Search for physical structures on the boundary by optimal analytic continuation from a finite set of interior data points

S. Ciulli
Dublin Institute for Advanced Studies, Dublin 4, Ireland

T. D. Spearman
School of Mathematics, Trinity College, Dublin 2, Ireland
(Received 4 May 1982)

The object of this paper is to develop a method for obtaining information about the discontinuity function along the cuts (which is related to the positions and widths of resonances), from data—either experimental or theoretical—given at some points inside the holomorphy domain. This will be achieved by means of an analytic continuation which is required to be optimal under some specific boundary conditions. Once errors are present or the number of data points is finite, analytic continuation is no longer unique but highly unstable. To give a well defined continuation prescription, a stabilizing condition is essential, and the latter has to be chosen to suit the physical problem under consideration. It is shown how such a continuation procedure may be used (a) to ascertain whether the data can be said to require a particular type of structure on the boundary such as that which would arise from a nearby pole on the second Riemann sheet, as would be associated with a resonance, and (b) if so, to determine the parameters of such a resonance. Among the many applications of the method derived in this paper, one which is of some topical interest, is to use as input the result of perturbative calculations in some region of the complex plane where such expansions may be meaningful (e.g., asymptotic or negative energy in QCD) and to attempt to compute quantities of physical interest in a region where direct perturbative calculations are not valid.

I. INTRODUCTION

Analytic continuation is an extremely important and widely used technique in physics. Many functions of physical interest are known (or supposed) to have analytic properties. When expressed in terms of suitably chosen variables these functions will be holomorphic in some domain and possess certain singularities, typically poles and branch points. Some range of values in the variables will form the physical region for a particular process; within this region the values of the function may be determined experimentally. Elsewhere within the domain of holomorphy, but outside the physical region for the process initially considered, the values of the function may also be of considerable interest. They might, for example, be related to empirical data for a different process, or they might cast light on some theoretical features of the mechanism involved. The obvious question is: Given information about the function within some data region what can we deduce about its values elsewhere? Can we perform an analytic continuation from the data region to some other part of the holomorphy domain which is of interest?

This problem is a particularly familiar one in particle physics. Since 1954, when the first dispersion relations for scattering amplitudes were derived,[1] the analyticity properties of amplitudes, some proven and others conjectured, have been widely exploited and have played a central role both in the development of theory and also in the phenomenological analysis of experimental data. Two-particle scattering amplitudes may, for example, be analytic functions of the energy. The physical region will be along part of the positive real axis and there the amplitudes (or their moduli) may be determined experimentally. It is of interest to relate the values thus obtained to those of the amplitude elsewhere on the real axis. The values for negative real energy will be associated with another scattering process where independent data may be available. If poles can be located their positions and residues will be the masses of intermediate single-particle states and associated coupling constants.

The role of analytic continuation in particle physics is not limited to phenomenological applications. It has long been known, for example, that the calculation of individual graphs within a perturbation expansion can be simplified by using analyticity.[2] In recent years there has been a renewed interest in analytic continuation as an adjunct to theoretical computations. However, the emphasis has shifted from problems where information about the singularities (in particular, the imaginary part of the amplitude along a cut) was used as input, to those in which this information cannot be computed directly and instead of featuring as input it becomes the target of the continuation procedure. Indeed, various theoretical methods, for example, perturbation computations for asymptotically free field theories such as QCD, are valid only within a limited region of the complex plane, and hence one turns naturally to analytic continuation to extend these results to the cuts, as is necessary if one is to determine important parameters such as those defining resonances.

There has been considerable interest in the techniques introduced by Shifman, Vainshtein, and Zakharov[3] and others to extend the applicability of perturbation theory in QCD. Shifman *et al.* consider the process of e^+e^- annihilation into hadrons. Perturbative results from QCD are believed to be reliable for large Q^2 ($Q^2=-s$ where s is the invariant e^+e^- mass squared) but information about resonances comes from small negative Q^2.

