Solution of Equations and Systems of Equations

A. M. OSTROWSKI

Solution of Equations and Systems of Equations

A. M. OSTROWSKI
University of Basel, Basel, Switzerland

ACADEMIC PRESS • New York and London • 1960

ACADEMIC PRESS INC.
111 FIFTH AVENUE
NEW YORK 3, N. Y.

United Kingdom Edition
Published by
ACADEMIC PRESS INC. (LONDON) LTD.
17 OLD QUEEN STREET, LONDON S.W. 1

Library of Congress Catalog Card Number: 59-13827

PRINTED IN THE UNITED STATES OF AMERICA

Preface

The course from which this book has developed was primarily given at the invitation of the National Bureau of Standards in the summer 1952 at the American University, Washington, D.C. Notes of the lectures were taken by Mr. M. Ticson and reproduced in a small number of copies as the working paper CL–52–2 for the internal use of the NBS.

As this course contained a great deal of unpublished material, it appeared advisable to publish the whole as a book. For this purpose, the whole course was worked through anew and completely rewritten, much new material was incorporated in the lectures, and 11 appendices have been added to deal with problems which could not be treated in the main text of the book. The treatment is still far from being complete, however.

While the lectures as they were given were intended more or less for the undergraduate level, in the present shape the book uses methods and habits of argumentation which conform more to the graduate level. I think that particularly because the practical importance of numerical methods has grown to such a degree, it is worth while for the mathematician to use in full his technique in order to make completely clear the background of these methods. Thus, I hope that this book will help in a certain degree to bridge over the gap that still exists between "pure" and "practical" mathematics.

In finishing this book, I use the opportunity to extend my thanks to Dr. John H. Curtiss, then the chief of the Mathematics Division of the NBS, on whose initiative these lectures were originally given and who is certainly mainly responsible for the particularly auspicious atmosphere in which the research associates could develop their activity in the Mathematics Division. Conversations with Dr. Olga Taussky-Todd and Dr. John Todd were always particularly inspiring.

I thank further Mr. M. Ticson, who made the first draft of the notes of these lectures, and Mr. W. F. Cahill, who was my assistant at that time. During the preparation of the lectures I enjoyed much help from the computing staff of the NBS, both in Washington and in Los Angeles, a help that was extremely valuable for trying out in the computing practice different methods. Finally, I have to thank my assistants in Basel B. Marzetta, S. Christeller, T. Witzemann and R. Bürki and further Drs. E. V. Haynsworth and Wa. Gautschi of NBS for the great help they gave to me in the preparation of the manuscript of this book and to Dr. Pierre Banderet and Mr. Howard Bell for their valuable help in correcting the proofs.

June, 1960

A. M. OSTROWSKI

Contents

Appendices

1 Introduction. Remainder Terms of Interpolation Formulas

1. In this book we shall consider methods of numerical solution of equations; first, a single equation in one unknown, and then systems of equations with a corresponding number of unknowns.

Most of these methods use the theory of interpolation. We will first derive the "remainder terms" for interpolation formulas.

In the following discussion J_x will denote an interval on the x-axis. (J_x) will denote the *interior* of such an interval, i.e., the interval *excluding the endpoints*. (J_x) is an *open interval*.

LEMMA ON FUNCTIONS WITH SEVERAL ROOTS

2. Lemma 1.1. Let $F(x)$ be defined in J_x. Assume that $F(x_\nu) = 0$ $(\nu = 1, \ldots, n+1)$, $x_\nu \prec J_x$ *, and that $F^{(n)}(x)$ exists in J_x. Then there exists a $\xi \prec J_x$ such that $F^{(n)}(\xi) = 0$. ξ lies even in (J_x) unless all x_ν are concentrated in one of the endpoints.

Remarks: The roots of $F(x)$ may be *multiple*. Any multiple root must be counted a corresponding number of times. This lemma is a generalization of Rolle's theorem. As an illustration of the exceptional case mentioned in the last sentence of the lemma, consider

$$F(x) = x^{n+1} \qquad J_x: 0 \leqq x \leqq 1;$$

then $\xi = 0$ and does not lie in (J_x).

