

A combined graphical and iterative approach to the problem of finding zeros of functions in the complex plane

By F. M. Larkin*

The paper describes how an automatic graph plotter, used as a digital computer output device, is capable of providing useful regional and global information about functions in the complex plane. This information, in the form of the approximate locations and types of zeros, singularities and branch cuts, enables one to choose starting points for zero finding iterations with a high degree of confidence in their convergence. A number of contour graphs are presented which illustrate general features of complex functions near typical zeros, poles and essential singularities.

1. Introduction

The numerical evaluation of those values of z which satisfy

$$f(z) = 0, \quad (1.1)$$

where $f(z)$ is a suitably single-valued, continuous and differentiable function of the complex variable z , is one of the classic problems of numerical analysis, and many methods exist for systematically improving estimates of one or more such values. Some of these methods, such as those due to Graeffe or Bairstow, are designed specifically to deal with polynomial functions, whereas others, such as Newton's method and Muller's method, are in principle applicable to the general case.

Leaving aside special methods, which can be constructed for special functions (such as polynomials), and which do not require an initial estimate of the intended root, the common difficulty with iteration procedures is that of ensuring convergence. Generally speaking the best that can be said of any practical procedure is that it may be "guaranteed to converge to the required root provided the initial estimate is accurate enough," and the deliberate vagueness is by no means intended as disparagement.

The object of this paper is to show how an automatic graph plotter may be used to obtain an approximate general picture of the properties of the function under consideration, and hence provide starting values for iterations from which convergence to the required zero may be unconditionally guaranteed. The technique described is only suitable for implementation on an automatic computer, since hand computation would be unbearably tedious, and it does necessitate some degree of human intervention. However, this apparent liability turns out in many practical cases to be a most rewarding imposition simply because the intervener, or any other interested party, can find at a glance the approximate location and type of the function's zeros, singularities and branch cuts in the complex plane. This information may not only make it possible to avoid wasting computer time on divergent iterations, but may also provide the heuristic basis upon which to build a fruitful analytic approximation.

2. The origin of convergence troubles

Iteration procedures commonly used for finding zeros of a general function $f(z)$ may be regarded as operating by fitting to $f(z)$ a function $w(z)$, over a local region of the z plane, and then taking an easily computed root of $w(z)$ as the current estimate of the required root of $f(z)$.

For example, if z_0 is an estimate of a root, Newton's method gives the new estimate z_1 as

$$z_1 = z_0 - \frac{f(z_0)}{f'(z_0)} \quad (2.1)$$

which is actually the result of using f and its first derivative to fit the linear function

$$w(z) = f(z_0) + (z - z_0) \cdot f'(z_0) \quad (2.2)$$

to $f(z)$ at the point z_0 , and then taking the zero of $w(z)$ as the next estimate of the desired zero of $f(z)$. Muller's technique avoids the necessity for computing a derivative of f , and fits a quadratic in z to function values of f at three distinct points in the z plane; one of the zeros of this quadratic is then taken as the next estimate of the zero of $f(z)$. Another method, which also avoids the difficulty of choosing between two zeros of a quadratic whilst still retaining convergence characteristics similar to Muller's method, is to fit the bilinear form

$$w(z) = \frac{z - a}{bz + c} \quad (2.3)$$

to function values of f at three distinct points in the z plane, taking a as the next estimate of the required root of $f(z)$. This process involves the elimination of b and c from three simultaneous linear equations, but has the virtues of only requiring function values and of being unambiguous.

One can see roughly that methods of this general type will be useful only if the required zero of $f(z)$ happens to lie within the region of the complex plane where the fitted function $w(z)$ is a fair approximation to $f(z)$, i.e. if the initial estimate of the zero is good enough. The principal source of difficulty with this kind of iteration is that in general it can utilize only *local* information about $f(z)$ whereas some *global*, or at least *regional*,

* U.K.A.E.A. Research Group, Culham Laboratory, Nr. Abingdon, Berks.

information is required in order to decide upon a suitable starting point.

3. Convergence regions

Consider the conformal transformation

$$w = F(z). \tag{3.1}$$

Points in the complex plane which remain invariant under this transformation obviously satisfy

$$z = F(z). \tag{3.2}$$

If z_0 is a root of equation (3.2) and $|F'(z_0)| < 1$ then, not only will z_0 be invariant, but a region containing z_0 will be contracted around z_0 by the transformation. To see this we simply set $z = z_0 + a$, $w = z_0 + b$, where $|a|$ is small, then

$$z_0 + b \simeq F(z_0) + a.F'(z_0),$$

so that

$$\frac{|b|}{|a|} \simeq |F'(z_0)|. \tag{3.3}$$

Obviously if $|F'(z_0)| > 1$ a region containing z_0 will be expanded around z_0 . The case where $|F'(z_0)| = 1$ in general leaves z_0 as a point common to an even number of regions, half of which are contracted towards z_0 and the rest expanded away from z_0 . The number of these regions is determined by the order of the first non-zero derivative of $F(z)$ at the point z_0 . The case of equality is unimportant for the purpose of the present discussion.

