ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 4, Number 2, Spring 1974

RATIONAL APPROXIMANTS FOR INVERSE FUNCTIONS OF TWO VARIABLES

C. L. CRITCHFIELD AND J. L. GAMMEL

ABSTRACT. A comparatively unknown method of finding the least root of an equation in one variable is generalized to deriving approximations to inverse functions of two variables in terms of ratios of polynomials. Convergence in the case of one variable has been proved rigorously (Koenig's theorem). A method of derivation is presented that makes the generalization possible. Simple examples are given and in all cases tried the rational approximants are more useful and less arduous than inverse double series using the same information.

1. **Introduction.** The work reported in this note grew out of a need to have available a convenient formalism for inverting functions of two variables. Let z and w be two functions of the independent variables x and y; the problem at hand is to find expressions for $x(z, w)$ and $y(z, w)$ from known $z(x, y)$ and $w(x, y)$. The latter are assumed to be analytic in the neighborhood of $x = y = 0$. Therefore each can be expressed as double series of positive powers of x and y and their products. Then x and y can be expressed also as double power series in z and w. By substituting the series for x and y with unknown coefficients into those for z and w with known coefficients one arrives at identities from which the unknowns can be determined. However, the process is very arduous when carried out for more than a few terms. Moreover, experience with one variable indicates that approximation by ratios of polynomials is generally superior to summation of truncated series using the same information. It is with this in mind that we present a process for forming rational approximants to inverse functions of two variables. In addition the connection between this process and theory of locating roots of functions of one variable and to Padé approximants is noted. Finally, certain simple examples are presented.

2. **Inversion of Double Series by Rational Fractions.** Let us absorb the constant terms into the dependent variables and write

$$Z \equiv z - \sum_{n=1}^{\infty} \sum_{k=0}^{n} a_{n-k,k} x^{n-k} y^k = 0, \tag{1}$$

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

$$W \equiv w - \sum_{n=1}^{\infty} \sum_{k=0}^{n} b_{n-k,k} x^{n-k} y^k = 0. \tag{2}$$

We propose to derive two approximately linear equations from these. Consider the two polynomials of degree N

$$G_N = 1 + \sum_{m=1}^{N} \sum_{\ell=0}^{m} g_{m-\ell,\ell}\, x^{m-\ell} y^{\ell}, \tag{3}$$

$$H_N = \sum_{m=1}^{N} \sum_{\ell=0}^{m} h_{m-\ell,\ell}\, x^{m-\ell} y^{\ell}, \tag{4}$$

in which there is a total of $N(N+3)$ unknown coefficients, g_{ik} and h_{ik}. In order to fix the values of these we impose the condition

$$\begin{aligned} G_N Z + H_N W = z &+ (g_{10} z + h_{10} w - a_{10}) x \\ &+ (g_{01} z + h_{01} w - a_{10}) y + x^2 0 + \cdots + R_N = 0. \end{aligned} \tag{5}$$

By dropping the remainder we have a linear equation for x and y. The coefficients g_{10}, g_{01}, h_{10} and h_{01} are calculated from the $N(N+3)$ simultaneous linear equations

$$K_{\mu\nu} \equiv \sum_{i,j} (a_{ij} g_{\mu-i,\nu-j} + b_{ij} h_{\mu-i,\nu-j}) - z g_{\mu\nu} - w h_{\mu\nu} = 0, \tag{6}$$

where i and j are integers. In (6), $g_{00} = 1$ and $h_{00} = 0$ and any negative subscript gives a null term. Moreover $g_{\mu\nu}$ and $h_{\mu\nu}$ vanish for $\mu + \nu > N$. These equations are shown in tabular form in Table I for the case of $N = 3$. Each row is a vector and the eighteen equations are obtained by taking the ordinary scalar products of the top row with the other eighteen and setting them equal to zero. Obviously one needs to solve only for g_{10}, g_{01}, h_{10} and h_{01}, and these are rational functions of z and w.