3. Proof. Let $g(x) \equiv F'(x)$. We begin by showing that $g(x)$ has at least n roots in J_x. Suppose first that the x_ν $(\nu = 1, \ldots, n+1)$

---

* The sign $\prec$ is the *symbol of inclusion*; the formula $a \prec K$ is to be read: a is contained in K. In the same way, the formula $K \succ a$ is to be read: K contains a.

are all distinct. If $F(x)$ has two distinct roots, then by Rolle's theorem, $g(x)$ has a root *between* these two. More generally, if $F(x)$ has k distinct roots, then $g(x)$ has $k - 1$ roots separating those k roots of $F(x)$. Clearly, if all x_ν are simple, then $g(x)$ has n distinct roots.

Assume now more generally that $F(x)$ has k different multiple roots in J_x, where a *simple* root is considered a root of multiplicity *one*. If $F(x)$ has x_0 as a root of multiplicity ν, then $g(x)$ will have x_0 as a root of multiplicity $\nu - 1$. Thus, each multiple root loses one unit of its multiplicity in $g(x)$. If there are k multiple roots, then $g(x)$ has at least $(n + 1) - k + (k - 1) = n$ roots, where the term $(k - 1)$ is the number of "new", Rollian roots of $g(x)$. Hence, in any case, $g(x)$ has at least n roots in J_x.

4. We consider now two cases:

Case I. We have $x_1 = x_2 = \ldots = x_{n+1}$. Then x_1 is a root of $F^{(n)}(x)$, i.e., $\xi = x_1$ and the lemma is true.

Case II. None of the x_ν has the multiplicity $n + 1$. Then, if $n = 1$, Rolle's theorem can be applied and the lemma is true. Assume the lemma is true for all smaller values of n. Then the lemma is true for $g(x)$ and, since $g(x)$ has at least one root in (J_x), there exists a $\xi \prec (J_x)$ such that $g^{(n-1)}(\xi) = F^{(n)}(\xi) = 0$, Q.E.D.

THEOREM ON QUOTIENTS OF FUNCTIONS WITH COMMON ROOTS

5. Theorem 1.1. Let $f(x)$, $g(x)$ be defined and n times differentiable in J_x. Assume that there exist n common roots $x_\nu \prec J_x$ $(\nu = 1, \ldots, n)$ of $f(x)$ and $g(x)$, where, if a root is counted with the multiplicity k, it must have at least the multiplicity k both for $f(x)$ and $g(x)$. Assume further that $g^{(n)}(x)$ does not vanish in J_x. Then for any $x_0 \neq x_\nu$ from J_x there exists a $\xi \prec (J_x)$ such that

$$\frac{f(x_0)}{g(x_0)} = \frac{f^{(n)}(\xi)}{g^{(n)}(\xi)} \qquad \xi \prec (J_x)\,(x_0 \neq x_\nu;\ \nu = 1, \ldots, n). \tag{1.1}$$

6. Proof. First, x_0 is not a root of $g(x)$. For, otherwise, since $x_0 \neq x_\nu$, this would mean that $g(x)$ has $n + 1$ roots in J_x and by Lemma 1.1 there exists a $\xi \prec J_x$ such that $g^{(n)}(\xi) = 0$, contrary to our hypothesis.

Let $\lambda = f(x_0)/g(x_0)$ and consider $F(x) = f(x) - \lambda g(x)$. $F(x)$ satisfies the hypothesis of Lemma 1.1, for x_ν $(\nu = 1, \ldots, n)$ and x_0 are roots of $F(x)$. Furthermore, since $g^{(n)}(x)$ and $f^{(n)}(x)$ exist in J_x, so does $F^{(n)}(x)$. Since $x_0 \neq x_\nu$ $(\nu = 1, \ldots, n)$, $F(x)$ has at least two *distinct* roots and by Lemma 1.1 there exists a ξ in (J_x) such that

$$F^{(n)}(\xi) = f^{(n)}(\xi) - \lambda g^{(n)}(\xi) = 0$$

or

$$\frac{f^{(n)}(\xi)}{g^{(n)}(\xi)} = \lambda = \frac{f(x_0)}{g(x_0)}, \qquad \text{Q.E.D.}$$

In the above theorem and in Lemma 1.1 it would be sufficient to assume the differentiability of $F(x)$, $f(x)$, $g(x)$ only in (J_x) and to require at the endpoints belonging to J_x only the continuity unless there are multiple roots at the endpoints of J_x belonging to J_x. In this last case, it is sufficient to assume as many derivatives in the corresponding endpoints as are implied by the multiplicity of the roots.