If $|F'(z_0)| < 1$ we define the "convergence region" for the point z_0 as the aggregate of points in the complex plane which can be transported arbitrarily close to z_0 by repeated application of the conformal transformation (3.1). Clearly a convergence region will not exist if $|F'(z_0)| > 1$.

If repeated application of transformation (3.1) is to iterate, starting from some given point, to a zero of a function $f(z)$, $F(z)$ must depend upon the form of $f(z)$. In particular, the required zero z_0 of $f(z)$ must also satisfy equation (3.2). Given a rule (for example, Newton's method) for constructing $F(z)$ from $f(z)$ we can envisage the function $f(z)$ as demarking a number of convergence regions in the complex plane, and, if the conformal transformation (3.1) then forms a useful iteration procedure, at least one convergence region will be associated with a zero of $f(z)$.

Clearly it would be of interest to establish the convergence regions of a function under a given iteration before embarking upon the somewhat uncertain procedure of the iteration itself. In general this is difficult, although it is simple enough to isolate graphically every zero of $f(z)$ within as small a region of the z plane as desired. However, it is possible to establish the convergence regions for any function under a steepest-descent type of iteration, equivalent to the limiting case of Newton's method with a very small step length, and this is described in Section 6.

4. The scope of graphical display

The following simple technique may be used to obtain useful information about $f(z)$ in any prescribed, finite region of the complex plane.

$$\text{Let } f(z) = u(x, y) + iv(x, y) \tag{4.1}$$

$$\text{where } z = x + iy \tag{4.2}$$

and regard the single-valued functions $u(x, y)$ and $v(x, y)$ as defining surfaces in a three-dimensional space. The zero-height contours of the functions u and v , drawn in the (x, y) plane, must certainly intersect at the zeros of $f(z)$, and an n -fold zero will be crossed n times in each surface. Moreover, all contours of u and v , of all heights, pass through all the poles, although, of course, u and v are not defined actually at the poles. To see this, suppose z_0 is a pole of order n , then the dominant term of the Laurent series of $f(z)$, expanded about z_0 , is

$$\frac{a}{(z - z_0)^n}, \quad a = \text{constant}. \tag{4.3}$$

$$\text{Let } z - z_0 = Re^{i\theta}$$

$$a = Ae^{i\phi}, \text{ where } A \text{ is not zero,}$$

so that (4.3) becomes

$$\frac{Ae^{i(\phi - n\theta)}}{R^n} = \frac{A \cos(\phi - n\theta)}{R^n} + \frac{Ai \sin(\phi - n\theta)}{R^n}. \tag{4.4}$$

Now consider the contour $u(x, y) = U$. Choose any R , greater than zero, such that $\frac{A}{R^n} \geq U$ and we see that there must be exactly $2n$ values of θ for which

$$\frac{A \cos(\phi - n\theta)}{R^n} = U. \tag{4.5}$$

Thus the contour height U in the $u(x, y)$ surface intersects a circle, with arbitrarily small radius, centred on the pole, exactly $2n$ times. In fact these intersection points are spaced equidistantly around the circumference of the circle, and the effect is that the contour appears to pass through the pole n times at angular spacings of $\frac{\pi}{n}$.

A similar argument applies to the imaginary part of $v(x, y)$.

If $u(x, y)$ and $v(x, y)$ are computed at the nodal points of a mesh spanning the finite region of interest this information may then be processed in order to compute the curves $u(x, y) = 0$, $v(x, y) = 0$ in a form suitable for presentation to an automatic graph plotter. Also, if a few extra contours of positive and negative heights are added, the resulting picture gives a clear indication of the approximate positions and types of zeros, singularities and branch cuts in the region of the complex plane considered.

Fig. 1 illustrates contours in the u and v surfaces for the function

$$f(z) = \frac{(z - 1)(z - i)}{(z + 1)(z + i)^2}. \tag{4.6}$$

The location of the zeros and poles is clear at a glance, and the order of either is determined by counting the number of times a zero-height contour passes through the root or pole.

If $f(z)$ is a rational function, having only a finite number of zeros and poles, these can all be located by plotting two contour graphs as follows.

- (i) Plot contours of the real and imaginary parts of $f(z)$ inside the unit circle in the z plane. This finds all zeros and poles inside the unit circle.
- (ii) Make the conformal transformation $w = 1/z$ and plot contours of the real and imaginary parts of $f(1/w)$ inside the unit circle in the w plane. The values of w at the zeros and poles which now appear in the w plane can then be inverted to find the positions of the zeros and poles of $f(z)$ which lie outside the unit circle in the z plane.

However, if $f(z)$ is a transcendental function, having an infinite number of poles or zeros and one or more essential singularities, at least one of the above two pictures will be very confused, due to the fact that the discrete mesh will be incapable of resolving the fine structure. For such functions one can obtain rather crude general information by the above method, and then proceed by examining selected regions of the z or w planes in more detail.