A standard approach to problems of this kind is to use Borel summation[3] or to follow an analogous procedure using some suitable analytic weight factor.[4] Provided some specific conditions are met it is possible to use the information contained in the numerical values of the expansion coefficients to reconstruct the function of interest not in the form of the original series, which diverges, but in that of a suitable integral representation. However it is not clear that these necessary conditions are satisfied in the case of QCD. If they are satisfied then, at least in principle, one could compute the result with arbitrary accuracy, but if these conditions are not met then one has to set a more limited objective and adopt a modified procedure.

A practical way of proceeding, particularly when looking for information about resonances which one believes may exist and therefore contribute, is to take moments and appropriate ratios of moments in the dispersion relations for the function of interest—for example, the vacuum-polarization tensor in QCD. Bell and Bertlmann[5] have refined this procedure and have shown (they repeated the computations in the context of a simple potential model where the final result may be checked) how this may be done to determine the positions of the first resonances. Basically the procedure is one of analytic continuation; it must be clear, however, that since a truncated perturbation expansion is itself an analytic function an exact continuation to the resonance will yield just the perturbation result itself which is known to be inaccurate there. However in the input region (large Q^2, in QCD) the perturbation result is close to the actual values. So the problem is to start with functions which are quite close to the approximate input in the large-Q^2 region and find an adequate continuation procedure which will (i) filter out any functions having unsuitable properties (such as wrong threshold behavior) which the truncated perturbation expansion might have, and (ii) ensure the appropriate physics (resonance structure) on the cuts.

In this paper we shall address problems of just the type described—making an analytic continuation from a set of approximate data with associated errors, and selecting, out of the range of solutions to that problem, the function which best meets certain specified conditions such as resonant type behavior within some region. Functional analysis provides powerful techniques for solving such problems. The key results are obtained in a compact and elegant closed form.

II. OUTLINE OF THE METHOD

One way of performing the analytic continuation between two regions of interest is to construct a representation of the amplitude, such as a dispersion relation, to which the data may be fitted. Such a representation or model will incorporate the analytic properties which the amplitude is believed to possess, but it will also include various other assumptions such as asymptotic behavior and some form of parametrization over parts of the axis where data may not be available. An alternative procedure is to take a function as it is defined by the empirical or theoretical values (with associated errors) at a set of discrete data points within the physical region, and to perform an analytic continuation which will attribute values to the function at other points within the domain of holomorphy. These values will be determined directly by the input data, but this procedure will also require some further conditions if it is to be well defined, since there is certainly not a unique analytic continuation from a discrete set of points. Even if the data set could be regarded as a continuum, so that a unique analytic continuation would appear to be given, one must consider the question of stability; whenever there are nonzero errors (experimental or theoretical) associated with the initial values, the mode of propagation of these errors must be taken into account. Indeed, one finds that without some further assumptions (for example, boundedness or smoothness) the continuation is not at all stable with respect to the initial input as the errors may be enhanced by arbitrarily large factors.[6]

This problem of stabilization of analytic continuation has been studied in some detail by Cutkosky *et al.*[7] and by one of the present authors (S.C.) and his collaborators.[8,9]

In practice, of course, data is obtained for a finite number of discrete values of the relevant variables, so that the data set is a finite set of discrete points. Consequently we focus our attention on the problem of performing an analytic continuation directly from a discrete set of points. As we have already pointed out, the more familiar procedure is to first fit the data within the data region to some interpolating function (for example, a polynomial or similar expansion) and to use this as the starting point for the analytic continuation.[10] The choice of this interpolating function may already incorporate some assumptions about the function which we seek to construct. We prefer to proceed directly from the data—this allows us to separate and make explicit the further conditions to be imposed in order to discriminate between the possible functions which could be obtained by analytic continuation alone. It also turns out that the continuation problem from a discrete set of n points, subject to stabilizing conditions of a certain type, may be expressed as an n-dimensional problem which can be solved explicitly; this is a major simplification and allows the results to be expressed in a very convenient closed form.

