Methods for Fitting Rational Approximations, Parts II and III

HANS J. MAEHLY

*Prepared posthumously by Christoph Witzgall**

Preface. Dr. Hans J. Maehly died on the 16th of November, 1961. Just six weeks before his untimely death, Dr. Maehly had joined the Applied Mathematics Division of Argonne National Laboratory. Before turning to new problems, he planned to complete the publication of his results on rational approximations. These results were obtained from 1958 to 1960 under Contract Nonr 2406(00) of the Bureau of Ships and its Applied Mathematics Laboratory, David Taylor Model Basin, with Princeton University.

In particular, he planned a series of publications on "Methods for Fitting Rational Approximations." This series was to consist of three parts: Part I, Telescoping Procedures for Continued Fractions; Part II, Direct Methods; and Part III, Indirect Methods. Part I was published in this Journal [8]. Dr. Maehly had just started to write Part II when death interrupted his work.

Dr. Maehly's interest in Chebyshev-approximations had been stimulated by the work of C. Hastings [3], F. D. Murnaghan and J. W. Wrench [12,13]. Simultaneously with W. Barth [2] and others, these authors gave versions of E. Remes' [14] "second algorithm" (1934) for the approximation by polynomials. Inspired by the great convergence power of continued fractions, Dr. Maehly was mainly concerned with developing algorithms for approximation by general rational functions. Following a suggestion of his scientific associate, Dr. K. Arbenz, he developed in 1959 a new approach which deviates from Remes' scheme of characterizing the error curve by its extrema (Part II, Section 9). Other significant contributions by Dr. Maehly to the subject of rational approximations include "telescoping" of continued fractions (Part I), and the "indirect" methods (Part III), which use single-precision arithmetic to produce multi-precision approximations.

The writer is honored by the invitation of Argonne National Laboratory to prepare the posthumous publication of Parts II and III of the series mentioned above, and he wishes to acknowledge the generous support he received while undertaking this task. In preparing the manuscript he enjoyed the help of Dr. R. F. King, assistant director of the Applied Mathematics Division. The writer is indebted to Dr. H. C. Thacher, Jr. and C. Mesztenyi for their valuable criticisms and suggestions.

The source material on which this publication is based consists of a series of Internal Reports concerning the Princeton-ONR project and a rough draft by Dr. Maehly of the introduction and first section of Part II. In addition, the writer had the benefit of many discussions with Dr. Maehly while working with him at Princeton on rational approximations. These discussions conveyed a fair idea of how he planned to write his series. Although some of his later refinements cannot be retrieved for lack of documentation, it is hoped that the following presentation provides a reasonably complete account of Dr. Maehly's important work on methods for rational approximations.

II. DIRECT METHODS

II-Introduction

In Part I of this series [8] we described the telescoping of a continued fraction, corresponding to the telescoping of a power series as described by C. Lanczos

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[4]. Telescoping methods adjust the coefficients of a truncated continued fraction or power series so as to nearly minimize the maximum error on a given interval. In other words, one obtains an approximation which is nearly, but in general not exactly, a Chebyshev approximation. Moreover, telescoping methods require that the function to be approximated be given in the form of a continued fraction or power series.

By contrast, the methods which we call "direct" can be applied to functions given in any form suitable for accurate numerical evaluation.

The term *direct methods* is meant to indicate that the coefficients of the Chebyshev-approximant are computed directly. On the other hand, the *indirect methods* to be treated in Part III determine the corrections required to modify a fixed approximant, for instance, the Padé-approximant, in order to get the Chebyshev-approximant.

We describe two different direct methods for rational Chebyshev approximation. For simplicity, we call them the "*First Direct Method*" and the "*Second Direct Method.*" The first is an extension of the method which F. D. Murnaghan and J. W. Wrench [12,13] developed for polynomial approximation (also known as the "Second Algorithm" of E. Remes [14]). For reasons to be explained in this paper, we preferred the Second Direct Method for our work at Princeton. It was coded by C. Mesztenyi.

The numbering of sections is a continuation of that used in Part I [8].

6. *Definitions and Notations*

Let $\mathcal{R}(l, m)$ be the set of rational functions

$$\frac{P_l(x)}{Q_m(x)} = \frac{p_0 + p_1 x + \cdots + p_l x^l}{q_0 + q_1 x + \cdots + q_m x^m} \tag{6.1}$$

with real coefficients p_i, q_k. The sum $n = l + m$ will be called the *degree* of the set $\mathcal{R}(l, m)$. We shall use the notation $R_n(x)$ to designate an element of $\mathcal{R}(l, m)$.

Let $f(x)$ be a given function, continuous on the interval $[a, b]$, and let $g(x)$ be a given *weight function*, continuous and positive in $[a, b]$. That rational function $R_n^*(x) \in \mathcal{R}(l, m)$ for which[1]

$$\max_{[a,b]} \frac{|R_n(x) - f(x)|}{g(x)} = \min \tag{6.2}$$

is called the *Chebyshev-approximant*, or *best-fit* rational function with respect to the weight function $g(x)$.

The weight function allows us the option of minimizing either the absolute or the relative error for $g(x) \equiv 1$ and $g(x) \equiv |f(x)|$, respectively. Another possibility is to minimize the absolute error of the reciprocal:

$$\max_{[a,b]} \left| \frac{1}{R_n(x)} - \frac{1}{f(x)} \right| = \min.$$

[1] "= min" stands for *assumes its infimum* or *is minimized.*

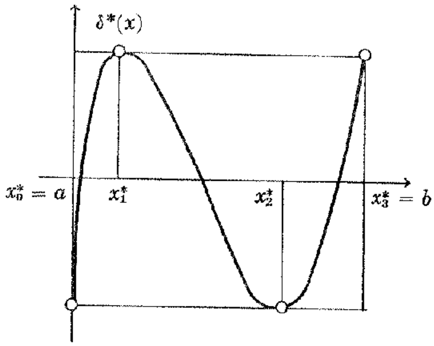

FIG. 1

Here we choose[2] $g(x) \equiv f^2(x)$. Moreover, weight functions may be employed to treat the approximation of odd functions (Part III, Section 14).

Each rational function $R_n(x)$ gives rise to an *error curve* or *error function*:

$$\delta(x) = \frac{R_n(x) - f(x)}{g(x)}. \tag{6.3}$$

The error curve of the best-fit approximant will be written with an asterisk:

$$\delta^*(x) = \frac{R_n^*(x) - f(x)}{g(x)}. \tag{6.4}$$

Extending the theorem of Chebyshev to fractional approximants, N. Achieser [1] has shown that $R_n^*(x)$ is uniquely characterized by $\delta^*(x)$ assuming its maximum absolute value sufficiently often with alternating signs. Arguments x_i^* for which the maximum absolute value is assumed are called *critical points*.

In most of the practical applications, the error curve will have *standard form*; that is, it will meet the following additional requirements:

(i) there are exactly $n+2$ critical points $x_0^* < \cdots < x_{n+1}^*$;

(ii) x_0^* and x_{n+1}^* coincide with the endpoints, i.e. $x_0 = a, \quad x_{n+1}^* = b$;

(iii) $\dfrac{d}{dx}\delta^*(x)$ is continuous, and vanishes only for $x = x_i^*$, $i = 1, \cdots, n$.

Figure 1 shows a standard form error curve for $n = 2$. If an error curve $\delta(x)$ has standard form, then it is necessarily the optimal error curve $\delta^*(x)$, that is, it corresponds to the Chebyshev-approximant $R_n^*(x)$. However, the optimal error curve need not have standard form (compare [9, 10, 19]). Every method for fitting rational approximations discussed in this part is based on the assumption that $\delta^*(x)$ has standard form. Nonstandard error curves, in particular the case of odd functions, are discussed in Part III, Section 14.

[2] Actually, one should use $g(x) \equiv f^*(x)R_n^*(x)$ where $(R_n^*(x))^{-1}$ is the Chebyshev-approximant of $f^{-1}(x)$. In practical computation, however, the replacement of $R_n^*(x)$ by $f(x)$ is permissible since it affects the error curve $\delta(x)$ less than roundoff does.