The above theorem affords us a means of studying the *remainder terms* of interpolating formulas.

INTERPOLATING FUNCTIONS

7. Let $f(x)$ be defined in n interpolation points in J_x, x_ν $(\nu = 1, \ldots, n)$. Let $T(x)$ be defined throughout J_x such that

$$f(x_\nu) = T(x_\nu) \qquad (\nu = 1, \ldots, n). \tag{1.2}$$

Then $T(x)$ is called an *interpolating function* for $f(x)$ with the interpolation points x_ν.

$T(x)$ could be chosen from many different classes of functions, e.g., polynomials, trigonometric functions, rational functions of x, etc. In practice we choose for $T(x)$ functions with very familiar properties.

The values x_ν may be partly *equal*, for instance, $x_\nu = a$ $(\nu = 1, \ldots, m)$; i.e., $T(x)$ may be "equal to $f(x)$ at the place a, m times." This means that then also the first $m - 1$ derivatives of $f(x)$ are equal to the corresponding first $m-1$ derivatives of $T(x)$ at $x_\nu = a$, i.e., $f(x) - T(x)$ has a multiple root of the multiplicity m at a and the corresponding equations (1.2) are replaced by

$$f^{(\mu)}(a) = T^{(\mu)}(a) \qquad (\mu = 0, 1, \ldots, m-1). \tag{1.3}$$

E.g., if both functions are "equal to $f(a)$ at a particular point a three times", $x_1 = x_2 = x_3 = a$, and if the remaining $n-3$ interpolation points are distinct, then (1.2) is replaced by

$$f^{(\mu)}(a) = T^{(\mu)}(a) \qquad (\mu = 0, 1, 2),$$

$$f(x_\nu) = T(x_\nu) \qquad (\nu = 4, 5, \ldots, n).$$

In the case of equal interpolation points, (1.2) must be always understood in the sense of this convention.

When the n interpolation points are all equal and $T(x)$ is a polynomial of degree $n-1$, then $T(x)$ is the well-known Taylor polynomial.

REMAINDER TERM IN GENERAL INTERPOLATION

8. Let $f(x)$ be defined and n times differentiable in J_x. Let $T(x)$ be an interpolating *polynomial* of degree $n-1$ for $f(x)$ with n interpolation points $x_1, \ldots, x_n$ from J_x, i.e., $f(x_\nu) = T(x_\nu)$ $(\nu = 1, \ldots, n)$. Let $g(x)$ be n times differentiable in J_x and such that $g(x_\nu) = 0$ $(\nu = 1, \ldots, n)$ and $g^{(n)}(x) \neq 0$ for each $x \prec J_x$.

Replace $f(x)$ by $f(x) - T(x)$ in Theorem 1.1. Then for each $x \neq x_\nu$ $(\nu = 1, \ldots, n)$, $x \prec J_x$, there exists a ξ depending on x, but contained in (J_x) so that

$$\frac{f(x) - T(x)}{g(x)} = \frac{f^{(n)}(\xi)}{g^{(n)}(\xi)} \qquad \xi \prec (J_x),$$

$$f(x) - T(x) = \frac{f^{(n)}(\xi)}{g^{(n)}(\xi)} g(x) \qquad \xi \prec (J_x). \tag{1.4}$$

Now put $g(x) = \Pi_{\nu=1}^{n} (x - x_\nu)$ and (1.4) becomes

$$f(x) - T(x) = \frac{f^{(n)}(\xi)}{n!} \prod_{\nu=1}^{n} (x - x_\nu). \tag{1.5}$$

When $x_1 = x_2 = \ldots = x_n$, (1.5) is the remainder term of the Taylor series.

When the interpolation points are all distinct, (1.5) becomes the remainder term of the n-point Lagrangian interpolation formula.