In a similar way consider the double polynomials

$$P_N = 1 + \sum_{m=1}^{N} \sum_{\ell=0}^{m} p_{m-\ell,\ell}\, x^{m-\ell} y^{\ell}, \tag{7}$$

$$Q_N = \sum_{m=1}^{N} \sum_{\ell=0}^{m} q_{m-\ell,\ell}\, x^{m-\ell} y^{\ell}, \tag{8}$$

and choose the $N(N+3)$ coefficients p_{ij} and q_{ij} so that

$$\begin{aligned} Q_N Z + P_N W = w &+ (p_{10} w + q_{10} z - b_{10}) x \\ &+ (p_{01} w + q_{01} z - b_{01}) y + x^2 0 + \cdots + R_N' = 0. \end{aligned} \tag{9}$$

The values of p_{10}, p_{01}, q_{10}, and q_{01} are determined from the simultaneous linear equations

$$K'_{\mu\nu} \equiv \sum_{i,j} (b_{ij}p_{\mu-i,\nu-j} + a_{ij}q_{\mu-i,\nu-j}) - wp_{\mu\nu} - zq_{\mu\nu} = 0, \tag{10}$$

using the same rules for subscripts and $p_{00} = 1$, $q_{00} = 0$. Table I can then be used after the following substitutions

$$\begin{aligned} a_{ik} &\leftrightarrows b_{ik}, \\ g_{ik} &\rightarrow p_{ik}, \\ h_{ik} &\rightarrow q_{ik}. \end{aligned} \tag{11}$$

In order to annihilate all terms in the product polynomials, (5) or (8), of degree n and less except for degree 1 and 0 one needs $\frac{1}{2}(n+1)(n+2) - 3$ unknowns in the polynomials. Setting this equal to $N(N+3)$ and solving we get

$$n' = ((8N^2 + 24N + 25)^{1/2} - 3)/2. \tag{12}$$

Only rarely will n' be integral for integral N. Asymptotically $n \rightarrow (2)^{1/2}N$ and we see that the remainders R_N and R_N' are of the order

$$R_N \sim R_N' \sim O(x^2 + y^2)^{N/\sqrt{2}}, \; N \rightarrow \infty.$$

For any conceivable application only rather small values of N are of interest. The lowest four integral values of n' in (12) are

$N =$	3	8	25	54
$n' =$	5	12	36	77
No. of Eq. $=$	18	88	700	3078

The No. of Eq. refers to the number of simultaneous equations in each of (6) and (10).

In practice one may not know all terms up to given degree so that the matching process might be quite different. Also, it may happen that the x dependence is more important than the y dependence in one or both z and w in the sense that

$$a_{n-k,k}x^{n-k}y^k \gg a_{n-k-1,\,k+1}x^{n-k-1}y^{k+1},$$

for all k and for $n = \text{int}\, n'$. Then one might use the 'left over' unknowns to annihilate the higher powers of x in terms of degree $n+1$. In the following section we consider a least squares treatment in the case $N = 1$.

3. **Completion of Equations Using Least Squares.** When $N = 1$ we have

$$G_1 = 1 + g_{10}x + g_{01}y \quad \text{and} \quad H_1 = h_{10}x + h_{01}y;$$

these four coefficients can be chosen to annihilate the quadratic terms in $G_1Z + H_1W$ and satisfy one additional condition. The three equations from (6) are

$$\begin{aligned} &a_{20} + a_{10}g_{10} + b_{10}h_{10} = 0, \\ (13) \qquad &a_{11} + a_{10}g_{01} + b_{10}h_{01} + a_{01}g_{10} + b_{01}h_{10} = 0, \\ &a_{02} + a_{01}g_{01} + b_{01}h_{01} = 0, \end{aligned}$$

and the term of degree three in $G_1Z + H_1W$ is

$$\begin{aligned} \tau = &- (a_{30} + a_{20}g_{10} + b_{20}h_{10})x^3 \\ &- (a_{21} + a_{20}g_{01} + b_{20}h_{01} + a_{11}g_{10} + b_{11}h_{10})x^2y \\ &- (a_{12} + a_{11}g_{01} + b_{11}h_{01} + a_{02}g_{10} + b_{02}h_{10})xy^2 \\ &- (a_{03} + a_{02}g_{01} + b_{02}h_{01})y^3. \end{aligned}$$

Let us average τ^2 over a unit circle in the (x, y) plane.

