $x_0$  is in  $J$ , all the  $x_i$  are in  $J$ . To guarantee that (2-1) has a solution we must assume the continuity of  $\varphi$ . We have

LEMMA 2-1. Let  $\varphi$  be a continuous function from  $J = [a, b]$  to  $J$ . Then there exists an  $\alpha$ ,  $a \leq \alpha \leq b$ , such that  $\varphi(\alpha) = \alpha$ .

PROOF. Since the function is from  $J$  to  $J$ ,

$$\varphi(a) \geq a, \quad \varphi(b) \leq b.$$

Let  $h(x) = \varphi(x) - x$ . Then

$$h(a) \geq 0, \quad h(b) \leq 0,$$

and the intermediate value theorem guarantees the existence of  $\alpha$  such that  $h(\alpha) = 0$ . The result follows immediately.

In order to draw additional conclusions we must impose an additional condition on  $\varphi$ . Let

$$|\varphi(s) - \varphi(t)| \leq L|s-t|, \quad 0 \leq L < 1, \quad (2-4)$$

for arbitrary points  $s$  and  $t$  in  $J$ . Observe that this Lipschitz condition implies continuity. We can now show that the solution of  $\varphi(x) = x$  is unique.

## 2.1-3

LEMMA 2-2. Let  $\varphi$  be a function from  $J$  to  $J$  which satisfies the Lipschitz condition (2-4). Then  $\varphi(x) = x$  has at most one solution.

PROOF. Assume that there are two distinct solutions,  $\alpha_1$  and  $\alpha_2$ . Then

$$\alpha_1 = \varphi(\alpha_1), \quad \alpha_2 = \varphi(\alpha_2)$$

and

$$\alpha_1 - \alpha_2 = \varphi(\alpha_1) - \varphi(\alpha_2).$$

An application of (2-4) yields

$$|\alpha_1 - \alpha_2| = |\varphi(\alpha_1) - \varphi(\alpha_2)| \leq L|\alpha_1 - \alpha_2| < |\alpha_1 - \alpha_2|,$$

which is a contradiction.

The existence and uniqueness of a solution  $\alpha$  of the equation  $\varphi(x) = x$  has now been verified. It is a simple matter to show that the sequence defined by  $x_{i+1} = \varphi(x_i)$  converges to this solution.

THEOREM 2-1. Let  $J$  be a closed, bounded interval and let  $\varphi$  be a function from  $J$  to  $J$  which satisfies the Lipschitz condition

## 2.1-4

$$|\varphi(s) - \varphi(t)| \leq L|s-t|, \quad 0 \leq L < 1,$$

for arbitrary points  $s$  and  $t$  in  $J$ . Let  $x_0$  be an arbitrary point of  $J$  and let  $x_{i+1} = \varphi(x_i)$ . Then the sequence  $\{x_i\}$  converges to the unique solution of  $\varphi(x) = x$  in  $J$ .

PROOF. Lemmas 2-1 and 2-2 assure us of the existence and uniqueness of a solution  $\alpha$  of the equation  $\varphi(x) = x$ . This solution is the candidate for the limit of the sequence defined by  $x_{i+1} = \varphi(x_i)$ . Thus

$$x_{i+1} - \alpha = \varphi(x_i) - \alpha = \varphi(x_i) - \varphi(\alpha).$$

From the Lipschitz condition we conclude that

$$|x_{i+1} - \alpha| \leq L|x_i - \alpha|, \quad L < 1.$$

Thus

$$|x_i - \alpha| \leq L^i |x_0 - \alpha|$$

and  $x_i \rightarrow \alpha$ .

If one did not know that  $\varphi(x) = x$  had a unique solution, one would have to show that the  $x_i$  satisfy the Cauchy criterion. A proof based on the Cauchy criterion may be found in Henrici [2.1-8, Chap. 4].

An estimate of the error of the  $p$ th approximant which depends only on the first two approximants and the Lipschitz constant may be derived as follows. From the Lipschitz condition we can conclude that for all  $i$ ,

$$|x_{i+1} - x_i| \leq L^i |x_1 - x_0|. \quad (2-5)$$

Let  $p$  and  $q$  be arbitrary positive integers. Then

$$x_{p+q} - x_p = (x_{p+q} - x_{p+q-1}) + (x_{p+q-1} - x_{p+q-2}) + \dots + (x_{p+1} - x_p)$$

and

$$|x_{p+q} - x_p| \leq |x_{p+q} - x_{p+q-1}| + |x_{p+q-1} - x_{p+q-2}| + \dots + |x_{p+1} - x_p|.$$

An application of (2-5) shows that

$$|x_{p+q} - x_p| \leq L^p(1+L+\dots+L^{q-1}) |x_1 - x_0|,$$

or

$$|x_{p+q} - x_p| \leq \frac{L^p(1-L^q)}{1-L} |x_1 - x_0|. \quad (2-6)$$

Let  $q \rightarrow \infty$ . Then

$$|x_p - \alpha| \leq \frac{L^p}{1-L} |x_1 - x_0|.$$

This gives us a bound on the error of the  $p$ th approximant.

Observe that if  $L$  is close to unity, convergence may be very slow. A good part of our effort from here on in will be devoted to the construction of I.F. which converge rapidly.

To do this we shall have to impose additional conditions on  $\phi$ .

2.2-1

## 2.2 Linear and Superlinear Convergence

In Section 2.1 the analytical problem,  $\varphi(x) = x$ , suggested the iterative procedure  $x_{i+1} = \varphi(x_i)$ . This is not typical of the problems on which we will work. In general we will be given the equation  $f = 0$  and we will want to construct a functional  $\varphi$  which will be used as an I.F. In Section 2.1 we showed that if  $\varphi$  satisfies a Lipschitz condition, then the iteration will converge to the unique solution of the fixed point problem. From now on, rather than seeking weak sufficient conditions under which the iterative sequence converges, we will be interested in constructing I.F. such that the sequence converges rapidly. We will be quite ready to impose rather strong conditions. In particular, we shall assume that  $f$  has a zero in the interval in which we work.

2.21 Linear convergence. Let us assume that  $\varphi'$  is continuous in a neighborhood of  $\alpha$ . We will certainly insist that

$$\varphi(\alpha) = \alpha. \quad (2-7)$$

For if  $x_1 \rightarrow \alpha$  and  $\varphi$  is continuous, then

$$\alpha = \lim_{i \rightarrow \infty} x_{i+1} = \lim_{i \rightarrow \infty} \varphi(x_i) = \varphi\left(\lim_{i \rightarrow \infty} x_i\right) = \varphi(\alpha).$$

Now

$$x_{i+1} = \varphi(x_i) = \varphi(\alpha) + \varphi'(\xi_i)(x_i - \alpha),$$

where  $\xi_i$  lies in the interval determined by  $x_i$  and  $\alpha$ . An application of (2-7) yields

$$x_{i+1} - \alpha = \varphi'(\xi_i)(x_i - \alpha). \quad (2-8)$$

We shall require that  $|\varphi'(\xi_i)| \leq K < 1$  in the neighborhood of  $\alpha$ . Since  $\varphi'$  is continuous it will be sufficient to require that  $|\varphi'(\alpha)| \leq K < 1$ . Then there is a neighborhood of  $\alpha$  such that

$$|\varphi'(x)| \leq L, \quad 0 \leq L < 1$$

and

$$|x_{i+1} - \alpha| \leq L|x_i - \alpha|.$$

Then

$$|x_i - \alpha| \leq L^i |x_0 - \alpha|$$

and  $x_i \rightarrow \alpha$ .

If  $\alpha$  is such that for every starting point  $x_0$  in a sufficiently small neighborhood of  $\alpha$  the sequence  $\{x_i\}$  converges to  $\alpha$ , then  $\alpha$  is called a point of attraction. This terminology is due to J. F. Ritt. See Ostrowski [2.2-1, p. 26] for a discussion of this and the definition of point of repulsion. In this terminology we can state our result as follows. If  $\varphi'$  is continuous in a neighborhood of  $\alpha$  and if  $\varphi(\alpha) = \alpha$  and  $|\varphi'(\alpha)| \leq L < 1$ , then  $\alpha$  is a point of attraction.

We can conclude something more. Let  $\varphi'(\alpha)$  be different from zero. Then  $\varphi'$  does not vanish in a neighborhood of  $\alpha$ . Let  $e_1 = x_1 - \alpha$ . Then (2-8) may be written as

$$e_{i+1} = \varphi'(\xi_i)e_i.$$

It is clear that if  $e_0$  is not zero and  $\varphi'$  does not vanish, then  $e_i$  does not vanish for any finite  $i$ . That is, the iteration cannot converge in a finite number of steps as long as the iterants lie in the neighborhood of  $\alpha$  where  $\varphi'$  does not vanish. Hence

$$\frac{e_{i+1}}{e_i} = \varphi'(\xi_i)$$

and

$$\frac{e_{i+1}}{e_i} \rightarrow \varphi'(\alpha).$$

This is the case of linear or first-order convergence.

2.2-4

In the case of superlinear convergence it is not necessary to require that  $|\phi'(\alpha)| < 1$ ; the iteration always converges in some neighborhood of  $\alpha$ .

2.22 Superlinear convergence. We now assume that  $\varphi^{(p)}$ ,  $p \geq 1$ , is continuous. As before we require that  $\varphi(\alpha) = \alpha$ . Then

$$x_{i+1} = \varphi(x_i) = \alpha + \varphi'(\alpha)(x_i - \alpha) + \dots + \frac{\varphi^{(p)}(\xi_i)}{p!} (x_i - \alpha)^p,$$

where  $\xi_i$  lies in the interval determined by  $x_i$  and  $\alpha$ . It should be clear that  $\varphi$  is of order  $p$  only if  $\varphi^{(j)}(\alpha) = 0$ ,  $j = 1, 2, \dots, p-1$ , and if  $\varphi^{(p)}(\alpha)$  is nonzero. Hence we impose the following conditions on  $\varphi$ :

$$\varphi(\alpha) = \alpha; \quad \varphi^{(j)}(\alpha) = 0, \quad j = 1, 2, \dots, p-1; \quad \varphi^{(p)}(\alpha) \neq 0. \quad (2-9)$$

Then

$$e_{i+1} = \frac{\varphi^{(p)}(\xi_i)}{p!} e_i^p,$$

where  $e_i = x_i - \alpha$ . Since  $\varphi^{(p)}(\alpha)$  is nonzero,  $\varphi^{(p)}$  does not vanish in some neighborhood of  $\alpha$ . Then the algorithm cannot converge in a finite number of steps provided that  $e_0$  is nonzero and that the iterants lie in the neighborhood of  $\alpha$  where  $\varphi^{(p)}$  does not vanish. (See also the beginning of Section 2.23.) Furthermore,

$$\frac{e_{i+1}}{e_i^p} \rightarrow \frac{\varphi^{(p)}(\alpha)}{p!}. \quad (2-10)$$

## 2.2-6

Hartree [2.2-2] defines the order of an I.F. by the conditions (2-9). This definition of order cannot be used for one-point I.F. with memory. Hence we prefer to state these conditions as part of a theorem. We summarize the results in

THEOREM 2-2. Let  $\varphi$  be an I.F. such that  $\varphi^{(p)}$  is continuous in a neighborhood of  $\alpha$ . Let  $e_1 = x_1 - \alpha$ . Then  $\varphi$  is of order  $p$  if and only if

$$\varphi(\alpha) = \alpha; \quad \varphi^{(j)}(\alpha) = 0, \quad j = 1, 2, \dots, p-1; \quad \varphi^{(p)}(\alpha) \neq 0.$$

Furthermore,

$$\frac{e_{i+1}}{e_1^p} \rightarrow \frac{\varphi^{(p)}(\alpha)}{p!}.$$

Recall that we assign an order to  $\varphi$  whether or not the sequence generated by  $\varphi$  converges. Sufficient conditions for a sequence to converge are derived below.

EXAMPLE 2-1. Let  $\varphi = x - f/f'$ , (Newton's I.F.).

Let  $f'''$  be continuous. A direct calculation reveals that  $\varphi(\alpha) = \alpha$ ,  $\varphi'(\alpha) = 0$ ,  $\varphi''(\alpha) = f''(\alpha)/f'(\alpha) \neq 0$ . Hence Newton's I.F. is second order and

$$\frac{e_{i+1}}{e_i^2} \rightarrow A_2(\alpha), \quad A_2 = \frac{f''}{2f'}.$$

We require that  $f'''$  be continuous in order to satisfy the hypothesis of Theorem (2-2) that  $\varphi''$  be continuous. It will be shown in Chapter 4 that we only need  $f''$  continuous.

We observed in Section 2.21 that a sequence formed by a linear I.F. need not converge in any neighborhood of  $\alpha$ . We shall show that a sequence formed by a superlinear I.F. always converges in some neighborhood of  $\alpha$ . In the following analysis the cases of linear and superlinear convergence are handled jointly.

We start our analysis from

$$e_{i+1} = M_i e_i^p, \quad M_i = \frac{\varphi^{(p)}(\xi_i)}{p!}. \quad (2-11)$$

Let

$$J = \left\{ x \mid |x - \alpha| \leq \Gamma \right\}$$

(the notation is specified at the end of Section 1.21) and let  $\varphi^{(p)}$  be continuous on  $J$ . Let  $x_0 \in J$  and let

$$\frac{|\varphi^{(p)}(x)|}{p!} \leq M$$

for all  $x \in J$ .

Since  $x_0 \in J$ ,

$$|M_0| \leq M, \quad |e_0| \leq \Gamma.$$

Hence

$$|e_1| = |M_0| |e_0|^p \leq M |e_0|^{p-1} |e_0| \leq M \Gamma^{p-1} \Gamma.$$

Let

$$M \Gamma^{p-1} = L < 1. \quad (2-12)$$

Then

$$|e_1| \leq L \Gamma < \Gamma.$$

Hence  $x_1 \in J$ . We now proceed to prove by induction that if (2-12) holds, then for all  $i$ ,

$$x_i \in J, \quad |e_i| \leq L^i \Gamma. \quad (2-13)$$

Assume that (2-13) holds. Then

$$|e_{i+1}| = |M_i| |e_i|^p \leq M |e_i|^{p-1} |e_i|$$

$$\leq M \Gamma^{p-1} |e_i| \leq L L^i \Gamma = L^{i+1} \Gamma < \Gamma.$$

2.2-9

Hence (2-13) holds for  $i + 1$  which completes the proof by induction. Since

$$|e_i| \leq L^i \Gamma, \quad L < 1,$$

we conclude that  $x_i \rightarrow \alpha$ . We have proven

THEOREM 2-3. Let  $\varphi^{(p)}$  be continuous on the interval  $J$ ,

$$J = \{x \mid |x - \alpha| \leq \Gamma\}.$$

Let  $x_0 \in J$  and let

$$\frac{|\varphi^{(p)}(x)|}{p!} \leq M$$

for all  $x \in J$ . Let

$$M\Gamma^{p-1} < 1.$$

Then for all  $i$ ,  $x_i \in J$  and  $x_i \rightarrow \alpha$ .

### 2.23 The advantages of higher order iteration

functions. If  $\phi'$  is continuous in a neighborhood of  $\alpha$  and if  $\phi$  is first order, then the sequence generated by  $\phi$  will converge to  $\alpha$  if and only if  $|\phi'(\alpha)| < 1$  unless one of the  $x_i$  becomes equal to  $\alpha$ . That the last restriction is necessary is shown by the example

$$\phi(x) = \begin{cases} 2x & \text{for } |x| \leq 1 \\ 0 & \text{for } |x| > 1 \end{cases}.$$

Here  $\phi'$  is certainly continuous in the neighborhood of zero and  $\phi'(0) = 2$  but the iteration converges for all  $x_0$  in the unit interval. Barring such cases,  $\phi'(\alpha)$  will have to be less than unity for convergence. If, on the other hand,  $p > 1$  and  $\phi^{(p)}$  is continuous in the neighborhood of  $\alpha$ , then there is always a neighborhood of  $\alpha$  where the iteration is guaranteed to converge.

This is one of a number of advantages of superlinear convergence over linear convergence. Perhaps the most important advantage is roughly summarized by the statement that  $x_{i+1}$  agrees with  $\alpha$  to  $p$  times as many significant figures as  $x_i$ . See Appendix C.

The higher order I.F. will often require fewer total samples of  $f$  and its derivatives. Recall that the informational efficiency of an I.F. is given by the ratio of

the order to the number of pieces of new information required per iteration. As will be shown in Chapter 5, there exist I.F. of all orders whose informational efficiency is unity. We defined such I.F. as optimal. To simplify matters, let  $\phi$  and  $\psi$  be optimal I.F. of orders 2 and 3 respectively. Let  $x_3$  be generated from  $x_0$  by the application of  $\phi$  three times and let  $y_2$  be generated from  $x_0$  by the application of  $\psi$  twice. Although each process requires six pieces of information,  $x_3 - \alpha = \underline{O}[(x_0 - \alpha)^8]$  whereas  $y_2 - \alpha = \underline{O}[(x_0 - \alpha)^9]$ . Ehrmann [2.2-3] considers a certain family of I.F. whose members have arbitrary order. He discusses which order is "best" under certain assumptions.

The main drawbacks to the use of high order one-point I.F. are the increasing complexity of the formulas and the need for evaluating higher derivatives of  $f$ . In Chapters 8 and 9 we shall see how these disadvantages may be overcome by using multipoint I.F.

Observe that if  $f$  satisfies a differential equation it may be cheaper to calculate the derivatives of  $f$  from the differential equation than from  $f$  itself. In particular, let  $f$  satisfy a second order differential equation. After  $f$  and  $f'$  have been calculated,  $f''$  is available from the differential equation. By differentiating the differential equation one may calculate  $f'''$ . This process may be continued as long as it is feasible.

It might appear that it is difficult to generate I.F. of higher order for all  $f$ . Such is not the case and algorithms for constructing I.F. of arbitrary order have been given by many authors. See, for example, Bodewig [2.2-4], Ehrmann [2.2-5], Ludwig [2.2-6], E. Schröder [2.2-7], Schwerdtfeger [2.2-8], and Zajta [2.2-9]. Many of these algorithms depend on the construction of I.F. such that

$$\varphi(\alpha) = \alpha, \quad \varphi^{(j)}(\alpha) = 0, \quad j = 1, 2, \dots, p-1. \quad (2-14)$$

We shall propose numerous techniques for generating I.F. of arbitrary order which also satisfy certain other criteria. It will turn out that it is not necessary to use (2-14) to construct these I.F.

From two I.F. of first order, it is possible to construct an I.F. of second order by the technique of Steffensen-Householder-Ostrowski. The reader is referred to Appendix D.

In many branches of numerical analysis one constructs approximations by insisting that the approximation be exact for a certain number of powers of  $x$ . There is one class of I.F., those generated by direct interpolation (Section 4.23), for which the I.F. yield the solution  $\alpha$  in one step for all polynomials of less than a certain degree. This is not a general property of I.F. In fact, powers of  $x$  could not be used if for no other reason than that  $x^m$  has a root of multiplicity  $m$ .

2.2-13

EXAMPLE 2-2. Let

$$\phi = x - \frac{f}{f'}.$$

Then  $\phi$  is second order and yields  $\alpha$  in one step if  $f$  is any linear polynomial. Let

$$\psi = x - \frac{f}{f'} + f^2.$$

Then  $\psi$  is second order. Let  $f = x$ . Then  $\psi$  will not yield the answer in one step and will not converge if  $|x_0| > 1$ .

### 2.3-1

#### 2.3 The Iteration Calculus

We will develop a calculus of I.F. The results to be proven in this section will play a key role in much of the later work. They will not apply to I.F. with memory. The nature of our hypotheses will be such that the order will be integral. No assumptions will be made as to the structure of the I.F. Theorems 8-1 and 8-2 also belong to the iteration calculus but are deferred to Chapter 8 because their applications are in that Chapter. Some of these results, for the case of simple roots, may be found in the papers by Zajta [2.3-1] and Ehrmann [2.3-2].

## 2.3-2

2.31 Preparation. In this section we do not restrict ourselves to simple zeros. The multiplicity of the zero is denoted by  $m$  and we indicate that  $\phi$  is a member of the class of I.F. whose order is  $p$  by writing  $\phi \in I_p$ . We always insist that  $\phi$  is of order  $p$  for all functions  $f$  whose zeros have a certain multiplicity. However  $\phi$  may be of order  $p$  for simple zeros but of a different order for multiple zeros. Hence our hypotheses contain phrases such as "Let  $\phi \in I_p$  for some set of values of  $m$ ."

EXAMPLE 2-3. Let  $u = f/f'$  and let

$$\varphi_1 = x - u, \quad \varphi_2 = x - mu, \quad \varphi_3 = x - \frac{u}{u'}.$$

Then  $\varphi_1 \in I_2$  for simple zeros whereas  $\varphi_1 \in I_1$  for multiple zeros. On the other hand,  $\varphi_2 \in I_2$  for zeros of multiplicity  $m$  and  $\varphi_3 \in I_2$  for zeros of arbitrary multiplicity. See Chapter 7 for details.

Let  $\phi^{(p)}$  be continuous in a neighborhood of  $\alpha$ .

If  $\phi \in I_p$ , then from Theorem 2-2,

$$\phi(\alpha) = \alpha; \quad \phi^{(j)}(\alpha) = 0 \quad \text{for } j = 1, 2, \dots, p-1; \quad \phi^{(p)}(\alpha) \neq 0. \quad (2-15)$$

2.3-3

The expansion of  $\varphi(x)$  in a Taylor series about  $\alpha$  yields

$$\varphi(x) - \alpha = \frac{\varphi^{(p)}(\xi)}{p!} (x-\alpha)^p, \quad (2-16)$$

where  $\xi$  lies in the interval determined by  $\alpha$  and  $x$ . Now,  $\xi$  is not a continuous function of  $x$  but  $\varphi^{(p)}(\xi)$  may be defined as a continuous function of  $x$  as follows. Let

$$V(x) = \frac{\varphi(x) - \alpha}{(x-\alpha)^p} \quad \text{for } x \neq \alpha, \quad V(\alpha) = \frac{\varphi^{(p)}(\alpha)}{p!}.$$

Then  $V(x)$  is continuous wherever  $\varphi(x)$  is continuous and  $x \neq \alpha$ . A p-fold application of L'Hospital's rule reveals that

$$\lim_{x \rightarrow \alpha} V(x) = \frac{\varphi^{(p)}(\alpha)}{p!}.$$

Hence

$$\varphi(x) - \alpha = V(x)(x-\alpha)^p, \quad (2-17)$$

where  $V(x)$  is a continuous function of  $x$  and  $V(\alpha)$  is nonzero.

Thus one may characterize I.F. of order  $p$  by the condition that  $\varphi(x) - \alpha$  has a zero of multiplicity  $p$  at  $\alpha$ .

By comparing (2-17) with (1-14) we observe that

$$C = V(\alpha)$$

where  $C$  is the asymptotic error constant.

## 2.3-4

Equation (2-17) may be put into a form which is more convenient for many applications. Let  $f'$  be continuous in a neighborhood of  $\alpha$  and let  $\alpha$  be a simple root. Then  $f'(\alpha)$  is nonzero and

$$f(x) = f'(\alpha)(x-\alpha).$$

Let

$$\lambda(x) = \frac{f(x)}{x-\alpha} \quad \text{for } x \neq \alpha, \quad \lambda(\alpha) = f'(\alpha).$$

Then

$$f(x) = \lambda(x)(x-\alpha), \quad (2-18)$$

where  $\lambda(x)$  is continuous if  $f$  is continuous and  $\lambda(\alpha) \neq 0$ .

From (2-17) and (2-18) we conclude that

$$\varphi(x) - \alpha = T(x)f^p(x), \quad (2-19)$$

where  $T(x) = V(x)/\lambda^p(x)$ . Then  $T(x)$  is continuous wherever  $\varphi^{(p)}(x)$  and  $f'(x)$  are continuous and  $f'(x)$  does not vanish.

Furthermore,

$$T(\alpha) = \frac{\varphi^{(p)}(\alpha)}{p! [f'(\alpha)]^p} \neq 0.$$

If the multiplicity  $m$  is greater than unity, then  $f$  is no longer proportional to  $x - \alpha$  and so we cannot obtain an expression like (2-19) which involves  $f$ . However,

2.3-5

$u = f/f'$  has only simple zeros and hence (2-19) is easily generalized. Let  $f^{(m)}$  be continuous and let  $\alpha$  have multiplicity  $m$  with  $m \geq 1$ . Proceeding as before we conclude the existence of continuous functions  $\lambda_1(x)$  and  $\lambda_2(x)$  such that

$$f(x) = \lambda_1(x)(x-\alpha)^m, \quad f'(x) = \lambda_2(x)(x-\alpha)^{m-1},$$

with

$$\lambda_1(\alpha) = \frac{f^{(m)}(\alpha)}{m!}, \quad \lambda_2(\alpha) = \frac{f^{(m)}(\alpha)}{(m-1)!} \neq 0.$$

Hence

$$u(x) = \frac{f(x)}{f'(x)} = \rho(x)(x-\alpha), \quad (2-20)$$

with

$$\rho(\alpha) = \frac{1}{m}.$$

Finally, from (2-17) and (2-20),

$$\phi(x) - \alpha = w(x)u^p(x), \quad (2-21)$$

and  $w(x)$  continuous and with

$$w(x) = \frac{v(x)}{\rho^p(x)}, \quad w(\alpha) = \frac{m^p \phi^{(p)}(\alpha)}{p!} \neq 0.$$

The importance of (2-17), (2-19), and (2-21) in our future work cannot be overestimated. In particular, we will have occasion to use (2-21) in the form

$$\alpha = \phi(x) - w(x)u^p(x). \quad (2-22)$$

2.32 The theorems of the iteration calculus. The hypotheses of the theorems to be proven below will not be weighed down by the statement of continuity conditions on the derivatives of  $\phi$  and  $f$ ; they should be clear from the theorems. The notation is the same as in Section 2.31.

**THEOREM 2-4.** Let  $\phi_1(x) \in I_{p_1}$ ,  $\phi_2(x) \in I_{p_2}$  for some set of values of  $m$  and let  $\phi_3(x) = \phi_2[\phi_1(x)]$ . Then for these values of  $m$ ,  $\phi_3(x) \in I_{p_1 p_2}$ .

**PROOF.** From (2-17),

$$\phi_1(x) = \alpha + v_1(x)(x-\alpha)^{p_1}, \quad \phi_2(x) = \alpha + v_2(x)(x-\alpha)^{p_2}.$$

Then

$$\phi_3(x) = \phi_2[\phi_1(x)] = \alpha + v_2[\phi_1(x)][\phi_1(x) - \alpha]^{p_2},$$

$$\phi_3(x) - \alpha = v_2[\phi_1(x)]v_1^{p_2}(x)(x-\alpha)^{p_1 p_2}.$$

Let

$$v_3(x) = v_2[\phi_1(x)]v_1^{p_2}(x).$$

The fact that

$$c_3 = v_3(\alpha) = v_2(\alpha)v_1^{p_2}(\alpha) = c_2 c_1^{p_2} \neq 0 \quad (2-23)$$

completes the proof.

NOTE. Observe that (2-23) gives the formula for the asymptotic error constant of the composite I.F. in terms of the asymptotic error constants of the constituent I.F.

COROLLARY. Let

$$\varphi_\ell(x) \in I_{p_\ell}, \quad \ell = 1, 2, \dots, k$$

for some set of values of m. Let

$$\varphi(x) = \varphi_{j_1}(\varphi_{j_2}(\dots(\varphi_{j_k}(x))\dots))$$

where the  $j_\ell$  are any permutation of the numbers 1, 2, ..., k.

Then for these values of m,  $\varphi(x) \in I_{p_1 p_2 \dots p_k}$ . In particular, if  $p_\ell = p$  for  $\ell = 1, 2, \dots, k$ , then  $\varphi(x) \in I_p$ .

EXAMPLE 2-4. Let

$$\varphi_1(x) = \varphi_2(x) = x - u(x), \text{ (Newton's I.F.)}.$$

Then

$$\varphi_1 \in I_2, \quad \varphi_2 \in I_2, \quad c_1 = c_2 = A_2(\alpha), \quad A_2 = \frac{f''}{2f'}.$$

Hence

$$\varphi_3(x) = \varphi_2[\varphi_1(x)] \in I_4, \quad c_3 = [A_2(\alpha)]^3.$$

THEOREM 2-5. Let  $\phi(x) \in I_p$ , with  $p > 1$ , for some set of values of  $m$ . Then for these values of  $m$  there exists a function  $H(x)$  such that

$$\phi(x) - x = -u(x)H(x), \quad H(\alpha) \neq 0.$$

PROOF. Since  $\phi(\alpha) = \alpha$ , there exists a function  $G(x)$  with  $G(\alpha) = 0$  such that  $\phi(x) = x - G(x)$ . Then for  $p > 1$ ,

$$\phi'(\alpha) = 0 = 1 - G'(\alpha).$$

Therefore  $G(x)$  has a simple zero at  $\alpha$  and  $G(x) = u(x)H(x)$  for some  $H(x)$  with  $H(\alpha) \neq 0$ .

Theorem 2-5 may be rephrased as follows. If  $\alpha$  is a  $p$ -fold zero,  $p > 1$ , of the function  $\phi(x) - \alpha$ , then it is a simple zero of the function  $\phi(x) - x$ . Since  $\phi_1(x) = x - f(x)$  and  $\phi_2(x) = x - f^2(x)$  are both of order one, the theorem is false when  $p = 1$ .

EXAMPLE 2-5. Since  $\phi(x) - x$  has only simple zeros if the order of  $\phi(x)$  is greater than unity, any I.F. which is of order  $p$  for functions  $f(x)$  with simple zeros will be of order  $p$  if  $f(x)$  is replaced by  $\phi(x) - x$ . Thus if  $f(x)$  is replaced by  $\phi(x) - x$  in Newton's I.F.,

2.3-9

$$\psi(x) = x - \frac{\varphi(x) - x}{\varphi'(x) - 1}$$

and  $\psi(x) \in I_2$ . Let  $\varphi(x)$  be Newton's I.F. Then

$$\psi(x) = x - \frac{u(x)}{u'(x)}.$$

This generalization of Newton's I.F. is of order two for roots of arbitrary multiplicity; it will be discussed in Chapter 7.

**THEOREM 2-6.** Let  $\varphi_1(x) \in I_{p_1}$  and  $\varphi_2(x) \in I_{p_2}$  for some set of values of  $m$ . Then for these values of  $m$  there exists a function  $U(x)$ , with  $U(x)$  bounded at  $\alpha$  and  $U(\alpha) \neq 0$ , such that

$$\varphi_2(x) = \varphi_1(x) + U(x)u^p(x),$$

with

$$p = \min[p_1, p_2], \quad \text{if } p_1 \neq p_2;$$

$$p = p_1, \quad \text{if } p_1 = p_2 \quad \text{and} \quad \varphi_1^{(p_1)}(\alpha) \neq \varphi_2^{(p_1)}(\alpha);$$

$$p > p_1, \quad \text{if } p_1 = p_2 \quad \text{and} \quad \varphi_1^{(p_1)}(\alpha) = \varphi_2^{(p_1)}(\alpha).$$

PROOF. Case 1.  $p_1 \neq p_2$ . Assume in particular that  $p_1 > p_2$ . Then

$$\varphi_1(x) = \alpha + w_1(x)u^{p_1}(x), \quad \varphi_2(x) = \alpha + w_2(x)u^{p_2}(x),$$

$$\varphi_2(x) - \varphi_1(x) = u^{p_2}(x) \left[ w_2(x) - w_1(x)[u(x)]^{p_1-p_2} \right].$$

Define

$$U(x) = w_2(x) - w_1(x)[u(x)]^{p_1-p_2}.$$

The fact that  $U(\alpha) = w_2(\alpha) \neq 0$  completes the proof of Case 1.

Case 2.  $p_1 = p_2$  and  $\varphi_1^{(p_1)}(\alpha) \neq \varphi_2^{(p_1)}(\alpha)$ .

$$\varphi_2(x) - \varphi_1(x) = u^{p_1}(x)[w_2(x) - w_1(x)].$$

Define  $U(x) = w_2(x) - w_1(x)$ . The fact that

$$U(\alpha) = w_2(\alpha) - w_1(\alpha) = \frac{m}{p_1!} \left[ \varphi_2^{(p_1)}(\alpha) - \varphi_1^{(p_1)}(\alpha) \right] \neq 0$$

completes the proof of Case 2.

Case 3.  $p_1 = p_2$  and  $\varphi_1^{(p_1)}(\alpha) = \varphi_2^{(p_1)}(\alpha)$ . Unless  $\varphi_1(x) \equiv \varphi_2(x)$ , there exists an integer  $q$ ,  $q > p_1$ , such that  $\varphi_1^{(q)}(\alpha) \neq \varphi_2^{(q)}(\alpha)$  and the proof may be completed as above.

The case of most interest for later applications is  
 $p_1 = p_2$ ,  $\varphi_1^{(p_1)}(\alpha) \neq \varphi_2^{(p_1)}(\alpha)$ . We prove a converse to Theorem 2-6.

THEOREM 2-7. Let  $\varphi_1(x) \in I_{p_1}$ , and let  
 $\varphi_2(x) = \varphi_1(x) + U(x)u^p(x)$  for some set of values of  $m$ . Then  
for these values of  $m$ ,  $\varphi_2(x) \in I_{p_2}$ , where

$$p_2 = \min[p, p_1], \quad \text{if } p \neq p_1;$$

$$p_2 = p, \quad \text{if } p_1 = p \quad \text{and} \quad U(\alpha) \neq -\frac{\varphi_1^{(p)}(\alpha)m^p}{p!};$$

$$p_2 > p, \quad \text{if } p_1 = p \quad \text{and} \quad U(\alpha) = -\frac{\varphi_1^{(p)}(\alpha)m^p}{p!}.$$

PROOF. Case 1.  $p \neq p_1$ . Assume  $p > p_1$ . Then

$$\varphi_1(x) = \alpha + w_1(x)u^{p_1}(x), \quad \varphi_2(x) = \varphi_1(x) + U(x)u^p(x).$$

Hence

$$\varphi_2(x) - \alpha = u^{p_1}(x) \left[ w_1(x) + U(x)[u(x)]^{p-p_1} \right].$$

Define  $w_2(x) = w_1(x) + U(x)[u(x)]^{p-p_1}$ . The fact that  
 $w_2(\alpha) = w_1(\alpha) \neq 0$  completes the proof of Case 1.

2.3-12

Case 2.  $p = p_1$  and  $U(\alpha) \neq -\varphi_1^{(p)}(\alpha)m^p/p!$ . Then

$$\varphi_2(x) - \alpha = u^p(x)[W_1(x) + U(x)].$$

Define  $w_2(x) = w_1(x) + U(x)$ . The fact that

$$w_2(\alpha) = w_1(\alpha) + U(\alpha) = \frac{\varphi_1^{(p)}(\alpha)m^p}{p!} + U(\alpha) \neq 0$$

completes the proof of Case 2.

Case 3.  $p = p_1$  and  $U(\alpha) = -\varphi_1^{(p)}(\alpha)m^p/p!$ . This case may be completed in a similar manner.

The "comparison theorem" just proved will permit us to deduce the order of a given I.F. if it differs by terms of order  $u^p$  from an I.F. whose order is known.

The following theorem permits the calculation of the asymptotic error constant of an I.F. of order  $p$  if the asymptotic error constant of another I.F. of order  $p$  is known. Recall that the asymptotic error constant  $C$  is defined by

$$C = \lim_{x \rightarrow \alpha} \frac{\varphi(x) - \alpha}{(x-\alpha)^p}. \quad (2-24)$$

Absolute values are not required in (2-24) since  $p$  is an integer.

## 2.3-13

THEOREM 2-8. Let  $\varphi_1(x) \in I_p$  and  $\varphi_2(x) \in I_p$  for some set of values of  $m$ . Let

$$G(x) = \frac{\varphi_2(x) - \varphi_1(x)}{(x-\alpha)^p}, \quad x \neq \alpha.$$

Let  $C_1$  and  $C_2$  be the asymptotic error constants of  $\varphi_1$  and  $\varphi_2$ . Then for these values of  $m$ ,

$$C_2 = C_1 + \lim_{x \rightarrow \alpha} G(x).$$

PROOF. A  $p$ -fold application of L'Hospital's rule yields

$$\lim_{x \rightarrow \alpha} G(x) = \frac{\varphi_2^{(p)}(\alpha) - \varphi_1^{(p)}(\alpha)}{p!} = C_2 - C_1.$$

EXAMPLE 2-6. Let

$$\varphi_1(x) = x - mu(x), \quad \varphi_2(x) = x - \frac{u(x)}{u'(x)}.$$

Then  $\varphi_1(x) \in I_2$  and  $\varphi_2(x) \in I_2$ , for  $m$  arbitrary. As will be shown in Section 7.41

$$mu(x) = x - \alpha - \frac{a_{m+1}(\alpha)}{ma_m(\alpha)} (x-\alpha)^2 + \underline{O}[(x-\alpha)^3],$$

2.3-14

where  $a_m(x) = f^{(m)}(x)/m!$ . Then it is easy to verify that

$$c_1 = \frac{a_{m+1}}{ma_m}, \quad c_2 = -\frac{a_{m+1}}{ma_m}, \quad \lim_{x \rightarrow a} G(x) = -\frac{2a_{m+1}}{ma_m},$$

as predicted by Theorem 2-8. Let

$$\varphi(x) = \frac{1}{2}[\varphi_1(x) + \varphi_2(x)] = x - \frac{1}{2}u(x)\left[m + \frac{1}{u'(x)}\right].$$

Since  $c_1$  and  $c_2$  are equal in magnitude but opposite in sign, it is clear that  $\varphi(x) \in I_3$  for all  $m$ .