HERMITE INTERPOLATING POLYNOMIAL

9. Let $H(x)$ be an interpolating polynomial of degree $n-1$. Let our n interpolation points be distributed so that $x_1 \neq x_2 \neq \ldots \neq x_k$, and x_1 occurs m_1 times, x_2 occurs m_2 times, ..., x_k occurs m_k times, where $n = m_1 + m_2 + \ldots + m_k$. We obtain the following array of values belonging to $f(x)$:

$$\begin{array}{l} x_1 : a_0^{(1)}, a_1^{(1)}, \ldots, a_{m_1-1}^{(1)} \\ x_2 : a_0^{(2)}, a_1^{(2)}, \ldots, a_{m_2-1}^{(2)} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ x_k : a_0^{(k)}, a_1^{(k)}, \ldots, a_{m_k-1}^{(k)} \end{array}$$

where $f^{(\mu)}(x_\kappa) = H^{(\mu)}(x_\kappa) = a_\mu^{(\kappa)}$ $(\mu = 0, 1, \ldots, m_\kappa - 1;\ \kappa = 1, 2, \ldots, k)$.

In this case (1.5) becomes

$$f(x) - H(x) = \frac{f^{(n)}(\xi)}{n!}(x-x_1)^{m_1}(x-x_2)^{m_2}\ldots(x-x_k)^{m_k}. \qquad (1.6)$$

The interpolating polynomial $H(x)$, called the *Hermite interpolating polynomial*, is uniquely determined. For, assume another such polynomial H_1. Then $H - H_1$ is of degree not greater than $n-1$. Furthermore, $H - H_1$ has x_1 as a root of multiplicity m_1, x_2 as a root of multiplicity $m_2, \ldots, x_k$ as a root of multiplicity m_k. Thus $H - H_1$ contains $(x-x_1)^{m_1}(x-x_2)^{m_2}\ldots(x-x_k)^{m_k}$ as a factor, i.e., is divisible by a polynomial of degree n. Since $H - H_1$ is of degree $\leqq n-1$, it must be identically zero.

10. If in particular in the Hermite interpolating polynomial, all multiplicity indices m_k have the value 1, the Hermite interpolating polynomial becomes the Lagrangian interpolating polynomial, which can be written down explicitly in the following way. Let

$$F(x) = \prod_{\nu=1}^{n} (x - x_\nu). \tag{1.7}$$

Then

$$L(x) = \sum_{\nu=1}^{n} f(x_\nu) \frac{F(x)}{(x - x_\nu)F'(x_\nu)} \tag{1.8}$$

is the corresponding Lagrangian interpolating polynomial.

2 Inverse Interpolation. Derivatives of the Inverse Function. One Interpolation Point

THE CONCEPT OF INVERSE INTERPOLATION

1. If a number a is used as an approximation for a number x, we write $a \doteq x$; the same notation is used for approximating functions.

One may approach the problem of finding the roots of a function $f(x)$ in two different ways. We may equate an interpolating function $T(x)$ of $f(x)$ to zero, $T(x) = 0$, and find the roots of this equation. The question arises whether the roots so obtained will be approximations of the roots of $f(x) = 0$. It is well-known that by changing the coefficients of an algebraic equation very slightly, we may get roots which differ considerably from the roots of the original equation (see Appendix A). This problem will later be discussed in greater detail, and we shall establish conditions under which the roots of $T(x) = 0$ are approximations of the roots of $f(x) = 0$ (see Appendices A, B and K).

The second approach to the problem of finding the roots of $f(x) = 0$ is by using the *inverse function.*

Let $y = f(x)$ be defined in J_x and given in n interpolation points x_ν $(\nu = 1, 2, \ldots, n)$:

$$f(x_\nu) = y_\nu. \tag{2.1}$$

Let $x = \phi(y)$ be the inverse function of $y = f(x)$. Then $\phi(y_\nu) = x_\nu$. The problem of finding a root of $f(x)$ now becomes the problem of evaluating $\phi(0)$.