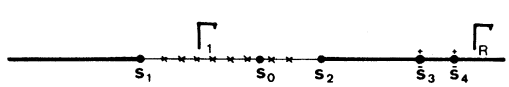

FIG. 1. Domain of holomorphy of the function $X(s)$: the cuts Γ_R extend from the branch points s_1 and s_2 along the real axes to $-\infty$ and $+\infty$, respectively; s_3, s_4,... are further branch points on the real axis; Γ_1 is the data region.

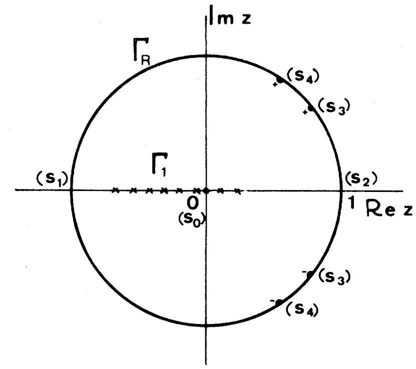

FIG. 2. Result of the conformal mapping which takes the cut s plane of Fig. 1 into the interior of the unit disk, $s=s_0$ onto $z=0$, the cuts Γ_R mapping, as shown, onto the circle $|z|=1$. If

$$u=[(s-s_0)(s_1+s_2-2s_0)-2(s_1-s_0)(s_2-s_0)]\times[(s_2-s_1)(s-s_0)]^{-1},$$

then

$$z(s)=u-(u-1)^{1/2}(u+1)^{1/2},$$

where the square roots are defined to have right-hand cuts and non-negative imaginary parts.

The problem, then, which we wish to solve is to find that analytic function which both adequately fits the data, given on the discrete set of data points, and also satisfies a specified stabilizing condition. The adequacy of the fit to the data is measured in terms of a χ^2 fit in the standard way. The stabilizing conditions which we consider relate to the smoothness or boundedness of the function defined in some particular way over a region of the complex plane. In practice we specify a value of χ^2 and for any such value construct that analytic function whose fit to the data has that value of χ^2 and which is optimal in the sense of best meeting the stabilizing conditions. In this paper we shall consider two particular types of stabilizing conditions; both are defined on the cuts which form the boundary to the domain of holomorphy.

In order to give a standard specification of these conditions it is convenient to first perform a mapping of the complex plane to the unit disk. The procedure is as follows.

The holomorphy domain typically has the form shown in Fig. 1. The complex plane has cuts Γ_R extending from branch points s_1, s_2 to $\pm\infty$ as shown. For the problem which we wish to consider the data are given in a region Γ_1 lying on the real axis between the two branch points, where the function of interest $X(s)$ is holomorphic. We shall be particularly interested here in analytic continuation from Γ_1 to the cuts Γ_R which form the boundary of the holomorphy domain and on which the stabilizing conditions are defined.

The cut complex plane of Fig. 1 then is mapped onto the z unit disk as shown in Fig. 2. The data region Γ_1 remains on the real axis, the cuts Γ_R now form the unit circle. Two types of stabilizing conditions will be considered, namely,

Type A:

$$\frac{1}{2\pi}\int_0^{2\pi} |\operatorname{Re}X(e^{i\phi})|^2\sigma(\phi)d\phi < \text{bound} \tag{1}$$

and

Type B:

$$\frac{1}{2\pi}\int_0^{2\pi} \left|\frac{d}{d\phi}\operatorname{Im}X(e^{i\phi})\right|^2 \sigma(\phi)d\phi < \text{bound} , \tag{2}$$

where $\sigma(\phi)$ is a real weight function which has to be strictly positive ($>\epsilon,\epsilon>0$), and even $[\sigma(-\phi)\equiv\sigma(2\pi-\phi)=\sigma(\phi)]$. Only real analytic functions, which satisfy the condition $X(\bar{z})=\bar{X}(z)$, are considered. A consequence of this is that $X(z)$ is real on the real axis between s_1 and s_2 (but not on the cuts Γ_R). Thus the data, which are given on the real axis within the holomorphy domain, must be real.[11] The function $\sigma(\phi)$, once the above conditions are satisfied, may be chosen freely. So it might be constructed in such a way as to filter out the unwanted truncated perturbative solution [see point (i) in the Introduction]. This might be achieved simply by choosing it so that the integrals occurring on the left-hand sides of the inequalities (1) or (2) should diverge for the perturbative function, thus automatically ensuring that the latter will stay outside the set of functions defined by the stabilization conditions A or B.