A more useful result than the one obtained in Theorem 2-8 is given by

**THEOREM 2-9.** Let  $\varphi_1(x) \in I_p$  and let  $\varphi_2(x) \in I_p$  for some set of values of  $m$ . Let

$$H(x) = \frac{\varphi_2(x) - \varphi_1(x)}{u^p(x)}, \quad x \neq a, \quad u = \frac{f}{f'}$$

Let  $C_1$  and  $C_2$  be the asymptotic error constants of  $\varphi_1(x)$  and  $\varphi_2(x)$ . Then for the values of  $m$ ,

$$C_2 = C_1 + \frac{1}{m^p} \lim_{x \rightarrow a} H(x).$$

2.3-15

PROOF. It is easy to show that

$$\lim_{x \rightarrow \alpha} \frac{u(x)}{x-\alpha} = \frac{1}{m}.$$

From the previous theorem,

$$\lim_{x \rightarrow \alpha} H(x) = \lim_{x \rightarrow \alpha} G(x) \left[ \frac{x-\alpha}{u(x)} \right]^p = (c_2 - c_1)m^p,$$

and the result follows.

EXAMPLE 2-7. Let  $m = 1$ . Then

$$\varphi_1(x) = x - u(x) - A_2(x)u^2(x), \quad \varphi_2(x) = x - \frac{u(x)}{1 - \frac{u(x)}{A_2(x)u(x)}}$$

are of third order and

$$c_1 = 2A_2^2(\alpha) - A_3(\alpha), \quad A_1 = \frac{f^{(1)}}{1!f'}.$$

Since

$$\varphi_2(x) = x - u(x) \left[ 1 + A_2(x)u(x) + A_2^2(x)u^2(x) \right] + \underline{o}[u^4(x)],$$

2.3-16

$$\lim_{x \rightarrow \alpha} H(x) = -A_2^2(\alpha).$$

Hence  $C_2 = A_2^2(\alpha) - A_3(\alpha)$ .

THEOREM 2-10. Let  $m = 1$  and let  $U(x)$  be an arbitrary function which may depend on  $f(x)$  and its derivatives and which is bounded at  $\alpha$ . Let  $\varphi_1(x) \in I_p$  and let

$$U(\alpha) \neq -\frac{\varphi_1^{(p)}(\alpha)}{p! [f'(\alpha)]^p}.$$

Then

$$\varphi(x) = \varphi_1(x) + U(x)f^p(x) \quad (2-25)$$

is the most general I.F. of order exactly  $p$ .

PROOF. From (2-19),

$$\varphi_1(x) = \alpha + T_1(x)f^p(x), \quad T_1(\alpha) = \frac{\varphi_1^{(p)}(\alpha)}{p! [f'(\alpha)]^p}.$$

By hypothesis,

$$\varphi(x) = \varphi_1(x) + U(x)f^p(x).$$

Then

$$\varphi(x) = \alpha + f^p(x)[T_1(x) + U(x)]. \quad (2-26)$$

2.3-17

Define

$$T(x) = T_1(x) + U(x).$$

Since  $T(\alpha)$  is nonzero,  $\varphi(x) \in I_p$ . The observation that  $U(x)f^p(x)$  is the most general addend to  $\varphi_1(x)$  which leads to (2-26) completes the proof.

Note that we do not assume that  $U(\alpha) \neq 0$ . It is clear that

$$\varphi(x) = \varphi_1(x) + *U(x)u^q(x), \quad q > p \quad (2-27)$$

is also of order  $p$ . Setting  $U(x) = *U(x)u^{q-p}(x)$  puts (2-27) into the form of (2-25). Note also that Theorem 2-10 need not hold if  $m > 1$ . For example,

$$\varphi(x) = \varphi_1(x) + U(x) \left[ \frac{f'(x)}{f''(x)} \right]^p$$

is also of order  $p$  if  $m > 1$ .

The following two theorems lead to another characterization of I.F. of order  $p$ .

THEOREM 2-11. Let  $\varphi(x) \in I_p$  for some set of values of  $m$ . Then for these values of  $m$ , there exists a function  $Q(x)$  such that  $Q(\alpha) \neq 0$  and  $f[\varphi(x)] = Q(x)f^p(x)$ .

PROOF. Let  $\alpha$  be a zero of multiplicity  $m$ . As was shown in Section 2.31, there exists a continuous function  $\lambda(x)$  such that

$$f(x) = \lambda(x)(x-\alpha)^m$$

and

$$\lambda(\alpha) = \frac{f^{(m)}(\alpha)}{m!} \neq 0.$$

Then

$$\begin{aligned} f[\varphi(x)] &= [\varphi(x) - \alpha]^m \lambda[\varphi(x)] \\ &= [V(x)(x-\alpha)^p]^m \lambda[\varphi(x)] \\ &= V^m(x) \lambda[\varphi(x)] [(x-\alpha)^m]^p. \end{aligned}$$

Since  $\lambda(\alpha) \neq 0$ ,  $\lambda(x)$  does not vanish in some neighborhood of  $\alpha$  and we may write

$$f[\varphi(x)] = V^m(x) \lambda[\varphi(x)] \lambda^{-p}(x) f^p(x).$$

Define

$$Q(x) = V^m(x) \lambda[\varphi(x)] \lambda^{-p}(x).$$

The fact that

$$Q(\alpha) = \left[ \frac{\varphi^{(p)}(\alpha)}{p!} \right]^m \left[ \frac{f^{(m)}(\alpha)}{m!} \right]^{1-p} \neq 0$$

completes the proof.

As a converse to Theorem 2-11, we have

THEOREM 2-12. Let  $f[\varphi(x)] = Q(x)f^p(x)$  for some set of values of  $m$  with  $Q(\alpha) \neq 0$ . Then for these values of  $m$ ,  $\varphi(x) \in I_p$ .

PROOF. Let  $\lambda(x)$  be defined as in the previous theorem. Then

$$f[\varphi(x)] = [\varphi(x) - \alpha]^m \lambda[\varphi(x)],$$

and from the hypothesis,

$$f[\varphi(x)] = Q(x)f^p(x) = Q(x)\lambda^p(x)(x-\alpha)^{mp}.$$

Define  $V(x)$  by

$$V^m(x) = \lambda^{-1}[\varphi(x)]\lambda^p(x)Q(x).$$

Then

$$\varphi(x) - \alpha = V(x)(x-\alpha)^p$$

and the fact that  $V(\alpha) \neq 0$  completes the proof.

2.3-20

Theorems 2-11 and 2-12 show that  $\varphi(x) \in I_p$  if and only if  $f[\varphi(x)] = Q(x)f^p(x)$  with  $Q(\alpha) \neq 0$ .

EXAMPLE 2-8. Let  $m = 1$  and let

$$\varphi(x) = x - u(x)H(x), \quad H(x) = \frac{1}{1 - A_2(x)u(x)}, \quad A_2(x) = \frac{f''}{2f'}.$$

This is Halley's I.F. which will be studied in Section 5.2.  
Then

$$f[\varphi(x)] = f(x) - u(x)H(x)f'(x) + \frac{1}{2}u^2(x)H^2(x)f''(x) + \underline{o}[u^3(x)],$$

$$H(x) = 1 + A_2(x)u(x) + A_2^2(x)u^2(x) + \underline{o}[u^3(x)].$$

Thus

$$\begin{aligned} f[\varphi(x)] &= f(x) - u(x)f'(x) - A_2(x)u^2(x)f'(x) + \frac{1}{2}u^2(x)f''(x) + \underline{o}[u^3(x)] \\ &= \underline{o}[u^3(x)] = \underline{o}[f^3(x)]. \end{aligned}$$

Therefore Halley's I.F. is third order for  $m = 1$ .

THEOREM 2-13. Let  $m = 1$  and  $\varphi(x) \in I_p$ . Then

$$\left. \frac{d^p f[\varphi(x)]}{dx^p} \right|_{x=\alpha} = f'(\alpha) \varphi^{(p)}(\alpha).$$

2.3-21

PROOF. Let  $\psi(x) = x - f(x)$ . Clearly  $\psi(x) \in I_1$ .

Let

$$\Psi(x) = \psi[\varphi(x)] = \varphi(x) - f[\varphi(x)].$$

By Theorem 2-4,  $\Psi(x) \in I_p$ . Hence

$$\varphi(x) - f[\varphi(x)] = \alpha + V_1(x)(x-\alpha)^p. \quad (2-28)$$

Since  $\varphi(x) \in I_p$ ,

$$\varphi(x) = \alpha + V(x)(x-\alpha)^p.$$

Therefore

$$f[\varphi(x)] = [V(x) - V_1(x)](x-\alpha)^p. \quad (2-29)$$

A second expression for  $f[\varphi(x)]$  may be derived as follows.

Define  $\lambda(x)$  by

$$f(x) = \lambda(x)(x-\alpha), \quad \lambda(\alpha) = f'(\alpha).$$

Then

$$f[\varphi(x)] = \lambda[\varphi(x)][\varphi(x) - \alpha] = \lambda[\varphi(x)]V(x)(x-\alpha)^p. \quad (2-30)$$

From (2-29) and (2-30),

$$\lambda[\varphi(x)]V(x) = V(x) - V_1(x). \quad (2-31)$$

From (2-28),

$$v_1(\alpha) = \frac{1}{p!} \left[ \varphi^{(p)}(\alpha) - \frac{d^p f[\varphi(x)]}{dx^p} \Big|_{x=\alpha} \right] = v(\alpha) - \frac{1}{p!} \frac{d^p f[\varphi(x)]}{dx^p} \Big|_{x=\alpha}.$$

Taking  $x = \alpha$  in (2-31) yields the desired result.

We turn to a generalization of the previous theorem.

THEOREM 2-14. Let  $m = 1$  and  $\varphi(x) \in I_p$ . Then

$$\frac{d^j f[\varphi(x)]}{dx^j} \Big|_{x=\alpha} = f'(\alpha) \varphi^{(j)}(\alpha), \quad 0 < j < 2p.$$

PROOF. Let  $D^j \equiv d^j/dx^j$ . It is clear from Theorem 2-11 that  $D^j f[\varphi(x)]|_{x=\alpha} = 0$  for  $0 < j < p$ . Since  $\varphi^{(j)}(\alpha) = 0$  for  $0 < j < p$ , the theorem is proved for these values of  $j$ .

Let  $p \leq j < 2p$ . Now

$$f(x) = f'(\alpha)(x-\alpha) + \tau(x)(x-\alpha)^2,$$

$$\varphi(x) = \alpha + v(x)(x-\alpha)^p.$$

2.3-23

It is clear that if  $f$  and  $\phi$  possess a sufficient number of continuous derivatives, then  $\tau$  and  $V$  may be defined so as to possess as many continuous derivatives as required. We have

$$f[\phi(x)] = f'(\alpha)[\phi(x) - \alpha] + \tau[\phi(x)][\phi(x) - \alpha]^2 = f'(\alpha)V(x)(x-\alpha)^p + S(x).$$

Then

$$D^j f[\phi(x)] = f'(\alpha) \sum_{k=0}^j C[j,k] D^{j-k} V(x) D^k (x-\alpha)^p + D^j S(x),$$

where  $C[j,k]$  denotes a binomial coefficient. Hence

$$D^j f[\phi(x)]|_{x=\alpha} = f'(\alpha)p!C[j,p]D^{j-p}V(x)|_{x=\alpha}.$$

It is easy to show that

$$D^{j-p}V(x)|_{x=\alpha} = \frac{(j-p)!}{j!} \phi^{(j)}(\alpha),$$

which completes the proof.

### 3.0-1

## CHAPTER 3

### THE MATHEMATICS OF DIFFERENCE RELATIONS

In this chapter we shall lay the mathematical foundations prerequisite to a careful analysis of the convergence and order of one-point I.F. and one-point I.F. with memory. The two theorems of Section 3.4 will be basic for that analysis.

It is assumed that the reader is familiar with the elementary aspects of difference equation theory. If this is not the case, Hildebrand [3.0-1, Chap. 3], Jordan [3.0-2, Chap. 11], Milne-Thomson [3.0-3], and Nörlund [3.0-4] may be used as references.

### 3.1-1

#### 3.1 Convergence of Difference Inequalities

LEMMA 3-1. Let  $\gamma_0, \gamma_1, \dots, \gamma_n$  be nonnegative integers and let  $q = \sum_{j=0}^n \gamma_j$ . Let  $M$  be a positive constant and let the  $\delta_i$  be a sequence of nonnegative numbers such that

$$\delta_{i+1} \leq M \prod_{j=0}^n (\delta_{i-j})^{\gamma_j}$$

and such that

$$\delta_i \leq \Gamma, \quad i = 0, 1, \dots, n. \quad (3-1)$$

Then

$$M\Gamma^{q-1} < 1$$

implies that  $\delta_i \rightarrow 0$ .

PROOF. Let

$$L = M\Gamma^{q-1}. \quad (3-2)$$

Let  $t$  be the first subscript for which  $\gamma_j$  is nonzero. Then

$$\delta_{n+1} \leq M\delta_{n-t}^{\gamma_t} \prod_{j=t+1}^n (\delta_{n-j})^{\gamma_j}$$

$$\leq M\Gamma^{q-1}\delta_{n-t} = L\delta_{n-t}.$$

We now prove by induction that

$$\delta_{i+1} \leq L\delta_{i-t}, \quad L < 1, \quad (3-3)$$

for all  $i$ . Note that (3-1) and (3-3) imply  $\delta_{i+1} \leq \Gamma$ .

3.1-2

Let (3-3) hold for  $i = 0, 1, \dots, k-1$ . Then

$$\begin{aligned}\delta_{k+1} &\leq M\delta_{k-t}^{\gamma_t} \prod_{j=t+1}^n (\delta_{k-j})^{\gamma_j} \\ &\leq M\Gamma^{q-1} \delta_{k-t} = L\delta_{k-t}.\end{aligned}$$

This completes the induction. From (3-3) we conclude that  $\delta_1 \rightarrow 0$ .

LEMMA 3-2. Let  $\gamma_0$  be a positive integer and let  $\gamma_1, \gamma_2, \dots, \gamma_n$  be nonnegative integers. Let  $q = \sum_{j=0}^n \gamma_j$ . Let  $M$  and  $N$  be positive constants and let  $\delta_i$  be a sequence of nonnegative numbers such that

$$\delta_{i+1} \leq M\delta_i^{\gamma_0} \prod_{j=1}^n (\delta_i + \delta_{i-j})^{\gamma_j} + N\delta_i^{\gamma_0+1},$$

and such that

$$\delta_i < \Gamma, \quad i = 0, 1, \dots, n. \quad (3-4)$$

Then

$$2^{q-\gamma_0} M\Gamma^{q-1} + N\Gamma^{\gamma_0} < 1$$

implies that  $\delta_1 \rightarrow 0$ .

### 3.1-3

PROOF. Let

$$L = 2^{q-\gamma_0} M\Gamma^{q-1} + N\Gamma^{\gamma_0}. \quad (3-5)$$

Then

$$\delta_{n+1} \leq M\delta_n^{\gamma_0} \prod_{j=1}^n (\delta_n + \delta_{n-j})^{\gamma_j} + N\delta_n^{\gamma_0+1},$$

$$\delta_{n+1} \leq \delta_n \left[ M\Gamma^{\gamma_0-1} \prod_{j=1}^n (2\Gamma)^{\gamma_j} + N\Gamma^{\gamma_0} \right]$$

$$= \delta_n \left[ 2^{q-\gamma_0} M\Gamma^{q-1} + N\Gamma^{\gamma_0} \right] = L\delta_n.$$

We now prove by induction that

$$\delta_{i+1} \leq L\delta_i, \quad L < 1, \quad (3-6)$$

for all  $i$ . Note that (3-4) and (3-6) imply  $\delta_{i+1} \leq \Gamma$ .

Let (3-6) hold for  $i = 0, 1, \dots, k-1$ . Then

$$\delta_{k+1} \leq \delta_k \left[ M\Gamma^{\gamma_0-1} \prod_{j=1}^n (2\Gamma)^{\gamma_j} + N\Gamma^{\gamma_0} \right] = L\delta_k.$$

This completes the induction. From (3-6) we conclude that

$$\delta_i \rightarrow 0.$$

3.2 A Theorem on the Solutions of Certain Inhomogeneous Difference Equations

Let

$$\sum_{j=0}^n \kappa_j \sigma_{1+j} = 0, \quad \kappa_N = 1 \quad (3-7)$$

be a homogeneous linear difference equation with constant coefficients. The indicial equation corresponding to (3-7) is the algebraic equation

$$\sum_{j=0}^N \kappa_j t^j = 0. \quad (3-8)$$

Now, if all the roots of (3-8) are simple, the general solution of (3-7) is given by

$$\sigma_1 = \sum_{j=1}^N c_j \rho_j^1$$

where the  $\rho_j$  are the roots of (3-8) and where the  $c_j$  are constants determined by the initial conditions. It is obvious that if all the  $\rho_j$  have moduli less than unity, then all solutions of (3-7) converge to zero.

We wish to extend this result to the case of a linear difference equation whose homogeneous part has constant coefficients and whose inhomogeneous part is a sequence converging to zero. Hence we shall consider

## 3.2-2

$$\sum_{j=0}^n \kappa_j \sigma_{i+j} = \beta_i, \quad \kappa_N = 1 \quad (3-9)$$

where  $\beta_i \rightarrow 0$ .

Let us first consider the special case of the first-order difference equation

$$\sigma_{i+1} + \kappa_0 \sigma_i = \beta_i. \quad (3-10)$$

Observe that the solution of the indicial equation is given by  $t = -\kappa_0 = \rho_1$ . We shall assume that  $\rho_1$  is less than unity and that  $\beta_i \rightarrow 0$ . The difference equation may be "summed" as follows. Let  $\ell$  and  $n$  be arbitrary with  $n - 1 \geq \ell$ . Then

$$\sigma_{\ell+1} + \kappa_0 \sigma_\ell = \beta_\ell,$$

$$\sigma_{\ell+2} + \kappa_0 \sigma_{\ell+1} = \beta_{\ell+1},$$

.....

.....

$$\sigma_{n-1} + \kappa_0 \sigma_{n-2} = \beta_{n-2},$$

$$\sigma_n + \kappa_0 \sigma_{n-1} = \beta_{n-1}.$$

Multiplying the equation whose leading term is  $\sigma_{n-j}$  by  $(-\kappa_0)^j$  or  $\rho_1^j$  and adding all the equations yields

$$\sigma_n = \rho_1^{n-\ell} \sigma_\ell + \sum_{i=\ell}^{n-1} \beta_i \rho_1^{n-i-1}. \quad (3-11)$$

### 3.2-3

Hence

$$|\sigma_n| \leq |\rho_1|^{n-\ell} |\sigma_\ell| + \sum_{i=\ell}^{n-1} |\beta_i| |\rho_1|^{n-i-1}. \quad (3-12)$$

Since  $\beta_1 \rightarrow 0$  we can fix  $\ell$  so large that the sum has magnitude less than  $\epsilon/2$ . Since  $\rho_1$  is less than unity we can then take  $n$  so large that the first term on the right side of (3-12) has magnitude less than  $\epsilon/2$ . We conclude that  $\sigma_1 \rightarrow 0$ . This is the desired result for the case of a first-order difference equation.

We shall now derive a generalization of (3-11) for the case of an  $N$ th order difference equation. Thus we start with

$$\sum_{j=0}^N \kappa_j \sigma_{1+j} = \beta_1, \quad \kappa_N = 1.$$

Let  $\rho_j$ ,  $j = 1, 2, \dots, N$  be the roots of the indicial equation. Let  $d_{ni}$  be constants whose values we shall choose at our later convenience. Let  $\ell$  and  $n$  be arbitrary integers such that  $n - N \geq \ell$ . Then

$$\sum_{i=\ell}^{n-N} d_{ni} \sum_{j=0}^N \kappa_j \sigma_{1+j} = \sum_{i=\ell}^{n-N} d_{ni} \beta_1,$$

3.2-4

or

$$\sum_{i=\ell}^{n-N} d_{ni} \sum_{s=i}^{i+N} \kappa_{s-i} \sigma_s = \sum_{i=\ell}^{n-N} d_{ni} \beta_i.$$

Changing the order of summation yields

$$\begin{aligned} \sum_{i=\ell}^{n-N} d_{ni} \beta_i &= \sum_{s=\ell}^{\ell+N-1} \sigma_s \sum_{i=\ell}^s d_{ni} \kappa_{s-i} \\ &\quad + \sum_{s=\ell+N}^{n-N} \sigma_s \sum_{i=s-N}^s d_{ni} \kappa_{s-i} + \sum_{s=n-N+1}^n \sigma_s \sum_{i=s-N}^{n-N} d_{ni} \kappa_{s-i} \\ &= 1 + 2 + 3. \end{aligned}$$

Consider 2. This is zero if we choose  $d_{ni}$  as a solution of the homogeneous difference equation. It is convenient to define  $d_{ni}$  by

$$d_{ni} = \sum_{j=1}^N D_j \rho_j^{n-i-1} \quad (3-13)$$

where the constants  $D_j$  will be chosen later.

Consider 3. Using (3-13) we can show that

$$\sum_{s=n-N+1}^n \sigma_s \sum_{i=s-N}^{n-N} d_{ni} \kappa_{s-i} = \sum_{\lambda=0}^{N-1} \sigma_{n-\lambda} \sum_{r=N-\lambda}^N \kappa_r \sum_{j=1}^N D_j \rho_j^{\lambda+r-1}.$$

## 3.2-5

We have the  $N$  coefficients  $D_1, D_2, \dots, D_N$  at our disposal. We choose them so that the coefficient of  $\sigma_n$  is unity and the coefficients of the  $\sigma_{n-\lambda}$  are zero for  $\lambda > 0$ . Thus, recalling that  $\kappa_N = 1$ ,

$$\sum_{j=1}^N D_j \rho_j^{N-1} = 1, \quad (3-14)$$

$$\sum_{r=N-\lambda}^N \kappa_r \sum_{j=1}^N D_j \rho_j^{\lambda+r-1} = 0, \quad \lambda = 1, 2, \dots, N-1.$$

Using the fact that

$$\sum_{r=0}^N \kappa_r \rho_j^r = 0, \quad j = 1, 2, \dots, N,$$

it is not difficult to prove that the system (3-14) is equivalent to the system

$$\sum_{j=1}^N D_j \rho_j^r = \delta_{r,N-1}, \quad r = 0, 1, \dots, N-1 \quad (3-15)$$

where  $\delta_{r,N-1}$  is the Kronecker symbol. The determinant of this system is a Vandermonde determinant. Hence  $D_j$  satisfying (3-15), and consequently (3-14), exist and are indeed the ratios of certain Vandermonde determinants. With this choice of the  $D_j$ ,  $\beta$  reduces simply to  $\sigma_n$ . We conclude that

3.2-6

$$\sigma_n = \sum_{i=\ell}^{n-N} d_{ni} \beta_i - \sum_{s=\ell}^{\ell+N-1} \sigma_s \sum_{i=\ell}^s d_{ni} \kappa_{s-i}.$$

We summarize our results in

LEMMA 3-3. Let

$$\sum_{j=0}^N \kappa_j \sigma_{i+j} = \beta_i, \quad \kappa_N = 1$$

be a linear difference equation whose homogeneous part has constant coefficients. Let the roots of the indicial equation be simple. Let

$$d_{ni} = \sum_{j=1}^N D_j \rho_j^{n-i-1} \quad (3-16)$$

where the  $\rho_j$  are the roots of the indicial equation and where the  $D_j$  are determined by

$$\sum_{j=1}^N D_j \rho_j^r = \delta_{r,N-1}, \quad r = 0, 1, \dots, N-1;$$

3.2-7

$\delta_{r,N-1}$  is the Kronecker symbol. Then

$$\sigma_n = \sum_{i=\ell}^{n-N} d_{ni} \beta_i - \sum_{s=\ell}^{\ell+N-1} \sigma_s \sum_{i=\ell}^s d_{ni} \kappa_{s-i}, \quad (3-17)$$

where  $\ell$  and  $n$  are arbitrary integers such that  $n - N \geq \ell$ .

We are now ready to prove

THEOREM 3-1. Let

$$\sum_{j=0}^N \kappa_j \sigma_{1+j} = \beta_1$$

be a linear difference equation whose homogeneous part has constant coefficients. Let the roots of the indicial equation be simple and have moduli less than unity. Let  $\beta_1 \rightarrow 0$ . Then  $\sigma_1 \rightarrow 0$  for all sets of initial conditions.

PROOF. Let

$$G_{j\ell} = D_j \sum_{s=\ell}^{\ell+N-1} \sigma_s \sum_{i=\ell}^s \kappa_{s-i} \rho_j^{-i-1}. \quad (3-18)$$

3.2-8

Using (3-16), (3-17), and (3-18) we can write

$$\sigma_n = \sum_{i=\ell}^{n-N} d_{ni} \beta_i - \sum_{j=1}^N G_{j\ell} \rho_j^n.$$

The notational adjustments which would be necessary if one of the roots of the indicial equation were equal to zero are obvious. Since

$$d_{ni} = \sum_{j=1}^N D_j \rho_j^{n-i-1},$$

and since the  $\rho_j$  having moduli less than unity, it is clear that there exists a constant A such that

$$\sum_{i=\ell}^{n-N} |d_{ni}| < A.$$

Let  $\epsilon > 0$  be preassigned and arbitrary. Since  $\beta_i \rightarrow 0$ , there exists a number L such that

$$|\beta_i| < \frac{\epsilon}{2A} \quad \text{for all } i \geq L.$$

Fix  $\ell$  such that  $\ell \geq L$ . Observe that  $G_{j\ell}$  is independent of n. With  $\ell$  fixed there exists a constant B such that

$$\sum_{j=1}^N |G_{j\ell}| \leq B.$$

3.2-9

Let

$$\rho = \max[|\rho_1|, |\rho_2|, \dots, |\rho_N|].$$

Let  $\eta$  be so large that

$$\rho^\eta < \frac{\epsilon}{2B}.$$

Then for all  $n > \eta$ , we have

$$|\sigma_n| \leq \sum_{i=\ell}^{n-N} |d_{ni}| |\beta_i| + \sum_{j=1}^N |g_{j\ell}| |\rho_j|^n,$$

$$|\sigma_n| < \frac{\epsilon}{2A} A + \frac{\epsilon}{2B} B = \epsilon$$

which completes the proof.

For the applications that we have in mind, the inhomogeneous part of the difference equation will converge to a nonzero constant. We wish to show that all solutions of the inhomogeneous equation will converge to a constant. This result follows easily from the above theorem. We have

COROLLARY. Let

$$\sum_{j=0}^N \kappa_j \sigma_{i+j} = \omega_i$$

be a linear difference equation whose homogeneous part has constant coefficients. Let the roots of the indicial equation be simple and have moduli less than unity. Let  $\omega_i \rightarrow \omega$ . Then

$$\sigma_i \rightarrow \frac{\omega}{\sum_{j=0}^N \kappa_j}$$

for all sets of initial conditions.

PROOF. Let

$$\sigma_i = \lambda_i + \frac{\omega}{\sum_{j=0}^N \kappa_j}.$$

Then

$$\sum_{j=0}^N \kappa_j \left[ \lambda_{i+j} + \frac{\omega}{\sum_{j=0}^N \kappa_j} \right] = \sum_{j=0}^N \kappa_j \lambda_{i+j} + \omega = \omega_i.$$

Let  $\beta_i = \omega_i - \omega$ . Then the conditions of the theorem apply and we conclude that  $\lambda_i \rightarrow 0$  and hence that

$$\sigma_i \rightarrow \frac{\omega}{\sum_{j=0}^N \kappa_j}.$$

### 3.3-1

#### 3.3 On the Roots of Certain Indicial Equations

The properties of the roots of the polynomial equation

$$g_{k,a}(t) = t^k - a \sum_{j=0}^{k-1} t^j = 0 \quad (3-19)$$

will be derived. This will turn out to be the indicial equation for certain families of difference equations which will be encountered in our later work. The case  $a = 1$  has been treated by Ostrowski [3.3-1, pp. 87-90].

## 3.3-2

3.31 The properties of the roots. If  $k = 1$ , the only root of (3-19) is  $t = a$ . For the remainder of this section we shall assume  $k \geq 2$ . We shall permit  $a$  to be any real number such that

$$ka > 1. \quad (3-20)$$

Since

$$g_{k,a}(1) = 1 - ka,$$

$t = 1$  is not a solution of (3-19). It is convenient to define

$$G_{k,a}(t) = (t-1)g_{k,a}(t) = t^{k+1} - (a+1)t^k + a. \quad (3-21)$$

Hence  $G_{k,a}(t)$  has a root at  $t = 1$  and roots at the roots of  $g_{k,a}(t)$ .

By Descartes' rule,  $g_{k,a}(t)$  has exactly one real positive simple root. This unique root will be labeled  $\beta_{k,a}$ . Since

$$g_{k,a}(a) = - \sum_{j=1}^{k-1} a^j, \quad g_{k,a}(1) = 1 - ka, \quad g_{k,a}(a+1) = 1$$

we have

LEMMA 3-4. The equation

$$g_{k,a}(t) = t^k - a \sum_{j=0}^{k-1} t^j = 0$$

### 3.3-3

has a unique real positive simple root,  $\beta_{k,a}$ , and

$$\max[1,a] < \beta_{k,a} < a + 1.$$

Thus  $\beta_{k,a}$  has modulus greater than one. We shall show later that all other roots have moduli less than one. In the next lemma we shall prove that  $\beta_{k,a}$  is a strictly increasing function of  $k$ .

LEMMA 3-5.  $\beta_{k-1,a} < \beta_{k,a}$ .

PROOF. This follows from the observation that the recurrence relation

$$tg_{k-1,a}(t) - a = g_{k,a}(t)$$

implies that  $g_{k,a}(\beta_{k-1,a})$  is negative.

The following inequality will be needed below.

LEMMA 3-6. Let  $ka > 1$ . Then

$$\frac{(a+1)^{k+1}}{a} > \left( k+1 \right) \left( 1 + \frac{1}{k} \right)^k.$$

## 3.3-4

PROOF. Let  $k$  be fixed and define

$$J(a) = \frac{(a+1)^{k+1}}{a}.$$

Then  $J'(a) = (a+1)^k(ka-1)/a^2$  and therefore  $J(a)$  is a strictly increasing function of  $a$  for  $ka > 1$ . The observation that

$$J\left(\frac{1}{k}\right) = (k+1)\left(1 + \frac{1}{k}\right)^k$$

completes the proof.

Descartes' rule shows that  $\beta_{k,a}$  is a simple root.

Furthermore

LEMMA 3-7. All the roots of  $g_{k,a}(t) = 0$  are simple.

PROOF. Define

$$G_{k,a}(t) = (t-1)g_{k,a}(t) = t^{k+1} - (a+1)t^k + a.$$

Then  $G'_{k,a}(t)$  has only one nonzero root,

$$v = \frac{k}{k+1}(a+1).$$

The fact that  $v$  is positive completes the proof.

## 3.3-5

In the following two lemmas we derive better bounds on  $\beta_{k,a}$ .

## LEMMA 3-8.

$$\frac{k}{k+1} (a+1) < \beta_{k,a} < a + 1.$$

Therefore, for a fixed

$$\lim_{k \rightarrow \infty} \beta_{k,a} = a + 1.$$

NOTE. Since  $ka > 1$  implies that  $(a+1)^k/(k+1) > 1$ , we need not write

$$\max\left[1, \frac{k}{k+1} (a+1)\right] < \beta_{k,a}.$$

PROOF. Let  $v = (a+1)^k/(k+1)$ . Then

$$G_{k,a}(v) = a - \frac{(a+1)^{k+1}}{k+1} \left(1 + \frac{1}{k}\right)^{-k}.$$

An application of Lemma 3-6 shows that  $G_{k,a}(v) < 0$ , which together with an application of Lemma 3-4 completes the proof.

## LEMMA 3-9.

$$a + 1 - \frac{ea}{(a+1)^k} < \beta_{k,a} < a + 1 - \frac{a}{(a+1)^k},$$

where e denotes the base of common logarithms.

3.3-6

PROOF. Let  $v = (a+1)k/(k+1)$ . The only positive root of the equation  $G''_{k,a}(t) = 0$  is  $t = v - (a+1)/(k+1)$ . Therefore  $G''_{k,a}(t)$  does not change sign in the interval  $v \leq t \leq a + 1$ . Since

$$G''_{k,a}(a+1) = 2k(a+1)^{k-1} > 0,$$

$G_{k,a}(t)$  is convex in the interval. Since the secant line through the points  $[v, G_{k,a}(v)]$ ,  $[a+1, G_{k,a}(a+1)]$  intersects the  $t$  axis at

$$t = a + 1 - \frac{a}{(a+1)^k} \left(1 + \frac{1}{k}\right)^k,$$

whereas the tangent line at  $[a+1, G_{k,a}(a+1)]$  intersects the  $t$  axis at

$$t = a + 1 - \frac{a}{(a+1)^k},$$

we conclude that

$$a + 1 - \frac{a}{(a+1)^k} \left(1 + \frac{1}{k}\right)^k < \beta_{k,a} < a + 1 - \frac{a}{(a+1)^k}.$$

An application of the well-known inequality  $(1 + 1/k)^k < e$  completes the proof.

## 3.3-7

We will next study the polynomial generated by dividing  $g_{k,a}(t)$  by the factor  $t - \beta_{k,a}$ . Define  $c_j$ ,  $0 \leq j \leq k-1$ , by

$$\frac{g_{k,a}(t)}{t - \beta_{k,a}} = \sum_{j=0}^{k-1} c_{k-1-j} t^j. \quad (3-22)$$

We first prove

LEMMA 3-10.

$$\sum_{j=0}^{k-1} c_j = \frac{ka-1}{\beta_{k,a}-1}.$$

PROOF. The result is obtained by setting  $t = 1$  in (3-22) and observing that  $g_{k,a}(1) = 1 - ka$ .

In the following three lemmas we shall prove that all the roots of the quotient polynomial have moduli less than unity. It will be convenient to sometimes abbreviate  $\beta_{k,a}$  by  $\beta$ .

## 3.3-8

LEMMA 3-11.  $c_j > 0$  for  $0 \leq j \leq k - 1$ .

PROOF. It is easy to show that

$$c_0 = 1, \quad c_j = \beta^j - a \sum_{\ell=0}^{j-1} \beta^\ell, \quad 1 \leq j \leq k - 1. \quad (3-23)$$

A second formula for the  $c_j$  may be derived by noting that

$$\beta^k - a \sum_{\ell=0}^{k-1} \beta^\ell = 0. \quad (3-24)$$

Bringing the last  $k - j$  terms of (3-24) to the right side of the equation and dividing by  $\beta^{k-j}$  yields

$$c_j = \beta^j - a \sum_{\ell=0}^{j-1} \beta^\ell = a \sum_{\ell=1}^{k-j} \beta^{-\ell}, \quad 1 \leq j \leq k - 1.$$

Let  $\theta = \beta^{-1}$ . Then

$$c_j = a\theta \frac{(1-\theta^{k-j})}{1-\theta}. \quad (3-25)$$

The fact that  $\theta \neq 1$  completes the proof.

LEMMA 3-12.

$$c_{j-1} c_{j+1} < c_j^2, \quad k > 2, \quad 1 \leq j \leq k - 2.$$

## 3.3-9

PROOF. Case 1.  $j = 1$ . We must show that  $c_0 c_2 < c_1^2$  or from (3-23),

$$\beta_{k,a}^2 - a\beta_{k,a} - a < (\beta_{k,a} - a)^2.$$

This simplifies to  $0 < 1 + a - \beta_{k,a}$  which is true from Lemma 3-4.

Case 2.  $j > 1$ . Then from (3-25), the result is equivalent to

$$(1-\theta^{k-j-1})(1-\theta^{k-j+1}) < (1-\theta^{k-j})^2,$$

or  $0 < 1 - 2\theta + \theta^2$  which holds since  $\theta \neq 1$ .

LEMMA 3-13. All the roots of the equation  $g_{k,a}(t) = 0$ , other than the root  $\beta_{k,a}$ , have moduli less than one.

PROOF. The proof depends on the following theorem which Ostrowski [3.3-2, p. 90] attributes to independent discoveries by Kakeya [3.3-3] and Eneström.

If in the equation  $g(x) = \sum_{j=0}^n b_{n-j}x^j$  all coefficients  $b_j$  are positive, then we have for each root  $\xi$  that

$$|\xi| \leq \max_{1 \leq j \leq n} b_j/b_{j-1}.$$

## 3.3-10

Applying this theorem to (3-22) we find, by virtue of Lemmas 3-11 and 3-12, that

$$|\xi| \leq \max_{1 \leq j \leq k-1} \frac{c_j}{c_{j-1}} = \frac{c_1}{c_0}.$$

Furthermore  $c_1/c_0 = \beta_{k,a} - a < 1$  from (3-23) and Lemma 3-4.

This completes the proof for the case  $k > 2$ . For  $k = 2$ ,

$$\frac{g_{2,a}(t)}{t-\beta_{2,a}} = t + \beta_{2,a} - a$$

and  $|\xi| = \beta_{2,a} - a < 1$ .