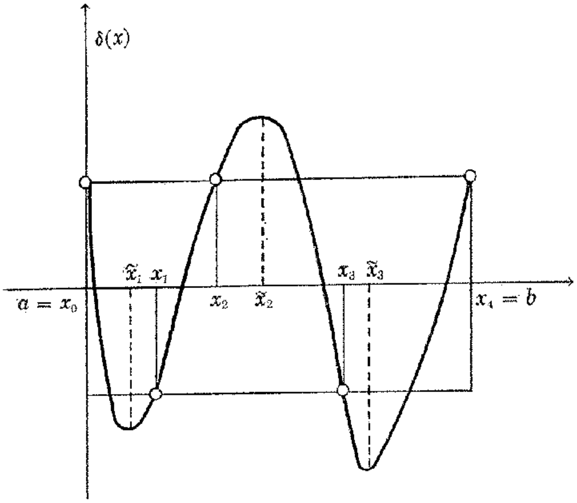

FIG. 2

Every standard error curve and its critical points constitute the unique solution $\delta(x) = \delta^*(x)$, $x_i = x_i^*$ of the following equations:

$$\delta(x_i) = (-1)^i \lambda, \quad i = 0, \cdots, n+1 \qquad \delta'(x_i) = 0, \quad i = 1, \cdots, n,$$

or

$$R_n(x_i) - f(x_i) - (-1)^i \lambda g(x_i) = 0, i = 0, \cdots, n+1 \quad (6.5)$$

$$(R_n'(x_i) - f'(x_i))g(x_i) - (R_n(x_i) - f(x_i))g'(x_i) = 0, \quad i = 1, \cdots, n. \quad (6.6)$$

There are $2n+2$ equations and the same number of unknowns, namely the interior critical points $x_1, \cdots, x_n$, the $n+1$ parameters determining $R_n(x)$, and the *error amplitude* λ. Thus our problem consists in solving the above system of nonlinear equations. Note that λ need not be positive.

7. *First Direct Method*

This method is a two-stage iteration method.

Stage I. A rational function $R_n(x) \in \mathfrak{R}(l, m)$ and an amplitude λ are calculated which, for a given set of arguments (guess at critical points) $a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b$, solve the equations (6.5).

Stage II. The error curve $\delta(x)$ corresponding to the $R_n(x)$ computed in Stage I will behave somewhat as indicated in Figure 2. The extrema of $\delta(x)$ then yield a new guess for the critical points $a = \tilde{x}_0 < \tilde{x}_1 < \cdots < \tilde{x}_n < \tilde{x}_{n+1} = b$, on which the subsequent Stage I step is based. Thus Stage II solves the equations (6.6) for given $R_n(x)$.

In general, the solution of the equations in each step of both stages will require an iterative procedure. We have iterations within an iteration. Characteristically for this kind of situation, the efficiency of our method will depend fundamentally on properly balancing these iterations. It would be wasteful to carry the inner iterations to a high accuracy at every step. On the other hand, the accuracy of the inner iterations must be high enough so as not to spoil the accuracy already attained in the overall iteration.

For Stage II we recommend a simple searching procedure. This procedure will also be used in the Second Direct Method, and will be described in Section 9. The remainder of this section will be devoted to Stage I. Although Stage II poses fewer theoretical problems, it requires more computational work than Stage I, because most of the evaluations of the function $f(x)$ occur in this stage.

The problem of the first stage is to determine a rational function $R_n(x) \in \mathcal{R}(l, m)$ and an amplitude λ such that for given arguments $x_0 < x_1 < \cdots < x_{n+1}$, ordinates $y_i = f(x_i)$ and weights $w_i = g(x_i) > 0$ the conditions

$$R_n(x_i) = y_i + (-1)^i \lambda w_i \quad i = 0, \cdots, n+1 \tag{7.1}$$

are fulfilled. We might call such a problem an "interpolation problem with weighted deviations." It is closely related to solving the system

$$P_l(x_i) - (y_i + (-1)^i \lambda w_i) Q_m(x_i) = 0 \quad i = 0, \cdots, n+1, \tag{7.2}$$

where $Q_m(x)$ and $P_l(x)$ denote the denominator and numerator, respectively, of $R_n(x)$. For each particular value of λ, this is a homogeneous linear system of $n+2$ equations for the $n+2$ coefficients of $P_l(x)$ and $Q_m(x)$. It has a nonzero solution if and only if its determinant vanishes. This determinant turns out to be a polynomial of degree $m+1$ in λ, and for each real root of this equation we get a real solution of (7.2). We choose that solution which corresponds to the real root of smallest absolute value.

We have not investigated all the questions which arise in connection with the problem (7.1). As with interpolation by rational functions (compare [10]) a solution of (7.2) does not necessarily yield a solution of (7.1). All solutions of (7.2) will lead to the same rational function, but there may be inaccessible points [10]. Also, the existence of real roots λ has not been established in general.

Moreover, a solution $R_n(x)$ *without poles* in $[a, b]$ is required for Stage II. We show that there can be at most one real solution of (7.1) without poles in $[a, b]$. Let $P_l(x)/Q_m(x)$, $S_l(x)/T_m(x)$ be two different solutions of (7.1) with the amplitudes λ and μ, respectively. λ and μ must be different, for otherwise $P_l(x)/Q_m(x)$ and $S_l(x)/T_m(x)$ would be equal, too. For x_i, $i=0, \cdots, n+1$, the difference

$$\frac{P_l(x)}{Q_m(x)} - \frac{S_l(x)}{T_m(x)} = \frac{P_l(x)T_m(x) - Q_m(x)S_l(x)}{Q_m(x)T_m(x)} \tag{7.3}$$

assumes the values $(-1)^i w_i(\lambda - \mu) \neq 0$, thus being subjected to at least $n+1$ changes of sign. The numerator $P_l(x)T_m(x) - Q_m(x)S_l(x)$ of (7.3) is a polynomial of degree not greater than n, and can account for at most n changes of sign. This leaves at least one change of sign to the denominator $Q_m(x)T_m(x)$, that is $P_l(x)/Q_m(x)$ or $S_l(x)/T_m(x)$ has a pole in $[a, b]$.[3]

[3] This proof and the following examples have been inserted by the writer, elaborating on short discussions with H. J. Maehly and J. Stoer.

It is probably quite safe to assume that in most applications there exists a bounded solution, and that it corresponds to the smallest real root λ. This is, however, not necessarily true, as is illustrated by the following examples.

If $l = 0$, $m = 1$, $y_0 = x_0 = -1$, $y_1 = x_1 = 0$, $y_2 = x_2 = 1$, $w_0 = w_1 = w_2 = 1$, then

$$R_1(x) \equiv \frac{1}{2x - \sqrt{2}}, \quad \lambda = \frac{1}{\sqrt{2}} \quad \text{and} \quad R_1(x) \equiv \frac{1}{2x + \sqrt{2}}, \quad \lambda = \frac{-1}{\sqrt{2}}$$

are the two solutions of (7.1). Both have poles in the interval $[-1, +1]$.

If $l = 0$, $m = 1$, $x_0 = 1$, $x_1 = 2$, $x_2 = 6$, $y_0 = 4$, $y_1 = 5$, $y_2 = -1$, $w_0 = w_1 = w_2 = 1$, then

$$R_1(x) \equiv \frac{6}{-x + 3}, \quad \lambda = -1 \qquad \text{and} \qquad R_1(x) \equiv \frac{6}{x}, \quad \lambda = 2$$

are the two solutions. The second solution is pole-free in $[1, 6]$, but the absolute value of its amplitude λ is larger.

If $m \geqq 2$, then the determination of the polynomial equation for λ poses some problems, too. Taking into account these facts, we must admit that our way of dealing with Stage I is not quite satisfactory. This is one of our reasons for preferring the Second Direct Method, described in Section 9.