In an earlier paper[12] we have shown how to solve the optimization problem of constructing the functions, holomorphic inside the unit circle, which assume specified precise values at the points of the data set Γ_1 and which solved the extremum problems A or B below:

Extremum problem A:

$$\frac{1}{2\pi}\int_0^{2\pi} |\operatorname{Re}X(e^{i\phi})|^2\sigma(\phi)d\phi \to \text{least} \equiv \delta_0^{A^2} , \tag{3}$$

Extremum problem B:

$$\frac{1}{2\pi}\int_0^{2\pi} \left|\frac{d}{d\phi}\operatorname{Im}X(e^{i\phi})\right|^2 \sigma(\phi)d\phi \to \text{least} \equiv \delta_0^{B^2} , \tag{4}$$

where the extremum is sought (see below) among all functions $X(z)=X^{(1)}(z)+M(z)$. Here $X^{(1)}(z)$ is some given real analytic function which assumes the values a_i at the points $z=z_i$, while the $M(z)$ are (any) real analytic functions which vanish at all data points z_i,

$$M(z_i)=0, \quad i=1 \text{ to } n , \tag{5}$$

so that all $X(z)$ will assume the same specified values a_i as $X^{(1)}(z)$ at the points $z=z_i$. As will be seen below, these extremum problems are of great importance for continuation procedures. The explicit results for the constants δ_0 as well as the extremum function $X^{(0)}(z)$, will be presented in Sec. III.

Experimental data (or in some cases theoretical input such as the QCD input discussed above) do not, of course, come as precise values: each number will have an associated error and to fit the data we must allow a variation about the median values subject to a measure, such as the standard χ^2 measure, of the quality of fit. In Sec. IV we show how the results of Ref. 12, based on a precise fit to certain data, may be extended to provide a solution to the problem of finding the holomorphic function which fits the data within a certain χ^2 limit and, subject to this, optimizes the above conditions.

The type of stabilizing condition which can be treated by our functional methods is restricted by the requirement that it should define some kind of norm on the function space under consideration. Conditions A and B, in particular, satisfy these requirements. In fact these two types of conditions are quite powerful: they have the flexibility of allowing one to choose a weight function $\sigma(\phi)$ which will emphasize particular parts of the region Γ_R to which we wish to make the continuation. The choice of norm will depend on the judgment by the user of which of the following, the magnitude of the amplitude or its variation, is more relevant to the underlying physics. Condition A is a boundedness constraint, whereas B puts a strong premium on smoothness. In particular, the integral in B would be very sensitive to resonance-type behavior, as would arise from a nearby pole on the second Riemann sheet. Indeed, in the discrepancy method described below, the structure to be detected manifests itself particularly through the variation of the imaginary part of the amplitude so that condition B is appropriate for the determination of resonance parameters.

These nearby second-sheet poles do, of course, frequently occur and, indeed, the interest in performing the analytic continuation from the data region is typically to ascertain whether such poles are required by the data. [Note that the higher Riemann sheets of the function in the original s variable (Fig. 1) are mapped to the exterior of the unit z disk of Fig. 2.] Clearly there is little point in applying condition B [Eq. (2)] directly to the function under consideration when this function is likely to have a pole close to the circle, since the existence of such a pole will give rise to large values of $[d(\mathrm{Im}X)/d\phi]^2$. But we can proceed as follows: we take a trial function $T_\kappa(z)$, normally defined by a conjugate pair of second-sheet poles, and we subtract from the data the value of the trial function appropriate to each data point z_i. The data thus modified define a new function which corresponds to the original function with a pole term subtracted from it. We call this new function the discrepancy function[13] and note that it depends parametrically on the position and residue of the trial pole. Now, unless the parameters κ of this trial function accurately represent the pole which the original data require, the imaginary part of the discrepancy function will necessarily have significant structure on the circle and the minimum value δ_0 of the integral (2) for the discrepancy function $D_\kappa(z)$,