The major results of this section are summarized in

THEOREM 3-2. Let

$$g_{k,a}(t) = t^k - a \sum_{j=0}^{k-1} t^j = 0.$$

If  $k = 1$ , this equation has the real root  $\beta_{1,a} = a$ . Assume  $k \geq 2$  and  $ka > 1$ . Then the equation has one real positive simple root  $\beta_{k,a}$  and

$$\max[1,a] < \beta_{k,a} < a + 1.$$

Furthermore,

$$a + 1 - \frac{ea}{(a+1)^k} < \beta_{k,a} < a + 1 - \frac{a}{(a+1)^k},$$

where e denotes the base of common logarithms. Hence

$\lim_{k \rightarrow \infty} \beta_{k,a} = a + 1$ . All other roots are also simple and have moduli less than one.

3.3-12

3.32 An important special case. A case of special interest occurs when  $a$  is a positive integer. (We will not be interested in nonintegral values of  $a$  until we investigate multiple roots in Chapter 7.) Lemma 3-4 may be simplified to read

$$a < \beta_{k,a} < a + 1, \quad k > 1.$$

Values of  $\beta_{k,a}$  for low values of  $k$  and  $a$  may be found in Table 3-1. Observe that  $\beta_{k,a}$  has almost attained its limit,  $a + 1$ , by the time  $k$  has attained the value 3 or 4. This is particularly true if  $a$  is large. This will have important consequences later.

3.3-13

TABLE 3-1. VALUES OF  $\beta_{k,a}$

$\downarrow k$	$\vec{a}$	1	2	3	4
1		1.000	2.000	3.000	4.000
2		1.618	2.732	3.791	4.828
3		1.839	2.920	3.951	4.967
4		1.928	2.974	3.988	4.994
5		1.966	2.992	3.997	4.999
6		1.984	2.997	3.999	5.000
7		1.992	2.999	4.000	5.000

3.4 The Asymptotic Behavior of the Solutions of Certain Difference Equations

3.41 Introduction. Consider the difference equation

$$e_{i+1} = K \prod_{j=0}^n e_{i-j}^s, \quad (3-26)$$

where  $K$  is a constant and  $s$  is a positive integer. Taking logarithms in (3-26) leads to a linear difference equation with constant coefficients whose indicial equation is

$$t^{n+1} - s \sum_{j=0}^n t^j = 0. \quad (3-27)$$

This is a special case of the equation studied in Section 3.3 with  $k = n + 1$  and  $a = s$ . The properties of the roots of (3-27), derived in the previous section, permit us to analyze the asymptotic behavior of the sequence  $\{e_i\}$ .

Now consider the difference equation

$$e_{i+1} = M_i \prod_{j=0}^n e_{i-j}^s, \quad (3-28)$$

where  $M_i \rightarrow K$ . Is the asymptotic behavior of the sequence generated by this equation the same as that generated by (3-26)? The answer turns out to be in the affirmative. It will turn out that the errors of certain important families of I.F. are governed by difference equations of the form specified by (3-28).

### 3.4-2

In Section 3.43 we shall show that the asymptotic behavior of the sequence

$$e_{i+1} = M_i e_i^s \prod_{j=1}^n (e_{i-j} - e_i)^s + N_i e_i^{s+1}, \quad (3-29)$$

where  $M_i \rightarrow K$ ,  $N_i \rightarrow L$ , is the same as the asymptotic behavior of the sequence generated by (3-26) or (3-28). Difference equations of the type defined by (3-29) will be encountered in Chapter 6. We shall designate difference equations of the type defined by (3-28) and (3-29) as of type 1 and type 2, respectively.

3.42 Difference equations of type 1. We shall study the asymptotic behavior of the solutions of the difference equation

$$e_{i+1} = M_1 \prod_{j=0}^n e_{i-j}^s, \quad (3-30)$$

where  $s$  is a positive integer. We shall show that if

$$M_1 \rightarrow K, \quad (3-31)$$

and if the magnitudes of  $e_0, e_1, \dots, e_n$  are sufficiently small, then the sequence of  $e_i$  converges to zero. Observe that if none of the  $M_i$  are zero and if none of  $e_0, e_1, \dots, e_n$  are zero, then  $e_i$  is not zero for any finite  $i$ . We shall prove that there exists a number  $p$  greater than unity such that  $|e_{i+1}|/|e_i|^p$  converges to a nonzero constant. It is easy to prove (see Section 1.23) that if such a number  $p$  exists, then it is necessarily unique.

Let

$$\delta_1 = |e_1|, \quad r = s(n+1), \quad (3-32)$$

Since convergent sequences are necessarily bounded, there exists a constant  $M$  such that

$$|M_1| \leq M$$

for all  $i$ . Then

$$\delta_{i+1} \leq M \prod_{j=0}^n \delta_{i-j}^s.$$

Let

$$\delta_i \leq r, \quad i = 0, 1, \dots, n.$$

An application of Lemma 3-1 with all the  $\gamma_j$  equal to  $s$  and with  $r$  replacing  $q$  shows that if

$$M r^{r-1} < 1,$$

then  $\delta_i \rightarrow 0$ .

Assume that  $M_1$  and  $K$  are nonzero. We will investigate how fast the sequence  $\{e_i\}$  converges to zero. Let

$$\sigma_i = \ln \delta_i = \ln |e_i|, \quad J_i = \ln |M_i|. \quad (3-33)$$

Then from (3-30), we have that

$$\sigma_{i+1} = J_i + s \sum_{j=0}^n \sigma_{i-j}. \quad (3-34)$$

Let  $t$  be an arbitrary parameter whose value will be chosen later. Let

$$c_0 = 1, \quad c_j(t) = t^j - s \sum_{\ell=0}^{j-1} t^\ell, \quad j = 1, 2, \dots, n+1,$$

(3-35)

## 3.4-5

and let

$$D_1(t) = \sigma_1 - t\sigma_{1-1}. \quad (3-36)$$

Then it is not difficult to see that (3-34) is identical with

$$\sum_{j=0}^n c_j(t) D_{1+j}(t) + c_{n+1}(t) \sigma_{1-n} = J_1 \quad (3-37)$$

for all values of  $t$ .

Consider the equation

$$c_{n+1}(t) = t^{n+1} - s \sum_{\ell=0}^n t^\ell = 0. \quad (3-38)$$

Observe that this is precisely the indicial equation of the difference equation derived from (3-26) by taking logarithms. It is a special case of the equation

$$g_{k,a}(t) = t^k - a \sum_{j=0}^{k-1} t^j = 0 \quad (3-39)$$

studied in Section 3.3 with  $k = n + 1$ ,  $a = s$ . The analysis of (3-39) was based on the assumption that  $ka > 1$ . Let  $r > 1$ . Since  $ka = (n+1)s = r$ , the conclusions of Section 3.3 apply. Hence all the roots of (3-38) are simple and there is one

## 3.4-6

real positive root greater than unity; all other roots have moduli less than unity. The real positive root is labeled  $\beta_{n+1,s}$ . We shall abbreviate  $\beta_{n+1,s}$  by  $p$  and choose the parameter  $t$  equal to  $p$ . Hence

$$c_{n+1}(p) = 0. \quad (3-40)$$

Let

$$c_j = c_j(p), \quad D_j = D_j(p).$$

Then (3-37) becomes

$$\sum_{j=0}^n c_j D_{i+1-j} = J_i. \quad (3-41)$$

The  $c_j$  are just the coefficients of the polynomial gotten by dividing  $c_{n+1}(t)$  by  $t - p$ . [See (3-23).] Hence the indicial equation corresponding to the difference equation (3-41) has only simple roots whose moduli are less than unity. Since  $M_i \rightarrow K$ , we conclude that  $J_i \rightarrow \ln|K|$ . All the conditions of the Corollary to Theorem 3-1 now apply and we conclude that

$$D_i \rightarrow \frac{\ln|K|}{\sum_{j=0}^n c_j}. \quad (3-42)$$

3.4-7

An application of Lemma 3-10, with  $ka = (n+1)s = r$  and with  $\beta_{k,a} = \beta_{n+1,s} = p$ , shows that

$$\sum_{j=0}^n c_j = \frac{r-1}{p-1}. \quad (3.43)$$

From (3-33), (3-36), (3-40), (3-42), and (3-43), we conclude that

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |K|^{(p-1)/(r-1)}.$$

We summarize these results in

**THEOREM 3-3.** Let

$$e_{i+1} = M_i \prod_{j=0}^n e_{i-j}^s$$

with

$$|e_i| \leq \Gamma, \quad i = 0, 1, \dots, n.$$

Let  $s$  be a positive integer and let  $r = s(n+1) > 1$ . Let  $M_i \rightarrow K$  and let  $|M_i| \leq M$  for all  $i$ . Let

$$M\Gamma^{r-1} < 1.$$

## 3.4-8

Then  $e_1 \rightarrow 0$ . Let  $p$  be the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

Let  $M_1$  and  $K$  be nonzero. Then

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |K|^{(p-1)/(r-1)}. \quad (3-44)$$

Table 3-1 gives values of  $p = p_{n+1,s}$  for low values of  $n$  and  $s$ . Note that no a priori assumption has been made of the existence of a number  $p$  such that

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow c \neq 0.$$

If the existence of such a number is assumed a priori, it is not difficult to prove that it must satisfy the indicial equation (3-38).

3.43 Difference equations of type 2. We shall study the asymptotic behavior of the solutions of the difference equation

$$e_{i+1} = M_i e_i^s \prod_{j=1}^n (e_{i-j} - e_i)^s + N_i e_i^{s+1}, \quad (3-45)$$

where

$$M_i \rightarrow K \neq 0, \quad N_i \rightarrow L, \quad (3-46)$$

and where  $n$  and  $s$  are positive integers. We shall show that if the magnitudes of  $e_0, e_1, \dots, e_n$  are sufficiently small, then  $e_i \rightarrow 0$  and there exists a number  $p$  greater than unity such that  $|e_{i+1}|/|e_i|^p$  converges to a nonzero constant. In fact, we shall demonstrate that the asymptotic behavior of the sequence generated by (3-45) is identical with the asymptotic behavior of the sequence generated by

$$e_{i+1} = M_i \prod_{j=0}^n e_{i-j}^s, \quad M_i \rightarrow K$$

which was studied in the previous section.

Let

$$\delta_i = |e_i|, \quad r = s(n+1) \quad (3-47)$$

and let

$$|M_i| \leq M, \quad |N_i| \leq N$$

3.4-10

for all  $i$ . Then

$$\delta_{i+1} \leq M\delta_i^s \prod_{j=1}^n (\delta_{i-j} + \delta_i)^s + N\delta_i^{s+1}.$$

Let

$$\delta_i \leq \Gamma, \quad i = 0, 1, \dots, n.$$

An application of Lemma 3-2, with all the  $\gamma_j$  equal to  $s$  and with  $q$  equal to  $r$ , shows that if

$$2^{r-s}M\Gamma^{r-1} + N\Gamma^s < 1,$$

then  $\delta_i \rightarrow 0$ .

Assume that  $e_i$  is nonzero for all finite  $i$ . We may write (3-45) as

$$e_{i+1} = \tau_i \prod_{j=0}^n e_{i-j}^s, \quad (3-48)$$

with

$$\tau_i = M_i \lambda_i + N_i \theta_i, \quad (3-49)$$

where

$$\lambda_i = \prod_{j=1}^n \left(1 - \frac{e_i}{e_{i-j}}\right)^s, \quad \theta_i = \frac{e_i}{\prod_{j=1}^n e_{i-j}^s}. \quad (3-50)$$

We shall demonstrate that  $\lambda_i \rightarrow 1$  and  $\theta_i \rightarrow 0$ .

To show that  $\lambda_i \rightarrow 1$  we note from (3-45) that  $e_{i+1}/e_i \rightarrow 0$ . Therefore  $e_i/e_{i-j} \rightarrow 0$  for all finite  $j$  and the result follows.

To show that  $\theta_i \rightarrow 0$  we proceed as follows. From (3-45) and (3-50),

$$e_i = M_{i-1} \lambda_{i-1} \prod_{j=0}^n e_{i-1-j}^s + N_{i-1} e_{i-1}^{s+1}.$$

Hence

$$\frac{e_i}{\prod_{j=1}^n e_{i-j}^s} = M_{i-1} \lambda_{i-1} e_{i-1-n}^s + N_{i-1} \frac{e_{i-1}}{\prod_{j=2}^n e_{i-j}^s}. \quad (3-51)$$

Repeat this process for the second term on the right side of (3-51). Carry out this reduction a total of  $n - 1$  times. The problem is reduced to proving that  $e_{i+1}/e_i^s \rightarrow 0$  and this is clearly true from (3-45).

Since  $\lambda_i \rightarrow 1$ ,  $\theta_i \rightarrow 0$ , and  $M_i \rightarrow K$ , we conclude that  $\tau_i \rightarrow K$ . Theorem 3-3 may now be applied to (3-48) and we arrive at

**THEOREM 3-4.** Let

$$e_{i+1} = M_i e_i^s \prod_{j=1}^n (e_{i-j} - e_i)^s + N_i e_i^{s+1}$$

3.4-12

with

$$|e_i| \leq r, \quad i = 0, 1, \dots, n.$$

Let  $s$  be a positive integer and let  $r = s(n+1) > 1$ . Let  $M_i \rightarrow K \neq 0$ ,  $N_i \rightarrow L$ , and let  $|M_i| \leq M$ ,  $|N_i| \leq N$  for all  $i$ .  
Let

$$2^{r-s} M r^{r-1} + N r^s < 1.$$

Then  $e_i \rightarrow 0$ . Let  $p$  be the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

Assume that  $e_i \neq 0$  for all finite  $i$ . Then

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |K|^{(p-1)/(r-1)}.$$

## 4.0-1

### CHAPTER 4

#### INTERPOLATORY ITERATION FUNCTIONS

In this chapter we shall study I.F. which are generated by direct or inverse hyperosculatory interpolation; such I.F. will be called interpolatory I.F. The major results concerning the convergence and order of interpolatory I.F. will be given in Theorems 4-1 and 4-3. A sweeping generalization of Fourier's conditions for monotone convergence to a solution will be given in Theorem 4-2.

## 4.1-1

### 4.1 Interpolation and the Solution of Equations

4.11 Statement and solution of an interpolation problem. The reader is referred to Appendix A for a discussion of hyperosculatory interpolation theory. Certain salient features of that discussion which are needed for the development of the theory of this chapter will be repeated here.

Consider the following rather general interpolation problem. We seek a polynomial  $P$  such that

$$P^{(k_j)}(x_{i-j}) = f^{(k_j)}(x_{i-j}) \quad \text{for } j = 0, 1, \dots, n; \quad (4-1)$$

$$k_j = 0, 1, \dots, \gamma_j - 1, \quad \gamma_j \geq 1; \quad x_{i-k} \neq x_{i-\ell} \quad \text{if } k \neq \ell.$$

That is, the first  $\gamma_j - 1$  derivatives of  $P$  are to agree with the first  $\gamma_j - 1$  derivatives of  $f$  at the  $n + 1$  points  $x_i, x_{i-1}, \dots, x_{i-n}$ . Let

$$q = \sum_{j=0}^n \gamma_j.$$

Then there exists a unique polynomial  $P_{n, \gamma_0, \gamma_1, \dots, \gamma_n}$  of degree  $q - 1$  which satisfies (4-1). For the sake of brevity we write

$$P_{n, \gamma} \equiv P_{n, \gamma_0, \gamma_1, \dots, \gamma_n}, \quad (4-2)$$

## 4.1-2

where  $\gamma$  signifies the vector  $\gamma_0, \gamma_1, \dots, \gamma_n$ . If, in particular,  $\gamma_j = s$  for all  $\gamma_j$ , then we write the interpolatory polynomial as  $P_{n,s}$ .

Let  $f^{(q)}(t)$  be continuous in the interval determined by  $x_1, x_{1-1}, \dots, x_{1-n}, t$ . Then

$$f(t) = P_{n,\gamma}(t) + \frac{f^{(q)}[\xi_1(t)]}{q!} \prod_{j=0}^n (t-x_{1-j})^{\gamma_j}, \quad (4-3)$$

where  $\xi_1(t)$  lies in the interval determined by  $x_1, x_{1-1}, \dots, x_{1-n}, t$ .

Let  $f'$  be nonzero and let  $f^{(q)}$  be continuous on an interval  $J$ . Let  $f$  map  $J$  into  $K$ . Then  $f$  has an inverse  $\mathfrak{F}$  and  $\mathfrak{F}^{(q)}$  is continuous on  $K$ . We can state the interpolation problem for  $\mathfrak{F}$  as follows. We seek a polynomial  $Q$  such that

$$Q^{(k_j)}(y_{1-j}) = \mathfrak{F}^{(k_j)}(y_{1-j}) \quad \text{for } j = 0, 1, \dots, n; \quad (4-4)$$

$$k_j = 0, 1, \dots, \gamma_j - 1, \quad \gamma_j \geq 1; \quad y_{1-k} \neq y_{1-\ell} \quad \text{if } k \neq \ell.$$

Then there exists a unique polynomial  $Q_n, \gamma_0, \gamma_1, \dots, \gamma_n$  of degree  $q - 1$  which satisfies (4-4). For the sake of brevity we write

$$Q_{n,\gamma} \equiv Q_{n,\gamma_0, \gamma_1, \dots, \gamma_n}. \quad (4-5)$$

If, in particular,  $\gamma_j = s$  for all  $\gamma_j$ , then we write the interpolatory polynomial as  $Q_{n,s}(t)$ .

#### 4.1-3

The error of the interpolation is given by

$$v(t) = Q_{n,\gamma}(t) + \frac{v^{(q)}[\theta_1(t)]}{q!} \prod_{j=0}^n (t-y_{1-j})^s, \quad (4-6)$$

where  $\theta_1(t)$  lies in the interval determined by

$y_1, y_{1-1}, \dots, y_{1-n}, t.$

#### 4.12 Relation of interpolation to the calculation

of roots. Let  $x_1, x_{1-1}, \dots, x_{1-n}$  be  $n + 1$  approximants to a zero  $\alpha$  of the function  $f$ . To calculate a new approximant  $x_{1+1}$ , it is reasonable to calculate the zero of the polynomial which interpolates  $f$  at the points  $x_1, x_{1-1}, \dots, x_{1-n}$ . The process is then repeated for the set  $x_{1+1}, x_1, \dots, x_{1-n+1}$ . One drawback of this procedure is that a polynomial equation must be solved at each step of the iteration. A second drawback is that the polynomial will have a number of zeros, some of which may be complex, and criteria are required to select one of these zeros as  $x_{1+1}$ . Once such criteria are established the point  $x_{1+1}$  is uniquely determined by the points  $x_1, x_{1-1}, \dots, x_{1-n}$ . We define  $\Phi_{n,\gamma}$  as the function which maps  $x_1, x_{1-1}, \dots, x_{1-n}$  into  $x_{1+1}$ .

The difficulties of having to solve a polynomial equation may be avoided by interpolating  $\tilde{f}$ , the inverse to  $f$ , at the points  $y_1, y_{1-1}, \dots, y_{1-n}$ , and evaluating the interpolatory polynomial at zero. The point  $x_{1+1}$  is uniquely determined; let the function which maps  $x_1, x_{1-1}, \dots, x_{1-n}$  into  $x_{1+1}$  be labeled  $\Phi_{n,\gamma}$ .

For either of the two processes described above,  $x_0, x_1, \dots, x_n$  must be available. One method for obtaining these starting values is described in Section 6.32.

#### 4.1-5

I.F. which are generated through direct or inverse hyperosculatory interpolation will be called interpolatory I.F. Hyperosculatory interpolation is not the only means by which I.F. may be generated. Other techniques will be studied in various parts of this book. It is a very useful technique however and gives a uniform method for deriving I.F. and of studying their properties and, in particular, their order. The most widely known I.F. are all examples of interpolatory I.F. A useful by-product of generating I.F. by hyperosculatory interpolation is the introduction of a natural classification scheme.

4.2 The Order of Interpolatory Iteration Functions

4.21 The order of iteration functions generated by inverse interpolation. Let  $x_1, x_{1-1}, \dots, x_{1-n}$  be  $n + 1$  approximations to a zero  $\alpha$  of  $f$ . Let  $Q_{n,\gamma}$  be the polynomial which interpolates  $\mathfrak{f}$  at the points  $y_1, y_{1-1}, \dots, y_{1-n}$  in the sense of (4-4). Define a new approximation to  $\alpha$  by

$$x_{1+1} = Q_{n,\gamma}(0).$$

Then repeat this procedure using the points  $x_{1+1}, x_1, \dots, x_{1-n+1}$ . We shall investigate how the error of  $x_{1+1}$  depends on the errors at the previous  $n + 1$  points.

We observed (4-6) that

$$\mathfrak{f}(t) = Q_{n,\gamma}(t) + \frac{\mathfrak{f}^{(q)}[\theta_1(t)]}{q!} \prod_{j=0}^n (t-y_{1-j})^{\gamma_j},$$

where  $\theta_1(t)$  lies in the interval determined by

$y_1, y_{1-1}, \dots, y_{1-n}, t$ , and where  $q = \sum_{j=0}^n \gamma_j$ . Set  $t = 0$ . Then

$$x_{1+1} - \alpha = - \frac{(-1)^q}{q!} \mathfrak{f}^{(q)}(\theta_1) \prod_{j=0}^n (y_{1-j})^{\gamma_j}, \quad (4-7)$$

where  $\theta_1 \equiv \theta_1(0)$ . Let  $e_{1-j} = x_{1-j} - \alpha$ . We can write (4-7) in terms of either the  $y_{1-j}$  or the  $e_{1-j}$ . Since

$$y_{1+1} = f(x_{1+1}) = f'(\eta_{1+1})e_{1+1},$$

4.2-2

where  $\eta_{i+1}$  lies in the interval determined by  $x_{i+1}$  and  $a$ , we conclude that

$$y_{i+1} = *M_i \prod_{j=0}^n y_{i-j}^s, \quad (4-8)$$

$$*M_i = - \frac{(-1)^q \vartheta^{(q)}(\theta_i)}{q! \vartheta'(\rho_{i+1})},$$

where  $\rho_{i+1} = f(\eta_{i+1})$ . Since

$$y_{i-j} = f(x_{i-j}) = f'(\eta_{i-j}) e_{i-j},$$

we can also conclude that

$$e_{i+1} = M_i \prod_{j=0}^n e_{i-j}^s, \quad (4-9)$$

$$M_i = - \frac{(-1)^q \vartheta^{(q)}(\theta_i)}{q! \prod_{j=0}^n [\vartheta'(\rho_{i-j})]^{\gamma_j}},$$

where  $\rho_{i-j} = f(\eta_{i-j})$ .

It is clear that if none of the set  $e_0, e_1, \dots, e_n$  is zero and if  $\vartheta^{(q)}$  does not vanish on the interval of iteration, then  $e_i$  is not equal to zero for any finite  $i$ .

## 4.2-3

We shall show, however, that if the initial approximations  $x_0, x_1, \dots, x_n$ , are sufficiently close to  $\alpha$ , then  $e_1 \rightarrow 0$ . We will use (4-9) for the proof. Let

$$J = \{x \mid |x - \alpha| \leq \Gamma\}.$$

Let  $f^{(q)}$  be continuous on  $J$  and let  $f'$  be nonzero on  $J$ . Let  $f$  map the interval  $J$  into the interval  $K$ . Then  $\tilde{f}^{(q)}$  is continuous on  $K$  and  $\tilde{f}'$  is nonzero there. Let

$$\frac{|\tilde{f}^{(q)}(y)|}{q!} \leq \lambda_1, \quad |\tilde{f}'(y)| \geq \lambda_2$$

for all  $y \in K$ ; that is, for all  $x \in J$ . Let

$$M = \frac{\lambda_1}{\lambda_2^q}.$$

Let  $x_0, x_1, \dots, x_n \in J$ . Then

$$\max[|e_0|, |e_1|, \dots, |e_n|] \leq \Gamma.$$

We shall show that if

$$M\Gamma^{q-1} \leq 1, \quad (4-10)$$

then all the  $x_i \in J$ .

4.2-4

Since  $x_0, x_1, \dots, x_n \in J$ ,  $|M_n| \leq M$ . Hence

$$|e_{n+1}| = |M_n| \prod_{j=0}^n |e_{1-j}|^{\gamma_j} \leq M\Gamma^q \leq \Gamma,$$

where the last inequality is due to (4-10). We proceed by induction. Let  $x_i \in J$ , for  $i = 0, 1, \dots, k$ . Hence  $|M_k| \leq M$  and

$$|e_{k+1}| = |M_k| \prod_{j=0}^n |e_{k-j}|^{\gamma_j} \leq M\Gamma^q \leq \Gamma,$$

which completes the induction. Since all the  $x_i \in J$ ,  $|M_i| \leq M$  for all  $i$ .

A modification of this proof could be used to show that  $e_i \rightarrow 0$ . Instead we note that

$$\delta_{i+1} \leq M \prod_{j=0}^n (\delta_{1-j})^{\gamma_j}, \quad \delta_{1-j} = |e_{1-j}|,$$

and use Lemma 3-1 to arrive at

LEMMA 4-1. Let  $q = \sum_{j=0}^n \gamma_j$ . Let

$$J = \left\{ x \mid |x - \alpha| \leq \Gamma \right\}$$

4.2-5

and let  $f^{(q)}$  be continuous and  $f'$  nonzero on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$ . Let

$$\frac{|\mathfrak{f}^{(q)}(y)|}{q!} \leq \lambda_1, \quad |\mathfrak{f}'(y)| \geq \lambda_2$$

for all  $x \in J$  and let  $M = \lambda_1/\lambda_2^q$ . Suppose that

$$M\Gamma^{q-1} < 1.$$

Then  $e_i \rightarrow 0$ .

Since  $e_i \rightarrow 0$ ,  $x_i \rightarrow a$ , and we can conclude that  
 $y_i \rightarrow 0$ .

4.22 The equal information case. We will now specialize to the case of interest for our future work. Let

$$\gamma_j = s, \quad j = 0, 1, \dots, n.$$

Then the same amount of information will be used at each point. Thus the first  $s - 1$  derivatives of the interpolatory polynomial are to agree with the first  $s - 1$  derivatives of  $\vartheta$  at  $y_1, y_{1-1}, \dots, y_{1-n}$ . Let

$$r = s(n+1).$$

The results of the previous section hold for this case if we replace  $q$  by  $r$  and  $\gamma_j$  by  $s$ . In particular,

$$y_{1+l} = *M_1 \prod_{j=0}^n y_{1-j}^s, \quad (4-11)$$

$$*M_1 = - \frac{(-1)^r \vartheta(r)(\theta_1)}{r! \vartheta'(\rho_{1+l})},$$

and

$$e_{1+l} = M_1 \prod_{j=0}^n e_{1-j}^s, \quad (4-12)$$

$$M_1 = - \frac{(-1)^r \vartheta(r)(\theta_1)}{r! \prod_{j=0}^n [\vartheta'(\rho_{1-j})]^s}.$$

Let the conditions of Lemma 4-1 hold with  $q$  replaced by  $r$  and  $\gamma_j$  replaced by  $s$ . Then  $|M_i| \leq M$  for all  $i$ ,  $e_1 \rightarrow 0$ , and  $M_i \rightarrow Y_r(\alpha)$ , where  $Y_r(x)$  was defined (1-8) as

$$Y_r(x) = - \frac{(-1)^r \mathfrak{g}^{(r)}(y)}{r! [\mathfrak{g}'(y)]^r} \Big|_{y=f(x)}.$$

All the conditions of Theorem 3-3 are now satisfied and we conclude that

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |Y_r(\alpha)|^{(p-1)/(r-1)}, \quad (4-13)$$

where  $p$  is the unique real positive root with magnitude greater than unity of the equation

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

Furthermore,  $*M_i \rightarrow -(-1)^r a_r(0)$ , where  $a_r(y)$  is defined (1-8) as

$$a_r(y) = \frac{\mathfrak{g}^{(r)}(y)}{r! \mathfrak{g}'(y)}.$$

Hence

$$\frac{|y_{i+1}|}{|y_i|^p} \rightarrow |a_r(0)|^{(p-1)/(r-1)}. \quad (4-14)$$

It is not difficult to see that one can pass from (4-14) to (4-13) by observing that  $y_{i-j} = f'(\eta_{i-j})e_{i-j}$ .

Our results concerning the convergence and order of I.F. generated by inverse interpolation are summarized in

THEOREM 4-1. Let

$$J = \left\{ x \mid |x - \alpha| \leq \Gamma \right\}.$$

Let  $r = s(n+1) > 1$ . Let  $f^{(r)}$  be continuous and let  $f', \tilde{f}^{(r)} \neq 0$  on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$  and let a sequence  $\{x_i\}$  be defined as follows: Let  $Q_{n,s}$  be an interpolatory polynomial for  $\tilde{f}$  such that the first  $s - 1$  derivatives of  $Q_{n,s}$  are equal to the first  $s - 1$  derivatives of  $\tilde{f}$  at the points  $y_1, y_{1-1}, \dots, y_{1-n}$ . Define

$$x_{i+1} = \varphi_{n,s}(x_i; x_{i-1}, \dots, x_{i-n}) = Q_{n,s}(0).$$

Let  $e_{i-j} = x_{i-j} - \alpha$ . Let

$$\frac{|\mathfrak{x}^{(r)}|}{r!} \leq \lambda_1, \quad |\mathfrak{x}'| \geq \lambda_2,$$

for all  $x \in J$ , and let  $M = \lambda_1/\lambda_2^r$ . Suppose that  $M\Gamma^{r-1} < 1$ .

Then,  $x_i \in J$  for all  $i$ ,  $e_i \rightarrow 0$ , and

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |y_r(\alpha)|^{(p-1)/(r-1)}, \quad (4-15)$$

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0, \quad (4-16)$$

and where

$$y_r(x) = - \left. \frac{(-1)^r \mathfrak{x}^{(r)}(y)}{r! [\mathfrak{x}'(y)]^r} \right|_{y=f(x)}.$$

Also

$$\frac{|y_{i+1}|}{|y_i|^p} \rightarrow |a_r(0)|^{(p-1)/(r-1)},$$

where

$$a_r(y) = \frac{\mathfrak{x}^{(r)}(y)}{r! \mathfrak{x}'(y)}.$$

We close this section with a number of comments.

We have proven that the order of any I.F. generated by inverse interpolation is given by a certain number  $p$  without making any a priori assumptions about the asymptotic behavior of the sequence of errors. If we had assumed a priori the existence of a number  $p$  such that  $|e_{i+1}|/|e_i|^p$  converges to a limit, then it would have been much easier to prove that this number  $p$  was determined by the indicial equation (4-16).

If  $n = 0$ , then  $\varphi_{n,s}$  is a one-point I.F.; if  $n > 0$ , then  $\varphi_{n,s}$  is a one-point I.F. with memory. The order  $p$  is an integer if and only if  $n = 0$ ; that is, if the I.F. has no memory.

Observe that the asymptotic error constant of the sequence  $\{y_i\}$  depends upon  $A_r$  whereas the asymptotic error constant of the sequence  $\{x_i\}$  depends upon  $Y_r$ . We shall see that whenever we deal with direct interpolation, the asymptotic error constant of the sequence  $\{x_i\}$  will depend upon  $A_r$ . Recall (1-8) that  $A_r$  is the same function of  $f$  as  $a_r$  is of  $\tilde{f}$ . Thus, in a certain sense,  $\{y_i\}$  plays the same role for inverse interpolation that  $\{x_i\}$  plays for direct interpolation.

Values of  $p$  for different values of  $n$  and  $s$  may be found in Table 3-1 with  $k = n + 1$  and  $a = s$ .

4.23 The order of iteration functions generated by direct interpolation. We now investigate the case where a new approximation to  $\alpha$  is generated by solving the polynomial which interpolates  $f$ . We immediately turn to the case where all the  $\gamma_j$  are equal to  $s$ .

Let  $x_i, x_{i-1}, \dots, x_{i-n}$  be  $n+1$  approximations to a zero  $\alpha$  of  $f$ . Let  $P_{n,s}$  be the polynomial whose first  $s-1$  derivatives are equal to the first  $s-1$  derivatives of  $f$  at  $x_i, x_{i-1}, \dots, x_{i-n}$ . Define a new approximation to  $\alpha$  by

$$P_{n,s}(x_{i+1}) = 0. \quad (4-17)$$

Then repeat this procedure for  $x_{i+1}, x_i, \dots, x_{i-n+1}$ .

Since  $P_{n,s}$  is a polynomial of degree  $r-1$ , where  $r = s(n+1)$ ,  $x_{i+1}$  will not generally be uniquely specified by (4-17). It is not even clear a priori that  $P_{n,s}$  has a real zero in the neighborhood of  $\alpha$ . We shall prove that under suitable conditions  $P_{n,s}$  does possess a real zero in the neighborhood of  $\alpha$ . Under certain hypotheses we shall, in fact, be able to prove much more.

Let  $J = \{x \mid |x-\alpha| \leq \Gamma\}$  and let  $x_1, x_{1-1}, \dots, x_{1-n} \in J$ . Let  $f'$  be nonzero on  $J$ . If the  $x_{1-j}$  bracket  $\alpha$ , then it is clear that  $P_{n,s}$  has a real zero in  $J$ . Hence it is sufficient to investigate the case where all the  $x_{1-j}$  lie on one side of  $\alpha$ . We shall first prove

LEMMA 4-2. Let

$$J = \{x \mid |x-\alpha| \leq \Gamma\}.$$

Let  $f^{(r)}$  be continuous on  $J$  and let  $f' \neq 0$  on  $J$ . Let  $x_1, x_{1-1}, \dots, x_{1-n} \in J$  and let  $x_1, x_{1-1}, \dots, x_{1-n}$  lie on one side of  $\alpha$ . Let

$$\frac{|f^{(r)}|}{r!} \leq v_1, \quad |f'| \geq v_2$$

for all  $x \in J$ . Suppose that

$$\frac{v_1}{v_2} \frac{(2\Gamma)^r}{\Gamma} < 1. \quad (4-18)$$

Then  $P_{n,s}$  has a real root,  $x_{1+1}$ , which lies in  $J$ .

PROOF. To prove the lemma it is sufficient to prove the result for the case that

$$x_{1-j} > \alpha, \quad j = 0, 1, \dots, n,$$

and  $f' > 0$ . Hence  $P_{n,s}(x_1) = f(x_1)$  is positive. We shall prove that  $P_{n,s}(\alpha-\Gamma)$  is negative. Since  $P_{n,s}$  interpolates  $f$ ,

$$P(t) = f(t) - \frac{f^{(r)}[\xi(t)]}{r!} \prod_{j=0}^n (t-x_{1-j})^s,$$

where  $\xi(t)$  lies in the interval determined by

$x_1, x_{1-1}, \dots, x_{1-n}, t$ . Then

$$P(\alpha-\Gamma) = f(\alpha-\Gamma) - \frac{f^{(r)}(\xi)}{r!} \prod_{j=0}^n (\alpha-\Gamma-x_{1-j})^s,$$

where  $\xi \equiv \xi(\alpha-\Gamma)$ . Furthermore,

$$P(\alpha-\Gamma) = -\Gamma f(\eta) - \frac{(-1)^r}{r!} f^{(r)}(\xi) \prod_{j=0}^n (\Gamma+x_{1-j}-\alpha)^s,$$

where  $\eta$  lies in  $(\alpha-\Gamma, \alpha)$ . Hence  $P(\alpha-\Gamma) < 0$  if

$$T = -\frac{(-1)^r}{r!} \frac{f^{(r)}(\xi)}{f'(\eta)\Gamma} \prod_{j=0}^n (\Gamma+x_{1-j}-\alpha)^s < 1.$$

Since

$$|T| \leq \frac{v_1}{v_2} \frac{(2r)^r}{r} < 1,$$

the proof is complete.

Hence  $P_{n,s}$  has a real root in  $J$  provided that (4-18) holds. We shall now show that under certain conditions we can prove a much stronger result without demanding that (4-18) holds.

LEMMA 4-3. Let

$$J = \{x \mid |x-\alpha| \leq r\}.$$

Let  $f^{(r)}$  be continuous on  $J$  and let  $f'f^{(r)} \neq 0$  on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$  and assume that these points all lie on one side of  $\alpha$ . Let these points be labeled such that  $x_n$  is the closest point to  $\alpha$ . Let

$$f(x_n)f^{(r)}(x_n) > 0, \quad r \text{ even}, \quad (4-19)$$

$$f'(x_n)f^{(r)}(x_n) < 0, \quad r \text{ odd}. \quad (4-20)$$

Then  $P_{n,s}$  has a real root  $x_{n+1}$  such that

$$\min[\alpha, x_n] < x_{n+1} < \max[\alpha, x_n].$$

PROOF. There are four possible cases depending on the signs of  $f(x_n)$  and  $f'$ . We shall prove the result for only two cases which will give the flavor of the proof; the other cases may be handled analogously.