There is another argument against the First Direct Method. Obviously, small changes of the arguments $x_1, \cdots, x_n$ do not affect the error curve $\delta(x)$ very much, since $x_1, \cdots, x_n$ are essentially stationary points of $\delta(x)$. Putting it the other way around, the computation of the coefficients of $R_n(x)$ from the arguments $x_1, \cdots, x_n$ is unstable.[4]

If $f(x)$ and $g(x)$ both have continuous second derivatives, then we may, at least theoretically, apply Newton's method to solve the equations (6.6)

$$\left(\frac{R_n - f}{g}\right)^1 (x_i) = 0 \qquad i = 1, \cdots, n,$$

thereby accomplishing Stage II. If we choose to solve also the equation for λ in Stage I by Newton's method, then the resulting algorithm will be quite close to the $(2n+2)$-dimensional Newton procedure applied to the system of equations (6.5) and (6.6). Since we have quadratic convergence in this case, we may expect the same for the First Direct Method. This expectation has been confirmed by F. D. Murnaghan and J. W. Wrench [12] and Veidinger [17] for approximation by polynomials. For fractionals it is still an open question. Whether there

[4] This argument was the subject of a controversy between H. J. Maehly and the writer. The writer admits that there is an instability, but he considers it a harmless one. The Chebyshev-approximant itself depends in an unstable way on the given data. In other words, two equally good approximants may possess quite different coefficients. The writer presumes that this is the same kind of instability which has been encountered above. This opinion is corroborated by the experiences W. Fraser and J. F. Hart [5] and J. Stoer [15] had with their recent realizations of the First Direct Method for general rational approximants. The writer is greatly pleased by the referee's presenting a neat theoretical argument pertaining to this instability problem. This argument is to be found at the end of this paper.

is quadratic convergence or not is actually of little practical interest. Convergence can be expected only if one starts with a reasonably good guess at the initial arguments $x_0, x_1, \cdots, x_{n+1}$. But then four or five iteration steps will yield a satisfactory approximant.

8. *Transformation of Fractional into Polynomial Approximation*

The First Direct Method described in the preceding section is an extension of the method of F. D. Murnaghan and J. W. Wrench for approximation by polynomials [12, 13]. In the polynomial case, Stage I presents no particular problem since the equation for the amplitude λ will be linear. Therefore, it might be useful to note that fitting a rational approximant $R_n^*(x)$ can be reduced to fitting a sequence of polynomials of the same degree. We give a short description of this technique, since it will be applied in Part III. Also, it should not be confused with related techniques employed by H. L. Loeb [6], L. Wittemeyer [18] and others.

We start out choosing two fixed *reference polynomials* $\bar{P}_l(x)$, $\bar{Q}_m(x)$. These polynomials should have no common factor of positive degree and at least one of them should actually reach the degree l or m, respectively. In other words, the representation of the rational function $\bar{P}_l(x)/\bar{Q}_m(x)$ in $\mathcal{R}(l, m)$ should be unique up to a constant multiple. We may write for the optimal error curve

$$\begin{aligned}\delta^*(x) &= \frac{1}{g(x)}\left[\left(\frac{P_l^*(x)}{Q_m^*(x)} - \frac{\bar{P}_l(x)}{\bar{Q}_m(x)}\right) - \left(f(x) - \frac{\bar{P}_l(x)}{\bar{Q}_m(x)}\right)\right] \\ &= \frac{(P_l^*(x)\bar{Q}_m(x) - Q_m^*(x)\bar{P}_l(x)) - Q_m^*(x)(f(x)\bar{Q}_m(x) - \bar{P}_l(x))}{g(x)Q_m^*(x)\bar{Q}_m(x)}\end{aligned} \tag{8.1}$$

where $R_n^*(x) = P_l^*(x)/Q_m^*(x)$ denotes the Chebyshev-approximant of $f(x)$ with the weight function $g(x)$. Hence $\delta^*(x) = (E_n^*(x) - F(x))/H(x)$ is also the error curve of the polynomial

$$E_n^*(x) = P_l^*(x)\bar{Q}_m(x) - Q_m^*(x)\bar{P}_l(x) \tag{8.2}$$

with respect to the function

$$F(x) = Q_m^*(x)(f(x)\bar{Q}_m(x) - \bar{P}_l(x)) \tag{8.3}$$

and the weight function

$$H(x) = g(x)Q_m^*(x)\bar{Q}_m(x). \tag{8.4}$$

Since $\delta^*(x)$ has standard form, $E_n^*(x)$ is the best-fit polynomial. Thus we have reduced our rational approximation problem to a problem of polynomial approximation, which may be attacked, for instance, by the method of Murnaghan and Wrench.

At the beginning of the iteration however, $Q_m^*(x)$ is not known. Hence we have to start with a guess for $Q_m^*(x)$: $Q_m(x) \approx Q_m^*(x)$. Then one step of the method of Murnaghan and Wrench is taken in the direction of the best-fit polynomial $E_n^*(x)$, depending on the particular choice of $Q_m(x)$. The resulting polynomial $E_n(x)$ determines new polynomials $P_l(x)$ and $Q_m(x)$ by virtue of the linear system

$$E_n(x) = P_l(x)\bar{Q}_m(x) - Q_m(x)\bar{P}_l(x). \tag{8.5}$$

The new $Q_m(x)$ so obtained enters the subsequent iteration step.

The linear system (8.5) has $n+1$ equations and $n+2$ unknowns. Hence there is one degree of freedom left. This is due to the fact that $P_l(x)$ and $Q_m(x)$ are determined only up to a common multiple. Therefore one has to add a scaling condition such as $Q_m(a) = 1$, or require that the constant term of $Q_m(x)$ should not be changed, etc. The linear system (8.5) is nonsingular if and only if the reference polynomials satisfy the above requirements, namely that the rational function $\bar{P}_l(x)/\bar{Q}_m(x)$ have a unique representation in $\mathcal{R}(l, m)$. By definition, the system (8.5) is nonsingular if and only if the solution space of the corresponding homogeneous system is one-dimensional. But if $R_l(x), S_m(x)$ solve $0 = R_l(x)\bar{Q}_m(x) - S_m(x)\bar{P}_l(x)$, then $R_l(x)/S_m(x) = \bar{P}_l(x)/\bar{Q}_m(x)$ holds. Thus nonsingularity of (8.5) and uniqueness of the representation $\bar{P}_l(x)/\bar{Q}_m(x)$ are equivalent.

Of course, the reference polynomial $\bar{Q}_m(x)$ should be positive throughout the interval $[a, b]$. Moreover, forming the difference $f(x)\bar{Q}_m(x) - \bar{P}_l(x)$ should not cause any substantial loss of accuracy, that is $\bar{P}_l(x)/\bar{Q}_m(x)$ should *not* be a good approximant of $f(x)$. This qualification will be unnecessary if some other way of evaluating the above difference is available. This will be the case for the Combined Methods procedure to be described in Part III.

9. *Second Direct Method*

If $\delta^*(x)$ has standard form, then it has exactly $n+1$ zeros $z_0^* < z_1^* < \cdots < z_n^*$ within the interval $[a, b]$. It may therefore be written in the form

$$\delta^*(x) = G(x) \prod_{k=0}^{n} (x - z_k^*). \tag{9.1}$$

Characterizing $\delta(x)$ by its zeros rather than by its extrema, avoiding the instability mentioned in the preceding section, has been suggested by K. Arbenz[5] (compare [7]). We call the resulting method for fitting rational approximations the *Second Direct Method.*

The Second Direct Method is again a two-stage iteration method.

Stage I. Let $a < z_0 < \cdots < z_n < b$ be a guess at the zeros of the optimal error curve $\delta^*(x)$. To this guess there corresponds an approximant, and therefore an error curve $\delta(x)$, by virtue of the condition

$$R_n(z_k) = f(z_k), \quad k = 0, \cdots, \text{n}. \tag{9.2}$$

Determining $R_n(x)$ requires only straightforward rational interpolation.