5. The electric field analogy

For the purposes of subsequent discussion it is convenient here to review one of the many physical inter-

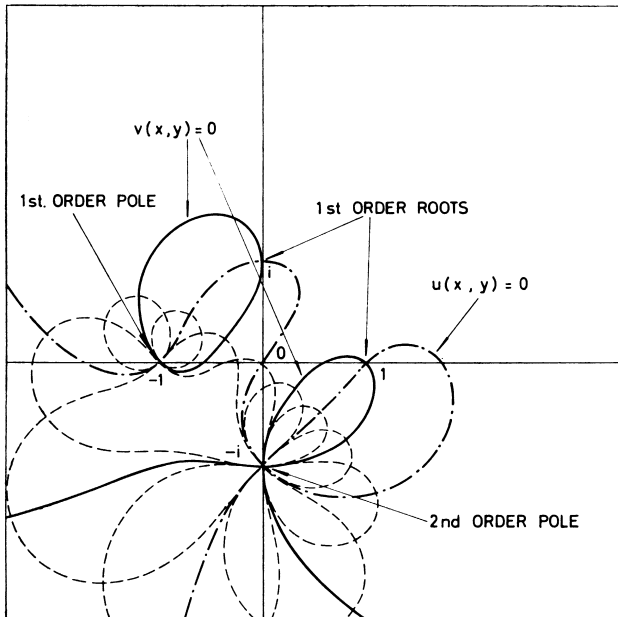


Fig. 1.—Contours of the real and imaginary parts of

$$f(z) = \frac{(z - 1)(z - i)}{(z + 1)(z + i)^2}$$

pretations assignable to functions of a complex variable.

If $f(z) = u(x, y) + iv(x, y)$

is a function of the complex variable z , then

$$\log [f(z)] = g(x, y) + ih(x, y), \tag{5.1}$$

where $g(x, y) = \log |f(z)|$ (5.2)

and $h(x, y) = \arg [f(z)]$ (5.3)

is also a function of z , and $g(x, y)$ and $h(x, y)$ both satisfy Laplace's equation

$$\nabla^2 g = 0 = \nabla^2 h \tag{5.4}$$

except at the singularities of $\log [f]$. This allows us to use $f(z)$ to construct a two-dimensional electric field in the following way.

At the position of each pole of $f(z)$ place an infinite positive line charge of magnitude equal to the order of the pole, and at the position of each zero place an infinite negative line charge of magnitude equal to the order of the zero. The equipotential lines of this electrostatic field are given by

$$\log |f(z)| = \text{constant} \tag{5.5}$$

and the electric field lines are given by the orthogonal curves

$$\arg [f(z)] = \text{constant.} \tag{5.6}$$

Thus a picture of the electric field associated with the function $f(z)$ is obtained simply by plotting contours in $g(x, y)$ and $h(x, y)$, the real and imaginary parts of $\log [f(z)]$, in exactly the same way as was described for $u(x, y)$ and $v(x, y)$.

The electric field vector is $(-\frac{\partial g}{\partial x}, -\frac{\partial g}{\partial y})$, and on appealing to the Cauchy-Riemann relations we see that the null points in the field occur where

$$\frac{d}{dz} [\log (f)] = 0, \tag{5.7}$$

i.e. at those points, other than zeros of $f(z)$, where $\frac{df}{dz} = 0$.

Fig. 2 illustrates the electrostatic field associated with the function defined in equation (4.6). The equipotentials plotted are given by

$$g(x, y) = -1, 0 \text{ and } +1,$$

whilst the electric field lines are given by

$$h(x, y) = -\pi + \frac{2n\pi}{7}, \quad n = 0, 1, \dots 6.$$

The branch cuts in the plane are due to the discontinuity in the arctan function as the arguments increase from $[\cos(\pi - \epsilon), \sin(\pi - \epsilon)]$ to $[\cos(\pi + \epsilon), \sin(\pi + \epsilon)]$, ϵ being a small positive number. Since seven lines of constant $\arg(f)$ are plotted, the order of a zero or pole

is found simply by dividing the number of lines issuing from a singularity in $\log(f)$ by seven.

6. Convergence regions for a simple descent method

Consider the iteration formula

$$z_1 = z - \lambda \frac{f(z)}{f'(z)}, \tag{6.1}$$

which reduces to Newton's method when $\lambda = 1$. It is easy to show (e.g. Lance, 1960) that one stage in the iteration represents a step in the complex plane, of finite length, in a direction tangential to the line of steepest descent in the $\log|f|$ surface at the point z . The real, positive number λ simply fixes the length of the correction step. If the step length is made small by making λ small the iteration will proceed along a line of steepest descent in the $\log|f|$ surface, i.e. along a field line, away from a pole and towards a root, in the associated electric field.