Case 1.  $f(x_n) > 0, f' > 0$ . We need only prove that  $P_{n,s}(\alpha) < 0$ . Since  $P_{n,s}$  interpolates  $f$ ,

$$P_{n,s}(t) = f(t) - \frac{f^{(r)}[\xi(t)]}{r!} \prod_{j=0}^n (t-x_{1-j})^s. \quad (4-21)$$

Hence

$$P_{n,s}(\alpha) = - \frac{(-1)^r}{r!} f^{(r)}(\xi) \prod_{j=0}^n (x_{1-j}-\alpha)^s,$$

where  $\xi \equiv \xi(\alpha)$ . For this case  $x_{1-j} - \alpha > 0$ . Hence  $P_{n,s}(\alpha)$  is negative if  $(-1)^{r-1} f^{(r)}$  is negative. The proof of Case 1 may now be easily completed.

Case 2.  $f(x_n) < 0, f' > 0$ . We need only prove that  $P_{n,s}(\alpha) > 0$ . From (4-21),

$$P_{n,s}(\alpha) = - \frac{f^{(r)}(\xi)}{r!} \prod_{j=0}^n (\alpha-x_{1-j})^s.$$

4.2-16

For this case  $\alpha - x_{i-j} > 0$ . Hence  $P_{n,s}(\alpha)$  is positive if  $-f'(r)$  is positive; that is if

$$f(x_n)f^{(r)}(x_n) > 0, \quad f'(x_n)f^{(r)}(x_n) < 0.$$

Let the hypotheses of the preceding lemma hold.

We can conclude that if  $\alpha < x_n$ , then

$$\alpha < x_{i+1} < x_i, \quad i = n, n+1, \dots . \quad (4-22)$$

If  $x_n < \alpha$ , then

$$x_i < x_{i+1} < \alpha, \quad i = n, n+1, \dots . \quad (4-23)$$

Let (4-22) hold. Then the sequence  $\{x_i\}$  is monotone decreasing and bounded from below. Hence it has a limit and

$$x_{i+1} - x_{i-j} \rightarrow 0, \quad j = 0, 1, \dots, n. \quad (4-24)$$

Label the limit  $\xi$ . Let  $t = x_{i+1}$  in (4-21). Then

$$f(x_{i+1}) = \frac{f^{(r)}(\xi)}{r!} \prod_{j=0}^n (x_{i+1} - x_{i-j})^s,$$

where  $\xi = \xi(x_{i+1})$  and where we have used the fact that  $P_{n,s}(x_{i+1}) = 0$ . Let  $i \rightarrow \infty$ . An application of (4-24) yields  $f(\xi) = 0$ . Since  $f'$  is assumed to be nonzero,  $\xi = a$  and hence  $x_i \rightarrow a$ . The same conclusion would have been obtained if we had started with (4-23). We summarize our results in

THEOREM 4-2. Let

$$J = \{x \mid |x-a| \leq r\}.$$

Let  $r = s(n+1) > 1$ . Let  $f^{(r)}$  be continuous on  $J$  and let  $f'f^{(r)} \neq 0$  on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$  and assume that these points all lie on one side of  $a$ . Suppose that

$$f(x_1)f^{(r)}(x_1) > 0, \quad r \text{ even} \quad (4-25)$$

$$f'(x_1)f^{(r)}(x_1) < 0, \quad r \text{ odd} \quad (4-26)$$

where  $i$  is any of  $0, 1, \dots, n$ . Define a sequence  $\{x_i\}$  as follows: Let  $P_{n,s}$  be an interpolatory polynomial for  $f$  such that the first  $s - 1$  derivatives of  $P_{n,s}$  are equal to the first  $s - 1$  derivatives of  $f$  at the points  $x_i, x_{i-1}, \dots, x_{i-n}$ . Let  $x_{i+1}$  be a point (whose existence we have verified in the preceding discussion) such that  $x_{i+1}$  is real,  $P_{n,s}(x_{i+1}) = 0$ , and

$$\min[a, x_i] < x_{i+1} < \max[a, x_i].$$

Then the sequence  $\{x_i\}$  converges monotonically to  $a$ .

This result is well known for the special case  $n = 0, s = 2$  which is Newton's I.F. For this case the conditions of Theorem 4-2 are known as Fourier conditions (Fourier [4.2-1]). Theorem 4-2 gives a sweeping generalization of Fourier's result. Although the sufficiency of the Fourier conditions are geometrically self-evident for Newton's method, they are not self-evident in the general case.

Observe that the hypotheses of Theorem 4-2 do not place any restrictions on the size of the interval where monotone convergence is guaranteed other than that  $f'f^{(r)} \neq 0$ . Note that the condition that  $x_0, x_1, \dots, x_n$  all lie on one side of  $J$  is automatically satisfied if  $n = 0$ . If (4-25) and (4-26) do not hold, then we shall have to place restrictions on the size of the interval where convergence is guaranteed. We have already seen in the proof of Lemma 4-2 that in the general case we require a condition on the size of the interval in order to assure that  $P_{n,s}$  has a real zero in the interval.

To prove convergence in the general case we start again with

$$P_{n,s}(t) = f(t) - \frac{f^{(r)}[\xi_1(t)]}{r!} \prod_{j=0}^n (t-x_{1-j})^s.$$

Then

$$P_{n,s}(\alpha) = -\frac{(-1)^r}{r!} f^{(r)}(\xi_i) \prod_{j=0}^n e_{i-j}^s, \quad e_{i-j} = x_{i-j} - \alpha, \quad \xi_i \equiv \xi_i(\alpha).$$

Let the conditions of Lemma 4-2 hold. Then there exists a real  $x_{i+1} \in J$  such that  $P_{n,s}(x_{i+1}) = 0$ . Furthermore,

$$P_{n,s}(\alpha) = (\alpha - x_{i+1}) P'(\eta_{i+1}),$$

where  $\eta_{i+1}$  lies in the interval determined by  $\alpha$  and  $x_{i+1}$ .

Then

$$e_{i+1} P'_{n,s}(\eta_{i+1}) = \frac{(-1)^r}{r!} f^{(r)}(\xi_i) \prod_{j=0}^n e_{i-j}^s.$$

Assume that  $P'_{n,s}$  does not vanish in the interval determined by  $\alpha$  and  $x_{i+1}$ . Then

$$e_{i+1} = H_{i+1} \prod_{j=0}^n e_{i-j}^s,$$

(4-27)

$$H_{i+1} = \frac{(-1)^r}{r!} \frac{f^{(r)}(\xi_i)}{P'_{n,s}(\eta_{i+1})}.$$

Observe that if none of the set  $e_0, e_1, \dots, e_n$  is zero and if  $f^{(r)}$  does not vanish on the interval of iteration, then  $e_i$  is not equal to zero for any finite  $i$ .

We shall show, however, that if the initial approximations  $x_0, x_1, \dots, x_n$ , are sufficiently close to  $\alpha$ , then  $e_i \rightarrow 0$ . Let

$$\frac{|f^{(r)}|}{r!} \leq v_1$$

for all  $x \in J$ . Let

$$|P'_{n,s}| \geq \mu_2$$

for all  $x$  in the interval determined by  $\alpha$  and  $x_{i+1}$ . Let  $H = v_1/\mu_2$ . Then  $|H_{i+1}| \leq H$  and

$$\delta_{i+1} \leq H \prod_{j=0}^n \delta_{i-j}^s, \quad \delta_{i-j} = |e_{i-j}|.$$

An application of Lemma 3-1 shows that if

$$H^r < 1$$

then  $e_i \rightarrow 0$ .

Observe that the argument used in the proof of Lemma 4-1 could not be used because  $H_{i+1}$  depends upon  $x_{i+1}$ .  
 Since

$$P'_{n,s}(x_{i+1}) \rightarrow f'(\alpha),$$

we conclude that

$$H_{i+1} \rightarrow (-1)^r A_r(\alpha), \quad A_r = \frac{f^{(r)}(\alpha)}{r! f'}. \quad (4-27)$$

All the conditions of Theorem 3-3 are now satisfied and we conclude that

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_r(\alpha)|^{\frac{(p-1)}{(r-1)}} \quad (4-28)$$

where  $p$  is the unique real positive root with magnitude greater than unity of the equation

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

Our results concerning the convergence and order of I.F. generated by direct interpolation are summarized in

THEOREM 4-3. Let

$$J = \{x \mid |x-a| \leq r\}.$$

Let  $r = s(n+1) > 1$ . Let  $f^{(r)}$  be continuous and let  $f', f^{(r)} \neq 0$  on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$ . Let

$$\frac{|f^{(r)}|}{r!} \leq v_1, \quad |f'| \geq v_2$$

for all  $x \in J$ . Suppose that

$$\frac{v_1}{v_2} 2^r \Gamma^{r-1} < 1. \quad (4-29)$$

Define a sequence  $\{x_i\}$  as follows: Let  $P_{n,s}$  be an interpolatory polynomial for  $f$  such that the first  $s-1$  derivatives of  $P_{n,s}$  are equal to the first  $s-1$  derivatives of  $f$  at the points  $x_1, x_{i-1}, \dots, x_{i-n}$ . Let  $x_{i+1}$  be a point such that  $x_{i+1}$  is real,  $P_{n,s}(x_{i+1}) = 0$ , and  $x_{i+1} \in J$ . The existence of such a point is assured by (4-29). Define  $\Phi_{n,s}$  by

$$x_{i+1} = \Phi_{n,s}(x_i; x_{i-1}, \dots, x_{i-n}).$$

Assume that  $|P'_{n,s}| \geq \mu_2$  for all  $x$  in the interval determined by  $a$  and  $x_{i+1}$ . Let  $H = v_1/\mu_2$  and suppose that  $H\Gamma^{r-1} < 1$ . Let  $e_{i-j} = x_{i-j} - a$ .

Then,  $x_i \in J$  for all  $i$ ,  $e_i \rightarrow 0$ , and

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_r(\alpha)|^{(p-1)/(r-1)}, \quad (4-30)$$

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0,$$

and where

$$A_r = \frac{f(r)}{r! f'}$$

NOTE. The point  $x_{i+1}$  may not be uniquely defined by the hypotheses of the theorem. Additional criteria must then be imposed in order to make  $\Phi_{n,s}$  a single-valued function.

The hypotheses of this theorem may seem rather strong; in the cases of greatest practical interest, however, some of the conditions are automatically satisfied. See the examples of the next section.

The form of (4-30) is strikingly similar to the form of (4-15). As before, the only parameters that appear are  $r$  and  $p$ .

If  $n = 0$ , then  $\Phi_{n,s}$  is a one-point I.F.; if  $n > 0$ , then  $\Phi_{n,s}$  is a one-point I.F. with memory.

### 4.3-1

#### 4.3 Examples

In all the examples of this section  $J$  will denote the interval defined by

$$J = \{x \mid |x - \alpha| \leq r\}.$$

We shall find that both the Newton I.F. and the secant I.F. may be derived from both direct and inverse interpolation. For a given I.F., the conditions sufficient for convergence may vary depending on the method of generation of the I.F. This should not be surprising since Theorems 4-1 and 4-3 were derived for general families of I.F. Hence the conditions sufficient for convergence, for a special case which is covered by both theorems, need not agree. The notation in this section is the same as in the previous section. The  $y_j$  are defined by (1-8).

EXAMPLE 4-1. We perform inverse interpolation with  $n = 0$  and  $s = 2$ . Hence  $r = 2$  and

$$Q_{0,2}(t) = \tilde{v}_1 + (t - y_1)\tilde{v}'_1, \quad \tilde{v}_1 = \tilde{v}(y_1).$$

Then

$$x_{i+1} = \varphi_{0,2}(x_i) = Q_{0,2}(0) = x_i - u_i, \quad u_i = f_i/f'_i.$$

## 4.3-2

This is Newton's I.F. Then

$$e_{1+1} = -\frac{1}{2} \frac{\tilde{v}''(\theta_1)}{[\tilde{v}'(\rho_1)]^2} e_1^2 = \frac{1}{2} \frac{f''(\xi_1)}{[f'(\xi_1)]^3} [f'(\eta_1)]^2 e_1^2, \quad (4-31)$$

where  $\theta_1 = f(\xi_1)$ ,  $\rho_1 = f(\eta_1)$ . Let  $x_0 \in J$  and let  $f''$  be continuous and  $f' \tilde{v}'' \neq 0$  on  $J$ . Let

$$\frac{1}{2} |\tilde{v}''| \leq \lambda_1, \quad |\tilde{v}'| \geq \lambda_2$$

for all  $x \in J$ . Let  $M = \lambda_1/\lambda_2^2$  and suppose that  $M\Gamma < 1$ . Then  $e_1 \rightarrow 0$  and

$$\frac{e_{1+1}}{e_1^2} \rightarrow y_2(\alpha). \quad (4-32)$$

No absolute value signs are required in (4-32) since the method is of integral order.

EXAMPLE 4-2. We perform inverse interpolation with  $n = 0$  and  $s = 3$ . Hence  $r = 3$  and

$$Q_{0,3}(t) = \tilde{v}_1 + (t-y_1)\tilde{v}'_1 + \frac{1}{2}(t-y_1)^2\tilde{v}''_1.$$

## 4.3-3

Then

$$x_{i+1} = \varphi_{0,3}(x_i) = \varphi_{0,3}(0) = x_i - u_i - A_2(x_i)u_i, \quad A_2 = \frac{f''}{2f'},$$

and

$$e_{i+1} = \frac{\vartheta'''(\theta_i)}{6[\vartheta'(\rho_i)]^3} e_i^3.$$

Let  $x_0 \in J$  and let  $f'''$  be continuous and  $f'\vartheta''' \neq 0$  on  $J$ .

Let  $\frac{1}{6} |\vartheta'''| \leq \lambda_1$ ,  $|\vartheta'| \geq \lambda_2$  for all  $x \in J$ . Let  $M = \lambda_1/\lambda_2^3$  and suppose that  $M\Gamma^2 < 1$ . Then  $e_i \rightarrow 0$  and

$$\frac{e_{i+1}}{e_i^3} \rightarrow v_3(a).$$

The I.F. of the form  $\varphi_{0,s}$  are of such importance that they are given the special designation  $E_s$ . They will be studied in considerable detail in Chapter 5; their properties form the basis for the study of all one-point I.F.

4.3-4

EXAMPLE 4-3. We perform inverse interpolation with  $n = 1$ ,  $s = 1$ . Therefore  $r = 2$  and

$$Q_{1,1}(t) = \tilde{y}_1 + (t - y_1) \left[ \frac{\tilde{y}_1 - \tilde{y}_{1-1}}{y_1 - y_{1-1}} \right],$$

where we have used the Newtonian form of the interpolatory polynomial as given in Appendix A. Then

$$x_{1+1} = \varphi_{1,1}(x_1; x_{1-1}) = Q_{1,1}(0) = x_1 - f_1 \left[ \frac{x_1 - x_{1-1}}{f_1 - f_{1-1}} \right].$$

This is the secant I.F. Then

$$e_{1+1} = -\frac{1}{2} \frac{\tilde{y}''(\theta_1)}{\tilde{y}'(\rho_1)\tilde{y}'(\rho_{1-1})} e_1 e_{1-1}.$$

Let  $x_0, x_1 \in J$  and suppose that the other conditions of Example 4-1 hold. Then  $e_1 \rightarrow 0$  and

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |Y_2(\alpha)|^{p-1},$$

where  $p = \frac{1}{2}(1+\sqrt{5}) \sim 1.62$ .

## 4.3-5

EXAMPLE 4-4. We perform direct interpolation with  $n = 0$ ,  $s = 2$ . Therefore  $r = 2$  and

$$P_{0,2}(t) = f_i + (t-x_i)f'_i.$$

Then

$$x_{i+1} = \Phi_{0,2}(x_i) = x_i - e_i.$$

This is Newton's I.F. again. Since  $P_{0,2}$  is a linear polynomial,  $x_{i+1}$  is always uniquely specified. Since  $P'_{0,2} = f'$ , the hypothesis of Theorem 4-3 which is concerned with the nonvanishing of  $P'_{n,s}$  is automatically satisfied by the condition on the nonvanishing of  $f'$ . Also (4-27) becomes

$$e_{i+1} = \frac{f''(\xi_i)}{2f'(x_i)} e_i^2. \quad (4-33)$$

Let  $x_0 \in J$  and let  $f''$  be continuous and  $f'f'' \neq 0$  on  $J$ . Let  $f(x_0)f''(x_0) > 0$ . From Theorem 4-2, we conclude that if  $x_0 < \alpha$ , then  $x_i$  converges to  $\alpha$  monotonically from below while if  $x_0 > \alpha$ , then  $x_i$  converges to  $\alpha$  monotonically from above. Furthermore

$$\frac{e_{i+1}}{e_i^2} \rightarrow A_2(\alpha). \quad (4-34)$$

Observe that since  $y_2(\alpha) \equiv A_2(\alpha)$ , (4-32) and (4-34) do not contradict each other.

#### 4.3-6

If we perform direct interpolation with  $n = 1$  and  $s = 1$  we will again derive the secant I.F. The I.F. of the form  $\Phi_{0,s}$  will be studied in Section 5.3 with emphasis on the study of  $\Phi_{0,3}$ . The case  $n = 2, s = 1$  will be studied in Section 10.2.

## 5.0-1

### CHAPTER 5

#### ONE-POINT ITERATION FUNCTIONS

In this chapter we shall study the theory of one-point I.F. These I.F. are of integral order. One particular basic sequence,  $E_s$ , will be studied in considerable detail in Section 5.1. By using  $E_s$  as a comparison sequence, we draw conclusions about certain properties of arbitrary one-point I.F. We consider Theorem 5-3 to be the "fundamental theorem of one-point I.F."

### 5.1-1

#### 5.1 The Basic Sequence $E_s$

We recall our definition of basic sequence. A basic sequence of I.F. is an infinite sequence of I.F. such that the  $p$ th member of the sequence is of order  $p$ . Techniques for generating basic sequences, some of which are equivalent, are due to Bodewig [5.1-1], Curry [5.1-2], Ehrmann [5.1-3], E. Schröder [5.1-4], Schwerdtfeger [5.1-5], and Whittaker [5.1-6], among others. See also Durand [5.1-7], Householder [5.1-8, Chap. 3], Korganoff [5.1-9, Chap. 3], Ludwig [5.1-10], Ostrowski [5.1-11, Appendix J], and Zajta [5.1-12]. From Theorem 2-6 we know that two I.F. of order  $p$  can differ only by terms proportional to  $u^p$ . Hence if the properties of one I.F. of order  $p$  are known, many of the properties of arbitrary I.F. of order  $p$  may be deduced. If the properties of a basic sequence are known, then many of the properties of arbitrary I.F. of any order may be deduced. In Section 5.11 we shall study a basic sequence whose simplicity of structure makes it useful as a comparison sequence.

5.11 The formula for  $E_s$ . The I.F.  $\varphi_{n,s}$  were defined and studied in Section 4.22. If  $n = 0$ , these I.F. are without memory. Because of their importance, the  $\varphi_{0,s}$  are given the special designation  $E_s$ . The conditions for the convergence of a sequence generated by  $E_s$  were derived in Section 4.22; we assume that these conditions hold. In contrast with the careful analysis which is required in the general case, we shall find that the proof that  $E_s$  is of order  $s$  is almost trivial.

In order that the material on  $E_s$  be self-contained, we start anew. Let  $f'$  be nonzero in a neighborhood of  $a$  and let  $f^{(s)}$  be continuous in this neighborhood. Then  $f$  has an inverse  $\tilde{g}$ , and  $\tilde{g}^{(s)}$  is continuous in a neighborhood of zero. Let  $Q_{0,s}$  be the polynomial whose first  $s - 1$  derivatives agree with  $\tilde{g}$  at the point  $y = f(x)$ . Then

$$\tilde{g}(t) = Q_{0,s}(t) + \frac{\tilde{g}^{(s)}[\theta(t)]}{s!} (t-y)^s,$$

and

$$Q_{0,s}(t) = \sum_{j=0}^{s-1} \frac{\tilde{g}^{(j)}}{j!} (t-y)^j,$$

where  $\theta(t)$  lies in the interval determined by  $y$  and  $t$ . Define

$$E_s = Q_{0,s}(0).$$

5.1-3

Hence

$$E_s = \sum_{j=0}^{s-1} \frac{(-1)^j}{j!} \mathfrak{f}^{(j)} f^j \quad (5-1)$$

or

$$E_s = x - \sum_{j=1}^{s-1} \frac{(-1)^{j-1}}{j!} \mathfrak{f}^{(j)} f^j. \quad (5-2)$$

Furthermore,

$$\alpha = E_s + \frac{(-1)^s}{s!} \mathfrak{f}^{(s)}(\theta) f^s, \quad (5-3)$$

where  $\theta \equiv \theta(0)$ .

In (5-2),  $E_s$  is expressed as a power series in  $f^j$ . For some applications it is more useful to express  $E_s$  as a power series in  $u$  where  $u = f/f'$ . Hence we write

$$E_s = x - \sum_{j=1}^{s-1} \frac{(-1)^{j-1} \mathfrak{f}^{(j)}}{j! [\mathfrak{f}']^j} u^j.$$

In practice  $\mathfrak{f}$  is not known and we must express  $E_s$  in terms of  $f$  and its derivatives. With the definition

$$Y_j(x) = \left. \frac{(-1)^{j-1} \mathfrak{f}^{(j)}(y)}{j! [\mathfrak{f}'(y)]^j} \right|_{y=f(x)},$$

5.1-4

we may write

$$E_s(x) = x - \sum_{j=1}^{s-1} Y_j u^j \quad (5-4)$$

and

$$\alpha = E_s + \frac{(-1)^s \mathfrak{F}^{(s)}(\theta)}{s! [\mathfrak{F}']^s} u^s. \quad (5-5)$$

Thus

$$\alpha = E_s + \underline{0}[u^s]. \quad (5-6)$$

We may write formally that

$$\alpha = x - \sum_{j=1}^{\infty} Y_j u^j. \quad (5-7)$$

The structure of the  $Y_j$  will be investigated in Section 5.13.

Assume that  $\mathfrak{F}^{(s)}$  does not vanish in an interval about zero. From (5-5),

$$\frac{E_s - \alpha}{(x-\alpha)^s} = \frac{(-1)^{s-1} \mathfrak{F}^{(s)}(\theta)}{s! [\mathfrak{F}']^s} \left(\frac{u}{x-\alpha}\right)^s.$$

Since

$$\frac{u}{x-\alpha} \rightarrow 1,$$

5.1-5

we conclude that

$$\frac{E_s - \alpha}{(x - \alpha)^s} \rightarrow Y_s(\alpha). \quad (5-8)$$

Let  $x_1 = x$ ,  $x_{1+1} = E_s$ ,  $e_1 = x_1 - \alpha$ . Then (5-8) may be written as

$$\frac{e_{1+1}}{e_1^s} \rightarrow Y_s(\alpha).$$

Hence  $E_s$  is of order  $s$  and has  $Y_s(\alpha)$  as its asymptotic error constant. Since the informational usage of  $E_s$  is  $s$ , its informational efficiency is unity. Hence  $E_s$  is an optimal basic sequence. (These terms are defined in Section 1.24.)

It is easy to see that

$$\frac{y_{1+1}}{y_1^s} \rightarrow (-1)^{s-1} a_s(0),$$

where

$$y_1 = f(x_1), \quad a_s(y) = \frac{\vartheta^{(s)}(y)}{s! \vartheta'(y)}.$$

5.1-6

Bodewig [5.1-13] attributes  $E_s$  to Euler [5.1-14].

In the Russian literature, these formulas are credited to Chebyshev who wrote a student paper entitled "Calcul des racines d'une équation" for which he was awarded a silver medal. This paper which was written in 1837 or 1838 has not been available to me.

We show that the formula of E. Schröder [5.1-15] is equivalent to  $E_s$ . Observe that

$$\frac{dy}{dx} = \frac{1}{f'(x)} \frac{d}{dx}, \quad g'(y) = \frac{1}{f'(x)}.$$

Then from (5-2),

$$E_s = x + \sum_{j=1}^{s-1} \frac{(-1)^j}{j!} f^j(x) \left[ \frac{1}{f'(x)} \frac{d}{dx} \right]^{j-1} \frac{1}{f'(x)},$$

which is Schröder's formula. Compare with the formula for a Bürmann series given by Hildebrand [5.1-16, p. 25].

## 5.1-7

5.12 An example. In certain cases one can show that the formal infinite series for  $\alpha$  converges to  $\alpha$ . A series solution of a quadratic equation is studied by E. Schröder [5.1-17].

EXAMPLE 5-1. Consider  $f(x) = x^n - A$ , with  $n$  an integer. If  $n \geq 2$  this leads to a formula for  $n$ th roots, while if  $n = -1$  this leads to a formula for the reciprocal of  $A$ . If  $f(x) = x^n - A$ , then

$$\vartheta(y) = (A+y)^{1/n}$$

and

$$\vartheta^{(j)}(y) = x^{1-jn} \prod_{k=0}^{j-1} \left(\frac{1}{n} - k\right).$$

Then

$$E_s = x + x \sum_{j=1}^{s-1} \frac{1}{j!} \left(\frac{A-x^n}{nx^n}\right)^j \prod_{k=0}^{j-1} (1-kn). \quad (5-9)$$

In particular,

$$E_2 = \frac{x}{n} \left[ n - 1 + \frac{A}{x^n} \right].$$

5.1-8

If  $n = 2$ , this is Heron's method for the approximation of square roots. The formula (5-9) was derived by Traub [5.1-18] using the binomial expansion. If  $n = -1$ ,

$$E_s = x \sum_{j=0}^{s-1} (1-Ax)^j, \quad (5-10)$$

and

$$\alpha = x \sum_{j=0}^{\infty} (1-Ax)^j.$$

This geometric series converges to  $1/A$  if  $|1-Ax| < 1$ .

Rabinowitz [5.1-19] points out that (5-10) may be used to carry out multiple-precision division.

See Durand [5.1-20, pp. 9-14] for the approximation of  $\ln A$ ,  $\sin^{-1}A$ , and  $\tan^{-1}A$ .

5.13 The structure of  $E_s$ . We defined  $E_s$  as

$$E_s = x - \sum_{j=1}^{s-1} Y_j u^j, \quad Y_j(x) = \frac{(-1)^{j-1} \mathfrak{g}(j)(y)}{j! [\mathfrak{g}'(y)]^j} \Big|_{y=f(x)}.$$

It is easy to show that  $Y_j$  satisfies the difference-differential equation,

$$jY_j(x) - 2(j-1)A_2(x)Y_{j-1}(x) + Y'_{j-1}(x) = 0, \quad Y_1(x) = 1, \quad A_2(x) = \frac{f''(x)}{2f'(x)}. \quad (5-11)$$

The first few  $Y_j$  may be calculated directly from this equation.

An explicit formula for the  $Y_j$  may be derived from the formula for the derivative of the inverse function derived in Appendix B. We have

$$\mathfrak{g}(j) = [f']^{-j} \sum (-1)^r (j+r-1)! \prod_{i=2}^j \frac{(A_i)^{\beta_i}}{\beta_i!}, \quad (5-12)$$

with the sum taken over all nonnegative integers  $\beta_i$  such that

$$\sum_{i=2}^j (i-1)\beta_i = j - 1, \quad (5-13)$$

and where  $r = \sum_{i=2}^j \beta_i$ . For  $j = 1$ ,  $\beta_i = 0$  for all  $i$ . From the definition of  $Y_j$  and (5-12) we have

THEOREM 5-1. Let

$$Y_j(x) = \frac{(-1)^{j-1} \mathfrak{F}'(j)(y)}{j! [\mathfrak{F}'(y)]^j} \Big|_{y=f(x)}, \quad A_j(x) = \frac{f^{(j)}(x)}{j! f'(x)},$$

where  $f$  and  $\mathfrak{F}$  are inverse functions. Then

$$Y_j = \frac{(-1)^{j-1}}{j!} \sum (-1)^r (j+r-1)! \prod_{i=2}^j \frac{(A_i)^{\beta_i}}{\beta_i!},$$

with the sum taken over all nonnegative integers  $\beta_i$  such that

$$\sum_{i=2}^j (i-1)\beta_i = j - 1,$$

and where  $r = \sum_{i=2}^j \beta_i$ .

The first few  $Y_j$  are given in Table 5-1. Observe that  $Y_j$  is independent of  $f$ ; it depends only on the derivatives of  $f$ . We conclude that

$$E_s = x - \sum_{j=1}^{s-1} \frac{(-1)^{j-1}}{j!} u^j \sum (-1)^r (j+r-1)! \prod_{i=2}^j \frac{(A_i)^{\beta_i}}{\beta_i!}, \quad (5-14)$$

where the inner sum is taken over all nonnegative integers  $\beta_i$  such that  $\sum_{i=2}^j (i-1)\beta_i = j - 1$  and where  $r = \sum_{i=2}^j \beta_i$ .

5.1-11

TABLE 5-1. FORMULAS FOR  $Y_j$

$$Y_1 = 1$$

$$Y_2 = A_2$$

$$Y_3 = 2A_2^2 - A_3$$

$$Y_4 = 5A_2^3 - 5A_2A_3 + A_4$$

$$Y_5 = 14A_2^4 - 21A_2^2A_3 + 6A_2A_4 + 3A_3^2 - A_5$$

## 5.1-12

By replacing the upper limit of the first sum by  $\infty$ , we obtain a formal infinite series formula for  $\alpha$ .

The first few  $E_s$  are given by:

$$E_2 = x - u,$$

$$E_3 = E_2 - A_2 u^2,$$

$$E_4 = E_3 - (2A_2^2 - A_3)u^3, \quad (5-15)$$

$$E_5 = E_4 - (5A_2^3 - 5A_2 A_3 + A_4)u^4.$$

The following corollaries follow easily from Theorem 5-1.

COROLLARY a.  $y_j$  is a polynomial in  $A_2, A_3, \dots, A_j$ .

COROLLARY b.  $A_j$  is the same polynomial in  $y_2, y_3, \dots, y_j$  that  $y_j$  is in  $A_2, A_3, \dots, A_j$ .

COROLLARY c. The sum of the coefficients in this polynomial is unity.

PROOF. The polynomial is an identity in  $x$ . Let  $f(x) = (1-x)^{-1}$ . At  $x = 0$ ,  $A_j = 1$  for all  $j$ . The inverse to  $f(x)$  is  $\tilde{f}(y) = 1 - y^{-1}$  and when  $x = 0$ ,  $y = 1$ . A short calculation shows that at  $y = 1$ ,  $\tilde{Y}_j = 1$  for all  $j$ .

COROLLARY d.  $\tilde{Y}_j$  depends on  $A_j$  only as  $(-1)^j A_j$ .

For some applications it is convenient to work with

$$z_j = \tilde{Y}_j u^j. \quad (5-16)$$

Then

$$E_s = x - \sum_{j=1}^{s-1} z_j. \quad (5-17)$$

We have

COROLLARY e. The form of  $z_j$  is given by

$$z_j = \sum c_{\beta_0, \beta_1, \dots, \beta_j} (f)^{\beta_0} (f')^{\beta_1} \dots (f^{(j)})^{\beta_j}$$

where  $\sum_{i=0}^j \beta_i = 0$ ,  $\sum_{i=1}^j i \beta_i = -1$ . Thus  $z_j$  is a polynomial, homogeneous of degree zero and isobaric of weight -1.

It is easy to show that  $Z_j$  satisfies the difference-differential equation,

$$jZ_j(x) - (j-1)Z_{j-1}(x) + u(x)Z'_{j-1}(x) = 0, \quad Z_1(x) = u(x). \quad (5-18)$$

In terms of the forward difference operator  $\Delta$ , this equation may be written as

$$\Delta[jZ_j(x)] + Z_1(x)Z'_j(x) = 0.$$

In Lemmas 5-1 and 5-2 and in Theorem 5-2, we shall make use of the fact that  $E_s$  is of order  $s$ . The following lemma enables us to calculate the asymptotic error constant of an arbitrary I.F. of order  $p$ .

LEMMA 5-1. Let  $\phi$  be an I.F. of order  $p$  and let  $C$  be its asymptotic error constant. Let

$$G(x) = \frac{\phi(x) - E_p(x)}{(x-a)^p}, \quad x \neq a.$$

Then

$$C = Y_p(a) + \lim_{x \rightarrow a} G(x).$$

5.1-15

PROOF. Take  $\varphi_2 \equiv \varphi$  and  $\varphi_1 \equiv E_p$  in Theorem 2-8 and observe that the asymptotic error constant of  $E_p$  is  $Y_p(\alpha)$ .

A more useful result is given in

LEMMA 5-2. Let  $\varphi$  be an I.F. of order  $p$  and let  $C$  be its asymptotic error constant. Let

$$H(x) = \frac{\varphi(x) - E_p(x)}{u^p(x)}, \quad x \neq \alpha.$$

Then

$$C = Y_p(\alpha) + \lim_{x \rightarrow \alpha} H(x).$$

PROOF. From the previous lemma,

$$\lim_{x \rightarrow \alpha} H(x) = \lim_{x \rightarrow \alpha} G(x) \left[ \frac{x-\alpha}{u(x)} \right]^p = C - Y_p(\alpha),$$

since

$$\lim_{x \rightarrow \alpha} \frac{u(x)}{x-\alpha} = 1.$$

## 5.1-16

EXAMPLE 5-2. Let

$$\varphi = x - \frac{u}{1-A_2 u}.$$

This is Halley's I.F. which will be derived in Section 5.21.

Since  $E_3 = x - u[1+A_2 u]$ , it is easy to show that  
 $\lim H(x) = -A_2^2(\alpha)$ . Then

$$c = Y_3(\alpha) - A_2^2(\alpha) = A_2^2(\alpha) - A_3(\alpha).$$

The order and asymptotic error constant of an I.F. with integer-valued order may be calculated using the results of Theorem 2-2. It is awkward however to apply this theorem for all but the simplest I.F. The following theorem permits the calculation of the order and asymptotic error constant of an I.F. by comparing it with  $E_{p+1}$ . This procedure turns out to be particularly useful in the development of Chapter 9.

THEOREM 5-2.  $\varphi$  is of order  $p$  if and only if

$$\lim_{x \rightarrow \alpha} \left[ \frac{\varphi(x) - E_{p+1}(x)}{u^p(x)} \right]$$

5.1-17

exists and is nonzero. Furthermore

$$C = \lim_{x \rightarrow a} \left[ \frac{\varphi(x) - E_{p+1}(x)}{u^p(x)} \right],$$

where  $C$  is the asymptotic error constant of  $\varphi$ .

PROOF. From (5-6),

$$\alpha = E_{p+1}(x) + O[u^{p+1}(x)].$$

Therefore,

$$\begin{aligned} \frac{\varphi(x) - \alpha}{(x-a)^p} &= \frac{\varphi(x) - E_{p+1}(x) + O[u^{p+1}(x)]}{u^p(x)} \left( \frac{u(x)}{x-a} \right)^p \\ &= \frac{\varphi(x) - E_{p+1}(x)}{u^p(x)} \left( \frac{u(x)}{x-a} \right)^p + \left( \frac{u(x)}{x-a} \right)^p O[u(x)]. \end{aligned}$$

Since

$$\lim_{x \rightarrow a} \frac{u(x)}{x-a} = 1,$$

we conclude that

$$\lim_{x \rightarrow a} \left[ \frac{\varphi(x) - \alpha}{(x-a)^p} \right] = \frac{\varphi(x) - E_{p+1}(x)}{u^p(x)},$$

which completes the proof.

5.1-18

Observe that Lemmas 5-1 and 5-2 each involve two I.F. of the same order; Theorem 5-2 involves two I.F. whose orders differ by unity.

The next lemma is used in Section 5.51.

LEMMA 5-3.

$$E_{s+1}(x) = E_s(x) - \frac{u(x)}{s} E'_s(x).$$

PROOF. From (5-18),

$$\sum_{j=2}^s j z_j(x) - \sum_{j=2}^s (j-1) z_{j-1}(x) + u(x) \sum_{j=2}^s z'_{j-1}(x) = 0.$$

This telescopes to

$$s z_s(x) - z_1(x) + u(x) \sum_{j=2}^s z'_{j-1}(x) = 0.$$

Since  $z_1(x) = u(x)$ ,

$$s z_s(x) = u(x) \left[ 1 - \sum_{j=2}^s z'_{j-1}(x) \right] = u(x) E'_s(x).$$

5.1-19

The fact that

$$E_{s+1}(x) - E_s(x) = -Z_s(x)$$

completes the proof.

EXAMPLE 5-3. Let  $s = 2$ . Then

$$E_3 = E_2 - \frac{1}{2}uE'_2 = x - u - \frac{1}{2}u(1-u'),$$

and

$$E_3 = x - u - \frac{1}{2}u(2A_2u) = x - u - A_2u^2.$$

## 5.2-1

### 5.2 Rational Approximations to $E_s$ .

It is common knowledge that rational functions are often preferable to polynomials for the approximation of functions. The rational function approximations to a function may be arranged into a two-dimensional array indexed by the degrees of the polynomials of the numerator and denominator. Such an array is called a Padé table. See Koberliantz [5.2-1], Kopal [5.2-2, Chap. IX], and Wall [5.2-3, Chap. 20].

Since  $E_s$  is a polynomial in  $u$  with coefficients depending on the derivatives of  $f$ , we extend the usual procedure and form a Padé table of I.F. The rational approximations to  $E_s$  are constructed so as to be order-preserving. For each  $s$  we obtain  $s - 1$  optimal I.F. of order  $s$ . In particular we obtain the often rediscovered Halley's I.F. Most of the material of this section first appeared in Traub [5.2-4].