[5] The proposal of Dr. Arbenz as to how to correct the zeros was, however, not quite satisfactory and entirely different from Dr. Maehly's proposal described in this section.

Stage II. The interior extrema $x_1 < \cdots < x_n$ are computed and used to correct the zeros of $\delta(x)$. These corrected zeros $a < \tilde{z}_0 < \cdots < \tilde{z}_n < b$ then reenter Stage I.

It is a decisive advantage of this method that Stage I does not present any theoretical problems: any method for rational interpolation may be chosen. Hence we shall concentrate on the description of Stage II.

We describe a method for determining the extrema x_i later in this section, but consider first the main question: how to correct the zeros z_k using the extrema x_i. For a short variational argument let us denote the error curve by $\epsilon(x)$ instead of $\delta(x)$. Then (9.1) gives (note $d(\ln|w|)/dw = 1/w$ for $w \neq 0$):

$$\delta \ln | \epsilon | = \sum_{k=0}^{n} \left(\frac{\partial \ln | G |}{\partial z_k} - \frac{1}{x - z_k} \right) \delta z_k + \frac{\partial \ln | \epsilon |}{\partial x} \delta x. \qquad (9.3)$$

We are particularly interested in the variation of the extreme values of $\epsilon(x)$. These are assumed either at the ends of our interval, whence $\delta x = 0$, or they are relative extrema, whence $\partial|\epsilon|/\partial x = 0$. In each case, we get

$$\delta \ln | \epsilon_i | = \sum_{k=0}^{n} \left(\frac{\partial \ln | G |}{\partial z_k} - \frac{1}{x - z_k} \right) \delta z_k , \qquad i = 0, \cdots , n+1,$$

where $\epsilon_i = \epsilon(x_i)$.

The correction we propose is based on the assumption that *if the error curve $\epsilon(x)$ is almost optimal, then the function $G(x)$ does not depend very much on the zeros z_i*, at least not in the neighborhood of an extremum x_i. This leads to the following expressions:

$$\delta \ln | \epsilon_i | = - \sum_{k=0}^{n} \frac{\delta z_k}{x_i - z_k}, \qquad i = 0, \cdots , n+1. \qquad (9.4)$$

We want the absolute values of the extrema $\epsilon_i = \epsilon(x_i)$ to be equal. Therefore we put

$$\ln|\epsilon_i| + \delta \ln|\epsilon_i| = \ln |\lambda|, \qquad i = 0, \cdots , n+1.$$

Substituting this result in (9.4) we obtain the following system of $n+2$ linear equations for the $n+2$ unknowns $\ln|\lambda|, \delta z_0 , \cdots , \delta z_n$:

$$\ln | \lambda | + \sum_{k=0}^{n} \frac{\delta z_k}{x_i - z_k} = \ln | \epsilon_i |, \qquad i = 0, \cdots , n+1. \qquad (9.5)$$

Eliminating $|\lambda|$ by subtracting (9.5) with $i = 0$ yields

$$\sum_{k=0}^{n} \left(\frac{1}{x_i - z_k} - \frac{1}{x_0 - z_k} \right) \delta z_k = \ln \left| \frac{\epsilon_i}{\epsilon_0} \right|, \qquad i = 1, \cdots , n+1,$$

and using the approximation

$$\ln \left| \frac{\epsilon_i}{\epsilon_0} \right| \cong 2 \frac{| \epsilon_i | - | \epsilon_0 |}{| \epsilon_i | + | \epsilon_0 |}$$

we arrive at the system

$$\sum_{k=0}^{n} \frac{(x_0 - x_i)\delta z_k}{(x_i - z_k)(x_0 - z_k)} = 2\frac{|\epsilon_i| - |\epsilon_0|}{|\epsilon_i| + |\epsilon_0|}, \qquad i = 1, \cdots, n+1. \tag{9.6}$$

Either (9.5) or (9.6) can be used for calculating the variations δz_k . Both systems are well-conditioned since they have their largest elements close to the diagonal.

We still have to describe the method we used for the determination of the extrema of the error curve $\delta(x)$, which inevitably carries "noise." This precludes the numerical computation of the derivatives $\delta'(x)$ and $\delta''(x)$. Hence Newton's Method cannot be applied. We settled for just searching in equal distances for the largest (smallest) value. The searching was stopped as soon as a value $\delta(\bar{x})$ was found which surpassed its neighbors $\delta(\bar{x} - h)$, $\delta(\bar{x} + h)$. Then these three points were interpolated by a parabola whose extremum x was chosen

$$\tilde{x} = \bar{x} - \left(\frac{\delta(\bar{x} + h) - \delta(\bar{x} - h)}{\delta(\bar{x} + h) - 2\delta(\bar{x}) + \delta(\bar{x} - h)}\right)\frac{h}{2}.$$

Since the bulk of the computation in the direct methods consists in evaluating the function $f(x)$, considerable care must be given to the selection of the searching distance h. In many practical applications, the error curve is approximately proportional to the $(n+1)$-th Chebyshev-polynomial transplanted from the interval $[-1, +1]$ to the interval $[a, b]$. In these cases, C. Mesztenyi chooses different increments h for each extremum during one Stage II step. These increments are varied in proportion to the distance of the two flanking Chebyshev-zeros.

II-Conclusions

Both direct methods can be regarded as modifications of interpolation algorithms, viz., the Second Direct Method as a modification of straightforward rational interpolation, the First Direct Method as a modification of "interpolation with weighted deviations." Since the latter interpolation problem seems still to await a completely satisfying solution, the Second Direct Method may be preferable. In numerical applications of the Second Direct Method, our scheme of correcting the zeros, based on the assumption of small variation of $G(x)$, has proved to be quite successful.

III. INDIRECT AND COMBINED METHODS

III-Introduction

The direct methods compute the error curve $\delta(x) = R_n(x) - f(x)$ of a rational approximant $R_n(x)$ "directly," that is, the functions $f(x)$ and $R_n(x)$ are evaluated independently and subtracted afterwards. Two or three digits should be saved for the determination of the error curve. Hence the accuracy required for fitting is necessarily two or three digits larger than the accuracy of the approximation obtained. In general, one needs double precision to effect a single-precision best fit.

This restriction can be avoided if a good approximant $\bar{R}_n(x)$ with single-precision coefficients is known beforehand. Then the idea is to compute corrections to these coefficients, again in single precision. The final addition of the corrections has to be carried out, of course, in multiple precision. We call methods which correct the coefficients of a given approximant *indirect methods.*

The indirect methods to be described in this part require representation of the function $f(x)$ to the full accuracy required by a finite continued fraction with single-precision coefficients:

$$f(x) = \frac{\alpha_0|}{|b_0} + \frac{\alpha_1 x|}{|b_1} + \cdots + \frac{\alpha_N x|}{|b_N}.$$

This function is to be approximated by a rational function $R_n^*(x)$ with $n < N$ in the sense of Chebyshev. Like the Telescoping Procedures treated in Part I [8], this can be regarded as a problem of "economizing continued fractions." However, as in Part II a "true" Chebyshev-approximant is computed.

The case $N = n+1$ is quite common. The original "Indirect Method" coded by C. Mesztenyi refers to it. Since a detailed description of the original Indirect Method can be found in [7], we restrict our treatment to the "combined" methods, which deal with the general case $N > n$. These methods employ any direct method to determine the correction to the Padé-approximant in order to get the Chebyshev-approximant. Thus combined methods may be viewed as modifications of direct methods. We shall distinguish the *First Combined Method* and the *Second Combined Method,* depending on whether the underlying method is the First– or the Second Direct Method. The First Combined Method for fractionals, which will be described in Section 12, has been coded by C. Witzgall. C. Mesztenyi programmed the Second Combined Method for polynomials.

Finally, we want to emphasize the fact that both the original Indirect Method and the Combined Methods were effectively used for large scale production of rational approximations [11].