The electrostatic field associated with $f(z)$ now provides a useful heuristic basis for discussion of the convergence of iteration formula (6.1) in the limiting case where λ is small, for we need only consider the disposition of field lines in relation to charges and null points. In general a field line starts (high potential end) and finishes (low potential end) either on a charge or at infinity. Also, apart from charges and null points, exactly one field line passes through every given point in the (x, y) plane. Field lines which pass through null points are termed "critical", and such a field line will only intersect charges of the same sign, whereas an ordinary field line will

start at a positive charge (pole) and end at a negative charge (zero), either of which may be at infinity. Thus from any starting point not on a critical field line the iteration *must* converge to a zero, either at infinity or in the finite part of the plane.

The convergence region for any particular zero is simply that region in the complex plane which exactly contains all the field lines which end at that zero. Moreover, the boundaries of the convergence regions are those critical field lines which do not pass through zeros, since it is impossible for the iteration to cross such a line. Thus in order to delineate the convergence regions we simply evaluate $\arg(f)$ at the points where

$$\frac{d}{dz} [\log(f)] = 0$$

and plot contours of $\arg(f)$ at these heights, in the region of interest.

Fig. 3 illustrates the convergence regions for the function defined in equation (4.6) under this small step limit of Newton's iteration, the picture corresponding exactly with Figs. 1 and 2. The points A, B and C are null points in the electric field, i.e. the zeros of $\frac{d}{dz} \{\log [f(z)]\}$. It will be seen that the convergence region for the zero at $z = i$ completely encloses the convergence region for the zero at $z = 1$, and this in turn is completely surrounded by the convergence region for the root at infinity. One curious consequence of the disposition of the convergence regions is that, from the starting point S, the iteration will converge to the zero

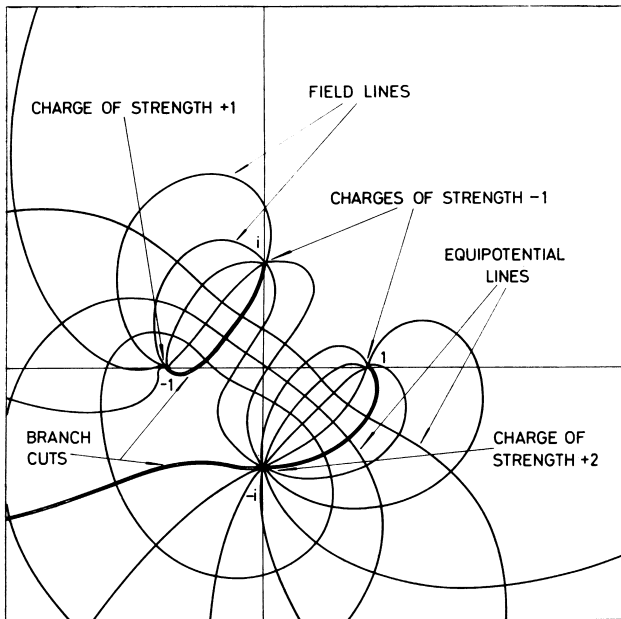


Fig. 2.—The electrostatic field associated with

$$f(z) = \frac{(z - 1)(z - i)}{(z + 1)(z + i)^2}$$

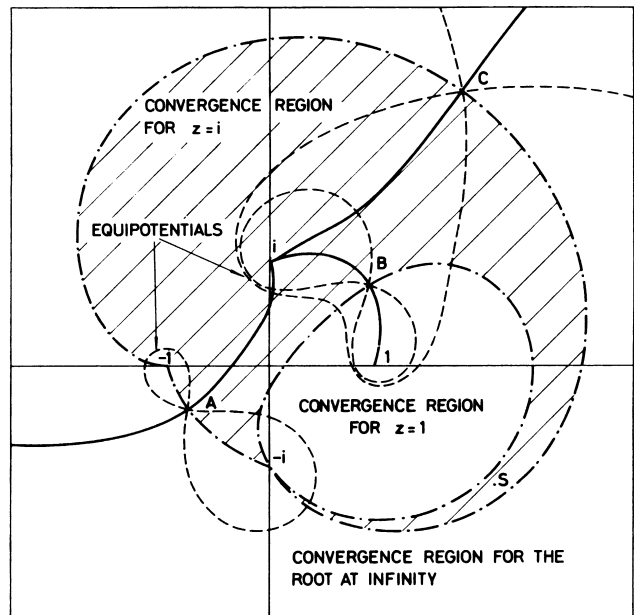


Fig. 3.—Critical field lines associated with

$$f(z) = \frac{(z - 1)(z - i)}{(z + 1)(z + i)^2}$$

at $z = i$, not to the nearer zero at $z = 1$, even though the point $z = 1$ lies between S and the point $z = i$. Notice also that the critical field lines which pass through zeros, but not poles, delineate the convergence regions for the function $\frac{1}{f(z)}$, and in this case dissect the plane into two infinite parts.