$$\delta_0=\inf\left\{\frac{1}{2\pi}\int_0^{2\pi}\left[\frac{d}{d\phi}[\mathrm{Im}D_\kappa(e^{i\phi})]\right]^2\sigma(\phi)d\phi\right\}^{1/2}, \tag{6}$$

will be larger than the value of δ_0 for the function defined by the unmodified data, as the effect of subtracting the trial pole will be to introduce more structure. However, if the parameters κ of the trial pole do coincide with those (κ_0) of the pole sought by the data, then the discrepancy function will be relatively smooth on the circle—certainly in the neighborhood of the pole, which can be emphasized within the norm integral by choosing a suitable $\sigma(\phi)$—and hence also $\delta_0\equiv\inf$ (norms) will be small. Assuming that the original data do require a pole, this value of δ_0 for the discrepancy function with the right pole parameters will be much less than the value of δ_0 computed from the unmodified data: the true values of the pole's parameters may hence be found by watching the minima of $\delta_0[D_\kappa]$.[14]

III. SOLUTION OF THE EXTREMUM PROBLEMS

In this section we review briefly the solution of the extremum problems A and B defined by Eqs. (3) and (4). We first establish the notation to be used, then outline the method of solution (which is described in detail in Ref. 12) and finally we presented the detailed results since these form the basis for the subsequent calculations. Problems A and B will be treated separately as there are some major differences even though the general method used is the same. The key to the method is the duality theorem[15] of functional analysis, a direct consequence of the well-known Hahn-Banach lemma.

Notation

We shall use capitals, e.g., $X(z)$, $Y(z)$, $M(z)\ldots$, to represent analytic functions, and correspondingly subscripted symbols $X_{\mathrm{Re}}(z)$, $X_{\mathrm{Im}}(z)$, . . . , to denote their real and imaginary parts. z' will be used to denote points on the unit circle $z'=e^{i\phi}$, and we shall frequently write

$$X_{\mathrm{Re}}(z')\equiv X_{\mathrm{Re}}(e^{i\phi})\equiv x(\phi) \tag{7}$$

using the lower-case letter to denote real functions of ϕ obtained as shown. The functions $x(\phi)$ will always be periodic so that $x(2\pi-\phi)=x(-\phi)$; since $X(z)=\bar{X}(\bar{z})$, $x(\phi)$ is even, $x(\phi)=x(-\phi)$. Further $\{z_i\}\equiv\{\mathrm{Re}z_i\}$ is the set of given points on the real axis and $\{a_i\}$ the real values specified.

The boundary value functions $x(\phi)$ are of central importance for defining both linear functionals and also norms for $X(z)$, each constructed in such a way as to suit the extremum problem under consideration. For instance, one may define a norm for $F(z)$ related to the L^2 norm of $f(\phi)$:

$$||F(z)||=\left\{\frac{1}{2\pi}\int_0^{2\pi}[f(\phi)]^2\sigma(\phi)d\phi\right\}^{1/2}, \tag{8}$$

where, following the above notation (Eq. 7)

$$f(\phi)=F_{\mathrm{Re}}(e^{i\phi})=\mathrm{Re}F(e^{i\phi}),$$

and where $\sigma(\phi)$ is a real, positive weight function, satisfying the condition

$$\sigma(\phi)=\sigma(-\phi). \tag{9}$$

We shall also have to deal with linear functionals Y^* acting on the analytic functions $X(z)$. Since the functions $X(z)$ may be expressed linearly in terms of their boundary values $x(\phi)$, the functionals Y^* may be seen as functionals y^* acting on these boundary value functions.[12] The Riesz theorem allows each such linear functional y^* to be associated with a real function $y(\phi)$ (which is even in ϕ) as follows:

$$\langle X,Y^*\rangle \equiv \langle x,y^*\rangle = \frac{1}{2\pi}\int_0^{2\pi} y(\phi)x(\phi)\sigma(\phi)d\phi \; , \tag{10}$$

where we have introduced the following notation for the functional Y^*:

$$Y^* \equiv \langle \cdot, Y^* \rangle \; .$$

Formulation of the problem

The objective is to construct a function $X(z)$ with the following properties:

(i) $X(z_i)=a_i$ where the points z_i are real and the values a_i are also real: $-1<z_i<+1$.