$$\begin{aligned} \overline{\tau^2} = &(a_{30} + a_{20}g_{10} + b_{20}h_{10})^2\overline{x^6} + (a_{03} + a_{02}g_{01} + b_{02}h_{01})^2\overline{y^6} \\ &+ [(a_{21} + a_{20}g_{01} + b_{20}h_{01} + a_{11}g_{10} + b_{11}h_{10})^2 \\ &+ 2(a_{30} + a_{20}g_{10} + b_{20}h_{10})(a_{12} + a_{11}g_{01} \\ &+ b_{11}h_{01} + a_{02}g_{10} + b_{02}h_{10})]\overline{x^4y^2} \\ &+ [(a_{12} + a_{11}g_{01} + b_{11}h_{01} + a_{02}g_{10} + b_{02}h_{10})^2 \\ &+ 2(a_{03} + a_{02}g_{01} + b_{02}h_{01})(a_{12} + a_{11}g_{01} \\ &+ b_{11}h_{01} + a_{02}g_{10} + b_{02}h_{10})]\overline{x^2y^4}. \end{aligned}$$

From (13) we find

$$\frac{\partial h_{10}}{\partial g_{10}} = -\frac{a_{10}}{b_{10}}, \quad \frac{\partial h_{01}}{\partial g_{10}} = -\frac{a_{01}}{b_{10}}, \quad \frac{\partial g_{01}}{\partial g_{10}} = \frac{b_{01}}{b_{10}},$$

and we use these relations to minimize $\overline{\tau^2}$ with respect to g_{10}. This then gives us a fourth equation. Let

$$\begin{aligned} &\alpha \equiv a_{20}b_{10} - a_{10}b_{20}, \quad \bar{\alpha} \equiv a_{02}b_{01} - a_{01}b_{02}, \\ &\beta \equiv (a_{20} + a_{02})b_{10} - a_{10}(b_{20} + b_{02}) + a_{11}b_{01} - a_{01}b_{11}, \\ &\bar{\beta} \equiv (a_{20} + a_{02})b_{01} - a_{01}(b_{20} + b_{02}) + a_{11}b_{10} - a_{10}b_{11}. \end{aligned}$$

Then the fourth equation becomes

$$
\begin{aligned}
&4(a_{30}\alpha + a_{03}\bar{\alpha}) + (a_{30} + a_{12})\beta + (a_{21} + a_{03})\bar{\beta} \\
&+ [4a_{20}\alpha + (a_{20} + a_{02})\beta + a_{11}\bar{\beta}]\, g_{10} \\
&+ [4b_{20}\alpha + (b_{20} + b_{02})\beta + b_{11}\bar{\beta}]\, h_{10} \\
&+ [4a_{02}\bar{\alpha} + (a_{20} + a_{02})\bar{\beta} + a_{11}\beta]\, g_{01} \\
&+ [4b_{02}\bar{\alpha} + (b_{20} + b_{02})\bar{\beta} + b_{11}\beta]\, h_{01} = 0.
\end{aligned} \tag{13a}
$$

Eqs. (13) and (13a) then serve to determine the g's and h's. An analogous set of equations for the p's and q's is obtained by making the substitutions indicated in (11). It is to be noted that these transformations merely change the sign for α, β, $\bar{\alpha}$ and $\bar{\beta}$. In fact the eliminant for the transformed set of equations is simply the negative of that for the g's and h's.

The use of least squares to determine 'left-over' unknowns can be generalized in an obvious way.

4. **Relationship to Theory of One Variable.** The process described above for making equations in two variables approximately linear is obviously related to the formation of the $[1/N]$ entry in the Padé table for functions of one variable [1]. Moreover the inversion of a function $f(x)$ so as to obtain $x(f)$ is equivalent to finding roots of

$$f(x) = f \tag{14}$$

where on the left-hand side f is a function of x and on the right-hand side it is an independent variable. The treatment of one variable is well established and rigorous [2]. Our treatment of two variables corresponds to finding the least root from the ratio of two Hankel determinants, $H_{1,n}/H_{1,n+1}$. The use of such determinants for finding roots dates back to 1860 [3]. Their derivation through multiplication by polynomials to make an equation approximately linear appeared in 1935 [4].