7. Some general remarks on convergence regions

If λ in equation (6.1) is a finite positive quantity the iteration will in general step onto a new field line at every stage. This means that an iteration starting from within a convergence region for the small λ case could well step out of that region either towards infinity or to a different root. It is difficult to be precise but intuitively one would expect this danger of “wandering” to be great if

(i) the iteration starts near to a convex part of the boundary of a convergence region for the small λ case,

or (ii) the iteration starts near to a root of $\frac{d}{dz} [\log (f)]$,
 or (iii) λ is large.

In particular a Newton iteration starting from the point S in Fig. 3 will diverge to infinity, and there exists a significant region surrounding the point B from which the iteration also diverges to infinity; moreover, the rate of divergence in both cases is considerably aggravated as λ is increased. Lance suggests a procedure of increasing λ up to N , the first integer for which

$$\left| f \left[z - (N + 1) \cdot \frac{f}{f'} \right] \right| > \left| f \left[z - N \cdot \frac{f}{f'} \right] \right| \quad (7.1)$$

given that

$$\left| f \left[z - N \frac{f}{f'} \right] \right| < |f(z)|.$$

However, in view of the example taken the author concludes that this is unwise as a general practice; it certainly appears that finite-sized convergence regions become smaller as λ is increased.

Now consider a zero of $f(z)$ surrounded by its associated convergence region under the iteration (6.1), where λ is finite. We can regard (6.1) as a conformal mapping which transforms the point z , inside the convergence region, into the point z_1 which in some sense is nearer the zero than z is. Moreover (6.1) maps the entire convergence region onto itself so that, in particular, the boundary of the region is a curve which is invariant under the conformal mapping

$$w = z - \lambda \frac{f(z)}{f'(z)}. \quad (7.2)$$

The points which transform into themselves under (7.2) are given by

$$\frac{f(z)}{f'(z)} = 0, \quad (7.3)$$

i.e. they are the poles and zeros of $f(z)$. Clearly the zeros of $f(z)$ must be invariant under (7.2) or the iteration would never converge; the fact that the poles are also invariant indicates that it will be inefficient to start the iteration from near to a pole. Similarly, the convergence regions for an iteration procedure of the general type,

$$w = F(z) \quad (7.4)$$

must be mapped onto themselves by the conformal transformation.

An illustration of the concept is given by the application of Newton’s method to the function

$$f(z) = \frac{z}{z - 1}. \quad (7.5)$$

In this case the conformal mapping (7.2), with $\lambda = 1$, becomes

$$w = z^2 \quad (7.6)$$

which transforms the unit disc into itself. Thus, for Newton’s method, the unit circle marks the boundary of the convergence region associated with the zero of the function defined in equation (7.5). Fig. 4 shows the electrostatic field picture of this function, and the progress of two Newton iterations is marked, illustrating the role of the unit circle as the boundary of the convergence region.

8. Contours of some typical functions

The examples portrayed in Figs. 1, 2, 3 and 4 illustrate some general topographical properties of rational functions. As far as zeros and poles are concerned the

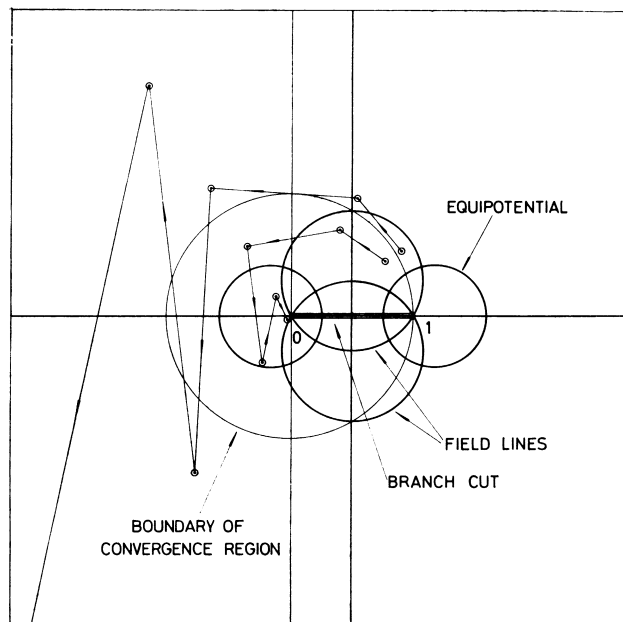


Fig. 4.—The electrostatic field associated with $f(z) = \frac{z}{z - 1}$ showing the progress of two Newton iterations

electrostatic field analogy seems to the author to be the easiest way of visualizing such a function. One interesting result of this analogy is the fact that Lucas's theorem, which states that the zeros of the derivative of a polynomial must lie within the smallest convex polygon which contains all the zeros of that polynomial, is physically obvious from the fact that no null points in the electric field can lie outside this polygon.