(ii) $X(z)$ is holomorphic in the unit disk, and $X(\bar{z})=\bar{X}(z)$.

(iii) Subject to (i) and (ii) above, $X(z)$ should satisfy the condition that $||X(z)||$ should have the least possible value. Here the norm $||X||$ is defined according to which of the conditions A or B we wish to implement.

The procedure we adopt is the following. A particular function $X^{(1)}(z)$ is constructed to possess properties (i) and (ii); this is an easy task, specifically we make the choice

$$X^{(1)}(z)=\sum_i a_i \frac{(z-z_1)(z-z_2)\cdots(z-z_i)_d\cdots(z-z_n)}{(z_i-z_1)(z_i-z_2)\cdots(z_i-z_i)_d\cdots(z_i-z_n)} \; , \tag{11}$$

with the convention that the factors with a subscript d are to be deleted. Now if $M(z)$ is any function which has value 0 at each of the points z_i and is holomorphic in the unit disk with $M(\bar{z})=\bar{M}(z)$, then the function $X^{(1)}(z)-M(z)$ also possesses properties (i) and (ii), and conversely, any function with those two properties may be represented in this form. So the function $X^{(0)}(z)$ possessing properties (i), (ii), and (iii) (which gives the solution to our problem) is

$$X^{(0)}(z)\equiv X^{(1)}(z)-M^{(0)}(z) \; , \tag{12}$$

where $M^{(0)}(z)$ is the solution to the minimization problem to be defined below.

At this stage it becomes necessary to be specific about the particular extremum problem to be solved. We shall first solve problem A and return later to problem B. Problem A is in fact a variant of a rather classical problem and can be approached in a variety of ways, see Refs. 16 and 17.

So, for the present we adopt the norm defined in Eq. (8) which is appropriate to the Dirichlet boundary condition of problem A. In terms of that norm the minimization problem for $M(z)$ is

$$\delta_0=\inf_M ||X^{(1)}(z)-M(z)|| \; , \tag{13}$$

the minimization being with respect to the class of functions $M(z)$ which satisfy

$$\begin{aligned} &M(z) \text{ holomorphic for } |z|<1 \; , \\ &M(\bar{z})=\bar{M}(z) \; , \\ &M(z_i)=0, \quad i=1,\ldots n \; . \end{aligned} \tag{14}$$

The duality theorem,[15] when applied to this minimization problem, yields

$$\delta_0=\inf_M ||X^{(1)}-M|| = \sup_{Y^*:\langle M,Y^*\rangle=0,||Y^*||=1} \langle X^{(1)},Y^*\rangle \; , \tag{15}$$

where the functional $\langle \cdot, Y^*\rangle$ is as defined in Eq. (10) and the extremum problem is now with respect to the class of functionals $\langle \cdot, Y^*\rangle$ which satisfy the conditions

$$\frac{1}{2\pi}\int_0^{2\pi}[y(\phi)]^2\sigma(\phi)d\phi=1 \; , \tag{16a}$$

$$\langle M,Y^*\rangle=0, \quad \text{for all } M(z) \text{ satisfying condition (14)} \; , \tag{16b}$$

$$y(\phi)=y(-\phi) \; . \tag{16c}$$

An explicit representation for the functionals $\langle \cdot, Y^* \rangle$

We need to identify the class of functionals satisfying conditions (16) but before doing this it is necessary to look more closely at the class of functions $M(z)$. Associated with each holomorphic function $M(z)$ there is a real function $m(\phi)$, which, following the notation we have established, is

$$m(\phi) \equiv M_{\mathrm{Re}}(e^{i\phi}) \equiv \mathrm{Re} M(e^{i\phi}) \ . \tag{17}$$