Let $T(y)$ be an interpolating function for $\phi(y)$, so that $T(y_\nu) = x_\nu$. We now evaluate $T(0) \doteq \phi(0)$. An estimate of the error involved here may be obtained from (1.5):

$$\phi(y) - T(y) = \frac{\phi^{(n)}(\eta)}{n!} \prod_{\nu=1}^{n} (y - y_\nu).$$

$$\phi(0) - T(0) = \frac{\phi^{(n)}(\eta)}{n!}(-1)^n y_1 y_2 \ldots y_n. \tag{2.2}$$

Notice that if all our interpolation points are close to the value of the root, then the error will be particularly small.

2. The above procedure has been known for many years, but mathematicians have generally been reluctant to use this approach because the problem of discussing the inverse function and its derivatives has been considered a difficult one. These difficulties are really only superficial. One is usually interested in the solution of $f(x) = 0$ in a certain interval. Generally within this interval $f'(x) \neq 0$. Otherwise, all well-known methods (inverse interpolation and so on) generally fail and one must resort to some special device. We shall therefore assume $f'(x) \neq 0$ in the considered interval, and we shall show that with this assumption the difficulties mentioned above are eliminated.

DARBOUX'S THEOREM ON VALUES OF $f'(x)$

3. We first prove a theorem which gives the right background for our hypotheses.

Theorem 2.1. (Darboux). Let $f(x)$ be defined and continuous in J_x: $a \leqq x \leqq b$. Suppose that $f'(x)$ exists in J_x and that $f'(a) = A$, $f'(b) = B$. Then all values between A and B are assumed by $f'(x)$ for $x \prec J_x$.

Remark: This property of $f'(x)$ is sometimes incorrectly given as a definition of a continuous function.

Proof of Darboux's Theorem. Without loss of generality we can assume $A < B$, since for $A > B$ it would be sufficient to replace $f(x)$ by $-f(x)$. Let C be any number satisfying $A < C < B$. We will prove that $f'(x)$ assumes the value C somewhere in (J_x). Consider $F(x) = f(x) - Cx$. Then

$$F'(x) = f'(x) - C, \qquad F'(a) = A - C < 0, \qquad F'(b) = B - C > 0.$$

We thus have a continuous function which has a negative derivative in a and a positive one in b. Hence, $F(x)$ assumes in (J_x)

values which are less than $F(a)$ and $F(b)$; $F(x)$ has therefore in J_x a minimum in a point ξ which is *interior* to J_x, and the derivative must vanish in ξ, $f'(\xi) - C = 0$, $f'(\xi) = C$, Q.E.D.

By Darboux's Theorem the assumption that $f'(x) \neq 0$ $(x \prec J_x)$ implies therefore that either $f'(x) > 0$ for all $x \prec J_x$ or $f'(x) < 0$ for all $x \prec J_x$.

DERIVATIVES OF THE INVERSE FUNCTION

4. Let $f(x)$ be defined in J_x. Assume that $f'(x)$ exists in J_x and is $\neq 0$. Then $f(x)$ is strictly monotonic in J_x and by the well-known existence theorems the inverse function $x = \phi(y)$ of $y = f(x)$ exists and has a derivative in the corresponding y-interval,

$$x = \phi(y), \qquad \phi'(y) = \frac{dx}{dy} = \frac{1}{y'} = \frac{1}{f'}. \tag{2.3}$$

If, moreover, $f''(x) = y''$ exists in J_x, it follows by differentiation of $\phi'(y)$:

$$\phi''(y) = \frac{-y''}{y'^3}. \tag{2.4}$$

We see that with the above assumptions, if $f(x)$ possesses first and second derivatives, so does the inverse function. The problem of finding workable expressions for higher derivatives of the inverse function can become quite complicated. We assume the existence of the first $n + 1$ $(n \geqq 0)$ derivatives of $f(x)$ and get a recurrence formula for obtaining the corresponding derivatives of $\phi(y)$.

5. Let

$$\phi^{(k)}(y) = \frac{X_k}{y'^{2k-1}} \qquad (k = 1, 2, \ldots, n + 1). \tag{2.5}$$

Here X_k is a *polynomial* in $y', y'', \ldots, y^{(k)}$. This is true for $n = 0, 1$. We have, in particular $X_1 = 1$, $X_2 = -y''$. Assume the truth of our assertion for the first n derivatives of $\phi(y)$. We write (2.5) with $k = n$ and get by differentiation, since $dy'/dy = y''/y'$,