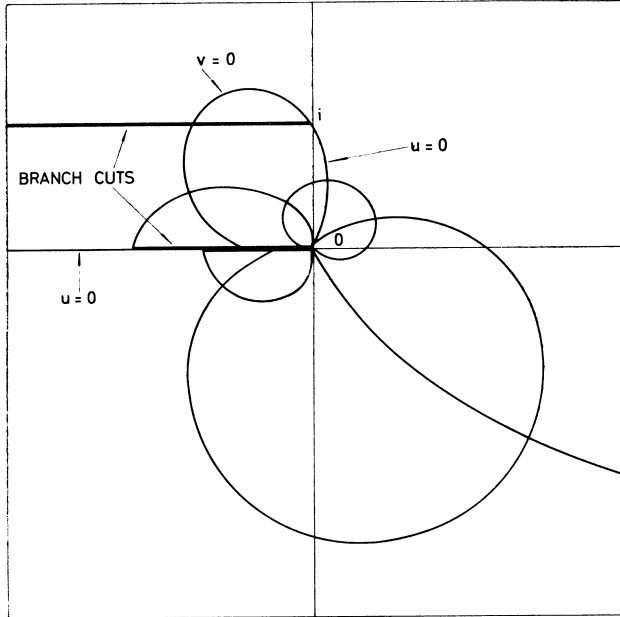


Fig. 5.—Contours of the real and imaginary part of

$$f(z) = \frac{(z - i)^{1/2}}{z^{2/3}}$$

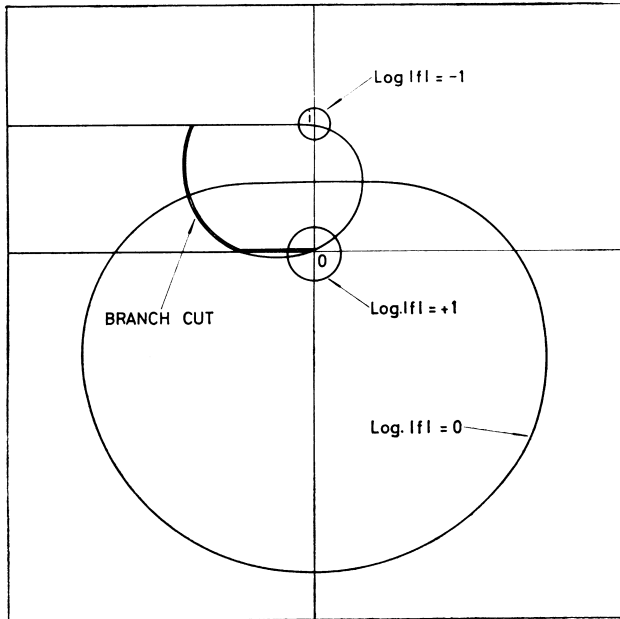


Fig. 6.—The electrostatic field associated with

$$f(z) = \frac{(z - i)^{1/2}}{z^{2/3}}$$

By way of further example, contours of some other common functions are shown in Figs. 5, 6, 7, 8, 9 and 10. Fig. 5 shows contours of the real and imaginary parts of

$$f(z) = \frac{(z - i)^{1/2}}{z^{2/3}} \quad (8.1)$$

and Fig. 6 shows corresponding contours for its logarithm, thus illustrating behaviour near zeros and poles of fractional order. In this case the position of the zero