5.21 Iteration functions generated by rational approximation to  $E_s$ . It is convenient to define a polynomial  $Y(u, s-1)$  by

$$Y(u, s-1) = \sum_{j=1}^{s-1} Y_j(x) u^j(x).$$

Then  $E_s = x - Y(u, s-1)$ . We will study rational approximations to  $Y(u, s-1)$  which are order-preserving. Define

$$\psi_{a,b} = x - \frac{R(u, s, a)}{Q(u, s, b)}, \quad a + b = s - 1, \quad a > 0,$$

where

$$R(u, s, a) = \sum_{j=1}^a R_{s,j}(x) u^j(x), \quad Q(u, s, b) = \sum_{j=0}^b Q_{s,j}(x) u^j(x), \quad Q_{s,0}(x) = 1.$$

Note that the "constant term" is absent in  $R(u, s, a)$  and present in  $Q(u, s, b)$ . This guarantees that  $\psi_{a,b}(a) = a$ . The  $a + b$  parameters

$$R_{s,j}(x), \quad j = 1, 2, \dots, a; \quad Q_{s,j}(x), \quad j = 1, 2, \dots, b,$$

are chosen so that

$$R(u, s, a) - Y(u, s-1) Q(u, s, b) = \underline{0}[u^s(x)]. \quad (5-19)$$

5.2-3

Then

$$E_s - \psi_{a,b} = O[u^s],$$

and by Theorem 2-7,  $\psi_{a,b}$  is of order  $s$ . Equivalently,  $\psi_{a,b}$  is of order  $a+b+1$ . The conditions imposed by (5-19) are satisfied if the  $R_{s,j}, Q_{s,j}$  are chosen so that

$$R_{s,\ell}(x)\omega_{\ell,a} - \sum_{j=0}^k Y_{\ell-j}(x)Q_{s,j}(x) = 0, \quad \ell = 1, 2, \dots, a+b,$$

where  $\omega_{\ell,a} = 1$ , for  $\ell \leq a$ ,  $\omega_{\ell,a} = 0$ , for  $\ell > a$ , and  $k = \min(\ell-1, b)$ . For parameters thus chosen,

$$c_{a,b} = \lim_{x \rightarrow a} \frac{\psi_{a,b}-a}{(x-a)^s} = Y_s(a) + \sum_{j=1}^b Y_{s-j}(a)Q_{s,j}(a).$$

Since the  $Y_j(x)$ ,  $1 \leq j \leq s-1$ , depend only on  $f^{(j)}$ ,  $1 \leq j \leq s-1$ , and since the  $R_{s,j}(x), Q_{s,j}(x)$  depend only on these  $Y_j(x)$ , we conclude that the  $\psi_{a,b}$  are all optimal I.F. A number of formulas of type  $\psi_{a,b}$ , together with their asymptotic error constants, may be found in Table 5-2. Note that  $\psi_{a,0} \equiv E_{s-1}$ .

For  $s$  fixed, which of the  $s-1$  I.F. generated by this process is to be preferred? There are indications (Frame [5.2-5] and Kopal [5.2-6]) that the I.F. which lie near the diagonal of the Padé table are best.

TABLE 5-2. FORMULAS FOR  $\psi_{a,b}$  AND  $C_{a,b}$

Formulas for $\psi_{a,b}$	Asymptotic Error Constants $C_{a,b} = \lim_{x \rightarrow a} \frac{\psi_{a,b} - a}{(x-a)^s}$
$s = 2: \psi_{1,0} = x - u$ , Newton	$C_{1,0} = Y_2(\alpha)$
$s = 3: \psi_{2,0} = x - u[1 + Y_2 u]$	$C_{2,0} = Y_3(\alpha)$
$\psi_{1,1} = x - \frac{u}{1 - Y_2 u}$ , Halley	$C_{1,1} = Y_3(\alpha) - Y_2^2(\alpha)$
$s = 4: \psi_{3,0} = x - u[1 + Y_2 u + Y_3 u^2]$	$C_{3,0} = Y_4(\alpha)$
$\psi_{2,1} = x - u \frac{[Y_2 + (Y_2^2 - Y_3)u]}{Y_2 - Y_3 u}$	$C_{2,1} = Y_4(\alpha) - \frac{Y_3^2(\alpha)}{Y_2(\alpha)}$
$\psi_{1,2} = x - \frac{u}{1 - Y_2 u + (Y_2^2 - Y_3)u^2}$	$C_{1,2} = Y_4(\alpha) - 2Y_2(\alpha)Y_3(\alpha) + Y_2^3(\alpha)$

5.2-5

EXAMPLE 5-4. If two I.F. have the same order, then a measure of which I.F. will converge faster is given by the relative magnitudes of the asymptotic error constants. Let  $f(x) = x^n - A$ ,  $\alpha = A^{1/n}$ . Define  $c_{a,b}$  by

$$\frac{\psi_{a,b}^{-\alpha}}{(x-\alpha)^s} \rightarrow c_{a,b}.$$

Then

$$s = 3: \quad c_{2,0} = \frac{(n-1)(2n-1)}{6\alpha^2},$$

$$c_{1,1} = \frac{n^2-1}{12\alpha^2},$$

$$s = 4: \quad c_{3,0} = \frac{(n-1)(2n-1)(3n-1)}{24\alpha^3},$$

$$c_{2,1} = \frac{(n^2-1)(2n-1)}{72\alpha^3},$$

$$c_{1,2} = \frac{n(n^2-1)}{24\alpha^3},$$

and

$$\lim_{n \rightarrow \infty} \frac{c_{2,0}}{c_{1,1}} = 4, \quad \lim_{n \rightarrow \infty} \frac{c_{3,0}}{c_{2,1}} = 9,$$

$$\lim_{n \rightarrow \infty} \frac{c_{1,2}}{c_{2,1}} = \frac{3}{2}.$$

## 5.2-6

The  $\psi_{a,b}$  are not the only optimal I.F. of rational form. Thus Kiss [5.2-7] suggests a fourth order formula which in our notation may be written as

$$\varphi = x - \frac{u(1-A_2 u)}{1-2A_2 u+A_3 u^2}. \quad (5-20)$$

Snyder [5.2-8] derives  $\psi_{1,1}$  (Halley's I.F.) by a "method of replacement," and a fourth order formula by a "method of double replacement." When Snyder's fourth order formula is translated into our notation, it is seen to be identical with (5-20) which Kiss derives by entirely different methods. Hildebrand [5.2-9, Sect. 9.12] studies I.F. generated by Thiele's continued-fraction expansions.

5.22 The formulas of Halley and Lambert. Perhaps the most frequently rediscovered I.F. in the literature is

$$\psi_{1,1} = x - \frac{u}{1-A_2 u}.$$

It has recently been rediscovered by Frame [5.2-10], Richmond [5.2-11], and H. Wall [5.2-12]. It was derived by E. Schröder [5.2-13, p. 352] in 1870. Salehov [5.2-14] investigates the convergence of the method which he calls the method of tangent hyperbolas. Zaguskin [5.2-15, p. 113] points out that  $\psi_{1,1}$  may be derived by the method of Domoryad. Bateman [5.2-16] points out that the method is due to Halley [5.2-17] (1694).

If  $f = x^n - A$ , Halley's I.F. becomes

$$\varphi = \frac{x[(n-1)x^n + (n+1)A]}{(n+1)x^n + (n-1)A}, \quad (5-21)$$

and

$$\frac{\varphi-\alpha}{(x-\alpha)^3} \rightarrow \frac{n^2-1}{12\alpha^2}, \quad \alpha = A^{1/n}.$$

In the current literature, (5-21) is often ascribed to Bailey [5.2-18] (1941). Equation (5-21) was derived by Uspensky [5.2-19] (1927). R. Newton [5.2-20] points out that Davies and Peck [5.2-21] (1876) call it Hutton's method but they give no reference. Dunkel [5.2-22] notes that the formula was known to Barlow [5.2-23] (1814). Kiss [5.2-24] and Müller [5.2-25] ascribe the method to Lambert [5.2-26] (1770).

### 5.3-1

#### 5.3 A Basic Sequence of Iteration Functions Generated by Direct Interpolation.

In Section 5.1 we studied the basic sequence  $E_s \equiv \phi_{0,s}$  generated by inverse interpolation at one point. We turn to the basic sequence  $\Phi_{0,s}$  generated by direct interpolation at one point. These latter I.F. have the drawback that for  $s > 2$ , a polynomial of degree  $s - 1$  must be solved at each iteration. They have the virtue of being exact for all polynomials of degree less than or equal to  $s - 1$ . In Section 5.33, we investigate a technique which reduces the degree of the polynomial to be solved but which preserves the order of the I.F. generated.

## 5.3-2

5.31 The basic sequence  $\Phi_{o,s}$ . The I.F.  $\Phi_{n,s}$  was defined and studied in Section 4.23. If  $n = 0$ , these I.F. are without memory. The conditions for the convergence of a sequence generated by  $\Phi_{o,s}$  were derived in Section 4.23; we assume that these conditions hold. In contrast with the careful analysis which is required in the general case, we shall find that the proof that  $\Phi_{o,s}$  is of order  $s$  is almost trivial.

In order that the material on  $\Phi_{o,s}$  be self-contained, we start anew. Let  $P_{o,s}$  be the polynomial whose first  $s - 1$  derivatives agree with  $f$  at the point  $x_1$ . Then

$$f(t) = P_{o,s}(t) + \frac{f^{(s)}[\xi_1(t)]}{s!} (t-x_1)^s, \quad (5-22)$$

and

$$P_{o,s}(t) = \sum_{j=0}^{s-1} \frac{f^{(j)}}{j!} (t-x_1)^j,$$

where  $\xi_1(t)$  lies in the interval determined by  $x_1$  and  $t$ .

Define  $x_{i+1}$  by

$$P_{o,s}(x_{i+1}) = 0. \quad (5-23)$$

Let a real root of (5-23) be chosen by some criteria. Let the function that maps  $x_i$  into  $x_{i+1}$  be labeled  $\Phi_{o,s}$ . Thus

$$x_{i+1} = \Phi_{o,s}(x_i).$$

5.3-3

The order and asymptotic error constant of  $\Phi_{o,s}$  are now easily obtained. Set  $t = \alpha$  in (5-22). Then

$$0 = P_{o,s}(\alpha) + \frac{(-1)^s}{s!} f^{(s)}(\xi_1) e_i^s,$$

where  $e_i = x_i - \alpha$  and where  $\xi_1 \equiv \xi_1(\alpha)$ . Since

$$P_{o,s}(\alpha) = - P'_{o,s}(\eta_{i+1}) e_{i+1},$$

where  $\eta_{i+1}$  lies in the interval determined by  $x_{i+1}$  and  $\alpha$ , we conclude that

$$P'_{o,s}(\eta_{i+1}) e_{i+1} = \frac{f^{(s)}(\xi_1)}{s!} e_i^s.$$

Let  $P'_{o,s}$  be nonzero in the interval determined by  $x_{i+1}$  and  $\alpha$  and let  $f^{(s)}$  be nonzero in the interval of iteration.

Observe that  $P'_{o,s} \rightarrow f'(\alpha)$ . Then

$$\frac{e_{i+1}}{e_i^s} \rightarrow (-1)^s A_s(\alpha), \quad A_s = \frac{f^{(s)}}{s! f'}. \quad (5-24)$$

Since  $s$  is an integer, it is not necessary to take absolute values in (5-24). We conclude that  $\Phi_{o,s}$  is a basic sequence.

## 5.3-4

5.32 The iteration function  $\Phi_{0,3}$ . The I.F.  $\Phi_{0,2}$  is Newton's I.F. We turn to  $\Phi_{0,3}$ . This I.F. was already studied by Cauchy [5.3-1]. See also Hitotumato [5.3-2]. We must solve

$$0 = f(x) + f'(x)(t-x) + \frac{1}{2}f''(x)(t-x)^2 = 0. \quad (5-25)$$

Then

$$\Phi_{0,3} = x - \frac{f'}{f''} \pm \frac{|f'|}{f''} (1-4A_2u)^{\frac{1}{2}}, \quad u = \frac{f}{f'}, \quad A_2 = \frac{f''}{2f'}.$$

$\alpha$  will be a fixed point of  $\Phi_{0,3}$  if and only if we take the + sign if  $f' > 0$  and the - sign if  $f' < 0$ . If this choice of sign is made, then  $\Phi_{0,3}$  differs only by terms of  $\underline{o}(u^3)$  from  $E_3$ . Thus

$$\Phi_{0,3} = x - \frac{f'}{f''} + \frac{f'}{f''} (1-4A_2u)^{\frac{1}{2}}. \quad (5-26)$$

For  $x$  close to  $\alpha$ , severe cancellation is bound to occur in (5-26). This may easily be avoided by observing that

$$\frac{-b + (b^2-4ac)^{\frac{1}{2}}}{2a} \equiv \frac{-2c}{b + (b^2-4ac)^{\frac{1}{2}}},$$

and hence taking

$$\Phi_{0,3} = x - \frac{2u}{1 + (1-4A_2u)^{\frac{1}{2}}}. \quad (5-27)$$

### 5.3-5

This last form is clearly best for computational purposes. Although generations of schoolboys have been drilled to rationalize the denominator, it is preferable, in this case, to irrationalize the denominator.

The generalized Fourier conditions of Theorem 4-2 assume a particularly simple form for  $\Phi_{o,3}$ . Let

$$f'f''' < 0$$

over the interval of iteration. Then if  $x_o > \alpha$ , the  $x_i$  converge to  $\alpha$  monotonically from above; if  $x_o < \alpha$ , the  $x_i$  converge to  $\alpha$  monotonically from below. Observe that in the case of Newton's method, monotone convergence is only possible from one side. This is because we demand that  $f(x_o)f''(x_o) > 0$  for Newton's I.F. and  $f$  must change sign as  $x$  goes through  $\alpha$ .

5.3-6

5.33 Reduction of degree. We observed that the use of  $\Phi_{0,s}$  requires the solution of an  $(s-1)$  degree polynomial at each iteration. We can effect certain degree-reducing changes which change the asymptotic error constant but which do not affect the order.

Consider the following change. Replace one of the  $t - x$  factors in  $(t-x)^{s-1}$  by  $E_2 - x = -u$ . ( $E_2 = x - u$  is Newton's I.F.) Define

$$R_s(t) = \sum_{j=0}^{s-2} \frac{f^{(j)}(x)}{j!} (t-x)^j + \frac{f^{(s-1)}(x)}{(s-1)!} (t-x)^{s-2}(E_2 - x). \quad (5-28)$$

We shall show that although  $R_s$  is a polynomial of degree  $s - 2$  in  $t - x$ , it still leads to an I.F. of order  $s$ . Observe that

$$P_{0,s}(t) - R_s(t) = \frac{f^{(s-1)}(x)}{(s-1)!} (t-x)^{s-2}(t-E_2)$$

and

$$P_{0,s}(\alpha) = R_s(\alpha) + \frac{f^{(s-1)}(x)}{(s-1)!} (\alpha-x)^{s-2}(\alpha-E_2).$$

Since

$$E_2 - \alpha = V(x)(x-\alpha)^2, \quad V(\alpha) = A_2(\alpha),$$

and

$$f(t) = P_{0,s}(t) + \frac{f^{(s)}[\xi(t)]}{s!} (t-x)^s,$$

5.3-7

we conclude that

$$0 = f(\alpha) = R_s(\alpha) + (-1)^s (x-\alpha)^s \left[ \frac{f^{(s)}(\xi)}{s!} - \frac{f^{(s-1)}(x)}{(s-1)!} V(x) \right],$$

where  $\xi$  lies in the interval determined by  $x$  and  $\alpha$ . Define  $x_{i+1}$  by

$$R_s(x_{i+1}) = 0.$$

Then

$$R_s(\alpha) = - R'_s(\eta_{i+1})(x_{i+1}-\alpha),$$

where  $\eta_{i+1}$  lies in the interval determined by  $x_{i+1}$  and  $\alpha$ . Assume that  $R'_s(\eta_{i+1})$  does not vanish. Then if  $e_i \rightarrow 0$ ,

$$\frac{e_{i+1}}{e_i^s} \rightarrow (-1)^s [A_s(\alpha) - A_{s-1}(\alpha)A_2(\alpha)]. \quad (5-29)$$

EXAMPLE 5-5. Let  $s = 3$ . Then we must solve  
 $R_3(t) = 0$ . From (5-28),

$$0 = f(x) + (t-x)[f'(x) + \frac{1}{2}f''(x)(-u)]$$

or

$$\varphi = x - \frac{u}{1-A_2u}.$$

## 5.3-8

This is Halley's I.F. which we have now generated by linearizing a second degree equation. From (5-29),

$$\frac{e_{i+1}}{e_i^3} = - \left[ A_3(\alpha) - A_2^2(\alpha) \right] \equiv Y_3(\alpha) - Y_2^2(\alpha),$$

as we found in Section 5.21.

**EXAMPLE 5-6.** Let  $s = 4$ . Then

$$0 = f(x) + f'(x)(t-x) + (t-x)^2 \left[ \frac{1}{2} f''(x) - \frac{1}{6} f'''(x)u \right],$$

or

$$0 = u + (t-x) + (t-x)^2 [A_2 - A_3 u]$$

and

$$\varphi = x - \frac{2u}{1 + [1 - 4u(A_2 - A_3 u)]^{\frac{1}{2}}}.$$

Then

$$\frac{e_{i+1}}{e_i^4} \rightarrow [A_4(\alpha) - A_2(\alpha)A_3(\alpha)]. \quad (5-30)$$

One could also arrive at (5-30) by expanding  $\varphi$  into a power series in  $u$  and applying Lemma 5-2.

## 5.3-9

There are numerous other ways by which the degree of  $P_{0,s}(t)$  can be lowered without changing the order. If one of the  $t - x$  terms in  $(t-x)^{s-1}$  were replaced by  $E_3 - x$ , then the degree would be lowered by one but neither the order nor the asymptotic error constant would be changed. We shall content ourselves with two more examples.

EXAMPLE 5-7. In

$$P_{0,3}(t) = f(x) + (t-x)f'(x) + \frac{1}{2}(t-x)^2f''(x),$$

replace  $(t-x)^2$  by  $(E_2 - x)^2 = u^2$ . Then

$$0 = f(x) + (t-x)f'(x) + \frac{1}{2}u^2f''(x)$$

or

$$\varphi = x - u - u^2 A_2,$$

which is just  $E_3$ .

EXAMPLE 5-8. In

$$P_{0,4}(t) = f(x) + (t-x)f'(x) + \frac{1}{2}(t-x)^2f''(x) + \frac{1}{6}(t-x)^3f'''(x),$$

## 5.3-10

replace one of the  $t - x$  in the quadratic term by  $E_3 - x$  and  
 replace  $(t-x)^2$  in the cubic term by  $(E_2 - x)^2 = u^2$ . Then

$$0 = f(x) + (t-x) \left[ f'(x) - \frac{1}{2} f''(x)(u+A_2u^2) + \frac{1}{6} f'''(x)u^2 \right],$$

$$\varphi = x - \frac{u}{1 + A_2u + (A_3 - A_2^2)u^2}.$$

This is  $\psi_{1,2}$  derived in Section 5.21.

### 5.4 The Fundamental Theorem of One-Point Iteration Functions

We review some of the terminology introduced in Section 1.24 which will be used in this section. The informational usage,  $d$ , of  $\varphi$  is defined as the number of new pieces of information required per iteration. If an I.F. belongs to the class of I.F. of order  $p$  and informational usage  $d$ , we write  $\varphi \in dI_p$ . The informational efficiency,  $EFF$ , of  $\varphi$  is defined by  $EFF = p/d$ . If  $EFF = 1$ ,  $\varphi$  is an optimal I.F. In this section we consider both simple and multiple zeros. If the order of  $\varphi$  is independent of the multiplicity  $m$  of the zero, then we say that its order is multiplicity-independent.

The reader familiar with I.F. has no doubt observed that one-point I.F. of order  $p$  depend explicitly on at least  $f$  and its first  $p - 1$  derivatives. Hence the informational usage of the I.F. is at least  $p$ . A theorem which gives a formal proof of this fact is given below. It is this theorem which causes us to label as optimal those I.F. whose informational efficiency is unity. The theorem is quite simple to prove and the result is in the "folklore" of numerical analysis. Its importance is that it makes us look for types of I.F. whose informational efficiency is greater than unity. Multipoint I.F. and I.F. with memory are not subject to the conclusions of this "fundamental theorem of one-point I.F." Recall that  $E_s$ , which was studied in Section 5.1, is of order  $p = s$ . This fact will be emphasized in this section by writing  $E_p$ . We give two equivalent formulations of

THEOREM 5-3. Let  $m = 1$ . Let  $\phi$  denote any one-point I.F. Then there exist  $\phi$  of all orders such that  $EFF(\phi) = 1$  and for all  $\phi$ ,  $EFF(\phi) \leq 1$ . Moreover  $\phi$  must depend explicitly on at least the first  $p - 1$  derivatives of  $f$ .

ALTERNATIVE FORMULATION. Let  $m = 1$ . Let  $\phi$  denote any one-point I.F. Then for all  $p$ , there exist  $\phi \in pI_p$  and if  $\phi \in dI_p$ , then  $d \geq p$ . Moreover  $\phi$  must depend explicitly on the first  $p - 1$  derivatives of  $f$ .

PROOF. Since  $E_p \in pI_p$ , there exist optimal I.F. of all orders. This disposes of the first part of the proof. Let  $\phi_1 \in I_p$ . From Theorem 2-10, the most general I.F. of order  $p$  is

$$\phi = \phi_1 + Uf^p$$

where  $U$  is any function bounded at  $a$ . Hence  $U$  cannot contain any terms in  $1/f$ . We take  $\phi_1 = E_p$ . The most convenient form to take for  $E_p$  is given by (5-2),

$$E_p = x - \sum_{j=1}^{p-1} \frac{(-1)^{j-1}}{j!} g^{(j)} f^j.$$

$E_p$  depends explicitly on  $f, f', \dots, f^{(p-1)}$ . Since the highest power appearing in  $E_p$  is  $f^{p-1}$ , none of its terms can be cancelled by  $Uf^p$ . Since  $\phi$  is a one-point I.F., none of

## 5.4-3

the  $f^{(j)}$  can be approximated by lower derivatives to within terms of  $\underline{O}(u^\ell)$ ,  $\ell > 0$ . Hence  $\phi$  must depend explicitly on  $f, f', \dots, f^{(p-1)}$  which completes the proof.

The restriction to one-point I.F. is essential as the following considerations show. Observe that

$$\frac{f[x - u(x)]}{f'(x)} = A_2(x)u^2(x) + \underline{O}[u^3(x)].$$

Recall that

$$E_3(x) = x - u(x) - A_2(x)u^2(x).$$

Since

$$\phi(x) = x - u(x) - \frac{f[x - u(x)]}{f'(x)}$$

and  $E_3$  differ only by terms of  $\underline{O}[u^3(x)]$ ,  $\phi$  is third order. That is, the second derivative appearing explicitly in  $E_3$  has been approximated in  $\phi$ . Observe that  $\phi$  uses information at  $x$  and at  $x - u(x)$  and is therefore an example of the multi-point I.F. which are studied in Chapters 8 and 9. Such approximation of derivatives is impossible for one-point I.F.

We turn to the case of multiple zeros.

COROLLARY. Let  $m$  be arbitrary and known. There exist  $\varphi$  of all orders, with  $\varphi$  depending explicitly on  $m$ , such that  $\text{EFF}(\varphi) = 1$  and for all  $\varphi$ ,  $\text{EFF}(\varphi) \leq 1$ . Moreover  $\varphi$  must depend explicitly on the first  $p - 1$  derivatives of  $f$ .

ALTERNATIVE FORMULATION. Let  $m$  be arbitrary and known. Then for all  $p$  there exist  $\varphi$  depending explicitly on  $m$  such that  $\varphi \in {}_p I_p$  and if  $\varphi \in {}_d I_p$ , then  $d \geq p$ . Moreover,  $\varphi$  must depend explicitly on the first  $p - 1$  derivatives of  $f$ .

PROOF. Define  $F = f^{1/m}$ . Then  $F$  has only simple zeros and  $F^{(j)}$  depends only on  $f^{(\ell)}$ ,  $\ell \leq j$ . An application of Theorem 5-3 completes the proof.

Note that this corollary assures us of the existence of optimal I.F. whose order is multiplicity-independent. A basic sequence of such I.F. will be explicitly given in Section 7.3. Observe that if we define  $G = f^{(m-1)}$  and insert  $G$  into any optimal one-point I.F. of order  $p$ , we obtain a one-point I.F. with informational efficiency equal to unity. This approach leads to I.F. which depend explicitly on  $f^{(m-1)}, f^{(m)}, \dots, f^{(m+p-2)}$ .

A case of greater interest is when  $m$  is not known. Note that  $u = f/f'$  has only simple zeros. Replacing  $f$  by  $u$  in any optimal one-point I.F. of order  $p$  leads to an I.F. which is of order  $p$ . We conclude that there exist I.F. of all orders such that  $\text{EFF}(\varphi) = p/(p+1)$  for zeros of all multiplicities. These I.F. do not contain  $m$  explicitly. These matters will be taken up in greater detail in Chapter 7.

## 5.5-1

5.5 The Coefficients of the Error Series of  $E_s$ 

We saw in Section 5.1 that

$$E_s - \alpha = Y_s(\alpha)(x-\alpha)^s + O[(x-\alpha)^{s+1}]. \quad (5-31)$$

In certain special cases, an explicit expression may be found for the error of  $E_s$ . Thus for the calculation of square roots,  $f = x^2 - A$  and

$$E_2 - \alpha = \frac{e^2}{2(\alpha+e)}, \quad \alpha = A^{\frac{1}{2}}, \quad e = x - \alpha.$$

In general, the expression for  $E_s - \alpha$ , is an infinite series whose leading term is given by the first term on the right side of (5-31). The coefficients of the infinite series are  $E_s^{(j)}(\alpha)/j!$ . A recursion formula for these coefficients will be found which does not involve differentiation. An interesting property of the coefficients will be proven in Section 5.52.

5.5-2

5.51 A recursion formula for the coefficients.

We recall the definitions of the following symbols which will be used frequently:

$$u(x) = \frac{f(x)}{f'(x)}, \quad a_j(x) = \frac{f^{(j)}(x)}{j!}, \quad A_j(x) = \frac{f^{(j)}(x)}{j!f'(x)}, \quad e = x - a.$$

The expansion of  $u(x)$  into a power series in  $e$  which will be needed below is derived now. Define  $v_\ell$  by

$$u(x) = f(x)/f'(x) = \sum_{\ell=1}^{\infty} v_\ell e^\ell. \quad (5-32)$$

Since

$$f(x) = \sum_{j=1}^{\infty} a_j e^j, \quad f'(x) = \sum_{j=1}^{\infty} j a_j e^{j-1},$$

and  $u(x)f'(x) \equiv f(x)$ , we obtain

$$a_\ell = \sum_{q=1}^{\ell} v_q (\ell+1-q) a_{\ell+1-q}, \quad \ell = 1, 2, \dots,$$

or

$$A_\ell = \sum_{q=1}^{\ell} v_q (\ell+1-q) A_{\ell+1-q}, \quad \ell = 1, 2, \dots, \quad (5-33)$$

5.5-3

and finally

$$v_\ell = A_\ell - \sum_{q=1}^{\ell-1} v_q (\ell+1-q) A_{\ell+1-q}, \quad \ell = 1, 2, \dots, \quad (5-34)$$

as a recursion formula for the calculation of the  $v_\ell$  with  $v_1 = 1$ . In (5-33) and (5-34), as throughout this section, the functions which occur are evaluated at  $\alpha$  unless otherwise indicated.

An explicit formula for the  $v_\ell$  may also be obtained. It is not difficult to prove that

$$v_\ell = A_\ell + \sum_{j=1}^{\ell-1} A_{\ell-j} \sum (-1)^r r! \prod_{i=1}^j \frac{[(i+1)A_{i+1}]^{\alpha_i}}{\alpha_i!}, \quad (5-35)$$

where  $r = \sum_{i=1}^j \alpha_i$  and where the inner sum is taken over all integers  $\alpha_i$  such that  $\sum_{i=1}^j i\alpha_i = j$ .

From either (5-34) or (5-35), the first few  $v_\ell$  may be calculated as

$$v_1 = 1,$$

$$v_2 = -A_2,$$

$$v_3 = 2A_2^2 - 2A_3,$$

$$v_4 = -4A_2^3 + 7A_2 A_3 - 3A_4.$$

## 5.5-4

We are ready to turn to the problem of finding the coefficients of the error series. Define  $\tau_{\ell,s}$  by

$$E_s(x) = \sum_{\ell=0}^{\infty} \tau_{\ell,s} e^{\ell}. \quad (5-36)$$

Since  $E_s(x) \in I_s$ , we expect  $\tau_{\ell,s} = 0$ , for  $0 < \ell < s$ , and  $\tau_{0,s} = a$ . This may be proven directly by induction on  $s$ . Let  $s = 1$ . Then

$$E_1(x) = x = a + (x-a) = a + e.$$

Now assume  $\tau_{\ell,s} = 0$ ,  $0 < \ell < s$ ,  $\tau_{0,s} = a$ . Substitute (5-36) into the formula of Lemma 5-3,

$$E_{s+1}(x) = E_s(x) - \frac{u(x)}{s} E'_s(x), \quad (5-37)$$

to find

$$\sum_{\ell=0}^{\infty} \tau_{\ell,s+1} e^{\ell} = a + \sum_{\ell=s}^{\infty} \tau_{\ell,s} e^{\ell} - \frac{u(x)}{s} \sum_{\ell=s}^{\infty} \ell \tau_{\ell,s} e^{\ell-1} = a + O(e^{s+1}),$$

which completes the induction.

Substituting (5-36) into (5-37) yields

$$\sum_{\ell=0}^{\infty} \tau_{\ell,s+1} e^{\ell} - \sum_{\ell=0}^{\infty} \tau_{\ell,s} e^{\ell} + \frac{u(x)}{s} \sum_{\ell=0}^{\infty} \ell \tau_{\ell,s} e^{\ell-1} = 0,$$

5.5-5

and using the expansion of  $u(x)$  given by (5-32), multiplying the series, and equating the coefficient of  $e^\ell$  to zero yields

$$s\tau_{\ell,s+1} + (\ell-s)\tau_{\ell,s} + \sum_{r=1}^{\ell-1} r\nu_{\ell+1-r}\tau_{r,s} = 0. \quad (5-38)$$

Since the  $\nu_\ell$  may be considered known, (5-38) can be used to determine the  $\tau_{\ell,s}$ . A number of  $\tau_{\ell,s}$  are given in Table 5-3. These results are summarized in

THEOREM 5-4. Let

$$E_s(x) = \sum_{\ell=0}^{\infty} \tau_{\ell,s} e^\ell, \quad u(x) = \sum_{\ell=1}^{\infty} \nu_\ell e^\ell.$$

Then

$$s\tau_{\ell,s+1} + (\ell-s)\tau_{\ell,s} + \sum_{r=1}^{\ell-1} r\nu_{\ell+1-r}\tau_{r,s} = 0,$$

with  $\tau_{0,s} = a$ ,  $\tau_{1,1} = 1$ ,  $\tau_{\ell,1} = 0$  for  $\ell > 1$ , and  $\tau_{\ell,s} = 0$  for  $0 < \ell < s$  and  $s > 1$ .

## 5.5-6

TABLE 5-3. FORMULAS FOR  $\tau_{\ell,s}$ 

$\downarrow \ell$	$\rightarrow s$	1	2	3	4
0		$\alpha$	$\alpha$	$\alpha$	$\alpha$
1		1			
2			$A_2$		
3			$-2A_2^2 + 2A_3$	$2A_2^2 - A_3$	
4			$4A_2^3 - 7A_2A_3 + 3A_4$	$-9A_2^3 + 12A_2A_3 - 3A_4$	$5A_2^3 - 5A_2A_3 + A_4$

5.5-7

Using Table 5-3 we find

$$E_2(x) - \alpha = A_2 e^2 + (-2A_2^2 + 2A_3)e^3 + (4A_2^3 - 7A_2 A_3 + 3A_4)e^4 + \underline{0}(e^5),$$

$$E_3(x) - \alpha = (2A_2^2 - A_3)e^3 + (-9A_2^3 + 12A_2 A_3 - 3A_4)e^4 + \underline{0}(e^5), \quad (5-39)$$

$$E_4(x) - \alpha = (5A_2^3 - 5A_2 A_3 + A_4)e^4 + \underline{0}(e^5).$$

Observe that for the cases worked out above, the coefficient of  $e^s$  in the expansion of  $E_s(x) - \alpha$  is  $Y_s(\alpha)$ , as expected.  
 (See Table 5-1 for the formulas of  $Y_s$ .)

5.52 A theorem concerning the coefficients. The following theorem may be used to check tables of the  $\tau_{\ell,s}$  and has a somewhat surprising corollary.

## THEOREM 5-5.

$$\sum_{s=1}^{\ell} \tau_{\ell,s} = A_{\ell}, \quad \ell = 1, 2, \dots . \quad (5-40)$$

PROOF. Note that since  $\tau_{\ell,s} = 0$ , for  $\ell < s$ , (5-40) is equivalent to  $\sum_{s=1}^{\infty} \tau_{\ell,s} = A_{\ell}$ . The proof is by induction on  $\ell$ . For  $\ell = 1$ ,  $\sum_{s=1}^{\ell} \tau_{\ell,s} = \tau_{1,1} = 1 = A_1$ . Let  $\sum_{s=1}^r \tau_{r,s} = A_r$ , for  $r = 1, 2, \dots, \ell-1$  with  $\ell \geq 2$ . Sum the recursion formula for  $\tau_{\ell,s}$  to obtain

$$\sum_{s=1}^{\ell} s\tau_{\ell,s+1} + \ell \sum_{s=1}^{\ell} \tau_{\ell,s} - \sum_{s=1}^{\ell} s\tau_{\ell,s} + \sum_{s=1}^{\ell} \sum_{r=1}^{\ell-1} r\nu_{\ell+1-r}\tau_{r,s} = 0,$$

or

$$(1-\ell) \sum_{s=1}^{\ell} \tau_{\ell,s} = \sum_{r=1}^{\ell-1} r\nu_{\ell+1-r} \sum_{s=1}^r \tau_{r,s},$$

5.5-9

where we have used the fact that  $\tau_{r,s} = 0$  for  $r < s$ . An application of the inductive hypothesis yields

$$(1-\ell) \sum_{s=1}^{\ell} \tau_{\ell,s} = \sum_{r=1}^{\ell-1} r v_{\ell+1-r} A_r = \sum_{q=1}^{\ell} (\ell+1-q) v_q A_{\ell+1-q} - \ell v_1 A_{\ell}.$$

Finally, the recursion formula for the  $v_\ell$ , (5-34), yields

$$(1-\ell) \sum_{s=1}^{\ell} \tau_{\ell,s} = (1-\ell) A_\ell$$

which completes the proof.

COROLLARY. Let  $k$  be an arbitrary positive integer.

Then

$$\alpha = \frac{1}{k} \sum_{s=1}^k E_s(x) - \frac{1}{k} \frac{f(x)}{f'(a)} + o(e^{k+1}). \quad (5-41)$$

PROOF.

$$\sum_{s=1}^k E_s(x) = \sum_{s=1}^k \sum_{\ell=0}^{\infty} \tau_{\ell,s} e^\ell$$

$$= k\alpha + \sum_{\ell=1}^k e^\ell \sum_{s=1}^k \tau_{\ell,s} + \sum_{\ell=k+1}^{\infty} e^\ell \sum_{s=1}^k \tau_{\ell,s}.$$

5.5-10

Since

$$\sum_{s=1}^k \tau_{\ell,s} = \sum_{s=1}^{\ell} \tau_{\ell,s}, \quad \ell \leq k,$$

an application of Theorem 5-5 yields

$$\sum_{s=1}^k E_s(x) = k\alpha + \sum_{\ell=1}^k A_\ell e^\ell + O(e^{k+1}).$$

The fact that

$$f(x) = \sum_{\ell=1}^{\infty} a_\ell e^\ell, \quad \frac{f(x)}{f'(\alpha)} = \sum_{\ell=1}^{\infty} A_\ell e^\ell,$$

and hence that

$$\frac{f(x)}{f'(\alpha)} = \sum_{\ell=1}^k A_\ell e^\ell + O(e^{k+1}),$$

yields the corollary.

Taking the limit as  $k \rightarrow \infty$  of (5-41) yields a special case of the general theorem that if a series is convergent, then it is also Cesàro summable.

## 6.0-1

### CHAPTER 6

#### ONE-POINT ITERATION FUNCTIONS WITH MEMORY

Two classes of one-point I.F. with memory are studied: interpolatory I.F. and derivative estimated I.F. These I.F. are always of nonintegral order. The structure of the main results, given by Theorems 6-1 to 6-4, is remarkably simple.

## 6.1-1

### 6.1 Interpolatory Iteration Functions

In Chapter 4 we developed the general theory of I.F. generated by direct or inverse hyperosculatory interpolation; such I.F. are called interpolatory I.F. If  $n = 0$ , the interpolatory I.F. are one-point I.F.; if  $n > 0$ , they are one-point I.F. with memory. The reader is referred to Theorems 4-1, 4-2, and 4-3 for results concerning the convergence and order of interpolatory I.F. The generalization to multiple zeros is handled in Chapter 7. In this section we limit ourselves to comments and examples. The notation is the same as in Chapter 4.