Conversely, once $m(\phi)$ is specified (a real, even [i.e., $m(\phi)=m(-\phi)$], square-integrable function on $[0,2\pi]$) the analytic function $M(z)$ is completely[15] determined and may be expressed as

$$M(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\phi}+z}{e^{i\phi}-z} m(\phi) d\phi \ . \tag{18}$$

Equation (18) is the Schwarz-Villat formula, which is simply the complex extension of the well-known Poisson integral by means of which harmonic functions $M_{\mathrm{Re}}(z)$ are constructed from their boundary values:

$$M_{\mathrm{Re}}(z) = \frac{1}{2\pi}\int_0^{2\pi} \mathscr{P}(z,z') m(\phi) d\phi \ . \tag{19a}$$

Here $\mathscr{P}(z,z')$ is the Poisson kernel

$$\mathscr{P}(z,e^{i\phi}) \equiv \mathrm{Re}\left[\frac{e^{i\phi}+z}{e^{i\phi}-z}\right] \equiv \frac{1-r^2}{1+r^2-2r\cos(\theta-\phi)} \ . \tag{19b}$$

Now consider the set of functionals $\langle \cdot, y^* \rangle$ defined by the following even special functions:

$$y(\phi) = \sum y_i \mathscr{P}(z_i, e^{i\phi})[\sigma(\phi)]^{-1} \ , \tag{20}$$

where the y_i are arbitrary real constants. One sees at once that[12]

$$\begin{aligned}\langle M, Y^* \rangle &\equiv \langle m, y^* \rangle \\ &= \sum_i y_i \frac{1}{2\pi}\int \mathscr{P}(z_i,e^{i\phi})[\sigma(\phi)]^{-1} \\ &\qquad \times m(\phi)\sigma(\phi)d\phi \\ &= \sum_i y_i M(z_i) \\ &= 0 \ ,\end{aligned} \tag{21}$$

since $M(z_i) = \mathrm{Re} M(z_i) = 0$. Hence any functional $\langle \cdot, y^* \rangle$, constructed by means of the functions $y(\phi)$ defined in Eq. (20), automatically satisfies the requirements (16b). In Ref. 12 it is demonstrated that *any* linear functional $\langle \cdot, y^* \rangle$ satisfying Eq. (16b) can be expressed in terms of a function $y(\phi)$ having the form (20). So the set of functions specified by Eq. (20) defines the class of functionals required for the extremum problem of Eq. (15), provided the normalization condition (16a) is satisfied. This condition becomes, in terms of the coefficients y_i of Eq. (20),

$$\sum_{ij} \alpha_{ij} y_i y_j = 1 \ , \tag{22}$$

where the constants α_{ij} are

$$\alpha_{ij} = \frac{1}{2\pi}\int_0^{2\pi} \mathscr{P}(z_i,e^{i\phi})\mathscr{P}(z_j,e^{i\phi})[\sigma(\phi)]^{-1} d\phi \ . \tag{23}$$

This integral may in fact be evaluated explicitly if we introduce a holomorphic function $S(z)$ whose real part has the value $[\sigma(\phi)]^{-1}$ when $z'=e^{i\phi}$. This may be done immediately using the Schwarz-Villat formula

$$S(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\phi}+z}{e^{i\phi}-z}\frac{1}{\sigma(\phi)} d\phi \ . \tag{24}$$

The weight function $\sigma(\phi)$ was restricted to be a strictly positive function so that $[\sigma(\phi)]^{-1}$ is bounded. It is also required to satisfy the condition $\sigma(\phi)=\sigma(-\phi)$ [Eq. (9)] from which it follows that

$$S(\bar{z}) = \bar{S}(z) \ ,$$

and in particular that the values $S(z_i)$ are real since the points z_i are real.