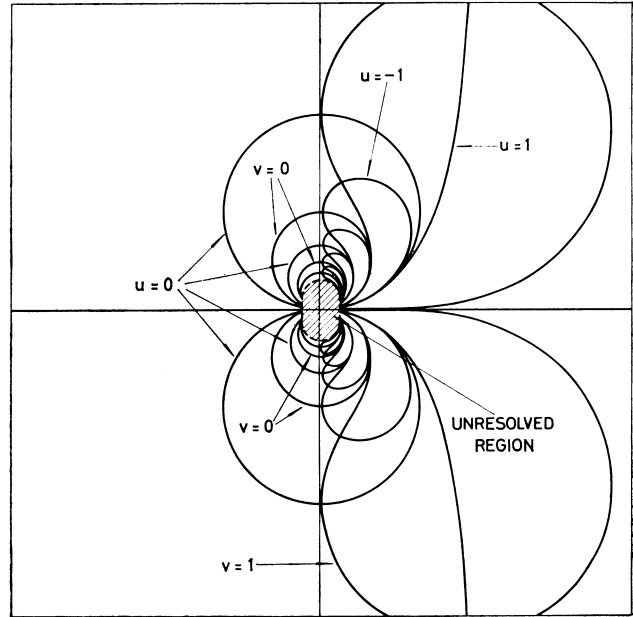


Fig. 7.—Contours of the real and imaginary parts of

$$f(z) = e^{1/z}$$

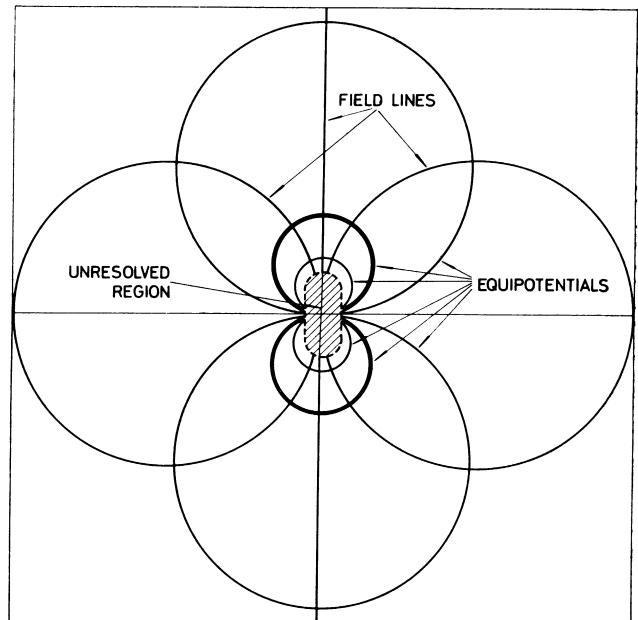


Fig. 8.—The electrostatic field associated with $f(z) = e^{1/z}$

is not clear from Fig. 5, but is quite plain in Fig. 6. Fig. 7 shows contours of the real and imaginary parts of

$$f(z) = e^{1/z} \quad (8.2)$$

near the isolated essential singularity at $z = 0$; as is well known, behaviour of the function near this point is seen to be very complex. However, the electric field analogy, portrayed in Fig. 8, is much easier to understand, and illustrates that (8.2) represents an electric dipole situated at $z = 0$. An essential singularity as a limit point of poles is exemplified by the function

$$f(z) = \operatorname{cosec}\left(\frac{1}{z}\right) \quad (8.3)$$

which has such a singularity at $z = 0$. Figs. 9 and 10 illustrate contours in the real and imaginary part of $\operatorname{cosec}\left(\frac{1}{z}\right)$ and its logarithm. Near the essential singularity the structure is too fine to be resolved by the finite mesh, but the general features are clear enough to establish the existence of a concentration of singularities near the origin.

9. Implementation of the technique

All the graphical results presented above were obtained using a computer program written by the author in a dialect of FORTRAN. The computer used was the IBM 7030, or Stretch, at A.W.R.E. Aldermaston, and the automatic graph plotter was the Bensen-Lehner Model J. The program is in general use at the Culham Laboratory of the U.K.A.E.A., only a minimal knowledge of computing being required. Normal operating procedure, given the function $f(z)$, is to compute contour

graphs over a finite rectangle in the complex plane, and thence proceed either to iterative zero-finding or to further contour graphs over a smaller rectangle contained in the original one. However, a number of program options provide a good deal of flexibility in operation and output, and no single procedure can be regarded as "the best" in all circumstances.

The contour graphs have been edited somewhat for the purpose of presentation, but this editing consists largely of annotation and is not normally required in everyday work. In practice automatic contour plotting has been found, over a period of two years, to be a very versatile form for presentation of computed results, and its use at the Culham Laboratory is by no means confined to the study of functions of a complex variable.

The iteration procedure available in the program is of the form

$$z' = z - \lambda \frac{f(z)}{g(z)} \quad (9.1)$$

where $g(z)$ is a 3-point approximation to $\frac{df}{dz}$. In Fig. 11

the points z_1, z_2 and z_3 are spaced equidistantly around the circle centred on z , and the approximation to the derivative is given by

$$\frac{df}{dz} \simeq g(z) = \frac{f(z_1) + e^{-\frac{2i\pi}{3}} f(z_2) + e^{-\frac{4i\pi}{3}} f(z_3)}{3(z_1 - z)} \quad (9.2)$$

Normally λ is set equal to unity but is reduced appropriately, after first trying a more accurate estimate of $\frac{df}{dz}$, if $|f(z')|$ is found to be greater than $|f(z)|$. The