10. *Combined Methods for Polynomials*

The combined methods for polynomial approximation are very simple and may serve as an introduction to the more complicated fractional case.

We assume that $f(x)$ is given as a high degree polynomial

$$f(x) = \sum_{k=0}^{N} c_k x^k \tag{10.1}$$

within the interval $[0, b]$. We are looking for the polynomial $P_n^*(x)$ which approximates the function $f(x)$ best in the sense of Chebyshev, that is, the polynomial for which

$$\max_{[0,b]} \frac{|P_n^*(x) - f(x)|}{g(x)} = \min$$

holds, where $g(x) > 0$ is the weight function.

If b is sufficiently small, then the nth Padé-polynomial $\bar{P}_n(x) = \sum_{k=0}^{n} c_k x^k$ is a good approximant for $f(x)$. Therefore the coefficients of $\Delta P_n(x) = P_n^*(x) - \bar{P}_n(x)$ will be considerably smaller in magnitude than the coefficients c_k, $k \leqq n$, of $P_n^*(x)$. Hence it will frequently be sufficient to compute $\Delta P_n(x)$ in single precision in order to achieve double precision for $P_n^*(x)$.

$P_n^*(x)$ is characterized by the error curve $\delta^*(x) = (P_n^*(x) - f(x))/g(x)$ having standard form. Now we may write as well

$$\delta^*(x) = \frac{\Delta P_n(x) - (f(x) - \bar{P}_n(x))}{g(x)} .$$

This shows that $\Delta P_n(x)$ is the Chebyshev-approximant of the function $f(x) - P_n(x)$ with respect to the weight function $g(x)$.

Now the crucial point is to find a method for evaluating the difference $f(x) = f(x) - \bar{P}_n(x)$ without substantial loss of accuracy. In the polynomial case, the solution of this problem is simple enough: $F(x)$ is just the "tail" of the polynomial (10.1). Hence

$$F(x) = \sum_{k=n+1}^{N} c_k x^k. \tag{10.2}$$

We may now employ any direct method for determining a single-precison best-fit polynomial $\Delta P_n(x)$ of the function $F(x)$ given by (10.2). $\Delta P_n(x)$ then is added to the Padé-Approximant $\bar{P}_n(x)$.

11. *Formula for $f(x) - R_n(x)$*

Consider a continued fraction

$$f(x) = \frac{\alpha_0|}{|b_0} + \frac{\alpha_1 x|}{|b_1} + \cdots + \frac{\alpha_N x|}{|b_N} .$$

Again, the combined methods consist in splitting off the nth Padé-approximant

$$\bar{R}_n(x) = \frac{\alpha_0|}{|b_0} + \frac{\alpha_1 x|}{|b_1} + \cdots + \frac{\alpha_n x|}{|b_n} .$$

As before, the crucial point will be to find for $\bar{R}_n(x) - f(x)$ a formula which can be evaluated without loss of accuracy.

To solve this problem we refer to the basic formulae derived in Part I, Section 3. There the convergents of a continued fraction

$$\frac{a_0|}{|b_0} + \frac{a_1|}{|b_1} + \cdots$$

were expressed in the form

$$\frac{P_\nu}{Q_\nu} = \frac{a_0|}{|b_0} + \frac{a_1|}{|b_1} + \cdots + \frac{a_\nu|}{|b_\nu} ;$$

where P_ν and Q_ν are determined by the recursion formulae (3.2)

$$P_\nu = a_\nu P_{\nu-2} + b_\nu P_{\nu-1}$$
$$Q_\nu = a_\nu Q_{\nu-2} + b_\nu Q_{\nu-1}$$

starting with (3.1): $P_{-2} = 1, \quad P_{-1} = 0, \quad Q_{-2} = 0, \quad Q_{-1} = 1.$

We proceed to derive an expression for the difference $P_n/Q_n - P_n/Q_n, \quad N > n,$ between two convergents of the same continued fraction. Putting

$$f_{n+1,N} = \frac{a_{n+1}\rvert}{\lvert b_{n+1}} + \frac{a_{n+2}\rvert}{\lvert b_{n+2}} + \cdots + \frac{a_N\rvert}{\lvert b_N},$$

we may write

$$\frac{P_N}{Q_N} = \frac{a_0\rvert}{\lvert b_0} + \frac{a_1\rvert}{\lvert b_0} + \cdots + \frac{a_n\rvert}{\lvert b_n} + \frac{f_{n+1,N}\rvert}{\lvert 1}. \tag{11.4}$$

Hence all we need is an expression for the difference between two successive convergents. According to (3.7) this expression reads

$$\frac{P_{\nu+1}}{Q_{\nu+1}} - \frac{P_\nu}{Q_\nu} = \frac{P_{\nu+1}\, Q_\nu - Q_{\nu+1}\, P_\nu}{Q_{\nu+1}\, Q_\nu}$$
$$= \frac{1}{Q_{\nu+1}\, Q_\nu} \prod_{k=0}^{\nu+1} (-a_k) = \frac{1}{Q_\nu \left(Q_{\nu-1} + \dfrac{b_{\nu+1}}{a_{\nu+1}}\, Q_\nu\right)} \prod_{k=0}^{\nu} (-a_k).$$

Applied to (11.4), this formula yields

$$\frac{P_N}{Q_N} - \frac{P_n}{Q_n} = \frac{1}{Q_n \left(Q_{n-1} + \dfrac{Q_n}{f_{n+1,N}}\right)} \prod_{k=0}^{n} (-a_k).$$

For

$$a_0 = \alpha_0, \qquad a_1 = \alpha_1\, x, \; \cdots, \; a_n = \alpha_n\, x,$$
$$f_{n+1,N}(x) = \frac{\alpha_{n+1}\, x\rvert}{\lvert b_{n+1}} + \frac{\alpha_{n+2}\, x\rvert}{\lvert b_{n+2}} + \cdots + \frac{\alpha_N\, x\rvert}{\lvert b_N},$$
$$S_l(x) = Q_{n-1}, \qquad \bar{Q}_m(x) = Q_n,$$

we get finally

$$f(x) - \bar{R}_n(x) = \frac{(-x)^{n+1}}{\bar{Q}_m(x)\left(S_l(x) + \dfrac{\bar{Q}_m(x)}{f_{n+1,N}(x)}\right)} \prod_{k=0}^{n} \alpha_k. \tag{11.5}$$

This is the desired expression.

12. *Combined Method for Fractional Approximations*

In this section we describe the combined method which arises from the First Direct Method. Except for a few details, this description applies also to other combined methods.

The Combined Method essentially amounts to choosing the polynomials $\bar{P}_l(x)$ and $\bar{Q}_m(x)$, whose quotient is the Padé-approximant $\bar{R}_n(x)$, as reference polynomials, and proceeding as outlined in Section 8. There the error curve $\delta^*(x) = R_n^*(x) - f(x))/g(x)$ was conceived as the error curve of the polynomial

$$E_n^*(x) = P_l^*(x)\bar{Q}_m(x) - Q_m^*(x)\bar{P}_l(x) \tag{12.1}$$

approximating $F(x) = Q_m^*(x)\ (f(x)Q_m^*(x) - \bar{P}_l(x))$ with $H(x) = g(x)Q_m^*(x) \cdot Q_m(x)$ as a weight function.

Recall that the direct approach described in Section 8 required that $\bar{P}_l(x)/\bar{Q}_m(x)$ *not* be a good approximation to $f(x)$. Now it is just the other way around: we insist that $\bar{P}_l(x)/\bar{Q}_m(x)$ be a good approximation. It is, of course, formula (11.5) that makes the difference by enabling us to avoid separate evaluation of $f(x)$ and $\bar{R}_n(x)$.

Substituting the correction polynomials $\Delta P_l(x) = P_l^*(x) - \bar{P}_l(x)$, $\Delta\bar{Q}_m(x) = Q_m^*(x) - \bar{Q}_m(x)$ into (12.1) we get

$$E_n^*(x) = \Delta P_l(x)\bar{Q}_m(x) - \Delta Q_m(x)\bar{P}_l(x). \tag{12.2}$$

Thus the corrections $\Delta P_l(x)$ and $\Delta Q_m(x)$ can be computed directly from (12.2).