In effect the Hankel determinants which are required are generated as coefficients in the power series expansion of:

$$\frac{1}{f - f(x)} = \sum_{k=0}^{\infty} C_k x^k.$$

For simplicity let us suppose this expansion to be meromorphic within a finite circle about the origin and that all poles are distinct. Koenig's theorem [2] then states that as $k \to \infty$, C_k/C_{k+1} approaches the least root of $f(x)$. Our extension to two variables is by analogy and it presumably can be generalized to n variables. The test of its usefulness is purely pragmatic and some examples are given in the following section.

5. **Some Particular Examples.** In order to illustrate the method and test its usefulness we have selected for the function z the simplest of (infinite) double power series

$$z = \frac{1}{(1-x)(1-y)} - 1 = x + y + x^2 + xy + \cdots. \tag{15}$$

For w we consider functions whose characteristics make the same angle with those of z at all points in the $X - Y$ plane. It is easily shown that all such functions are power series in

$$w = \frac{\lambda - 1}{2}x + \frac{\lambda + 1}{2}y + \frac{1}{4}(x^2 - y^2) - \frac{1}{2}\lambda xy \tag{16}$$

and we choose w to be the quadratic polynomial itself. In (16), λ is the cotangent of the angle between characteristics.

The numerical study is made as follows: a coordinate pair (x, y) is selected and values of z and w computed from the exact formulas, these values are then used in the approximation methods to compute values x' and y'. We then calculate

$$\rho = \left(\frac{(x - x')^2 + (y - y')^2}{x^2 + y^2}\right)^{1/2}. \tag{17}$$

Contours in the $X - Y$ plane for constant values of ρ are then drawn in order to compare various cases. Three types of comparisons are presented.

In Fig. 1 the contours for $\rho = .01$ are shown for the methods described above with $N = 3$ (where coefficients through degree five in the products are zero) and $N = 1$ (where quadratic terms are zero and cubic terms minimized by least squares). The circles in each case are drawn tangent to the nearest point of the contours. The area of the circle is a measure of the usefulness of the approximation. In this sense the $N = 3$ approximation is 8 times as useful as that with $N = 1$. The calculations were made with $\lambda = .75$.

In Fig. 2 the contour for $N = 1$ and $\rho = 0.1$ is compared with that obtained when the series are inverted and summed through third degree in z and w. The area of usefulness as defined above is 2.5 greater for the rational approximant. These calculations were made for $\lambda = 1$.

In Fig. 3 we plot the radius of the area of usefulness $r_{.1}$ (i.e., for $\rho = 0.1$) as a function of the angle α between characteristics. Again the curve labelled $N = 1$ is that for the corresponding rational approximant in which least squares are applied and that labeled SERIES is the sum of the inverted double series through terms of degree three.

The authors wish to thank Mary Menzel for her assistance on the computations. They are indebted also to Professor Householder for helpful discussions.

References

1. *The Padé Approximant in Theoretical Physics,* Edited by George A. Baker, Jr. and John L. Gammel, Academic Press, 1970.

2. Cf. A. S. Householder, *The Numerical Treatment of a Single Non-Linear Equation,* McGraw-Hill, 1970, p. 116, Chap. 3.1.

3. A. M. Ostrowski, *Solution of Equations and Systems of Equations,* Academic Press, 1960, p. 180 et seq.

4. E. Fürstenau, *New Method of Representing and Calculating Imaginary Roots of Algebraic Equations by Determinants of Their Coefficients,* Ges. Naturw. Schrift Marburg (1860) [Lib. Cong. QA212.F8] abstracted in *The Theory of Determinants,* T. Muir, vol. III, p. 423.

5. C. L. Critchfield and J. Beek, Jr., *Method for Finding the Roots of the Equation $f(x) = 0$ where f is Analytic,* Journ. Res. Nat. Bur. Stand., 14 (1935), 595–600, Research Paper RP 790.

Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico 87544

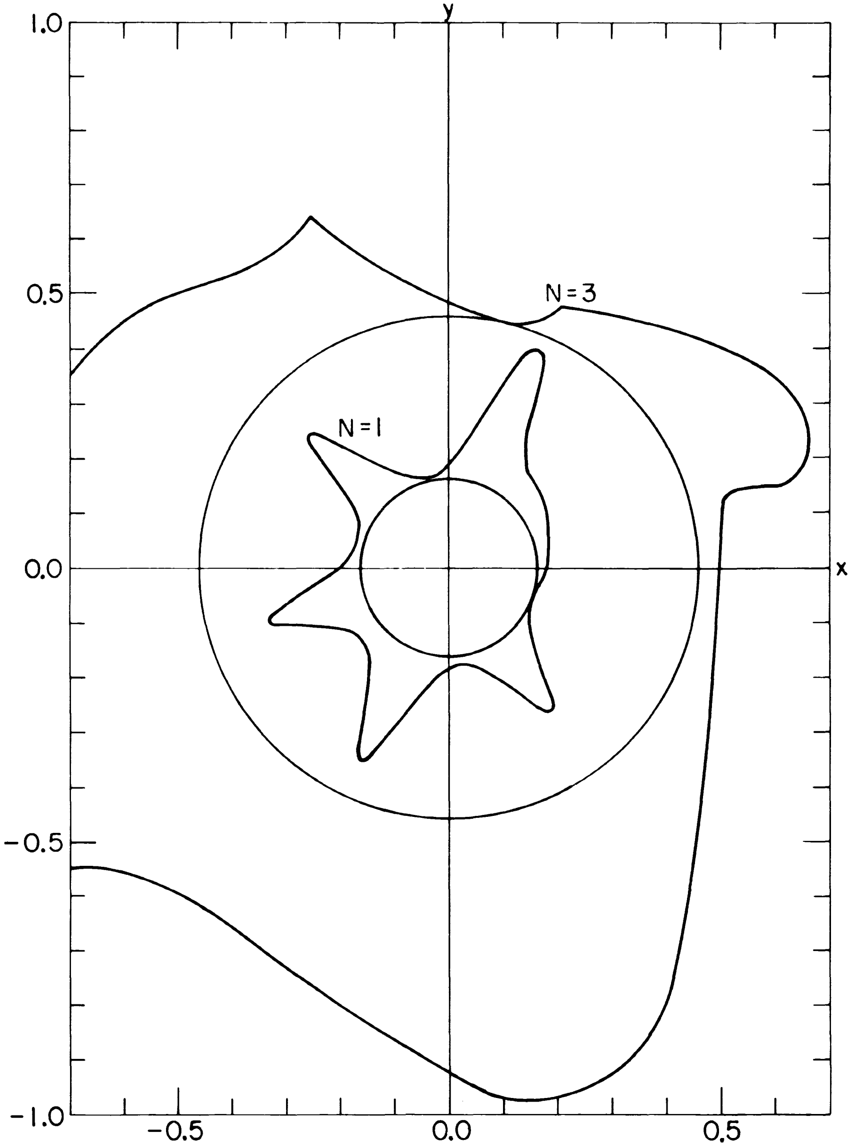

Figure 1. Contours for 1% accuracy and $\lambda = 0.75$. Curves labelled $N = 1$ and $N = 3$ are for the corresponding rational approximants.