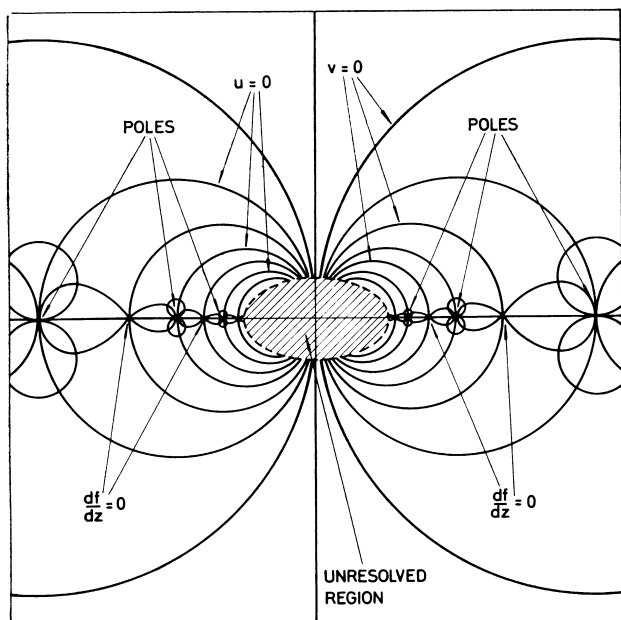


Fig. 9.—Contours of the real and imaginary parts of

$$f(z) = \operatorname{cosec}\left(\frac{1}{z}\right)$$

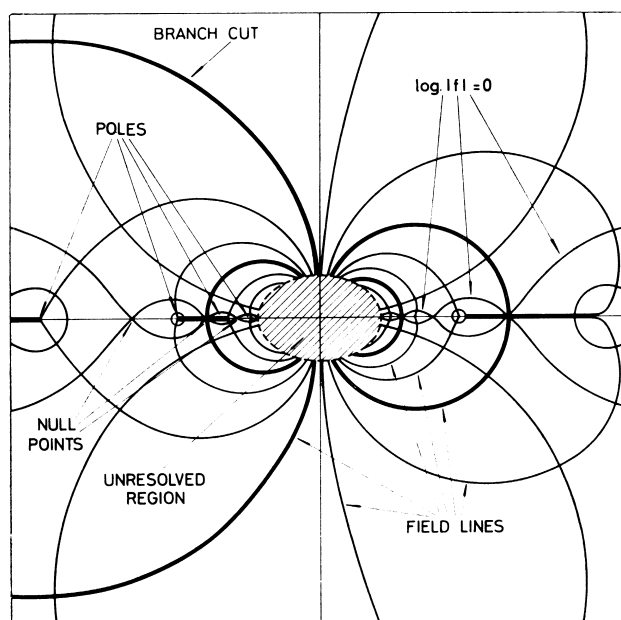


Fig. 10.—The electrostatic field associated with

$$f(z) = \operatorname{cosec}\left(\frac{1}{z}\right)$$

radius of the circle, on which z_1 , z_2 and z_3 lie, decreases as the iteration converges, as shown in Fig. 11. This method avoids the necessity for explicit evaluation of $\frac{df}{dz}$ whilst retaining quadratic convergence characteristics similar to those of Newton's method proper. A disadvantage is that $f(z)$ must be computed three times at every stage, but the method's inherent simplicity and reliability make it attractive for cases where $f(z)$ is easily computed. Moreover, this iteration is convergent in many cases where Newton's method proper is divergent, since the limiting case, when λ is made small, must be convergent.

10. Conclusions

An automatic graph plotter, used as an on-line or off-line digital computer output device, is capable of providing reasonably accurate regional and global information about functions of a complex variable. This information, in the form of the approximate locations and types of zeros, singularities and branch cuts in the complex plane, enables one to choose starting points for a zero-finding iteration so that convergence of the iteration may be guaranteed. In special cases the convergence region for a particular zero may be located exactly, but in any case a preliminary graphical study can isolate the zero well enough to inspire confidence in the convergence of a subsequent iteration. Moreover, the physical insight provided by the electrostatic-field interpretation of a function of a complex variable is helpful in understanding the function generally, as well as in guiding one's choice of a starting point for the iteration.

References

- Methods of NEWTON and BAIRSTOW: see, for example, *Modern Computing Methods*, published by the National Physical Laboratory, 1961.
 GRAEFFE's method: see, for example, *The Calculus of Observations*, by Whittaker and Robinson, Blackie & Son, 1952.
 LANCE, G. N.: *Numerical Methods for High Speed Computers*, Iliffe, 1960.
 MULLER's method: *ibid.*

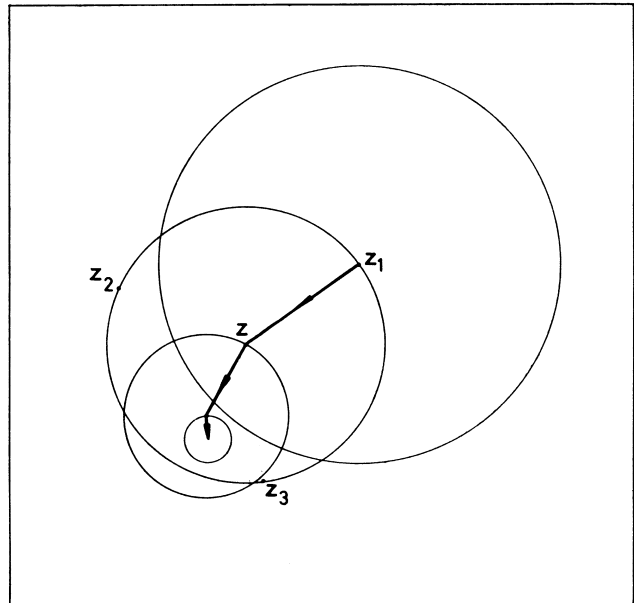


Fig. 11.—Progress of Newton-type iteration

The contour graphs presented illustrate how general features of a function may be recognized at a glance.

Acknowledgements

The author is indebted to a number of colleagues at the Culham Laboratory, both visitors and permanent staff, for helpful discussions on the above work and for programming assistance; also to the U.K.A.E.A. for permission to publish.