Equation (12.2) does not characterize $\Delta P_l(x)$ and $\Delta Q_m(x)$ completely. As we have seen in Section 8 there is one degree of freedom left, reflecting the fact that $P_l^*(x)$ and $Q_m^*(x)$ are determined only up to a common multiple by their quotient $R_n^*(x)$. Therefore we normalize $P_l^*(x)$ and $Q_m^*(x)$ by requiring that $Q_m^*(x)$ have the same constant term as the Padé-polynomial $\bar{Q}_m(x)$. This leads to the additional relation

$$\Delta Q_m(0) = 0, \tag{12.3}$$

fixing $\Delta Q_m(x)$ and $\Delta P_l(x)$.

Our algorithm involves a two-stage iteration, although the two iterations are fused together, so that most of the time the algorithm functions like a one-stage iteration.

Stage I. For a given guess $0 = x_0 < \cdots < x_{n+1} = b$ at the critical points, determine $R_n(x_i) = P_l(x_i)/Q_m(x_i)$ such that

$$R_n(x_i) = (f(x_i) - \bar{R}_n(x_i)) + (-1)^i\lambda g(x_i)$$

holds for all i. This is again our well-known problem of rational interpolation with weighted deviations described in Section 7. Here it is solved, according to the proposal of Section 8, by an iteration procedure involving interpolation with weighted deviations only for polynomials.

This iteration runs as follows: start with a guess at $Q_m(x)$. Compute $E_n(x)$ and λ by solving

$$E_n(x_i) = F(x_i) + (-1)^i\lambda H(x_i). \tag{12.4}$$

Determine $P_l'(x)$ and $Q_m'(x)$ such that

$$E_n(x) = \Delta P_l(x)\tilde{Q}_m(x) - \Delta Q_m(x)\tilde{P}_l(x). \tag{12.5}$$

Replace $Q_m(x)$ by $Q_m(x) + \Delta Q_m(x)$, and restart the iteration.

Stage II. The extrema of $\delta(x) = (E_n(x) - F(x))/H(x)$ are used as a new guess x_i at the critical points, entering Stage I (compare Section 9).

Initial Guess. If the interval $[0, b]$ is not too large, then the error-curve is close to the polynomial which arises from the Chebyshev-polynomial $T_{n+1}(\xi)$ if the interval $[-1, +1]$ is transformed linearly into the interval $[0, b]$ (compare [9, 19]). Therefore we choose the critical points of the transformed Chebyshev-polynomial as an initial guess:

$$x_i = \left(1 - \cos\frac{i\pi}{n+1}\right)\frac{b}{2}, \qquad i = 0, \cdots, n+1.$$

As an initial guess at $Q_m^*(x)$ we choose $\tilde{Q}_m(x)$, the denominator of the Padé-approximant $R_n(x)$.

Experience has shown our initial guess at $Q_m^*(x)$ to be much poorer than our initial guess at the critical points. This suggests the following iteration pattern: start with two or three steps of the Stage I iteration, keeping the initial x_i. By then $Q_m(x)$ will be close enough to $Q_m^*(x)$ so that for the sequel one step of the Stage I iteration may be combined with one step of the Stage II iteration. The combined iteration step then consists in determining $E_n(x)$ by (12.4), using the resulting error curve for a Stage II correction of x_i, and finally correcting $Q_m(x)$.

For the determination of $E_n(x)$ by (12.4) C. Witzgall suggested the following modification of Newton's interpolation method (for another method see [13]). Consider the two polynomials $C_{n+1}(x)$ and $D_{n+1}(x)$ which satisfy

$$\left.\begin{aligned} C_{n+1}(x_i) &= F(x_i) \\ D_{n+1}(x_i) &= (-1)^i H(x_i) \end{aligned}\right\} \qquad i = 0, \cdots, n+1. \tag{12.6}$$

Then we have

$$E_n(x) = C_{n+1}(x) + \lambda D_{n+1}(x),$$

where λ is uniquely determined since the highest powers of $C_{n+1}(x)$ and $D_{n+1}(x)$ must cancel.

Note that the two linear systems (12.6) for the coefficients of $C_{n+1}(x)$ and $D_{n+1}(x)$ differ only with respect to their right-hand sides. Hence both systems can be solved simultaneously. In order to improve the condition of the linear systems, and for the sake of a more stable evaluation, the polynomials are written in Newton form

$$C_{n+1}(x) = \hat{c}_0 + \hat{c}_1(x - x_0) + \cdots + \hat{c}_{n+1}\prod_{i=0}^{n}(x - x_i)$$

$$D_{n+1}(x) = \hat{d}_0 + \hat{d}_1(x - x_0) + \cdots + \hat{d}_{n+1}\prod_{i=0}^{n}(x - x_i).$$

This leads to a triangular linear system with two right-hand sides for the coefficients $\hat{d}_i$ and $\hat{c}_i$. Clearly $\lambda = -\hat{c}_{n+1}/\hat{d}_{n+1}$. The resulting polynomial $E_n(x)$ again has Newton form

$$E_n(x) = \hat{e}_0 + \hat{e}_1\,(x - x_0) + \cdots + \hat{e}_n \prod_{i=0}^{n-1} (x - x_i).$$

This form guarantees stable evaluation of the error curve. For the computation of the corrections $\Delta P_l(x)$, $\Delta Q_m(x)$, however, $E_n(x)$ has to be converted afterwards into the customary polynomial form.

The above method for determining the polynomials $C_{n+1}(x)$ and $D_{n+1}(x)$ is related to the familiar divided-differences algorithm. However, it uses fewer divisions, and is recommended if the division time is larger than 1.5 multiplication times. (For similar considerations compare J. W. Tukey and H. C. Thacher [16].)

13. *Linear System for the Correction Polynomials*

The coefficients of the linear system (12.5) are determined by the coefficients of the reference polynomials $P_l(x)$ and $Q_m(x)$. In most cases, these coefficients display different orders of magnitude. This poses quite a serious problem. We solve it by triangularizing the system (12.5) in such a way that it can be solved without loss of accuracy. This triangularization is an indispensable part of our algorithm and, in fact, of every combined method.

We refer again to the quantities $P_{\mu\nu}$, $Q_{\mu\nu}$ defined in (3.1) and (3.2) for a general continued fraction (11.2). We note that we have in addition to the formulas in Section 3:

$$P_\mu Q_\nu - P_\nu Q_\mu = \prod_{\lambda=0}^{\mu+1} (-a_\lambda) Q_{\mu+2,\,\nu}\,. \tag{13.1}$$

This is true for $\nu = \mu$ and $\nu = \mu+1$, and both sides of (13.1) follow the same recursion for stepping up ν.