## 6.1-2

6.11 Comments. Observe that interpolatory one-point I.F. with memory use  $s$  pieces of new information at  $x_1$  and reuse  $s$  pieces of old information at the  $n$  points  $x_{i-1}, x_{i-2}, \dots, x_{i-n}$ . Thus their informational usage is  $s$ . Their order is determined by the unique positive real root of the equation

$$t^{n+1} - s \sum_{j=0}^n t^j = 0. \quad (6-1)$$

As was shown in Section 3.3, bounds on this root are given by

$$s < \beta_{n+1,s} < s + 1. \quad (6-2)$$

Hence bounds on the informational efficiency are given by

$$1 < \text{EFF} < 1 + \frac{1}{s}.$$

Furthermore,

$$\lim_{n \rightarrow \infty} \beta_{n+1,s} = s + 1.$$

Theorem 5-3 states that the informational efficiency of any one-point I.F. is less than or equal to unity. Hence the increase in informational efficiency for interpolatory I.F. with memory is directly traceable to the reuse of old information. On the other hand, we conclude from (6-2) that the old information adds less than one to the order of an interpolatory I.F.

### 6.1-3

The dependence of the order on  $n$  and  $s$  may be seen from Table 3-1 with  $k = n + 1$  and  $a = s$ . Observe that the order approaches its limiting value,  $s + 1$ , quite rapidly as a function of  $n$ ; this is particularly true for large  $s$ . The case of most practical interest is  $n = 1$ . Then (6-1) may be solved exactly and

$$\beta_{1,s} = \frac{1}{2}[s + (s^2 + 4s)^{\frac{1}{2}}].$$

A drawback of interpolatory I.F. with memory is that multiple precision arithmetic may have to be used for at least part of the calculation. A drawback of interpolatory I.F. generated by direct interpolation is that a polynomial of degree  $r - 1$  must be solved for each iteration. A reduction of degree technique, demonstrated for one-point I.F. in Section 5.33, is also available for one-point I.F. with memory; an example is given in Section 10.21.

## 6.1-4

6.12 Examples. The reader is referred to Appendix A for the interpolation formulas used in the following examples. The notation is the same as in Chapter 4. We shall not give the conditions for convergence; such conditions may be found in the examples of Section 4.3. The first three examples use inverse interpolation; the next three examples use direct interpolation. We take  $y_1 = f_1 = f(x_1)$ ,  $\tilde{v}_1 = \tilde{v}(y_1)$ ,

$$f[x_1, x_{i-1}] = \frac{f_1 - f_{i-1}}{x_1 - x_{i-1}}, \quad \tilde{v}[y_1, y_{i-1}] = \frac{\tilde{v}_1 - \tilde{v}_{i-1}}{y_1 - y_{i-1}}. \quad (6-3)$$

EXAMPLE 6-1.  $n = 1$ ,  $s = 1$ , (secant I.F.). In the Newtonian formulation,

$$\varphi_{1,1} = \tilde{v}_1 - y_1 \tilde{v}[y_1, y_{i-1}] = x_1 - \frac{f_1}{f[x_1, x_{i-1}]}.$$

In the Lagrange-Hermite formulation,

$$\varphi_{1,1} = \frac{f_1 x_{i-1} - f_{i-1} x_1}{f_1 - f_{i-1}}.$$

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |y_2(\alpha)|^{p-1}, \quad p = \frac{1}{2}(1 + \sqrt{5}) \sim 1.62.$$

$$EFF = \frac{p}{s} \sim 1.62.$$

6.1-5

EXAMPLE 6-2.  $n = 2, s = 1.$

$$\varphi_{2,1} = \varphi_{1,1} + \frac{f_i f_{i-1}}{f_i - f_{i-2}} \left\{ \frac{1}{f[x_i, x_{i-1}]} - \frac{1}{f[x_{i-1}, x_{i-2}]} \right\}.$$

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |Y_3(\alpha)|^{\frac{1}{2}(p-1)}, \quad p \sim 1.84.$$

$$EFF = \frac{p}{s} \sim 1.84.$$

EXAMPLE 6-3.  $n = 1, s = 2.$

$$\varphi_{1,2} = \varphi_{0,2} + f_1^2 H,$$

$$\varphi_{0,2} = x_i - \frac{f_i}{f'_i},$$

$$H = \frac{1}{f_i - f_{i-1}} \left\{ \frac{1}{f'_i} - \frac{1}{f[x_i, x_{i-1}]} \right\} - \frac{f_{i-1}}{(f_i - f_{i-1})^2} \left\{ \frac{1}{f'_i} + \frac{1}{f'_{i-1}} - \frac{2}{f[x_i, x_{i-1}]} \right\}.$$

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |Y_4(\alpha)|^{\frac{1}{3}(p-1)}, \quad p = 1 + \sqrt{3} \sim 2.73.$$

$$EFF = \frac{p}{s} \sim 1.37.$$

6.1-6

EXAMPLE 6-4.  $n = 1, s = 1$ , (secant I.F.). In the Newtonian formulation,

$$P_{1,1}(t) = f_i + (t-x_i)f[x_i, x_{i-1}],$$

$$\Phi_{1,1} = x_i - \frac{f_i}{f[x_i, x_{i-1}]},$$

In the Lagrangian formulation,

$$P_{1,1}(t) = f_i\left(\frac{t-x_{i-1}}{x_i-x_{i-1}}\right) + f_{i-1}\left(\frac{t-x_i}{x_{i-1}-x_i}\right),$$

$$\Phi_{1,1} = \frac{f_i x_{i-1} - f_{i-1} x_i}{f_i - f_{i-1}}.$$

Observe that for the case  $n = 1, s = 1$ , the I.F. generated by direct and inverse interpolation are identical. This is the only case, for  $n > 0$ , where this is so.

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_2(\alpha)|^{p-1} = |Y_2(\alpha)|^{p-1},$$

$$p = \frac{1}{2}(1 + \sqrt{5}) \sim 1.62.$$

$$EFF = \frac{p}{s} \sim 1.62.$$

6.1-7

EXAMPLE 6-5.  $n = 2, s = 1$ . The second degree polynomial equation

$$P_{2,1}(t) = f_1 + (t-x_1)f[x_1, x_{1-1}] + (t-x_1)(t-x_1+h)f[x_1, x_{1-1}, x_{1-2}],$$

$$h = x_i - x_{i-1}, \quad f[x_1, x_{i-1}, x_{i-2}] = \frac{f[x_1, x_{i-1}] - f[x_{i-1}, x_{i-2}]}{x_i - x_{i-2}},$$

must be solved for  $t - x_1$ .

$$\frac{|e_{i+1}|}{|e_1|^p} \rightarrow |A_3(\alpha)|^{\frac{1}{2}(p-1)}, \quad p \sim 1.84.$$

$$EFF = \frac{p}{s} \sim 1.84.$$

This I.F. is discussed in Section 10.21.

EXAMPLE 6-6.  $n = 1, s = 2$ . The third degree polynomial equation

$$0 = P_{1,2}(t) = f_1 + (t-x_1)f'_1 + (t-x_1)^2 H,$$

$$H = \frac{f'_1 - f[x_1, x_{1-1}]}{x_i - x_{i-1}} + (t-x_1) \left[ \frac{f'_1 + f'_{i-1} - 2f[x_1, x_{i-1}]}{(x_i - x_{i-1})^2} \right]$$

must be solved for  $t - x_1$ .

6.1-8

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_4(\alpha)|^{\frac{1}{3}(p-1)}, \quad p = 1 + \sqrt{3} \sim 2.73.$$

$$EFF = \frac{p}{s} \sim 1.37.$$

## 6.2 Derivative Estimated One-Point Iteration Functions With Memory

6.21 The secant iteration function and its generalization. We have used direct and inverse interpolation to derive one-point I.F. with memory. We observed that the secant I.F. could be generated from either direct or inverse interpolation. We now give two additional derivations of the secant I.F.; the method of derivation suggests a second general technique for generating one-point I.F. with memory.

The secant I.F., together with slight modifications thereof, must share with Halley's I.F. (Section 5.2) the distinction of being the most often rediscovered I.F. in the literature. Discussions of the secant I.F. may be found in Bachmann [6.2-1], Collatz [6.2-2, Chap. III], Hsu [6.2-3], Jeeves [6.2-4], Ostrowski [6.2-5, Chap. 3], and Putzer [6.2-6]. Its order seems to have been first given by Bachmann [6.2-7] (1954). Let  $\tilde{g}$  be the inverse to  $f$ . One way of writing Newton's I.F. is

$$E_2 = \tilde{g}(y) - y\tilde{g}'(y). \quad (6-4)$$

Newton's I.F. is second order and uses two pieces of information. We estimate  $\tilde{g}'(y)$  by  $Q'_{1,1}$  where

$$Q'_{1,1}(t) = \tilde{g}(y) + (t-y) \left[ \frac{\tilde{g}(y) - \tilde{g}(y_{1-1})}{y - y_{1-1}} \right]$$

is the first degree polynomial which agrees with  $\tilde{y}$  at the points  $y$  and  $y_{i-1}$ . It is convenient to use the symbols  $y$  and  $y_i$  and the symbols  $x$  and  $x_i$  interchangeably. Replacing  $\tilde{y}'$  by  $Q'_{1,1}$  in (6-3) leads to

$$\varphi[x_i; x_{i-1}] = x_i - y_i \left[ \frac{x_i - x_{i-1}}{y_i - y_{i-1}} \right]$$

which is the secant I.F.

On the other hand, we may write

$$E_2 = x - \frac{f(x)}{f'(x)}$$

and estimate  $f'(x)$  by differentiating the first degree polynomial

$$P_{1,1}(t) = f(x) + (t-x) \left[ \frac{f(x) - f(x_{i-1})}{x - x_{i-1}} \right]$$

which agrees with  $f$  at the points  $x$  and  $x_{i-1}$ . Again the secant I.F. is generated.

In the following sections we will deal with two broad generalizations of these ideas.

- a. Rather than estimating the first derivative of an optimal I.F. of second order, we estimate the  $(s-1)$ st derivative of an optimal I.F. of order  $s$ .

### 6.2-3

- b. Rather than estimating the first derivative from two values of the function, we estimate the  $(s-1)$ st derivative from  $n + 1$  values (one new and  $n$  old) of the first  $s - 2$  derivatives.

In general the estimation of  $f^{(s-1)}$  and the estimation of  $\tilde{g}^{(s-1)}$  leads to different I.F. We investigate the former in Section 6.22 and the latter in Section 6.23. After studying the estimation of derivatives in the optimal basic sequence  $E_s$ , we will be able to deal with the case of arbitrary optimal I.F. The one-point I.F. with memory thus generated will be said to be derivative estimated.

6.22 Estimation of  $f^{(s-1)}$ . We first develop the theory of one-point I.F. with memory which are generated by estimating  $f^{(s-1)}$  in the I.F.  $E_s$ ; the corresponding theory for arbitrary optimal I.F. then follows easily. From Section 5.11,

$$E_s = x + \sum_{j=1}^{s-1} Y_j u^j, \quad (6-5)$$

$$Y_j(x) = \frac{(-1)^{j-1} \mathfrak{g}(j)(y)}{j! [\mathfrak{g}'(y)]^j} \Big|_{y=f(x)}, \quad u = \frac{f}{f'}.$$

$E_s$  uses  $s - 1$  derivatives and hence  $s$  pieces of information. We estimate  $f^{(s-1)}(x_1)$  from  $f^{(\ell)}(x_{1-j})$ , with  $j = 0, 1, \dots, n$ , and  $\ell = 0, 1, \dots, s-2$ . The symbols  $x$  and  $x_1$  will be interchangeably. The I.F. so generated use  $s - 1$  pieces of new information at the point  $x_1$  and reuse  $s - 1$  pieces of information from the previous  $n$  points. Hence  $s - 1$  will be of basic importance and we define

$$S = s - 1. \quad (6-6)$$

In place of  $r = s(n+1)$ , we introduce

$$R = S(n+1). \quad (6-7)$$

The advantage of using  $S$  and  $R$  instead of  $s$  and  $r$  will become evident below.

6.2-5

Let  $P_{n,S}(t)$  be the polynomial such that

$$P_{n,S}^{(\ell)}(x_{i-j}) = f^{(\ell)}(x_{i-j}), \quad j = 0, 1, \dots, n, \quad \ell = 0, 1, \dots, S-1. \quad (6-8)$$

We estimate  $f^{(S)}(x)$  by  $P_{n,S}^{(S)}(x)$ . Let

$$*f_n^{(S)}(x) \equiv P_{n,S}^{(S)}(x). \quad (6-9)$$

As shown in Appendix A,

$$f^{(S)}(x) - *f_n^{(S)}(x) = \frac{S!}{R!} f^{(R)}(\xi_1) \prod_{j=1}^n (x-x_{i-j})^S, \quad (6-10)$$

where  $\xi_1$  lies in the interval determined by  $x_i, x_{i-1}, \dots, x_{i-n}$ .

Let  $*E_{n,S}$  be generated from  $E_{S+1}$  by estimating  $f^{(S)}(x)$  by  $*f_n^{(S)}(x)$ . A new approximation to  $a$  is defined by

$$x_{i+1} = *E_{n,S}(x_i; x_{i-1}, \dots, x_{i-n}).$$

We derive the error equation for  $*E_{n,S}$ . Recall that  $y_j$  is independent of  $f^{(S)}$  for  $j < S$ . Let  $*Y_{n,S}$  be generated from  $Y_S$  by estimating  $f^{(S)}$  by  $*f_n^{(S)}$ . Then

$$*E_{n,S} = x + \sum_{j=1}^{S-1} y_j u^j + *Y_{n,S} u^S = E_S + *Y_{n,S} u^S. \quad (6-11)$$

6.2-6

On the other hand,

$$\alpha = E_S + Y_S u^S + \frac{(-1)^{S+1}}{(S+1)!} \mathfrak{g}^{(S+1)}(\theta_1) f^{S+1}, \quad (6-12)$$

where  $\theta_1$  lies in the interval determined by 0 and  $y_1$ . Hence

$$*E_{n,S} - \alpha = u^S [*Y_{n,S} - Y_S] - \frac{(-1)^{S+1}}{(S+1)!} \mathfrak{g}^{(S+1)}(\theta_1) f^{S+1}.$$

Since by corollary d of Theorem 5-1,  $Y_S$  depends on  $f^{(S)}$  only as  $(-1)^S A_S$ ,  $A_S = f^{(S)} / (S! f')$ , we conclude that

$$*E_{n,S} - \alpha = \frac{(-1)^S u^S}{S! f'} \left[ *f_n^{(S)} - f^{(S)} \right] - \frac{(-1)^{S+1}}{(S+1)!} \mathfrak{g}^{(S+1)}(\theta_1) f^{S+1}.$$

An application of (6-10) yields

$$*E_{n,S} - \alpha = - \frac{(-1)^R f^{(R)}(\xi_i)}{R! f'} u^S \prod_{j=1}^n (x_{i-j} - x)^S - \frac{(-1)^{S+1}}{(S+1)!} \mathfrak{g}^{(S+1)}(\theta_1) f^{S+1}.$$

Let

$$e_{i-j} = x_{i-j} - \alpha, \quad e_{i+1} = x_{i+1} - \alpha = *E_{n,S} - \alpha,$$

$$y_1 = f'(\eta_1) e_1 = \frac{e_1}{\mathfrak{g}'(\rho_1)},$$

where  $\eta_1$  lies in the interval determined by  $x_1$  and  $\alpha$  and  $\rho_1 = f(\eta_1)$ . Then

$$e_{i+1} = *M_i e_i^S \prod_{j=1}^n (e_{i-j} - e_i)^S + *N_i e_i^{S+1},$$

$$*M_i = - \frac{(-1)^{R_f(R)}(\xi_i)}{R! [f']^{S+1}} [f'(\eta_i)]^S, \quad (6-13)$$

$$*N_i = - \frac{(-1)^{S+1} \mathfrak{g}^{(S+1)}(\theta_i)}{(S+1)! [\mathfrak{g}'(\rho_i)]^{S+1}}.$$

This is the error equation for  $*E_{n,S}$ ; we use it to derive the conditions for convergence and the order. Assume that  $e_i$  is nonzero for all finite  $i$ . We shall show that if  $x_0, x_1, \dots, x_n$  are sufficiently close to  $\alpha$ , then  $e_i \rightarrow 0$ .

Let

$$J = \left\{ x \mid |x - \alpha| \leq \Gamma \right\}.$$

Let  $f^{(R)}$  be continuous on  $J$  and let  $f'$  be nonzero on  $J$ . Let  $f$  map the interval  $J$  into the interval  $K$ . Since  $n > 0$ ,  $R \geq S + 1$ . Hence we can conclude that  $\mathfrak{g}^{(S+1)}$  is continuous on  $K$ . Let

$$\frac{|f^{(R)}|}{R!} \leq v_1, \quad v_3 \geq |f'| \geq v_2$$

6.2-8

for all  $x \in J$ . Let

$$*M = \frac{v_1 v_3^S}{v_2^{S+1}}.$$

Let

$$\frac{|x^{(S+1)}|}{(S+1)!} \leq \lambda_1, \quad |x'| \geq \lambda_2$$

for all  $y \in K$ ; that is, for all  $x \in J$ . Let

$$*N = \frac{\lambda_1}{\lambda_2^{S+1}}.$$

Let  $x_0, x_1, \dots, x_n \in J$ . Then

$$\max[|e_0|, |e_1|, \dots, |e_n|] \leq \Gamma.$$

By an inductive argument analogous to one employed in Section 4.21, it can be shown that if

$$2^{R-S} *M \Gamma^{R-1} + *N \Gamma^S \leq 1,$$

then  $x_i \in J$  for all  $i$ . Hence  $|*M_i| \leq *M$ ,  $|*N_i| \leq *N$  for all  $i$ .

Then

$$\delta_{i+1} \leq *M \delta_i^S \prod_{j=1}^n (\delta_{i-j} + \delta_i)^S + *N \delta_i^{S+1}, \quad \delta_{i-j} = |e_{i-j}|.$$

An application of Lemma 3-2 shows that if

$$2^{R-S} *_{M\Gamma}^{R-1} + *_{N\Gamma}^S < 1,$$

then  $\delta_1 \rightarrow 0$ . Hence  $e_1 \rightarrow 0$ .

Observe that

$$*_{M_1} \rightarrow -(-1)^R A_R(\alpha), \quad *_{N_1} \rightarrow Y_{S+1}(\alpha).$$

All the conditions of Theorem 3-4 now apply and we conclude that

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |A_R(\alpha)|^{(p-1)/(R-1)}, \quad (6-14)$$

where  $p$  is the unique real positive root with magnitude greater than unity of the equation

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

We summarize our results in

**THEOREM 6-1.** Let

$$J = \left\{ x \mid |x-\alpha| \leq \Gamma \right\}.$$

6.2-10

Let  $R = S(n+1)$ ,  $n > 0$ . Let  $f^{(R)}$  be continuous and let  $f'$  be nonzero on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$  and let a sequence  $\{x_i\}$  be defined as follows: Let  $*f_n^{(S)}$  be defined by (6-8) and (6-9). Let  $*E_{n,S}$  be generated from  $E_{S+1}$  by estimating  $f^{(S)}$  by  $*f_n^{(S)}$ . Define

$$x_{i+1} = *E_{n,S}(x_1, x_{i-1}, \dots, x_{i-n}).$$

Let  $e_{i-j} = x_{i-j} - \alpha$ . Let

$$\frac{|f^{(R)}|}{R!} \leq v_1, \quad v_3 \geq |f'| \geq v_2,$$

$$\frac{|\vartheta^{(S+1)}|}{(S+1)!} \leq \lambda_1, \quad |\vartheta'| \geq \lambda_2$$

for all  $x \in J$  and let

$$*M = \frac{v_1 v_3^S}{v_2^{S+1}}, \quad *N = \frac{\lambda_1}{\lambda_2^{S+1}}.$$

Assume that  $e_i$  is nonzero for all finite  $i$ . Suppose that  $2^{R-S} *M^R R^{-1} + *N^S < 1$ .

Then,  $x_i \in J$  for all  $i$ ,  $e_i \rightarrow 0$ , and

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_R(\alpha)|^{(p-1)/(R-1)}, \quad (6-15)$$

6.2-11

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0,$$

and where  $A_R = f^{(R)}/(R!f')$ .

We turn to the case where  $f^{(S)}$  is estimated in an arbitrary optimal one-point I.F. It will turn out that the order and asymptotic error constant of the I.F. so generated are identical with the case where  $f^{(S)}$  is estimated in  $E_{S+1}$ . From Theorem 2-10,

$$\varphi_{S+1} = E_{S+1} + U f^{S+1} \quad (6-16)$$

is the most general I.F. of order  $S + 1$ . Let  $\varphi_{S+1}$  be an optimal one-point I.F. Let  $*\varphi_{n,S}$  be generated from  $\varphi_{S+1}$  by estimating  $f^{(S)}$  by  $*f_n^{(S)}$  in  $\varphi_{S+1}$ , and let  $*U_{n,S}$  be defined analogously. Then

$$*\varphi_{n,S} = *E_{n,S} + *U_{n,S} f^{S+1}.$$

Hence

$$*\varphi_{n,S} - \alpha = *E_{n,S} - \alpha + *U_{n,S} f^{S+1},$$

$$*\varphi_{n,S} - \alpha = e_{i+1} = *M_i e_i^S \prod_{j=1}^n (e_{i-j} - e_i)^S + \left\{ *N_i + \frac{*U_{n,S}}{[\delta'(\rho_i)]^{S+1}} \right\} e_i^{S+1}$$

(6-17)

where  $M_i$  and  $N_i$  were defined in (6-13). We can proceed as before to arrive at

THEOREM 6-2. Let

$$J = \left\{ x \mid |x - \alpha| \leq \Gamma \right\}.$$

Let  $R = S(n+1)$ ,  $n > 0$ . Let  $f^{(R)}$  be continuous and let  $f'$  be nonzero on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$  and let a sequence  $\{x_i\}$  be defined as follows: Let  $*f_n^{(S)}$  be defined by (6-8) and (6-9). Let  $\varphi_{S+1}$  be an arbitrary optimal one-point I.F.; let  $\varphi_{S+1} = E_{S+1} + U f^{S+1}$ . Let  $*\varphi_{n,S}$  be generated from  $\varphi_{S+1}$  by estimating  $f^{(S)}$  by  $*f_n^{(S)}$  in  $\varphi_{S+1}$ ; then

$$*\varphi_{n,S} = *E_{n,S} + *U_{n,S} f^{S+1}.$$

Define

$$x_{i+1} = *\varphi_{n,S}(x_i; x_{i-1}, \dots, x_{i-n}).$$

6.2-13

Let  $e_{i-j} = x_{i-j} - \alpha$ . Let

$$\frac{|f^{(R)}|}{R!} \leq v_1, \quad v_3 \geq |f'| \geq v_2,$$

$$\frac{|\mathfrak{z}^{(S+1)}|}{(S+1)!} \leq \lambda_1, \quad |\mathfrak{z}'| \geq \lambda_2, \quad |*U_{n,S}| \leq \mu,$$

for all  $x \in J$ . Let

$$*M = \frac{v_1 v_3^S}{v_2^{S+1}}, \quad \#N = \frac{\lambda_1 + \mu}{\lambda_2^{S+1}}.$$

Assume that  $e_i$  is nonzero for all finite  $i$ . Suppose that  $2^{R-S} *M^{\Gamma^{R-1}} + \#N^{\Gamma^S} < 1$ .

Then,  $x_i \in J$  for all  $i$ ,  $e_i \rightarrow 0$ , and

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_R(\alpha)|^{(p-1)/(R-1)}, \quad (6-18)$$

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0,$$

and where  $A_R = f^{(R)}/(R!f')$ .

6.2-14

6.23 Estimation of  $\mathfrak{F}^{(s-1)}$ . We turn to one-point I.F. with memory which are generated by estimating  $\mathfrak{F}^{(s-1)}$  in the I.F.  $E_s$ ; the corresponding theory for arbitrary optimal I.F. then follows easily.

We again take  $S = s - 1$ ,  $R = S(n+1)$ ; the symbols  $y$  and  $y_{i-1}$  are used interchangeably. Let  $Q_{n,S}$  be the polynomial such that

$$Q_{n,S}^{(\ell)}(y_{i-j}) = \mathfrak{F}^{(\ell)}(x_{i-j}), \quad j = 0, 1, \dots, n, \quad \ell = 0, 1, \dots, S-1. \quad (6-19)$$

We estimate  $\mathfrak{F}^{(S)}(y)$  by  $Q_{n,S}^{(S)}(y)$ . Let

$$\mathfrak{F}_n^{(S)}(y) \equiv Q_{n,S}^{(S)}(y). \quad (6-20)$$

As shown in Appendix A,

$$\mathfrak{F}^{(S)}(y) - \mathfrak{F}_n^{(S)}(y) = \frac{S!}{R!} \mathfrak{F}^{(R)}(\theta_i) \prod_{j=1}^n (y - y_{i-j})^S, \quad (6-21)$$

where  $\theta_i$  lies in the interval determined by  $y_i, y_{i-1}, \dots, y_{i-n}$ .

Let  $\mathbb{E}_{n,S}$  be generated from  $E_{S+1}$  by estimating  $\mathfrak{F}^{(S)}(y)$  by  $\mathfrak{F}_n^{(S)}(y)$ . A new approximation to  $\alpha$  is defined by

$$x_{i+1} = \mathbb{E}_{n,S}(x_i, x_{i-1}, \dots, x_{i-n}).$$

6.2-15

We derive the error equation for  $\mathbb{E}_{n,S}$ . We have

$$\begin{aligned}\mathbb{E}_{n,S} &= x + \sum_{j=1}^{S-1} \frac{(-1)^j}{j!} \mathfrak{f}^{(j)}_x + \frac{(-1)^S}{S!} \mathbb{E}_n^{(S)} f^S \\ &= E_S + \frac{(-1)^S}{S!} \mathbb{E}_n^{(S)} f^S,\end{aligned}\quad (6-22)$$

$$\alpha = E_S + \frac{(-1)^S}{S!} \mathfrak{f}^{(S)} f^S + \frac{(-1)^{S+1}}{(S+1)!} \mathfrak{f}^{(S+1)}(\theta_1) f^{S+1}, \quad (6-23)$$

where  $\theta_1$  lies in the interval determined by 0 and  $y_1$ . From (6-21), (6-22), and (6-23),

$$\mathbb{E}_{n,S} - \alpha = - \frac{(-1)^R}{R!} \mathfrak{f}^{(R)}(\theta_1) f^S \prod_{j=1}^n (y_{1+j} - y_1)^S - \frac{(-1)^{S+1}}{(S+1)!} \mathfrak{f}^{(S+1)}(\theta_1) f^{S+1}.$$

Since

$$y_{1+1} = f'(\eta_{1+1})(x_{1+1} - \alpha) = \frac{\mathbb{E}_{n,S} - \alpha}{\mathfrak{f}'(\rho_{1+1})},$$

where  $\eta_{1+1}$  lies in the interval determined by  $x_{1+1}$  and  $\alpha$  and  $\rho_{1+1} = f(\eta_{1+1})$ , we conclude that

$$\begin{aligned}y_{1+1} &= \mathbb{M}_{1+1} y_1^S \prod_{j=1}^n (y_{1+j} - y_1)^S + \mathbb{N}_{1+1} y_1^{S+1}, \\ \mathbb{M}_{1+1} &= - \frac{(-1)^R \mathfrak{f}^{(R)}(\theta_1)}{R! \mathfrak{f}'(\rho_{1+1})}, \\ \mathbb{N}_{1+1} &= - \frac{(-1)^{S+1} \mathfrak{f}^{(S+1)}(\theta_1)}{(S+1)! \mathfrak{f}'(\rho_{1+1})}.\end{aligned}\quad (6-24)$$

6.2-16

Because of the dependence of (6-24) on  $y_{i-j}$ , we shall focus our attention on  $H$ ,

$$H = \{y \mid |y| \leq \Lambda\}.$$

Let  $\varphi^{(R)}$  be continuous on  $H$  and let  $\varphi'$  be nonzero there.

Let  $y_0, y_1, \dots, y_n \in H$ . Since  $M_{i+1}$  and  $N_{i+1}$  depend on  $y_{i+1}$  through the parameter  $\rho_{i+1}$ , we cannot prove that all  $y_i \in H$  in a manner analogous to the proof of the last section. Instead we assume that  $y_i \in H$  for all  $i$  and that  $y_i$  is nonzero for all finite  $i$ . Let

$$\frac{|\varphi^{(R)}|}{R!} \leq \lambda_3, \quad |\varphi'| \geq \lambda_2, \quad \frac{|\varphi^{(S+1)}|}{(S+1)!} \leq \lambda_1,$$

for all  $y \in H$ . Let

$$\iota_M = \frac{\lambda_3}{\lambda_2}, \quad \iota_N = \frac{\lambda_1}{\lambda_2}.$$

Then

$$|y_{i+1}| \leq \iota_M |y_i|^S \prod_{j=1}^n (|y_{i-j}| + |y_i|)^S + \iota_N |y_{i+1}|^{S+1}.$$

An application of Lemma 3-2 shows that if

$$2^{R-S} \iota_M^{R-1} + \iota_N S < 1,$$

6.2-17

then  $y_{i+1} \rightarrow 0$ . Hence  $e_{i+1} \rightarrow 0$ . Observe that

$$L_{M_{i+1}} \rightarrow -(-1)^R a_R(0), \quad a_R(y) = \frac{\mathfrak{J}^{(R)}(y)}{R! \mathfrak{J}'(y)}.$$

All the conditions of Theorem 3-4 are now satisfied and we can conclude that

$$\frac{|y_{i+1}|}{|y_i|^p} \rightarrow |a_R(0)|^{(p-1)/(R-1)},$$

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

Since

$$y_{i-j} = f(x_{i-j}) = \frac{e_{i-j}}{\mathfrak{J}'(\rho_{i-j})},$$

where  $\rho_{i-j}$  lies in the interval determined by  $y_{i-j}$  and  $a$ , we have that

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |Y_R(a)|^{(p-1)/(R-1)}.$$

We summarize our results in

THEOREM 6-3. Let

$$H = \{y \mid |y| \leq \Lambda\}.$$

Let  $R = S(n+1)$ ,  $n > 0$ . Let  $\mathfrak{g}^{(R)}$  be continuous and let  $\mathfrak{g}'$  be nonzero on  $H$ . Let  $x_0, x_1, \dots, x_n$  be given and let a sequence  $\{x_i\}$  be defined as follows: Let  $\mathbb{E}_n^{(S)}$  be defined by (6-19) and (6-20). Let  $\mathbb{E}_{n,S}$  be generated from  $E_{S+1}$  by estimating  $\mathfrak{g}^{(S)}$  by  $\mathbb{E}_n^{(S)}$ . Define

$$x_{i+1} = \mathbb{E}_{n,S}(x_i; x_{i-1}, \dots, x_{i-n}).$$

Assume that  $y_i \in H$  for all  $i$  and that  $y_i$  is nonzero for all finite  $i$ . Let

$$\frac{|\mathfrak{g}^{(R)}|}{R!} \leq \lambda_3, \quad |\mathfrak{g}'| \geq \lambda_2, \quad \frac{|\mathfrak{g}^{(S+1)}|}{(S+1)!} \leq \lambda_1,$$

for all  $y \in H$  and let

$$\mathbb{E}_M = \frac{\lambda_3}{\lambda_2}, \quad \mathbb{E}_N = \frac{\lambda_1}{\lambda_2}.$$

Suppose that  $2^{R-S} \mathbb{E}_M R^{-1} + \mathbb{E}_N S < 1$ .

Then  $y_1 \rightarrow 0$  and

$$\frac{|y_{1+1}|}{|y_1|^p} \rightarrow |a_R(0)|^{(p-1)/(R-1)}, \quad (6-25)$$

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0,$$

and where

$$a_R(y) = \frac{\mathfrak{F}^{(R)}(y)}{R! \mathfrak{F}'(y)}.$$

Furthermore,

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |Y_R(\alpha)|^{(p-1)/(R-1)}, \quad (6-26)$$

where

$$Y_R(x) = - \left. \frac{(-1)^R \mathfrak{F}^{(R)}(y)}{R! [\mathfrak{F}'(y)]^j} \right|_{y=f(x)}.$$

The case where  $\mathfrak{F}^{(S)}$  is estimated in an arbitrary optimal one-point I.F. may be handled in a fashion which should by now be familiar. We summarize the results in

6.2-20

THEOREM 6-4. Let

$$H = \{y \mid |y| \leq \Lambda\}.$$

Let  $R = S(n+1)$ ,  $n > 0$ . Let  $\mathfrak{F}^{(R)}$  be continuous and let  $\mathfrak{F}'$  be nonzero on  $H$ . Let  $x_0, x_1, \dots, x_n$  be given and let a sequence  $\{x_i\}$  be defined as follows: Let  ${}^L\mathfrak{F}_n^{(S)}$  be defined by (6-19) and (6-20). Let  $\varphi_{S+1}$  be an arbitrary optimal one-point I.F.; let  $\varphi_{S+1} = E_{S+1} + Uf^{S+1}$ . Let  ${}^L\varphi_{n,S}$  be generated from  $\varphi_{S+1}$  by estimating  $\mathfrak{F}^{(S)}$  by  ${}^L\mathfrak{F}_n^{(S)}$  in  $\varphi_{S+1}$ ; then

$${}^L\varphi_{n,S} = {}^L\mathbb{E}_{n,S} + {}^L\mathbb{U}_{n,S} f^{S+1}.$$

Define

$$x_{i+1} = {}^L\varphi_{n,S}(x_i; x_{i-1}, \dots, x_{i-n}).$$

Assume that  $y_i \in H$  for all  $i$  and that  $y_i$  is nonzero for all finite  $i$ . Let

$$\frac{|{}^L\mathfrak{F}_n^{(R)}|}{R!} \leq \lambda_3, \quad |\mathfrak{F}'| \geq \lambda_2,$$

$$\frac{|{}^L\mathfrak{F}_n^{(S+1)}|}{(S+1)!} \leq \lambda_1, \quad |{}^L\mathbb{U}_{n,S}| \leq \kappa,$$

6.2-21

for all  $y \in H$ . Let

$$\gamma_M = \frac{\lambda_3}{\lambda_2}, \quad \gamma_N = \frac{\lambda_1 + \kappa}{\lambda_2}.$$

Suppose that  $2^{R-S} \gamma_M^{R-1} + \gamma_N^S < 1$ .

Then  $y_1 \rightarrow 0$  and

$$\frac{|y_{i+1}|}{|y_i|^p} \rightarrow |a_R(0)|^{(p-1)/(R-1)}, \quad (6-27)$$

where  $p$  is the unique real positive root of

$$t^{n+1} - s \sum_{j=0}^n t^j = 0.$$

Furthermore,

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |y_R(a)|^{(p-1)/(R-1)}. \quad (6-28)$$

6.24 Examples. The reader is referred to Appendix A for the approximate derivative formulas used in the following examples. We shall not give the conditions for convergence. The notation is the same as in Sections 6.22 and 6.23.