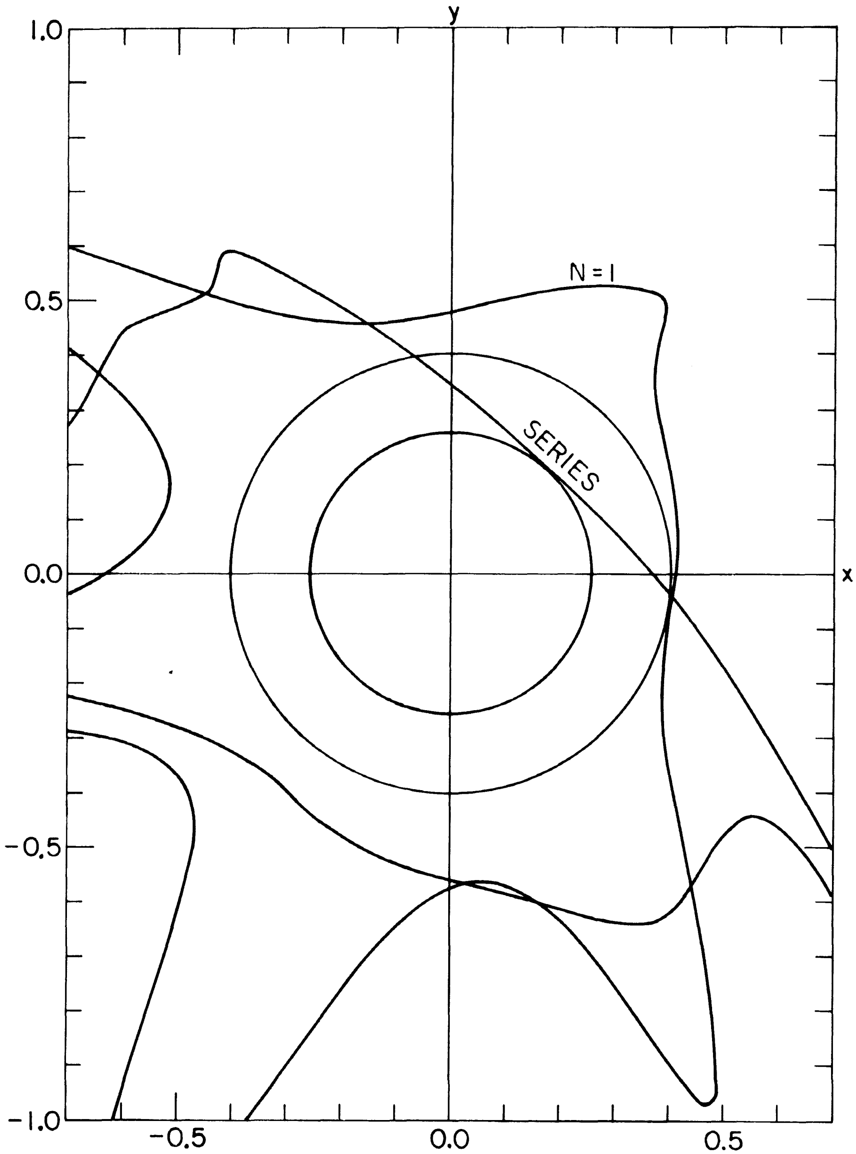

Figure 2. Contours for 10% accuracy ($\rho = 0.1$). Curve labelled $N = 1$ is for the corresponding rational approximant and that labelled SERIES is for inverted double series of degree three. Circles are tangent at nearest point in each case.