For the continued fraction (11.1) the quantities $P_{\mu\nu}$, $Q_{\mu\nu}$ become polynomials $P^{(\mu\nu)}(x)$, $Q^{\mu\nu}(x)$defined by

$$P^{(\mu,\,\mu-1)}(x) = 0 \qquad P^{(\mu,\,\mu)}(x) = \begin{cases} \alpha_\mu x & \text{for} \quad \mu > 0 \\ \alpha_0 & \text{for} \quad \mu = 0 \end{cases}$$

$$Q^{(\mu,\,\mu-1)}(x) = 0 \qquad Q^{(\mu,\,\mu)}(x) = b_\mu$$

and the recursion

$$P^{(\mu\nu)}(x) = \alpha_\nu x\, P^{(\mu,\nu-2)}(x) + b_\nu P^{(\mu,\,\nu-1)}(x)$$

$$Q^{(\mu\nu)}(x) = \alpha_\nu x Q^{(\mu,\,\nu-2)}(x) + b_\nu Q^{(\mu,\,\nu-1)}(x).$$

Putting $P^{(\nu)}(x) = P^{(0\nu)}(x)$, $Q^{(\nu)}(x) = Q^{(0\nu)}(x)$ we have for the polynomials characterizing the Padé-approximant

$$\bar{P}_l(x) = P^{(n)}(x), \qquad \bar{Q}_m(x) = Q^{(n)}(x).$$

In Part I, Section 2, the following representation of the correction polynomials in terms of Padé-polynomials was suggested (2.13a):

$$\Delta P_l(x) = \gamma_0 + x \sum_{k=1}^{n} \gamma_k P^{(k-2)}(x) \qquad \Delta Q_m(x) = x \sum_{k=1}^{n} \gamma_k Q^{(k-2)}(x). \quad (13.2)$$

This representation takes the relation $\Delta Q_m(0) = 0$ into account. Formula (13.1) yields

$$P^{(k-2)}(x)\tilde{Q}_m(x) - Q^{(k-2)}(x)\tilde{P}_l(x) = (-)^k x^{k-1} \prod_{\lambda=0}^{k-1} \alpha_\lambda Q^{(k,n)}(x)$$

for $k = 1, \cdots, n$. By virtue of these formulae we get a triangular linear system for the coefficients γ_k:

$$\sum_{k=0}^{n} \gamma_k \left[(-x)^k \prod_{\lambda=0}^{k-1} \alpha_\lambda Q^{(k,n)}(x) \right] = E_n(x). \qquad (13.3)$$

The γ_k are obtained in the order of their subscripts, that is γ_0 first, then γ_1, etc. For small intervals it follows from (2.13b) that this is also the direction of increasing modulus: $|\gamma_0| < |\gamma_1| < \cdots < |\gamma_n|$. The solving of the triangular system is therefore a stable operation, and may be carried out in single precision. The final computation of $\Delta P_l(x)$ and $\Delta Q_m(x)$, however, requires multi-precision.

14. *Nonstandard Error Curves. Even and Odd Functions*

Even and odd functions $f(x)$ require special consideration if they are approximated on a symmetric interval $[-b, b]$ with respect to an even weight function $g(x)$ (compare Part I, Section 4). In this case, the Chebyshev-approximant of an even function is even, and that of an odd function is odd; since, if $R_n^*(x)$ is the Chebyshev-approximant of an even function, then the same is true for $R^{**}(x) = R^*(-x)$. But the Chebyshev-approximant is unique, which implies $R^*(x) = R^*(-x)$. A similar argument shows that the Chebyshev-approximant of an odd function is odd.

As a consequence, the Chebyshev-approximant $R_n^*(x) \in \mathcal{R}(2\lambda, 2\mu)$ of an even function is also the Chebyshev element in the sets $\mathcal{R}(2\lambda + 1, 2\mu)$, $\mathcal{R}(2\lambda, 2\mu + 1)$, $\mathcal{R}(2\lambda + 1, 2\mu + 1)$. According to the theorem of Achiezer [1], the error curve $\delta^*(x) = (R_n^*(x) - f(x))/g(x)$ must assume its extreme deviation with alternating sign in at least one point more than expected. In other words, $\delta^*(x)$ usually will not have standard form. The same can be shown for the Chebyshev-approximants $R_n^*(x) \in \mathcal{R}(2\lambda + 1, 2\mu)$ of an odd function.

The case of an even function is easily taken care of by introducing the new variable $y = x^2$ and the functions $h(y) = f(x)$, $k(y) = g(x)$. Then $h(y)$ is approximated on $[0, b^2]$ with the weight function $k(y)$. Normally, the corresponding error curve will have standard form.

In the case of an odd function, we consider $h(y) = (1/x)f(x)$, $k(y) = g(x)$. There is no trouble minimizing the relative error of $f(x)$, since the zero of the weight function $g(x) = f(x)$ will cancel. However, minimizing the absolute

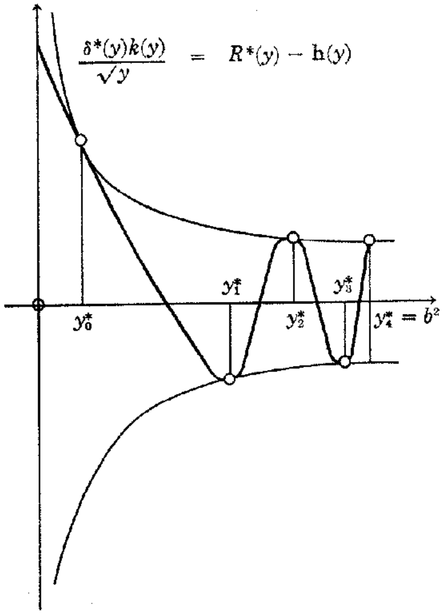

FIG. 3

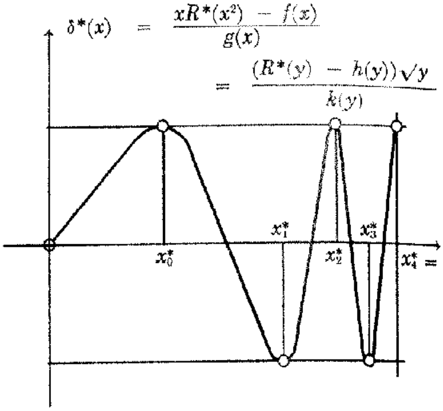

FIG. 4

error of $f(x)$ with respect to a strictly positive weight function $g(x)$ requires a modification of the fitting methods.

The simplest modification consists in choosing the unbounded weight function $(k(y))/\sqrt{y}$ for the approximation of $h(y)$ in [0, b^2]. In this case, the first extremum y_0 lies in the interior of the interval [0, b^2], its exact position unknown (compare Figures 3 and 4). Hence in the First Direct Method and its corresponding combined method, y_0 must be treated in the same way as the other extrema $y_1, \cdots, y_n$ of $\delta(x)$, that is, it must be corrected in Stage II. As an initial guess at the critical points $y_0^*, \cdots, y_{n+1}^*$, we use

$$y_i = \left(b \cos \frac{n + 2 + i}{2n + 3} \pi\right)^2, \qquad i = 0, \cdots, n + 1.$$

This modification was proposed and coded by C. Witzgall for the First Combined Method.

Another scheme does not use the transformation $y = x^2$. It works on the original nonstandard error curve $\delta^*(x) = (xR_n^*(x^2) - f(x))/g(x)$ in the interval $[0, b]$ (not $[-b, b]$). We shall roughly describe the resulting modifications (compare also Part I, Section 3).

In the First Direct Method a guess $0 < x_0 < \cdots < x_{n+1} = b$ at the critical points is used to determine an approximant $xR_n(x^2)$ by solving the following interpolation problem with weighted deviations (Section 7):

$$R_n(x_i^2) = \frac{f(x_i)}{x_i} + (-1)^i\lambda \frac{g(x_i)}{x_i} \qquad i = 0, \cdots, n + 1.$$

In the Second Direct Method one encounters the interpolation problem:

$$z_k R_n(z_k^2) = f(z_k) \qquad k = 0, \cdots, n.$$

In the combined methods, the function $f(x)$ is represented by the continued fraction

$$f(x) = \frac{\alpha_0 \, x \rvert}{\lvert b_0} + \frac{\alpha_1 \, x^2 \rvert}{\lvert b_1} + \cdots + \frac{\alpha_N \, x^2 \rvert}{\lvert b_N} .$$