EXAMPLE 6-7.  $n = 1, S = 1$ , (secant I.F.).

$$*f'_1 = f[x_1, x_{1-1}],$$

$$*E_{1,1} = x_1 - \frac{f_1}{*f'_1}.$$

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |A_2(\alpha)|^{p-1}, \quad p = \frac{1}{2}(1 + \sqrt{5}) \sim 1.62.$$

EXAMPLE 6-8.  $n = 2, S = 1$ .

$$*f'_2 = f[x_1, x_{1-1}] + f[x_1, x_{1-2}] - f[x_{1-1}, x_{1-2}],$$

$$*E_{2,1} = x_1 - \frac{f_1}{*f'_2}.$$

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |A_3(\alpha)|^{\frac{1}{2}(p-1)}, \quad p \sim 1.84.$$

6.2-23

EXAMPLE 6-9.  $n = 1, S = 2.$

$$*f''_1 = \frac{2}{x_i - x_{i-1}} \left\{ 2f'_1 + f'_{i-1} - 3f[x_i, x_{i-1}] \right\}.$$

$$*E_{1,2} = x_i - u_i - \frac{1}{2} \frac{u_i^2}{f'_1} *f''_1, \quad u_i = \frac{f'_1}{f'_1}.$$

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_4(\alpha)|^{\frac{1}{3}(p-1)}, \quad p = 1 + \sqrt{3} \sim 2.73.$$

EXAMPLE 6-10.  $n = 1, S = 2.$  We estimate  $f''$  by  $*f''_1$   
in Halley's I.F. rather than in  $E_3.$  Thus

$$*\varphi_{1,2} = x_i - \frac{u_i}{1 - \frac{1}{2} \left( u_i/f'_1 \right) *f''_1}.$$

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_4(\alpha)|^{\frac{1}{3}(p-1)}, \quad p = 1 + \sqrt{3} \sim 2.73.$$

6.2-24

EXAMPLE 6-11.  $n = 1, S = 1$ , (secant I.F.).

$${}^{\perp}\mathfrak{g}'_1 = \frac{1}{f[x_1, x_{1-1}]},$$

$${}^{\perp}E_{1,1} = x_1 - f_1 {}^{\perp}\mathfrak{g}'_1.$$

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |y_2(\alpha)|^{p-1}, \quad p = \frac{1}{2}(1 + \sqrt{5}) \sim 1.62.$$

EXAMPLE 6-12.  $n = 2, S = 1$ .

$${}^{\perp}\mathfrak{g}'_2 = \frac{1}{f[x_1, x_{1-1}]} + \frac{1}{f[x_1, x_{1-2}]} - \frac{1}{f[x_{1-1}, x_{1-2}]},$$

$${}^{\perp}E_{2,1} = x_1 - f_1 {}^{\perp}\mathfrak{g}'_2.$$

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |y_3(\alpha)|^{\frac{1}{2}(p-1)}, \quad p \sim 1.84.$$

6.2-25

EXAMPLE 6-13.  $n = 1, s = 2.$

$${}^{\perp}v''_1 = \frac{2}{f'_1 - f'_{1-1}} \left[ \frac{2}{f'_1} + \frac{1}{f'_{1-1}} - \frac{3}{f[x_1, x_{1-1}]} \right],$$

$${}^{\perp}E_{1,2} = x_1 - u_1 + \frac{1}{2} f_1^2 {}^{\perp}v''_1.$$

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow |Y_4(\alpha)|^{\frac{1}{3}(p-1)}, \quad p = 1 + \sqrt{3} \sim 2.73.$$

### 6.3 Discussion of One-Point Iteration Functions With Memory

6.31 A conjecture. The theory of interpolatory I.F. and derivative estimated I.F. has been developed in Sections 6.1 and 6.2. In the case of interpolatory I.F., the first  $s - 1$  derivatives of  $f$  are evaluated. Hence  $s$  pieces of new information are required for each iteration. For the case of derivative estimated I.F., the  $(s-1)$ st derivative of an optimal one-point I.F. is estimated from the first  $s - 2$  derivatives at  $n + 1$  points. If we set  $S = s - 1$ , then  $S$  pieces of new information are used for each iteration. For interpolatory I.F.,  $r = s(n+1)$  represents the product of the number of new pieces of information per iteration with the number of points at which information is used;  $R = S(n+1)$  plays the corresponding role for derivative estimated I.F. A glance at Theorems 4-1, 4-3, 6-1, 6-2, 6-3, and 6-4, reveals a remarkable regularity in the structure of the results. The only parameters which enter are  $p$  and  $r$  or  $p$  and  $R$ ;  $p$  depends only on  $s$  and  $n$  or  $S$  and  $n$ . When we deal with inverse interpolation the asymptotic error constant depends on  $Y_r$ ; when we deal with direct interpolation the asymptotic error constant depends on  $A_r$ . From Theorems 6-2 and 6-4 we can conclude that the asymptotic behavior of a sequence generated by a derivative estimated I.F. is independent of the optimal one-point I.F. in which the derivative is estimated.

### 6.3-2

Values of  $p$  for different values of  $n$  and  $s$  or  $n$  and  $S$  may be found in Table 3-1.

The effect of estimating the highest derivative of optimal one-point I.F. has been studied. The idea of estimating the two highest derivatives suggests itself. Calculations of orders of such methods indicate that such a procedure would not be profitable.

We have proved, for the case of interpolatory and derivative estimated I.F., that the old information adds less than unity to the order. We conjecture that this is true no matter how the old information is reused.

CONJECTURE. Let

$$\varphi = \varphi[x_1; x_{1-1}, x_{1-2}, \dots, x_{1-n}]$$

$$= G[x_1, f_1, f'_1, \dots, f_1^{(\ell-1)}; x_{1-1}, f_{1-1}, f'_{1-1}, \dots,$$

$$f_{1-1}^{(\ell-1)}, \dots, x_{1-n}, f_{1-n}, f'_{1-n}, \dots, f_{1-n}^{(\ell-1)}]$$

be any one-point I.F. with memory. (The semicolon shows that new information is used only at  $x_1$ .) Let  $\varphi \in \ell^I p$ . Then  $p < \ell + 1$ .

### 6.3-3

In particular, we conjecture that it is impossible to construct a one-point I.F. with memory which is of second order and which does not require the evaluation of derivatives. The fact that new information is used at only one point is critical; in Section 8.6 we give I.F. which are of order greater than two and which do not require the evaluation of any derivatives. Those I.F. are, however, multipoint I.F. with memory.

It must be emphasized that the conjecture does not apply to the case where the sequence of approximants generated by an I.F. is "milked" for more information. See Appendix D for a discussion of acceleration of convergence.

## 6.3-4

6.32 Practical considerations. One-point I.F. with memory typically contain terms which approach o/o as the approximants approach  $\alpha$ . Consider, for instance, the I.F. of Example 6-9:

$$*f_1'' = \frac{2}{x_i - x_{i-1}} \left\{ 2f'_i + f'_{i-1} - 3f[x_i, x_{i-1}] \right\}, \quad (6-29)$$

$$*E_{1,2} = x_i - u_i - \frac{1}{2} \frac{u_i^2}{f'_i} *f_1''.$$

In theory,  $*f_1'' \rightarrow f''(\alpha)$ . In practice, it approaches a o/o form which naturally poses computational difficulties. Note that  $*f_1''$  is multiplied by  $u_i^2$  which goes to zero quite rapidly. Note also that the last term of (6-29) may be regarded as a correction term to  $x_i - u_i$ . Hence  $*f_1''$  need not be known too accurately. It might be worthwhile to do at least part of the computation with multiple precision arithmetic; this matter has not been fully explored.

In order to use an I.F. such as  $\varphi_{n,s}$ ,  $n+1$  approximants to  $\alpha$  must be available. This suggests using  $\varphi_{1,s}$ , which requires but two approximants, followed successively by  $\varphi_{2,s}, \varphi_{3,s}, \dots, \varphi_{n,s}$ , at the beginning of a calculation.

## 6.3-5

6.33 Iteration functions which do not use all available information. All the I.F. studied in this chapter use all the old information available at  $n$  points. It is possible to construct I.F. which use only part of the old information available at  $n$  points. This may lead to simpler I.F. but ones which are not of as high an order. For example, the simplest estimate of  $f''$  is

$${}^t f''_1 = \frac{f'_1 - f'_{1-1}}{x_1 - x_{1-1}}.$$

Define  ${}^t E_{1,2}$  by

$${}^t E_{1,2} = x_1 - u_1 - \frac{1}{2} \frac{u_1^2}{f'_1} {}^t f''_1. \quad (6-30)$$

It may be shown that the indicial equation associated with this I.F. is  $t^2 - 2t - 1 = 0$ , with roots  $1 \pm \sqrt{2}$ . Thus the order of  ${}^t E_{1,2}$  is 2.41. On the other hand,  $*E_{1,2}$ , which estimates  $f''$  from  $f_1, f_{1-1}, f'_1, f'_{1-1}$ , is of order 2.73. If the evaluation of  $f$  and  $f'$  is expensive, it is preferable to use the slightly more complicated  $*E_{1,2}$ .

### 6.3-6

6.34 An additional term in the error equation. In Section 5.5 a difference equation was derived which permitted the recursive calculation of the coefficients of the error series of  $E_s$ . Generalization to I.F. with memory has not been attempted. The first two terms of the error series have been worked out for a number of I.F. and are given below. The leading term of these I.F. is, of course, given by Theorem 6-1.

$$*E_{1,1} - \alpha \approx A_2(\alpha)e_1e_{i-1} - [A_2^2(\alpha) - A_3(\alpha)]e_1e_{i-1}^2, \text{ secant I.F.,} \quad (6-31)$$

$$*E_{2,1} - \alpha \approx -A_3(\alpha)e_1e_{i-1}e_{i-2} + A_2(\alpha)e_1^2, \quad (6-32)$$

$$*E_{1,2} - \alpha \approx A_4(\alpha)e_1^2e_{i-1}^2 + [2A_2^2(\alpha) - A_3(\alpha)]e_1^3. \quad (6-33)$$

## 7.0-1

### CHAPTER 7

#### MULTIPLE ROOTS

In Section 7.2 we show that all  $E_s$  are of linear order for nonsimple zero. In Section 7.3 we study an optimal basic sequence  $\{e_s\}$  whose order is multiplicity-independent. Our results on I.F. generated by direct interpolation are extended to the case of multiple roots in Section 7.5. Theorem 7-6 gives a necessary and sufficient condition for an I.F. to be of second order for roots of arbitrary multiplicity. An I.F. of incommensurate order is studied in Section 7.8.

## 7.1-1

### 7.1 Introduction

A number of definitions which relate order to multiplicity are given in Section 1.23.

Three functionals are of particular interest:

$$u = \frac{f}{f'}, \quad G = f^{(m-1)}, \quad F = f^{1/m}. \quad (7-1)$$

Observe that  $u$  has only simple zeros;  $G$  and  $F$  have only simple zeros provided that the zero of  $f$  has multiplicity  $m$ . The multiplicity must be known a priori if  $G$  or  $F$  are to be used. If  $f$  and its derivatives are replaced by  $u$ ,  $G$ , or  $F$  and their derivatives in any I.F., then the entire theory which pertains to simple zeros may be applied. If, for example, we replace  $f$  by  $u$  in Newton's I.F., then we generate

$$\phi = x - \frac{u}{u'}$$

which is second order for zeros of all multiplicities and

$$\frac{\phi-\alpha}{(x-\alpha)^2} \rightarrow \frac{u''(\alpha)}{2u'(\alpha)}.$$

It is well known that Newton's I.F. is of linear order for all nonsimple roots. E. Schröder [7.1-1, p. 324] points out that

$$\phi = x - mu \quad (7-2)$$

7.1-2

is of second order for zeros of multiplicity  $m$ ; this fact has been often rediscovered. In Section 7.2 we will prove that all  $E_s$  are of linear order for nonsimple zeros. In Section 7.3 we will construct an optimal basic sequence of which (7-2) is the first member.

## 7.2-1

7.2 The Order of  $E_s$ 

$E_s$  was derived from the Taylor series expansion of  $\tilde{f}$  under the assumption that  $f$  has only simple zeros. We can, nevertheless, inquire as to the behavior of  $E_s$  when applied to the calculation of multiple zeros. We prove

THEOREM 7-1. The order of  $E_s$  is linear for all nonsimple zeros. Moreover,

$$\frac{E_{s+1}-\alpha}{x-\alpha} \rightarrow \frac{(-1)^s}{s!m^s} \prod_{\ell=1}^s (1-\ell m). \quad (7-3)$$

PROOF. From (5-16) and (5-17),

$$E_{s+1}(x) = x - \sum_{j=1}^s z_j(x), \quad z_j(x) = Y_j(x) u^j(x).$$

Hence

$$E_{s+1}(x) - \alpha = x - \alpha - \sum_{j=1}^s z_j(x).$$

Define  $\gamma_{\ell,j}$  by

$$z_j(x) = \sum_{\ell=1}^{\infty} \gamma_{\ell,j} (x-\alpha)^\ell. \quad (7-4)$$

7.2-2

Hence

$$E_{s+1}(x) - \alpha = \left[ 1 - \sum_{j=1}^s \gamma_{1,j} \right] (x-\alpha) + \underline{O}[(x-\alpha)^2]. \quad (7-5)$$

Since  $Z_j(x)$  satisfies the difference - differential equation

$$jZ_j(x) - (j-1)Z_{j-1}(x) + u(x)Z'_{j-1}(x) = 0, \quad Z_1(x) = u(x), \quad (7-6)$$

we find on substituting (7-4) into (7-6) and noting

$u(x) = (x-\alpha)/m + \underline{O}[(x-\alpha)^2]$ , that

$$j\gamma_{1,j} - (j-1)\gamma_{1,j-1} + \frac{1}{m}\gamma_{1,j-1} = 0, \quad \gamma_{1,1} = \frac{1}{m}. \quad (7-7)$$

Setting  $M = 1/m$  yields

$$\gamma_{1,j} = \left[ \frac{j-1-M}{j} \right] \gamma_{1,j-1}, \quad \gamma_{1,1} = M.$$

Hence

$$\gamma_{1,j} = (-1)^{j+1} C[M, j], \quad (7-8)$$

where  $C[M, j]$  denotes a binomial coefficient. Substituting (7-8) into (7-5) yields

$$E_{s+1}(x) - \alpha = (x-\alpha) \left[ \sum_{j=0}^s (-1)^j C[M, j] \right] + \underline{O}[(x-\alpha)^2] = (x-\alpha)(-1)^s C[M-1, s] + \underline{O}[(x-\alpha)^2],$$

7.2-3

where we have used the well-known identity

$$\sum_{j=0}^s (-1)^j C[M, j] = (-1)^s C[M-1, s].$$

Hence

$$E_{s+1}(x) - \alpha = (x-\alpha) \left[ \frac{(-1)^s}{s!} \prod_{\ell=1}^s \left( \frac{1}{m} - \ell \right) \right] + \underline{o}[(x-\alpha)^2].$$

Since  $m$  is a positive integer, the coefficient of  $(x-\alpha)$  is zero if and only if  $m = 1$ ; factoring out  $1/m$  from each term of the product completes the proof.

EXAMPLE 7-1.

$$E_2(x) - \alpha = \left( 1 - \frac{1}{m} \right) (x-\alpha) + \underline{o}[(x-\alpha)^2]$$

which is a well-known result.

The asymptotic error constant is given by the right side of (7-3). We have

COROLLARY. Let

$$G(m, s) = \frac{(-1)^s}{s! m^s} \prod_{\ell=1}^s (1-\ell m).$$

Then

$$G(m, s) < 1, \quad \lim_{m \rightarrow \infty} G(m, s) = 1.$$

PROOF. Since  $G(m, s)$  may be written as

$$G(m, s) = \prod_{\ell=1}^s \left(1 - \frac{1}{\ell^m}\right),$$

the result follows immediately.

If  $m > 1$  and if a sequence of approximants is formed by  $x_{i+1} = E_2(x_i)$ , then

$$x_{i+1} - \alpha = \left(1 - \frac{1}{m}\right) (x_i - \alpha) + \underline{o}[(x_i - \alpha)^2]. \quad (7-9)$$

Since the sequence converges linearly, we can apply Aitken's  $\delta^2$  formula (Appendix D) and estimate  $\alpha$  by

$$\alpha^* = x_{i+2} - \frac{(x_{i+2} - x_{i+1})^2}{x_{i+2} - 2x_{i+1} + x_i}.$$

Using  $\alpha^*$  as an estimate for  $\alpha$ ,

$$x_{i+1} - \alpha \approx \left(1 - \frac{1}{m}\right) (x_i - \alpha)$$

may be used to calculate an estimate for  $m$ . Note that  $m$  is an integer-valued variable. Once  $m$  is known, the second order I.F.,  $\varphi = x - mu$  may be used.

## 7.3-1

7.3 The Basic Sequence  $E_s$ 

7.31 Introduction. In Section 5.13, explicit formulas were derived for  $E_s$ ; these I.F. form an optimal basic sequence for simple zeros. In Section 7.2 we showed that the order of  $E_s$  is linear for nonsimple zeros and hence that  $\{E_s\}$  is not a basic sequence for  $m > 1$ . An optimal basic sequence is constructed below for the case where  $m$  is arbitrary but known.

Since the multiplicity of a zero is often not known a priori, the results are of limited value as far as practical problems are concerned. The study is, however, of considerable theoretical interest and leads to some surprising results. We will find that multiplication of the terms of  $E_s$  by certain polynomials in  $m$  leads to I.F. with the desired properties. These polynomials are found explicitly; their coefficients depend on Stirling numbers of the first and second kind.

One may derive these new I.F. by the following technique. Let  $\alpha$  be a zero of multiplicity  $m$ . Let

$$h(x) = f^{1/m}(x) = z.$$

Then  $h(x)$  has a simple zero at  $\alpha$ . Let  $H$  be the inverse to  $h$ . Then proceeding as in Section 5.11 it is clear that

$$\varphi_s = \sum_{j=0}^{s-1} \frac{(-1)^j}{j!} z^{J_H(j)}(z) \quad (7-10)$$

## 7.3-2

is of order  $s$  for all  $m$ . In particular,

$$\varphi_2 = H(z) - zH'(z) = x - \frac{f^{1/m}(x)}{h'(x)} = x - mu(x), \quad (7-11)$$

$$\varphi_3 = H(z) - zH'(z) + \frac{1}{2}z^2H''(z) = x - \frac{1}{2}m(3-m)u(x) - m^2A_2(x)u^2(x).$$

$\varphi_2$  was known to E. Schröder [7.3-1] (1870). See also Bodewig [7.3-2] and Ostrowski [7.3-3, Chap. 8].

Rather than using (7-10) to generate higher order I.F. of this type, we attack the problem from another point of view.

## 7.3-3

7.32 The structure of  $\mathcal{E}_s$ . It is advantageous to extend the notation for I.F. so that  $f$  and  $m$  appear explicitly as parameters. Thus we replace (5-17) by

$$E_{s+1}(x, f, l) = x - \sum_{j=1}^s z_j(x, f, l).$$

Let

$$F = f^{1/m}. \quad (7-12)$$

Observe that a zero of multiplicity  $m$  of  $f$  is a zero of multiplicity 1 of  $F$ . Clearly,  $\{E_{s+1}(x, F, l)\}$  is an optimal basic sequence for all  $m$ . (It is convenient to use  $E_{s+1}$  rather than  $E_s$  throughout this section.) Let

$$E_{s+1}(x, F, l) = \mathcal{E}_{s+1}(x, f, m),$$

$$z_j(x, F, l) = w_j(x, f, m).$$

Then (5-18) becomes

$$jw_j(x, f, m) - (j-1)w_{j-1}(x, f, m) + mu(x)w'_{j-1}(x, f, m) = 0, \quad (7-13)$$

$$w_1(x, f, m) = mu(x),$$

while Lemma 5-3 becomes

7.3-4

LEMMA 7-1.

$$e_{s+1}(x, f, m) = e_s(x, f, m) - \frac{mu(x)}{s} e'_s(x, f, m).$$

We seek coefficients  $\rho_{s,j}(m)$  such that

$$e_{s+1}(x, f, m) = x - \sum_{j=1}^s \rho_{s,j}(m) z_j(x, f, 1); \quad (7-14)$$

$$\rho_{s,j}(m) = 0, \quad s < j,$$

and such that  $\{e_{s+1}(x, f, m)\}$  is an optimal basic sequence.

Substituting (7-14) into the formula of Lemma 7-1 yields

$$x - \sum_{j=1}^s \rho_{s,j}(m) z_j(x, f, 1)$$

$$= x - \sum_{j=1}^{s-1} \rho_{s-1,j}(m) z_j(x, f, 1) - \frac{m}{s} u(x) \left[ 1 - \sum_{j=1}^{s-1} \rho_{s-1,j}(m) z'_j(x, f, 1) \right].$$

7.3-5

We use (5-18) to eliminate  $u(x)Z'_j(x, f, l)$  and find

$$\sum_{j=1}^s Z_j(x, f, l) [-s\rho_{s,j}(m) + s\rho_{s-1,j}(m) + mj\rho_{s-1,j-1}(m) - mj\rho_{s-1,j}(m)] = 0,$$

where we have taken  $\rho_{s,0}(m)$ , which may be considered as the coefficient of  $x$  in (7-14), equal to unity. Then,

$$s\rho_{s,j}(m) + (mj-s)\rho_{s-1,j}(m) - mj\rho_{s-1,j-1}(m) = 0, \quad (7-15)$$

with  $\rho_{s,0}(m) = 1$  for  $s \geq 0$  and  $\rho_{s,j}(m) = 0$ , for  $s < j$  as initial conditions. Equation (7-15) permits the recursive calculation of the  $\rho_{s,j}(m)$ . Equation (7-15) shows that  $\rho_{s,j}(m)$  is a polynomial in  $m$ ; an explicit formula is derived below.

Define the associated functions  $\sigma_{\ell,j}(m)$  by

$$w_\ell(x, f, m) = \sum_{j=1}^{\ell} \sigma_{\ell,j}(m) Z_j(x, f, l), \quad \ell < j. \quad (7-16)$$

The  $\sigma_{\ell,j}(m)$  were introduced by Zajta [7.3-4]. Since

$$\begin{aligned} e_{s+1}(x, f, m) &= x - \sum_{\ell=1}^s w_\ell(x, f, m) = x - \sum_{\ell=1}^s \sum_{j=1}^{\ell} \sigma_{\ell,j}(m) Z_j(x, f, l) \\ &= x - \sum_{j=1}^s Z_j(x, f, l) \left( \sum_{\ell=j}^s \sigma_{\ell,j}(m) \right), \end{aligned}$$

7.3-6

we have

$$\rho_{s,j}(m) = \sum_{\ell=j}^s \sigma_{\ell,j}(m). \quad (7-17)$$

To find a recursion formula for the  $\sigma_{\ell,j}(m)$  we substitute (7-16) into (7-13), and use (5-18) to eliminate  $u(x)z'_j(x,f,1)$ . Then

$$\sum_{j=1}^{\ell} z_j(x,f,1)[\ell\sigma_{\ell,j}(m) - (\ell-1)\sigma_{\ell-1,j}(m) - mj\sigma_{\ell-1,j-1}(m) + mj\sigma_{\ell-1,j}(m)] = 0.$$

Hence

$$(\ell+1)\sigma_{\ell+1,j}(m) + (mj-\ell)\sigma_{\ell,j}(m) - mj\sigma_{\ell,j-1}(m) = 0 \quad (7-18)$$

with  $\sigma_{0,0}(m) = 1$ ,  $\sigma_{\ell,0}(m) = 0$  for  $\ell > 0$ , and  $\sigma_{\ell,j}(m) = 0$  for  $\ell < j$ , as initial conditions. Equation (7-18) shows that  $\sigma_{\ell,j}$  is a polynomial in  $m$ .

We digress briefly to list some definitions from the Calculus of Finite Differences. The reader is referred to Jordan [7.3-5] or Riordan [7.3-6] for the standard theory. Our notation is not quite standard. We define:

## 7.3-7

$$[x]_\ell = \prod_{i=0}^{\ell-1} (x-i) \quad \text{"Falling Factorial"}$$

$$r[x]_\ell = \prod_{i=0}^{\ell-1} (x+i) \quad \text{"Rising Factorial"}$$

$$c[x, \ell] = [x]_\ell / \ell! \quad \text{"Falling Binomial Coefficient"}$$

$$c_r[x, \ell] = r[x]_\ell / \ell! \quad \text{"Rising Factorial Coefficient"}$$

$$[x]_\ell = \sum_{j=0}^{\ell} s_{\ell, j} x^j \quad \text{"Stirling Numbers of the First Kind"}$$

$$x^\ell = \sum_{j=0}^{\ell} t_{\ell, j} [x]_j \quad \text{"Stirling Numbers of the Second Kind"}$$

Stirling numbers of the first and second kind are often denoted by two different types of  $s$ ; for example, by  $s$  and  $S$ . We will use  $S$  and  $T$ . There is no danger of confusing the latter with the usual symbol for a Chebyshev polynomial.

We continue our study of  $\rho_{s, j}(m)$  and  $\sigma_{\ell, j}(m)$ . A generating function for the  $\sigma_{\ell, j}(m)$  may be derived by defining

$$h_j(x, m) = \sum_{\ell=0}^{\infty} \sigma_{\ell, j}(m) x^\ell.$$

7.3-8

It follows from (7-18) that  $h_j(x, m)$  satisfies

$$(1-x)h'_j(x, m) + mh_j(x, m) - mh_{j-1}(x, m) = 0, \quad (7-19)$$

where  $m$  is a parameter. A solution of (7-19) is

$$h_j(x, m) = [1 - (1-x)^m]^j.$$

It is not difficult to verify that the functions  $\kappa_{\ell, j}(m)$  which satisfy

$$[1 - (1-x)^m]^j = \sum_{\ell=0}^{\infty} \kappa_{\ell, j}(m)x^\ell$$

also satisfy (7-18) and its initial conditions, and hence

$$[1 - (1-x)^m]^j = \sum_{\ell=0}^{\infty} \sigma_{\ell, j}(m)x^\ell.$$

Observing that

$$1 - (1-x)^m = \sum_{r=1}^m (-1)^{r+1} C[m, r] x^r$$

and applying the multinomial theorem yields

$$\sigma_{\ell, j}(m) = j!(-1)^{\ell+j} \sum_{r=1}^{\ell} \prod_{r=1}^{\ell} \frac{(C[m, r])^{\alpha_r}}{\alpha_r!}, \quad (7-20)$$

with the sum taken over all nonnegative integers  $\alpha_r$  such that

$$\sum_{r=1}^{\ell} r\alpha_r = \ell, \quad \sum_{r=1}^{\ell} \alpha_r = j.$$

On the other hand,

$$\begin{aligned} [1 - (1-x)^m]^j &= \sum_{r=0}^j C[j, r](-1)^r(1-x)^{rm} \\ &= \sum_{\ell=j}^{mj} x^\ell \sum_{r=0}^j (-1)^r C[j, r](-1)^\ell C[rm, \ell]. \end{aligned}$$

Thus

$$\sigma_{\ell, j}(m) = \sum_{r=0}^j (-1)^r C[j, r](-1)^\ell C[rm, \ell]. \quad (7-21)$$

Since  $C[rm, \ell]$  is a polynomial in  $m$  of degree  $\ell$ , (7-21) exhibits an expansion of  $\sigma_{\ell, j}(m)$  in the polynomials  $C[rm, \ell]$ . Since

$$\begin{aligned} C[rm, \ell] &= \frac{1}{\ell!} \sum_{k=0}^{\ell} s_{\ell, k} r^k m^k, \\ \sigma_{\ell, j}(m) &= \sum_{r=0}^j (-1)^r C[j, r] \frac{(-1)^\ell}{\ell!} \sum_{k=0}^{\ell} s_{\ell, k} r^k m^k \\ &= \frac{(-1)^\ell}{\ell!} \sum_{k=0}^{\ell} s_{\ell, k} m^k \sum_{r=0}^j (-1)^r C[j, r] r^k. \end{aligned}$$

Using

$$T_{k,j} = \frac{(-1)^j}{j!} \sum_{r=0}^j (-1)^r c[j,r] r^k$$

yields

$$\sigma_{\ell,j}(m) = (-1)^{\ell+j} \frac{j!}{\ell!} \sum_{k=j}^{\ell} s_{\ell,k} T_{k,j} m^k. \quad (7-22)$$

Equation (7-22) exhibits  $\sigma_{\ell,j}(m)$  as a polynomial in  $m$ . It is not difficult to show that

$$c_r[mx, \ell] = \sum_{j=0}^{\ell} c_r[x, j] (-1)^{j+\ell} \frac{j!}{\ell!} \sum_{k=j}^{\ell} s_{\ell,k} T_{k,j} m^k$$

and hence that

$$c_r[mx, \ell] = \sum_{j=0}^{\ell} \sigma_{\ell,j}(m) c_r[x, j].$$

Thus  $c_r[mx, \ell]$  is a generating function for the  $\sigma_{\ell,j}(m)$  relative to the base functions  $c_r[x, j]$ . Since  $c_r[x, \ell] = c[x+\ell-1, \ell]$  we have equivalently that

$$c[mx+\ell-1, \ell] = \sum_{j=0}^{\ell} \sigma_{\ell,j}(m) c[x+j-1, j].$$

7.3-11

We now derive an explicit formula for the  $\rho_{s,j}(m)$ . Recalling that

$$\rho_{s,j}(m) = \sum_{\ell=j}^s \sigma_{\ell,j}(m),$$

and using (7-22) yields

$$\begin{aligned}\rho_{s,j}(m) &= \sum_{\ell=j}^s (-1)^{\ell+j} \frac{j!}{\ell!} \sum_{k=j}^{\ell} s_{\ell,k} t_{k,j} m^k \\ &= (-1)^j j! \sum_{k=j}^s t_{k,j} m^k \sum_{\ell=k}^s \frac{(-1)^\ell}{\ell!} s_{\ell,k}.\end{aligned}$$

Since

$$\sum_{\ell=k}^s \frac{(-1)^\ell}{\ell!} s_{\ell,k} = \frac{(-1)^s}{s!} s_{s+1,k+1},$$

$$\rho_{s,j}(m) = (-1)^{s+j} \frac{j!}{s!} \sum_{k=j}^s s_{s+1,k+1} t_{k,j} m^k.$$

Our results are summarized in

7.3-12

THEOREM 7-2. Define  $\rho_{s,j}(m)$  by

$$E_{s+1}(x, f, m) = x - \sum_{j=1}^s \rho_{s,j}(m) Z_j(x, f, 1); \quad \rho_{s,j}(m) = 0, \quad \text{for } s < j.$$

Define  $\sigma_{\ell,j}(m)$  by

$$W_\ell(x, f, m) = \sum_{j=1}^{\ell} \sigma_{\ell,j}(m) Z_j(x, f, 1); \quad \sigma_{\ell,j}(m) = 0, \quad \text{for } \ell < j.$$

Then

$$\rho_{s,j}(m) = \sum_{\ell=j}^s \sigma_{\ell,j}(m), \quad (7-23)$$

$$[1 - (1-x)^m]^j = \sum_{\ell=0}^{\infty} \sigma_{\ell,j}(m) x^\ell, \quad (7-24)$$

$$\sigma_{\ell,j}(m) = \sum_{r=0}^j (-1)^r C[j,r] (-1)^\ell C[r,m,\ell], \quad (7-25)$$

$$\sigma_{\ell,j}(m) = (-1)^{\ell+j} \frac{j!}{\ell!} \sum_{k=j}^{\ell} S_{\ell,k} T_{k,j} m^k, \quad (7-26)$$

$$C_r[mx, \ell] = \sum_{j=0}^{\ell} \sigma_{\ell,j}(m) C_r[x, j], \quad (7-27)$$

$$\rho_{s,j}(m) = (-1)^{j+s} \frac{j!}{s!} \sum_{k=j}^s S_{s+1,k+1} T_{k,j} m^k. \quad (7-28)$$

7.3-13

The following corollaries follow from the various parts of Theorem 7-2. This is not a complete list of the properties of  $\sigma_{\ell,j}(m)$  and  $\rho_{\ell,j}(m)$ .

COROLLARY a.  $\sigma_{\ell,j}(m)$  is a polynomial in  $m$  of degree  $\ell$ .

PROOF. This follows from (7-26).

COROLLARY b.  $\sigma_{\ell,\ell}(m) = m^\ell$ .

PROOF. This follows from (7-26), since

$$T_{\ell,\ell} = S_{\ell,\ell} = 1.$$

COROLLARY c.  $\sigma_{\ell,1}(m) = (-1)^{\ell+1} C[m, \ell]$ .

PROOF. This follows from (7-25).

COROLLARY d.  $\sigma_{\ell,j}(m) = 0$ , for  $\ell < j$  and  $\ell > mj$ .

PROOF. For  $\ell < j$ ,  $\sigma_{\ell,j}(m)$  was defined as zero.

For  $\ell > mj$ , this follows from (7-24).

7.3-14

COROLLARY e.

$$\sum_{\ell=j}^{mj} \sigma_{\ell,j}(m) = 1, \quad \sum_{\ell=0}^{\infty} \sigma_{\ell,j}(m) = 1.$$

PROOF. Set  $x = 1$  in (7-24) and use Corollary d.

COROLLARY f.  $\sigma_{\ell,j}(1) = \delta_{\ell,j}$  (Kronecker delta).

PROOF. Set  $m = 1$  in (7-24).

COROLLARY g.

$$\sum_{j=0}^{\ell} \sigma_{\ell,j}(m) = \sum_{j=0}^{\infty} \sigma_{\ell,j}(m) = C[m+\ell-1, \ell].$$

PROOF. Set  $x = 1$  in (7-27).

COROLLARY h. The leading coefficient of  $\sigma_{\ell,j}(m)$  is

$$\frac{(-1)^{\ell}}{\ell!} \sum_{i=0}^j (-1)^i C[j,i] i^{\ell}.$$

PROOF. This follows from (7-26) and noting that

$$T_{\ell,j} = \frac{(-1)^j}{j!} \sum_{i=0}^j (-1)^i C[j,i] i^{\ell}.$$

7.3-15

COROLLARY i.  $\rho_{s,j}(m)$  is a polynomial in  $m$  of degree  $s$ .

PROOF. This follows from (7-28).

COROLLARY j.  $\rho_{s,s}(m) = m^s$ .

PROOF. This follows from (7-28).

COROLLARY k.  $\rho_{s,1}(m) = 1 + (-1)^{s+1} C[m-1, s]$ .

PROOF. This follows from (7-23) and Corollary c.

COROLLARY l. The leading coefficient of  $\rho_{s,j}(m)$  is given by

$$\frac{(-1)^s}{s!} \sum_{i=0}^j (-1)^i C[j,i] i^s.$$

PROOF. This follows from (7-28) and

$$T_{s,j} = \frac{(-1)^j}{j!} \sum_{i=0}^j (-1)^i C[j,i] i^s.$$

7.3-16

COROLLARY m.  $\rho_{s,j}(m) = 1$  for every  $m$  such that  $s \geq mj$ .

PROOF. This follows from (7-23) and Corollaries d  
and e.

COROLLARY n.  $\rho_{s,j}(1) = 1$ .

PROOF. This follows from Corollary (7-28).

COROLLARY o.

$$\sum_{j=1}^s \rho_{s,j}(m)(-1)^{j+1} C[1/m, j] = 1, \quad s = 1, 2, 3, \dots .$$

PROOF.

$$e_{s+1}(x, f, m) = x - \sum_{j=1}^s \rho_{s,j}(m) Z_j(x, f, 1)$$

7.3-17

is to hold for arbitrary  $f$ . Take  $f(x) = x^m$ . Then  $F(x) = f^{1/m}(x) = x$  and hence  $E_{s+1}(x, x, 1) = \varepsilon_{s+1}(x, x^m, m) = 0$ , for  $s = 1, 2, 3, \dots$ . It is not difficult to show that  $Z_j[x, x^m, 1] = (-1)^{j+1} C[1/m, j] x$ . Thus

$$0 = x - \sum_{j=1}^s \rho_{s,j}(m)(-1)^{j+1} C[1/m, j] x,$$

and the result follows.

Corollary n shows that for  $m = 1$ ,

$$\varepsilon_{s+1}(x, f, m) = x - \sum_{j=1}^s \rho_{s,j}(m) Z_j(x, f, 1)$$

reduces as expected to

$$E_{s+1}(x) = x - \sum_{j=1}^s Z_j(x, f, 1).$$

Thus  $\varepsilon_{s+1}(x, f, 1) = E_{s+1}(x)$ .

Corollary e leads to an interesting result. From (7-14) and (7-23),

$$\varepsilon_{s+1}(x, f, m) = x - \sum_{j=1}^s \sum_{\ell=j}^s \sigma_{\ell,j}(m) Z_j(x, f, 1).$$

7.3-18

Then

$$\begin{aligned} \lim_{s \rightarrow \infty} E_{s+1}(x, f, m) &= x - \sum_{j=1}^{\infty} \sum_{\ell=j}^{\infty} \sigma_{\ell, j}(m) Z_j(x, f, l) \\ &= x - \sum_{j=1}^{\infty} Z_j(x, f, l) = \lim_{s \rightarrow \infty} E_s(x, f, l). \end{aligned}$$

7.3-19

7.33 Formulas for  $\epsilon_s$ . Table 7-1 lists some of the  $\sigma_{\ell,j}(m)$  expressed in terms of  $C[rm,\ell]$ , while Table 7-2 lists some of the  $\sigma_{\ell,j}(m)$  expressed in powers of  $m$ . Table 7-3 lists some of the  $\rho_{s,j}(m)$ . Using Table 7-3, recalling that  $Z_j(x) = Y_j(x)u^j(x)$ , and using the expressions for  $Y_j(x)$  given in Table 5-1, enables us to calculate a number of the  $\epsilon_{s+1}(x,f,m)$ :

$$\epsilon_2 = x - mu(x),$$

$$\epsilon_3 = x - mu(x) \left[ \frac{1}{2} (3-m) + mA_2(x)u(x) \right],$$

$$\epsilon_4 = x - mu(x) \left\{ \frac{1}{6} (m^2 - 6m + 11) + m(2-m)A_2(x)u(x) + m^2 \left[ 2A_2^2(x) - A_3(x) \right] u^2(x) \right\},$$

$$\epsilon_5 = x - mu(x) \left\{ -\frac{1}{24} (m^3 - 10m^2 + 35m - 50) + \frac{1}{12} m(7m^2 - 30m + 35)A_2(x)u(x) \right.$$

$$\left. + \frac{1}{2} m^2(5-3m) \left[ 2A_2^2(x) - A_3(x) \right] u^2(x) \right\}$$

$$+ m^3 \left[ 5A_2^3(x) - 5A_2(x)A_3(x) + A_4(x) \right] u^3(x) \right\}.$$

TABLE 7-1.  $\sigma_{\ell,j}(\mathbf{m}) = \sum_{r=0}^j (-1)^r c[j,r](-1)^\ell c[\mathbf{rm},\ell].$

$\downarrow \ell$	$\uparrow j$	1	2	3	4
1	$c[m,1]$				
2	$-c[m,2]$	$-2c[m,2] + c[2m,2]$			
3	$c[m,3]$	$2c[m,3] - c[2m,3]$	$3c[m,3] - 3c[2m,3] + c[3m,3]$		
4	$-c[m,4]$	$-2c[m,4] + c[2m,4]$	$-3c[m,4] + 3c[2m,4] - c[3m,4]$	$-4c[m,4] + 6c[2m,4] - 4c[3m,4] + c[4m,4]$	

7.3-21

TABLE 7-2.  $\sigma_{\ell,j}(m) = (-1)^{\ell+j} \frac{j!}{\ell!} \sum_{k=j}^{\ell} S_{\ell,k} T_{k,j} m^k$ .