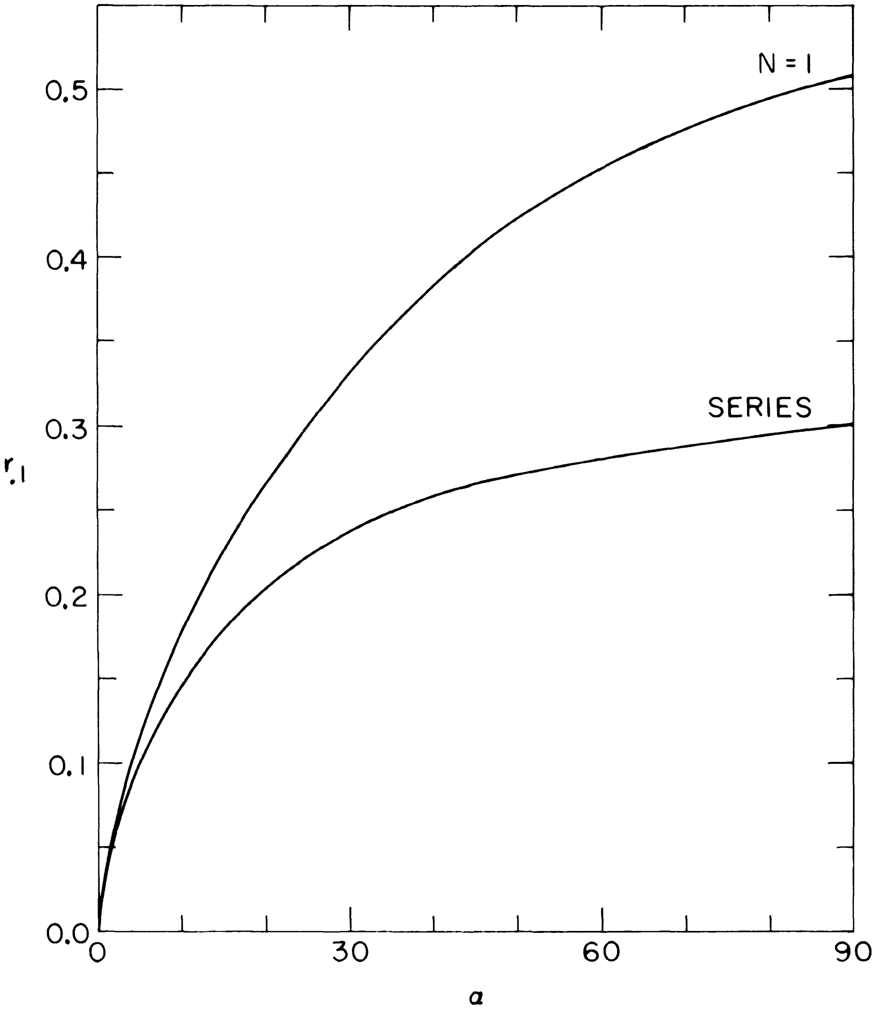

Figure 3. Radii of circles inscribed in contours for $\rho = 0.1$ as a function of α, the angle between characteristics.

Table I

Coefficients for 18 Linear Equations

1	g_{10}	h_{10}	g_{01}	h_{01}	g_{20}	h_{20}	g_{11}	h_{11}	g_{02}	h_{02}	g_{30}	h_{30}	g_{21}	h_{21}	g_{12}	h_{12}	g_{03}	h_{03}
a_{20}	a_{10}	b_{10}			$-z$	$-w$												
a_{11}	a_{01}	b_{01}	a_{10}	b_{10}			$-z$	$-w$										
a_{02}			a_{01}	b_{01}					$-z$	$-w$								
a_{30}	a_{20}	b_{20}			a_{10}	b_{10}					$-z$	$-w$						
a_{21}	a_{11}	b_{11}	a_{20}	b_{20}	a_{01}	b_{01}	a_{10}	b_{10}					$-z$	$-w$				
a_{12}	a_{02}	b_{02}	a_{11}	b_{11}			a_{01}	b_{01}	a_{10}	b_{10}					$-z$	$-w$		
a_{03}			a_{02}	b_{02}					a_{01}	b_{01}							$-z$	$-w$
a_{40}	a_{30}	b_{30}			a_{20}	b_{20}					a_{10}	b_{10}						
a_{31}	a_{21}	b_{21}	a_{30}	b_{30}	a_{11}	b_{11}	a_{20}	b_{20}			a_{01}	b_{01}	a_{10}	b_{10}				
a_{22}	a_{12}	b_{12}	a_{21}	b_{21}	a_{02}	b_{02}	a_{11}	b_{11}	a_{20}	b_{20}			a_{01}	b_{01}	a_{10}	b_{10}		
a_{13}	a_{03}	b_{03}	a_{12}	b_{12}			a_{02}	b_{02}	a_{11}	b_{11}					a_{01}	b_{01}	a_{10}	b_{10}
a_{04}			a_{03}	b_{03}					a_{02}	b_{02}							a_{01}	b_{01}
a_{50}	a_{40}	b_{40}			a_{30}	b_{30}					a_{20}	b_{20}						
a_{41}	a_{31}	b_{31}	a_{40}	b_{40}	a_{21}	b_{21}	a_{30}	b_{30}			a_{11}	b_{11}	a_{20}	b_{20}				
a_{32}	a_{22}	b_{22}	a_{31}	b_{31}	a_{12}	b_{12}	a_{21}	b_{21}	a_{30}	b_{30}	a_{02}	b_{02}	a_{11}	b_{11}	a_{20}	b_{20}		
a_{23}	a_{13}	b_{13}	a_{22}	b_{22}	a_{03}	b_{03}	a_{12}	b_{12}	a_{21}	b_{21}			a_{02}	b_{02}	a_{11}	b_{11}	a_{20}	b_{20}
a_{14}	a_{04}	b_{04}	a_{13}	b_{13}			a_{03}	b_{03}	a_{12}	b_{12}					a_{02}	b_{02}	a_{11}	b_{11}
a_{05}			a_{04}	b_{04}					a_{03}	b_{03}							a_{02}	b_{02}