The quantities $P_{\mu\nu}$, $Q_{\mu\nu}$ become polynomials in x: $P^{(\mu,\nu)}(x)$, $Q^{(\mu,\nu)}(x)$. Putting $P^{(\nu)}(x) = (1/x) \, P^{(0\nu)}(x)$, $Q^{(\nu)}(x) = Q^{(0\nu)}(x)$, $S_l(x^2) = Q^{(n-1)}(x)$, $\bar{Q}_m(x^2) = Q^{(n)}(x)$, $\bar{P}_l(x^2) = P^{(n)}(x)$, we have instead of (11.5):

$$f(x) - x\bar{R}_n(x^2) = \frac{(-1)^{n+1} x^{2n+1}}{\bar{Q}_m(x^2)\left(S_l(x^2) + \dfrac{\bar{Q}_m(x^2)}{f_{n+1,N}(x^2)}\right)} \prod_{k=0}^{n} \alpha_k ,$$

where

$$f_{n+1,N}(x^2) = \frac{\alpha_{n+1} \, x^2 \rvert}{\lvert b_{n+1}} + \frac{\alpha_{n+2} \, x^2 \rvert}{\lvert b_{n+2}} + \cdots + \frac{\alpha_N \, x^2 \rvert}{\lvert b_N} .$$

The polynomial $E(x)$ will be odd, that is, divisible by x. The system (13.3) then is to be replaced by

$$\sum_{k=0}^{n} \gamma_k \left[(-x^2)^k \prod_{\lambda=0}^{k-1} \alpha_\lambda \, Q^{(k,n)}(x) \right] = \frac{1}{x} E(x),$$

with (compare (4.8a))

$$\Delta P_l(x^2) = \gamma_0 + x^2 \sum_{k=2}^{n} \gamma_k \, P^{(k-2)}(x)$$

$$\Delta Q_m(x^2) = x^2 \gamma_1 + x^2 \sum_{k=2}^{n} \gamma_k \, Q(^{(k-2)}(x).$$

III-Conclusions

The indirect methods described in this part have proved to be quite effective for actual computation on a small computer such as the IBM 650. Using single precision only, they yielded up to triple-precision approximations. This is, however, possible only for small intervals, where the Padé-approximant is sufficiently accurate. As the interval gets larger, the combined methods, for instance, will adopt the behavior of direct methods.

The application of the combined methods and of the original Indirect Method requires that the functions $f(x)$ be given by a power series or a continued fraction whose coefficients can be expressed in single precision. Thus the indirect methods described in this part cannot replace the direct methods.

* * *

REFEREE'S NOTE. Footnote No. 4 in Section 7 expresses a difference of opinion between the author and the writer concerning the inherent instability of the First Direct Method. The following argument pertains to this controversy.

The number λ is a function of the trial critical points $(x_1, \cdots, x_n)$ via equation (6.5). We may therefore write $\lambda = \lambda(x_1, \cdots, x_n)$. By Achieser's Generalization of de la Vallée-Poussin's Theorem [1, p. 52], $\lambda(x_1^*, \cdots, x_n^*) \geqq \lambda(x_1, \cdots, x_n)$, and consequently the function $\lambda(x_1, \cdots, x_n)$ has an absolute maximum at $(x_1^*, \cdots, x_n^*)$. Therefore $\partial\lambda/\partial x = 0$ at $(x_1^*, \cdots, x_n^*)$. If a certain nondegeneracy assumption is made about the data, it can be proved that the coefficients of P/Q are *insensitive* to slight changes in $(x_1, \cdots, x_n)$. This proof goes as follows:

Equation 6.5 gives us $(0 \leqq i \leqq n+1)$

$$\sum_{j=0}^{l} p_j x_i^j = [(-1)^i \lambda g(x_i) + f(x_i)] \sum_{j=0}^{m} q_j x_i^j.$$

Differentiation gives us $(k \neq i, \quad 1 \leqq k \leqq n)$

$$\sum_{j=0}^{l} \frac{\partial p_j}{\partial x_k} x_i^j = [(-1)^i \lambda g(x_i) + f(x_i)] \sum_{j=0}^{m} \frac{\partial q_j}{\partial x_k} x_i^j + (-1)^i \frac{\partial \lambda}{\partial x_k} g(x_i) \sum_{j=0}^{m} q_j x_i^j.$$

Evaluating at $(x_1^*, \cdots, x_n^*)$ and remembering $\dfrac{\partial\lambda}{\partial x_k} = 0$ we see that the vector

$$\mathbf{U} = \left(\frac{\partial p_0}{\partial x_k}, \cdots, \frac{\partial p_l}{\partial x_k}, \frac{\partial q_0}{\partial x_k}, \cdots, \frac{\partial q_m}{\partial x_k}\right)$$

(in which the derivatives are evaluated at $(x_1^*, \cdots, x_n^*)$) is orthogonal to the $n+1$ vectors

$$\mathbf{V}_i = (1, x_i^*, (x_i^*)^2, \cdots, (x_i^*)^l, c_i, c_i x_i^*, \cdots, c_i(x_i^*)^m),$$

where $0 \leqq i \leqq n+1$, $i \neq \mathrm{k}$, $c_i = -[(-1)^i \lambda g(x_i^*) + f(x_i^*)]$. The set of vectors $\{\mathbf{V}_0, \cdots, \mathbf{V}_{n+1}\}$ is linearly dependent because of the choice of λ. Make the nondegeneracy assumption that every set of $n+1$ vectors $\mathbf{V}_i$ is *independent*. Then $\mathbf{U}$ is orthogonal to a set of vectors having rank $n+1$. If the approximation is *normalized*, say, by fixing one coefficient, then $\mathbf{U}$ is really a vector of just $n+1$ components, and must therefore vanish. The insensitivity of the coefficients to slight inaccuracies in $x_1, \cdots, x_n$ is obviously an *advantage*, numerically speaking.

REFERENCES

1. Achiezer, N. I. *Lektii po teorii approksimatsii* (Leningrad, 1947). Eng. Transl.: *Theory of Approximation*, Frederick Ungar Publ. Co., New York, 1956.
2. Barth, W. Ein Iterationsverfahren zur Approximation durch Polynome. *Z. angew. Math. Mech. 38* (1958), 258–260.
3. Hastings, C. *Approximations for Digital Computers*. Princeton University Press, 1955.
4. Lanczos, C. *Applied Analysis*, pp. 457–463. Prentice Hall, Englewood Cliffs, N. J., 1956.
5. Fraser, W.; and Hart, J. F. On the computation of rational approximations to continuous functions. *Com. ACM 5* (1962), 401–403, 414.

6. Loeb, H. L. A note on rational function approximation. Convair Astronautics Appl. Math. Ser. 27 (Sept. 1959).

7. Maehly, H. J. First interim progress report on rational approximations. Princeton University Tech. Report, June 23, 1958.

8. ——. Methods for fitting rational approximations, Part I: Telescoping procedures for continued fractions. *J. ACM 7* (1960), 150–162.

9. Maehly, H. J., and Witzgall, C. Tschebyscheff-Approximationen in kleinen Intervallen I, Approximation durch Polynome. *Numer. Math. 2* (1960), 142–150.

10. —— and ——. Tschebyscheff-Approximationen in kleinen Intervallen II, Stetigkeitssätze für gebrochen rationale Approximationen. *Numer. Math. 2* (1960), 293–307.

11. Cheney, E. W., Fraser, W., et al. *Handbook on Chebyshev Approximations.* In preparation.

12. Murnaghan, F. D., and Wrench, J. W., Jr. The approximation of differentiable functions by polynomials. David Taylor Model Basin Report 1175, April 1958.

13. —— and ——. The determination of the Chebyshev approximating polynomial for a differentiable function. *MTAC 13* (1959), 185–193.

14. Remes, E. Sur le calcul effectif des polynomes d'approximation de Tchebichef. *C. R. Acad. Sci. Paris. 199* (1934), 337–340.

15. Stoer, J. A direct method for the Chebychev approximation by rational functions. In preparation.

16. Tukey, J. W., and Thacher, H. C., Jr. Rational interpolation made easy by a recursive algorithm. In preparation.

17. Veidinger, L. On the numerical determination of the best approximation in the Chebyshev sense. *Numer. Math. 2* (1960), 99–105.

18. Wittemeyer, L. Rational approximation of empirical functions. *Nordisk Tidskrift for Informations-Behandlung 2* (1962), 53–60.

19. Witzgall, C. Tschebyscheff-Approximationen in kleinen Intervallen III, Approximation durch gebrochen rationale Functionen. In preparation.