$\downarrow \ell$	$\uparrow j$	1	2	3	4
1		$m$			
2			$- (m/2)(m-1)$	$m^2$	
3			$(m/6)(m-1)(m-2)$	$- m^2(m-1)$	$m^3$
4		$- (m/24)(m-1)(m-2)(m-3)$	$(m^2/12)(m-1)(m-11)$	$- (3/2)m^3(m-1)$	$m^4$

$$\text{TABLE 7-3. } \rho_{s,j}(m) = (-1)^{s+j} \frac{j!}{s!} \sum_{k=j}^s s_{s+1,k+1} T_{k,j} m^k.$$

	$\overset{\rightarrow}{j}$	$l^{\dagger}$	2	3	4
$\downarrow s$					
1		$m$			
2		$-(m/2)(m-3)$	$m^2$		
3		$(m/6)(m^2-6m+11)$	$-m^2(m-2)$	$m^3$	
4		$-(m/24)(m^3-10m^2+35m-50)$	$(m^2/12)(7m^2-30m+35)$	$-(m^3/2)(3m-5)$	$m^4$

<sup>†</sup>From Corollary k, the expressions in this column can be replaced by

$$\rho_{s,1}(m) = 1 + \frac{(-1)^{s+1}}{s!} \prod_{i=1}^s (m-i).$$

## 7.4-1

7.4 The Coefficients of the Error Series of  $\mathcal{E}_S$ 

The error series for  $E_S(x, f, 1)$  was studied in Section 5.5. We now derive an algorithm for calculating the coefficients of the error series of  $\mathcal{E}_S(x, f, m)$ .

Recall the definitions of the following symbols which will be used frequently:

$$u(x) = \frac{f(x)}{f'(x)}, \quad a_j(x) = \frac{f^{(j)}(x)}{j!}, \quad A_j(x) = \frac{a_j(x)}{a_1(x)}$$

$$B_{j,m}(x) = \frac{a_{j+m-1}(x)}{m a_m(x)}, \quad e = x - a.$$

Observe that  $B_{j,1}(x) \equiv A_j(x)$ .

Let  $a$  be a zero of multiplicity  $m$ . The expansion of  $u(x)$  into a power series in  $e$ , which will be needed below, is derived now. Throughout this section, all functions are evaluated at  $a$  unless otherwise indicated. Define  $\omega_\ell(m)$  by

$$mu(x) = \sum_{\ell=1}^{\infty} \omega_\ell(m) e^\ell. \quad (7-29)$$

Since

$$f(x) = \sum_{r=m}^{\infty} a_r e^r, \quad f'(x) = \sum_{r=m}^{\infty} r a_r e^{r-1},$$

and  $m\mu(x)f'(x) \equiv mf(x)$ , we obtain

$$ma_\ell = \sum_{q=1}^{\ell+1-m} \omega_q(m)(\ell+1-q)a_{\ell+1-q}, \quad \ell = m, m+1, \dots$$

or

$$\omega_\ell = mB_{\ell,m} - \sum_{q=1}^{\ell-1} \omega_q(m)(\ell+m-q)B_{\ell+1-q,m}, \quad \ell = 1, 2, \dots, \quad (7-30)$$

Since  $B_{\ell,1} = A_\ell$ , we observe that (7-30) reduces to (5-34) when  $m = 1$ , and hence that  $\omega_\ell(1) = v_\ell$  as expected.

It is not difficult to prove that an explicit formula for  $\omega_\ell(m)$  is given by

$$\omega_\ell(m) = mB_{\ell,m} + m \sum_{j=1}^{\ell-1} B_{\ell-j,m} \sum (-1)^r r! \prod_{i=1}^j \frac{[(m+i)B_{i+1,m}]^{\alpha_i}}{\alpha_i!}, \quad (7-31)$$

where  $r = \sum_{i=1}^j \alpha_i$  and where the inner sum is taken over all nonnegative integers  $\alpha_i$  such that  $\sum_{i=1}^j i\alpha_i = j$ . Observe that  $\omega_\ell(m)$  is the same function of the  $B_{i,m}$  as  $v_\ell$  is of the  $A_i$  except for coefficients which depend on  $m$  only. From either (7-30) or (7-31), the first few  $\omega_\ell(m)$  may be calculated as

7.4-3

$$\omega_1(m) = 1,$$

$$\omega_2(m) = -B_{2,m},$$

$$\omega_3(m) = (m+1)B_{2,m}^2 - 2B_{3,m},$$

$$\omega_4(m) = -(m+1)^2 B_{2,m}^3 + (3m+4)B_{2,m}B_{3,m} - 3B_{4,m}.$$

We turn to the problem of finding the coefficients of the error series. Define  $\lambda_{\ell,s}(m)$  by

$$e_s(x,f,m) = \sum_{\ell=0}^{\infty} \lambda_{\ell,s}(m) e^{\ell}. \quad (7-32)$$

Since  $e_s(x,f,m) \in I_s$ , we expect  $\tau_{\ell,s} = 0$  for  $0 < \ell < s$ , and  $\tau_{0,s} = \alpha$ . This may be proven directly by induction on  $s$ .

Let  $s = 1$ . Then

$$e_1(x,f,m) = x = \alpha + (x-\alpha) = \alpha + e.$$

Now assume  $\lambda_{0,s}(m) = \alpha$  and  $\lambda_{\ell,s}(m) = 0$ , for  $0 < \ell < s$ . Substitute (7-32) into the formula of Lemma 7-1,

$$e_{s+1}(x,f,m) = e_s(x,f,m) - \frac{mu(x)}{s} e'_s(x,f,m), \quad (7-33)$$

to find

$$\sum_{\ell=0}^{\infty} \lambda_{\ell, s+1}(m) e^{\ell} = \alpha + \sum_{\ell=s}^{\infty} \lambda_{\ell, s}(m) e^{\ell} - \frac{mu(x)}{s} \sum_{\ell=s}^{\infty} \ell \lambda_{\ell, s}(m) e^{\ell-1} = \alpha + O(e^{s+1}),$$

which completes the induction.

Substituting (7-32) into (7-33), using

$$mu(x) = \sum_{\ell=1}^{\infty} \omega_{\ell}(m) e^{\ell},$$

and equating to zero the coefficient of  $e^{\ell}$ , we arrive at

**THEOREM 7-3.** Let the multiplicity of  $\alpha$  be  $m$ . Let

$$s_s(x, f, m) = \sum_{\ell=0}^{\infty} \lambda_{\ell, s}(m) e^{\ell}, \quad mu(x) = \sum_{\ell=1}^{\infty} \omega_{\ell}(m) e^{\ell}.$$

Then

$$s \lambda_{\ell, s+1}(m) + (\ell-s) \lambda_{\ell, s}(m) + \sum_{r=1}^{\ell-1} r \omega_{\ell+1-r}(m) \lambda_{r, s} = 0, \quad (7-34)$$

with  $\lambda_{0, s}(m) = \alpha$ ,  $\lambda_{1, 1}(m) = 1$ ,  $\lambda_{\ell, 1}(m) = 0$  for  $\ell > 1$ , and  
 $\lambda_{\ell, s}(m) = 0$  for  $0 < \ell < s$  and  $s > 1$ .

TABLE 7-4.  $\lambda_{\ell,s}(m)$ 

$\downarrow \ell$	$\rightarrow s$	1	2	3	4
0	$\alpha$			$\alpha$	$\alpha$
1	1				
2			$B_{2,m}$		
3				$-\frac{1}{2}(m+3)B_{2,m}^2 - B_{3,m}$	$\left(-\frac{1}{2}(m+1)(2m+7)B_{2,m}^3 + \left(\frac{1}{3}(m^2+6m+8)B_{2,m}^3\right.\right.$
4				$-(m+1)B_{2,m}^2 + 2B_{3,m}$	$\left.\left.+ 3(m+3)B_{2,m}B_{3,m} - 3B_{4,m}\right)\right) - (m+4)B_{2,m}B_{3,m} + B_{4,m}$

7.4-6

Since the  $\omega_\ell(m)$  may be considered as known, (7-34) can be used to determine the  $\lambda_{\ell,s}(m)$ . Note that  $\lambda_{\ell,s}(m)$  is the same function of  $\omega_\ell(m)$  as  $\tau_{\ell,s}$  is of  $v_\ell$ . Some of the  $\lambda_{\ell,s}(m)$  are listed in Table 7-4. Using this table we find

$$E_2(x, f, m) - \alpha = B_{2,m} e^2 + \left[ - (m+1)B_{2,m}^2 + 2B_{3,m} \right] e^3$$

$$+ \left[ (m+1)^2 B_{2,m}^3 - (3m+4)B_{2,m} B_{3,m} + 3B_{4,m} \right] e^4,$$

$$E_3(x, f, m) - \alpha = \left[ \frac{1}{2}(m+3)B_{2,m}^2 - B_{3,m} \right] e^3$$

$$+ \left[ - \frac{1}{2}(m+1)(2m+7)B_{2,m}^3 + 3(m+3)B_{2,m} B_{3,m} - 3B_{4,m} \right] e^4,$$

$$E_4(x, f, m) - \alpha = \left[ \frac{1}{3}(m^2+6m+8)B_{2,m}^3 - (m+4)B_{2,m} B_{3,m} + B_{4,m} \right] e^4.$$

### 7.5-1

#### 7.5 Iteration Functions Generated By Direct Interpolation

In this section we generalize Theorem 4-3 to the case of multiple roots. We have to assume stronger conditions than in the case of simple roots in order to carry through our analysis.

## 7.5-2

7.51 The error equation. Let  $x_1, x_{1-1}, \dots, x_{1-n}$  be  $n+1$  approximants to a zero  $\alpha$  of multiplicity  $m$ . Let  $P_{n,s}$  be the polynomial whose first  $s-1$  derivatives are equal to the first  $s-1$  derivatives of  $f$  at  $x_1, x_{1-1}, \dots, x_{1-n}$ . Define a new approximant to  $\alpha$  by

$$P_{n,s}(x_{1+1}) = 0. \quad (7-35)$$

Then repeat this procedure using the points  $x_{1+1}, x_1, \dots, x_{1-n+1}$ . We assume that we can find a real root of  $P_{n,s}$  which satisfies (7-35). If  $P_{n,s}$  has a number of real roots, one of them is chosen as  $x_{1+1}$  by some criteria.

We have

$$f(t) = P_{n,s}(t) + \frac{f^{(r)}[\xi_1(t)]}{r!} \prod_{j=0}^n (t-x_{1-j})^s,$$

where  $\xi_1(t)$  lies in the interval determined by  $x_1, x_{1-1}, \dots, x_{1-n}, t$ , and where  $r = s(n+1)$ . Set  $t = x_{1+1}$ . Then

$$f(x_{1+1}) = \frac{f^{(r)}[\xi_{1+1}]}{r!} \prod_{j=0}^n (x_{1+1}-x_{1-j})^s,$$

7.5-3

where  $\xi_{i+1} \equiv \xi_i(x_{i+1})$ . Since the multiplicity of  $\alpha$  is  $m$ ,

$$f(x_{i+1}) = \frac{f^{(m)}(\eta_{i+1})}{m!} (x_{i+1} - \alpha)^m,$$

where  $\eta_{i+1}$  lies in the interval determined by  $x_{i+1}$  and  $\alpha$ .

Letting  $e_{i-j} = x_{i-j} - \alpha$ , we arrive at

$$e_{i+1}^m = (-1)^r \frac{m!}{r!} \frac{f^{(r)}(\xi_{i+1})}{f^{(m)}(\eta_{i+1})} \prod_{j=0}^n (e_{i-j} - e_{i+1})^s. \quad (7-36)$$

Equation (7-36) is the error equation for the case of multiple roots. We assume that

$$m < r = s(n+1). \quad (7-37)$$

Before analyzing this error equation we investigate the roots of an indicial equation.

## 7.5-4

7.52 On the roots of an indicial equation. In Section 3.3 we investigated the properties of the roots of the polynomial equation

$$g_{k,a}(t) = t^k - a \sum_{j=0}^{k-1} t^j = 0, \quad (7-38)$$

under the assumption that for  $k > 1$ ,

$$ka > 1. \quad (7-39)$$

The order of the I.F. which are being studied in Section 7.5 is determined by the roots of the equation

$$mt^{n+1} - s \sum_{j=0}^n t^j = 0, \quad (7-40)$$

which is of the form (7-38) with

$$k = n + 1, \quad a = \frac{s}{m}. \quad (7-41)$$

Since we demand that  $m < r = s(n+1)$ , (7-39) is satisfied and Theorem 3-2 becomes

## 7.5-5

THEOREM 7-4. Let

$$g_{n+1,s/m}(t) = t^{n+1} - \frac{s}{m} \sum_{j=0}^n t^j = 0.$$

If  $n = 0$ , this equation has the real root  $\beta_{1,s/m} = s/m$ .

Assume  $n \geq 1$  and  $m < r = s(n+1)$ . Then the equation has one real positive simple root  $\beta_{n+1,s/m}$  and

$$\max\left[1, \frac{s}{m}\right] < \beta_{n+1,s/m} < \frac{s}{m} + 1.$$

Furthermore,

$$\frac{s}{m} + 1 - \frac{e^{-s/m}}{\left(\frac{s}{m} + 1\right)^{n+1}} < \beta_{n+1,s/m} < \frac{s}{m} + 1 - \frac{s/m}{\left(\frac{s}{m} + 1\right)^{n+1}}$$

where  $e$  denotes the base of natural logarithms. Hence

$$\lim_{n \rightarrow \infty} \beta_{n+1,s/m} = \frac{s}{m} + 1.$$

All other roots are also simple and have moduli less than one.

Tables 7-5, 7-6, 7-7, and 7-8 give values of  $\beta_{k,a}$ ,  $a = s/m$ , for  $m = 1, 2, 3$ , and  $4$  respectively. Table 7-5 is identical with Table 3-1 and is repeated here to facilitate comparison with the other three tables. The cases in which the condition  $ka > 1$  is violated are left blank.

## 7.5-6

TABLE 7-5. VALUES OF  $\beta_{k,a}$ 

$\downarrow k$	$\vec{a} \ 1/1$	$2/1$	$3/1$	$4/1$
1		2.000	3.000	4.000
2	1.618	2.732	3.791	4.828
3	1.839	2.920	3.951	4.967
4	1.928	2.974	3.988	4.994
5	1.966	2.992	3.997	4.999
6	1.984	2.997	3.999	5.000
7	1.992	2.999	4.000	5.000

## 7.5-7

TABLE 7-6. VALUES OF  $\beta_{k,a}$ 

	$\vec{a} \ 1/2$	$2/2$	$3/2$	$4/2$
$\downarrow k$				
1			1.500	2.000
2		1.618	2.186	2.732
3	1.234	1.839	2.390	2.920
4	1.349	1.928	2.459	2.974
5	1.410	1.966	2.484	2.992
6	1.445	1.984	2.494	2.997
7	1.466	1.992	2.498	2.999

7.5-8

TABLE 7-7. VALUES OF  $\beta_{k,a}$

$\downarrow k$	$\vec{a} \ 1/3$	$2/3$	$3/3$	$4/3$
1				1.333
2		1.215	1.618	2.000
3		1.446	1.839	2.210
4	1.126	1.552	1.928	2.284
5	1.199	1.604	1.966	2.313
6	1.243	1.631	1.984	2.325
7	1.271	1.646	1.992	2.330

## 7.5-9

TABLE 7-8. VALUES OF  $\beta_{k,a}$ 

$\downarrow k$	$\rightarrow a$	$1/4$	$2/4$	$3/4$	$4/4$
1					
2				1.319	1.618
3			1.234	1.548	1.839
4			1.349	1.648	1.928
5	1.079		1.410	1.697	1.966
6	1.130		1.445	1.721	1.984
7	1.163		1.466	1.734	1.992

7.53 The order. We return to the analysis of the error equation,

$$e_{i+1}^m = M_i \prod_{j=0}^n (e_{i-j} - e_{i+1})^s, \quad (7-42)$$

$$M_i = (-1)^r \frac{m!}{r!} \frac{f^{(r)}(\xi_{i+1})}{f^{(m)}(\eta_{i+1})}, \quad m < r.$$

We shall assume that

$$\frac{e_{i+1}}{e_i} \rightarrow 0. \quad (7-43)$$

Hence  $e_i \rightarrow 0$ .

We may rewrite (7-42) as

$$e_{i+1}^m = \tau_i \prod_{j=0}^n e_{i-j}^s, \quad (7-44)$$

where

$$\tau_i = M_i \lambda_i, \quad \lambda_i = \prod_{j=0}^n \left(1 - \frac{e_{i+1}}{e_{i-j}}\right)^s. \quad (7-45)$$

From (7-43), we conclude that  $\lambda_i \rightarrow 1$ . Assume that  $e_i$  does not vanish for any finite  $i$ . Let

$$\sigma_i = \ln \delta_i = \ln |e_i|, \quad T_i = \ln |\tau_i|. \quad (7-46)$$

From (7-44),

$$m\sigma_{i+1} = T_i + s \sum_{j=0}^n \sigma_{i-j},$$

or

$$\sigma_{i+1} = \frac{T_i}{m} + \frac{s}{m} \sum_{j=0}^n \sigma_{i-j}. \quad (7-47)$$

Now, (7-47) is identical with (3-34), except that  $T_i/m$  replaces  $J_i$  and  $s/m$  replaces  $s$ . Observe that

$$T_i \rightarrow \ln \left| \frac{m!}{r!} \frac{f^{(r)}(\alpha)}{f^{(m)}(\alpha)} \right|$$

whereas

$$J_i \rightarrow \ln |K|.$$

Furthermore, (3-43) is replaced by

$$\sum_{j=0}^n c_j = \frac{\frac{s}{m} (n+1) - 1}{p-1}.$$

By methods analogous to those used in the proof of Theorem 3-3 one may show that

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow \left| \frac{m!}{r!} \frac{f^{(r)}(\alpha)}{f^{(m)}(\alpha)} \right|^{(p-1)/(r-m)}.$$

We summarize our results in

THEOREM 7-5. Let

$$J = \left\{ x \mid |x - \alpha| \leq r \right\}$$

where  $\alpha$  is a zero of multiplicity  $m$ . Let  $r = s(n+1)$  and assume that  $m < r$ . Let  $f^{(r)}$  be continuous and let  $f^{(m)} f^{(r)}$  be nonzero on  $J$ . Let  $x_0, x_1, \dots, x_n \in J$  and define a sequence  $\{x_i\}$  as follows: Let  $P_{n,s}$  be an interpolatory polynomial for  $f$  such that the first  $s - 1$  derivatives of  $P_{n,s}$  are equal to the first  $s - 1$  derivatives of  $f$  at the points  $x_1, x_{1-1}, \dots, x_{1-n}$ . Assume that there exists a real number,  $x_{1+1} \in J$ , such that  $P_{n,s}(x_{1+1}) = 0$ . Define  $\Phi_{n,s}$  by

$$x_{1+1} = \Phi_{n,s}(x_1; x_{1-1}, \dots, x_{1-n}).$$

Let  $e_{1-j} = x_{1-j} - \alpha$  and assume that  $e_1$  does not vanish for any finite  $i$  but that  $e_{1+1}/e_1 \rightarrow 0$ .

Then

$$\frac{|e_{1+1}|}{|e_1|^p} \rightarrow \left| \frac{m!}{r!} \frac{f^{(r)}(\alpha)}{f^{(m)}(\alpha)} \right|^{(p-1)/(r-m)}, \quad (7-48)$$

where  $p$  is the unique real positive root of

$$t^{n+1} - \frac{s}{m} \sum_{j=0}^n t^j = 0.$$

7.5-13

Note that (7-48) reduces to (4-30) if  $m = 1$ . For the case  $m = 1$ , we showed that  $e_1 \rightarrow 0$  provided the initial errors were sufficiently small. For the case  $m > 1$ , we assume  $e_{i+1}/e_i \rightarrow 0$ , which implies  $e_1 \rightarrow 0$ .

## 7.5-14

7.54 Discussion and examples. The results of Section 7.53 are very satisfying; the only parameters that appear in the conclusion of Theorem 7-5 are the order, the product of the number of new pieces of information with the number of points at which information is used, and the multiplicity of the zero. The effect of the multiplicity is to reduce the factor  $s$ , which appears linearly in the equation which determines the order, to  $s/m$ . For  $n$  fixed, the order depends only on the ratio of  $s$  to  $m$ . Thus the order is the same for  $n = 1$ ,  $s = 1$ ,  $m = 1$  (secant I.F.), as for  $n = 1$ ,  $s = 2$ ,  $m = 2$ . Furthermore, the limiting value of the order as  $n \rightarrow \infty$  is simply  $1 + s/m$ .

EXAMPLE 7-2.  $s = 3$ ,  $n = 0$ . This I.F. was discussed in Section 5.32 for the case of simple roots. If  $m = 1$ ,

$$\frac{e_{1+1}}{e_1^3} \rightarrow -A_3(\alpha).$$

If  $m = 2$ ,

$$\frac{|e_{1+1}|}{|e_1|^{1.5}} \rightarrow \left| \frac{1}{3} \frac{f'''(\alpha)}{f''(\alpha)} \right|^{.5}.$$

7.5-15

EXAMPLE 7-3.  $s = 1$ ,  $n = 2$ . This I.F. is discussed  
in Section 10.21 for the case of simple roots. If  $m = 1$ ,

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow |A_3(\alpha)|^{\frac{1}{2}(p-1)}, \quad p \sim 1.84.$$

If  $m = 2$ ,

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow \left| \frac{1}{3} \frac{f'''(\alpha)}{f''(\alpha)} \right|^{p-1}, \quad p \sim 1.23.$$

## 7.6-1

### 7.6 One-Point I.F. With Memory

A number of techniques for constructing one-point I.F. with memory for the case of multiple zeros are given in this section. Since these techniques are variations on earlier themes we shall pass over them lightly.

The I.F. studied in Section 7.5 are one-point with memory if  $n > 0$ . We showed that if  $m < r = s(n+1)$ , then bounds on the order  $p$  are given by

$$\max\left[1, \frac{s}{m}\right] < p < \frac{s}{m} + 1.$$

Since  $u = f/f'$  has only simple zeros, we can apply the theory which pertains to simple zeros by replacing  $f$  and its derivatives by  $u$  and its derivatives. As an example, we consider direct interpolation studied in Section 4.23. The conclusion of Theorem 4-3 states that

$$\frac{|e_{i+1}|}{|e_i|^p} \rightarrow \left| \frac{f^{(r)}(\alpha)}{r!f'(\alpha)} \right|^{(p-1)/(r-1)}.$$

Let  $\Phi_{n,s}(u)$  be the I.F. generated from  $\Phi_{n,s}$  by replacing  $f$  by  $u$ . Then

$$\frac{|\Phi_{n,s}(u) - \alpha|}{|x-\alpha|^p} \rightarrow \left| \frac{u^{(r)}(\alpha)}{r!u'(\alpha)} \right|^{(p-1)/(r-1)}.$$

7.6-2

In Section 7.4,  $\omega_\ell(m)$  was defined by

$$mu(x) = \sum_{\ell=1}^{\infty} \omega_\ell(m) e^\ell, \quad e = x - \alpha.$$

Hence

$$\frac{|\Phi_{n,s}(u) - \alpha|}{|x-\alpha|^p} \rightarrow |\omega_r(m)|^{(p-1)/(r-1)}.$$

In particular,

$$\Phi_{1,1}(u) = x_i - u_i \left[ \frac{x_i - x_{i-1}}{u_i - u_{i-1}} \right] \quad (7-50)$$

is of order  $\frac{1}{2}(1+\sqrt{5}) \sim 1.62$  for all  $m$ . It is not difficult to show that the first two terms of the error series are given by

$$\begin{aligned} \Phi_{1,1}(u) - \alpha &\approx -B_{2,m}(\alpha) e_i e_{i-1} + \left[ mB_{2,m}^2(\alpha) - 2B_{3,m}(\alpha) \right] e_i e_{i-1}^2, \quad B_{j,m} = \frac{a_{j+m-1}}{m a_m}, \\ B_{j,m} &= \frac{a_{j+m-1}}{m a_m}, \quad a_j = \frac{f^{(j)}}{j!}. \end{aligned} \quad (7-51)$$

$\Phi_{1,1}$  has an informational usage of two and an informational efficiency of .81 for all  $m$ .  $\Phi_{1,1}$  suffers from the drawback that as the approximants converge to  $\alpha$ ,  $u$  approaches a o/o form which may necessitate multiple precision arithmetic.

If  $m$  is known, a number of other techniques are available. Since  $f^{(m-1)}$  has only simple zeros, the theory which pertains to simple zeros is applicable to the I.F. generated by replacing  $f$  by  $f^{(m-1)}$ . Such I.F. have the advantage that no o/o forms occur; they have the disadvantage that rather high derivatives of  $f$  must be used.

We can also replace  $f$  by  $F = f^{1/m}$ . We can perform derivative estimation on the new I.F. For example, define  $*F'_n$  analogously to  $*f'_n$ . (See Section 6.22.) It is clear that the I.F.

$$*E_{n,1}(F) = x - \frac{F}{*F_n} \quad (7-52)$$

has orders ranging from  $\frac{1}{2}(1 + \sqrt{5})$  to 2, as a function of  $n$ .

As a final example, we perform derivative estimation on the second order I.F.,

$$e_2 = x - m \frac{f}{f'}.$$

## 7.6-4

Define

$$*\varepsilon_{n,1} = x - m \frac{f}{*f_n}. \quad (7-53)$$

It may be shown that the order of this I.F. ranges from 1 to  $\frac{1}{2}(1+\sqrt{5})$  as a function of  $n$  and  $m$ .  $*\varepsilon_{1,1}$  (secant I.F.) is of linear order for all nonsimple zeros. The essential difference between (7-53) and (7-52) is that the denominator on the right side of (7-53) is a linear combination of  $f_{1-j}$ ,  $0 \leq j \leq n$ , whereas the denominator on the right side of (7-52) is a linear combination of  $f_{1-j}^{1/m}$ ,  $0 \leq j \leq n$ .

### 7.7 Some General Results

The following statements typify the results which we have obtained in this chapter:

- a.  $E_s$ ,  $s = 2, 3, \dots$ , is of linear order for all non-simple zeros.
- b.  $\mathcal{E}_s$ ,  $s = 2, 3, \dots$ , is optimal; it depends explicitly on  $m$  and its order is multiplicity-independent.
- c. A nonoptimal multiplicity-independent I.F. may be generated from any I.F. by replacing  $f$  and its derivatives by  $u$  and its derivatives.
- d.  $\Phi_{n,s}$ , generated by direct interpolation, is multiplicity dependent but not of linear order if  $m < r = s(n+1)$ .

In this section we shall make some general remarks.

First we suggest the following

CONJECTURE. It is impossible to construct an optimal I.F. which does not depend explicitly on  $m$  and which is multiplicity-independent.

We have excluded explicit dependence on  $m$  because of  $b$  and we have restricted ourselves to optimal I.F. because of  $c$ .

The following incorrect discussion by Bodewig [7.7-1] illustrates the pitfalls which one may encounter in the study of multiple roots. Bodewig argues as follows: Since we require  $\varphi(\alpha) = \alpha$ , we take

$$\varphi(x) = x - f(x)H(x).$$

A necessary condition that  $\varphi$  be of second order is that

$$\varphi'(\alpha) = 0 = 1 - f'(\alpha)H(\alpha) - f(\alpha)H'(\alpha). \quad (7-54)$$

Since  $f(\alpha)$  and  $f'(\alpha)$  are both zero for  $m > 1$ , Bodewig reasons that the second and third terms in (7-54) vanish if  $\alpha$  is a multiple root. This ignores the possibility that  $H(x)$  has a singularity at  $x = \alpha$ . In Newton's I.F., for example,  $H(x) = 1/f'(x)$ . Hence  $f(\alpha)H'(\alpha) = (1-m)/m$  while  $f'(\alpha)H(\alpha) = 1$  and none of the terms of (7-54) vanish when  $m > 1$ . Taking  $H(x) = m/f'(x)$  or  $H(x) = 1/[f'(x)u'(x)]$  shows that Bodewig's conclusion that  $\varphi'(\alpha) \neq 0$  for multiple roots is incorrect.

The above reasoning led to difficulties because the condition  $\varphi(\alpha) = \alpha$ , which implies  $f(x)H(x) \rightarrow 0$ , permits  $H(x)$  to be of  $\underline{O}[(x-\alpha)^{1-m}]$ . Since  $u(x) = \underline{O}(x-\alpha)$ , we can avoid the difficulty by taking

7.7-3

$$\varphi(x) = x - u(x)h(x).$$

Then

$$\begin{aligned}\varphi(x) - \alpha &= x - \alpha - \left\{ \frac{x-\alpha}{m} + \underline{o}[(x-\alpha)^2] \right\} h(x) \\ &= (x-\alpha) \left[ 1 - \frac{h(x)}{m} \right] + h(x) \underline{o}[(x-\alpha)^2].\end{aligned}$$

Let  $h$  be continuous in a closed interval about  $\alpha$ . Then  $\varphi$  is of second order if and only if

$$h(x) - m = \underline{o}(x-\alpha). \quad (7-55)$$

In particular, we require

$$h(\alpha) = m. \quad (7-56)$$

If  $h'$  is continuous, then (7-56) implies (7-55). Equation (7-55) is equivalent to

$$\frac{h(x) - m}{u(x)} \rightarrow D \neq 0. \quad (7-57)$$

We summarize our results in

## 7.7-4

THEOREM 7-6. Let  $\varphi(x) = x - u(x)h(x)$ . Let  $\alpha$  be a root of multiplicity  $m$ . If  $h$  is continuous in a closed interval about  $\alpha$ , then  $\varphi$  is of second order if and only if

$$\frac{h(x) - m}{u(x)} \rightarrow D \neq 0.$$

If  $h'$  is continuous then  $\varphi$  is of second order if and only if  $h(\alpha) = m$ .

For I.F. of second order, the Conjecture which we stated earlier in this section is equivalent to the following: Let  $\varphi(x) = x - u(x)h(x)$ , where  $h$  does not depend upon any derivatives of  $f$  higher than the first and does not depend explicitly on  $m$ . Let  $h(x)$  be a continuously differentiable function of  $x$ . Then it is impossible that  $h(\alpha) = m$  for all  $f$  whose zeros have multiplicity  $m$ .

In Section 7.8 we construct an  $h$  which depends only on  $f$  and  $f'$  and for which  $h \rightarrow m$ . But this  $h$  is not continuously differentiable at  $\alpha$ .

7.8-1

7.8 An I.F. Of Incommensurate Order

LEMMA 7-2. Let  $f^{(m+1)}$  be continuous in the neighborhood of a zero  $\alpha$  of multiplicity  $m$ . Let

$$h(x) = \frac{\ln|f(x)|}{\ln|u(x)|}, \quad x \neq \alpha; \quad h(\alpha) = m.$$

Then  $h(x) \rightarrow m$ .

PROOF. Let

$$f(x) = (x-\alpha)^m g(x). \quad (7-58)$$

Then

$$g(x) \rightarrow g(\alpha) = \frac{f^{(m)}(\alpha)}{m!} \neq 0.$$

Let

$$G(x) = (x-\alpha) \frac{g'(x)}{g(x)}.$$

Then  $G(x) = \underline{o}(x-\alpha)$  and

$$u(x) = \frac{x-\alpha}{m+G}. \quad (7-59)$$

From (7-58) and (7-59)

$$h(x) = \frac{\ln|f(x)|}{\ln|u(x)|} = \frac{m \ln|x-\alpha| + \ln|g|}{\ln|x-\alpha| - \ln|m+G|} \rightarrow m.$$

## 7.8-2

LEMMA 7-3. Let  $f$  and  $h$  be as in Lemma 7-2. Then  $uh' \rightarrow 0$ .

PROOF. We shall assume for simplicity that  $f$  and  $f'$  are positive; the result is true in general. Then for  $x \neq a$ ,

$$h'(x) = \frac{1}{u \ln[u(x)]} - \frac{\ln f(x)u'(x)}{\{\ln[u(x)]\}^2 u(x)}.$$

Hence

$$u(x)h'(x) = \frac{1}{\ln[u(x)]} - \frac{h(x)u'(x)}{\ln[u(x)]}.$$

The fact that  $h \rightarrow m$ ,  $u' \rightarrow 1/m$ ,  $\ln[u] \rightarrow -\infty$  completes the proof.

LEMMA 7-4. Let  $f$  and  $h$  be defined as in Lemma 7-2.

Let

$$\varphi(x) = x - u(x)h(x). \quad (7-60)$$

Then  $\varphi'(x) \rightarrow 0$ .

PROOF. For  $x \neq a$ ,  $h$  and hence  $\varphi$  are differentiable and

$$\varphi'(x) = 1 - u'(x)h(x) - u(x)h'(x).$$

Note that  $h \rightarrow m$  from Lemma 7-2,  $uh' \rightarrow 0$  from Lemma 7-3, and  $u' \rightarrow 1/m$ ; the conclusion follows immediately.

7.8-3

We have shown that  $h$ , which depends only on  $f$  and  $f'$ , has the property that  $h \rightarrow m$ . Since  $\varphi' \rightarrow 0$ , one might hope to conclude that  $\varphi$  is second order but this is not the case. It may be shown that  $(h-m)/u \rightarrow \infty$ . Then an application of Theorem 7-6 shows that  $\varphi$  is not second order. Instead we have

THEOREM 7-7. Let  $h$  and  $f$  satisfy the conditions of Lemma 7-2 and let  $\varphi = x - uh$ . Then

$$\frac{\varphi-\alpha}{x-\alpha} \ln|x-\alpha| \rightarrow -\ln\left(m \left|\frac{f^{(m)}(\alpha)}{m!}\right|^{1/m}\right). \quad (7-61)$$

PROOF. Since

$$\varphi = x - uh = x - u \frac{\ln|f|}{\ln|u|},$$

we have

$$\frac{\varphi-\alpha}{x-\alpha} \ln|x-\alpha| = \ln|x-\alpha| \left[ 1 - \frac{\alpha}{x-\alpha} \frac{\ln|f|}{\ln|u|} \right].$$

Recall that

$$h = \frac{\ln|f|}{\ln|u|} = \frac{m \ln|x-\alpha| + \ln|g|}{\ln|x-\alpha| - \ln|m+G|}, \quad u = \frac{x-\alpha}{m+G},$$

where

$$f = (x-\alpha)^m g, \quad G = (x-\alpha) \frac{g'}{g}.$$

The fact that

$$g \rightarrow \frac{f^{(m)}(\alpha)}{m!}, \quad G = \underline{o}(x-\alpha)$$

permits the completion of the proof.

This theorem shows that the order of  $\varphi$  is incommensurate with the order scale defined by (1-14). Observe that if  $\varepsilon$  is an arbitrary preassigned positive number, then

$$\frac{\varphi-\alpha}{x-\alpha} \rightarrow 0, \quad \frac{|\varphi-\alpha|}{|x-\alpha|^{1+\varepsilon}} \rightarrow \infty.$$

We can write (7-61) as

$$\frac{\varphi-\alpha}{x-\alpha} \rightarrow C, \quad C = - \frac{\ln\left(m \left|\frac{f^{(m)}(\alpha)}{m!}\right|^{1/m}\right)}{\ln|x-\alpha|}.$$

Thus  $\varphi$  converges linearly and its asymptotic error constant goes logarithmically to zero. It is clear that the sequence of  $x_i$  generated by  $\varphi$  converges if  $x_0$  is sufficiently close to  $\alpha$ . In particular, let  $f(x) = (x-\alpha)^m$ . Then  $C = -\ln(m)/\ln|x-\alpha|$  and  $C < 1$  if  $|x-\alpha| < 1/m$ .

The following generalization suggests itself.

Recall that

$$h = \frac{\ln|f|}{\ln|u|} = \frac{m \ln|x-\alpha| + \ln|g|}{\ln|x-\alpha| - \ln|m+G|},$$

where  $\ln|m+G| \rightarrow \ln m$ . Since  $h \rightarrow m$ , this suggests that we can accelerate the convergence of  $h$  to  $m$  by defining

$$h_2 = \frac{\ln|f|}{\ln|u| + \ln|h_1|}, \quad h_1 = \frac{\ln|f|}{\ln|u|},$$

or more generally

$$h_{i+1} = \frac{\ln|f|}{\ln|u| + \ln|h_i|}, \quad i = 0, 1, \dots; \quad h_0 = 1.$$

This, in turn, suggests implicitly defining  $h$  by

$$h = \frac{\ln|f|}{\ln|u| + \ln|h|}. \quad (7-62)$$

For  $f = (x-\alpha)^m$ , (7-62) may be derived from another point of view. We have  $u = (x-\alpha)/m$  and

$$(mu)^m = (x-\alpha)^m = f,$$

and hence

$$m = \frac{\ln|f|}{\ln|u| + \ln m}.$$

The analysis of the order of I.F. of the form  $\varphi = x - uh_1$  and  $\varphi = x - uh$  has not been carried out.

7.8-6

After  $f$  and  $f'$  have been calculated, it is not expensive to calculate  $h = \ln|f|/\ln|u|$ . In a general root-finding routine it might be worthwhile to always calculate  $h$ . After the limit of  $h$  has been determined to the nearest integer, the routine can switch to  $\varepsilon_2 = x - mu$ .