The detailed evaluation of the integral in Eq. (23) is explained in Ref. 12. The results obtained are, for $i \neq j$,

$$\alpha_{ij} = 1/2\left\{\frac{S(z_i)-S(z_j)}{z_i - z_j}(z_i+z_j) + [S(z_i)+S(z_j)]\left[\frac{1+z_iz_j}{1-z_iz_j}\right]\right\} \tag{25a}$$

and for $i=j$,

$$\alpha_{ii} = z_i S'(z_i) + S(z_i)\left[\frac{1+z_i^2}{1-z_i^2}\right] \ , \tag{25b}$$

where $S' \equiv dS/dz$. Note that if the weight func-

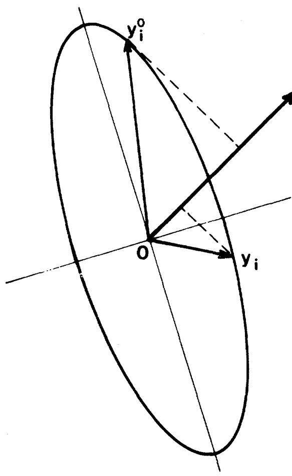

FIG. 3. The ellipsoid $\sum \alpha_{ij} y_i y_j = 1$, and the optimal vector $y_i^{(0)}$ whose projection on the data vector a_i is largest.

tion $\sigma(\phi)$ were a constant, which we could take to be 1 so that $S(z_i)=S(z_j)=1$, then the result for all i and j would be

$$\alpha_{ij} = \frac{1+z_i z_j}{1-z_i z_j} \ . \tag{26}$$

In this case, all the coefficients α_{ij} are clearly positive; in fact, since $\sum \alpha_{ij} y_i y_j$ is a *norm* and hence cannot vanish unless all y_i are identically zero, the matrix α_{ij} is always positive definite and $\sum \alpha_{ij} y_i y_j = 1$ represents an ellipsoid.

The extremum problem

We wish to determine δ_0, where

$$\delta_0 \equiv \inf_M ||X^{(1)} - M|| \ , \tag{27}$$

and we also want to know the function $M^{(0)}$ giving the minimum, so that we will have obtained the function $X^{(0)} \equiv X^{(1)} - M^{(0)}$. We saw from Eq. (15) that the extremum problem (27) could be replaced by

$$\delta_0 = \sup_{Y^*} \langle X^{(1)}, Y^* \rangle \ , \tag{28}$$

where the supremum is taken with respect to the set of functionals satisfying Eqs. (16a)–(16c). This set of functionals can be represented by the set of functions $y(\phi)$, defined in Eq. (20). When we substitute for $\langle \cdot, Y^* \rangle$ in Eq. (28) we see that [cf. Eqs. (10) and (20)]

$$\delta_0 = \sup_{y_i} \sum_i y_i \frac{1}{2\pi} \int_0^{2\pi} \mathscr{P}(z_i, e^{i\phi}) [\sigma(\phi)]^{-1} x^{(1)}(\phi) \sigma(\phi) d\phi = \sup_{y_i} \sum y_i a_i \ , \tag{29}$$

since the integrals appearing here yield by definition the values of $X^{(1)}(z)$ at $z = z_i$, i.e., the constants a_i. The coefficients y_i must also satisfy the condition (22)

$$\sum_{i,j} \alpha_{ij} y_i y_j = 1 \ , \tag{22}$$

where the constants α_{ij} are given by Eqs. (25a) and (25b).

In geometric terms the problem is illustrated in Fig. 3. Since α_{ij} is positive-definite, Eq. (22) represents an n-dimensional ellipsoid and we look for that vector y_i on the ellipsoid whose component in the direction a_i is a maximum.

We can solve the problem analytically using Lagrange multipliers. We set

$$\Phi = \sum_i y_i a_i - \lambda \left[\sum_{ij} \alpha_{ij} y_i y_j - 1 \right] \tag{30}$$

and differentiate, to get the following equations for the optimal vector $y^{(0)}$:

$$\frac{\partial \Phi}{\partial y_i} = a_i - 2\lambda \sum_j \alpha_{ij} y_j = 0 \ . \tag{31}$$

It follows that

$$y_i^0 = \frac{1}{2\lambda} \sum_j (\alpha^{-1})_{ij} a_j \ . \tag{32}$$

λ is determined from Eq. (22) by substituting from Eq. (32) for $y_i^{(0)}$; the result is

$$\lambda = \tfrac{1}{2} \left[\sum_{ij} (\alpha^{-1})_{ij} a_i a_j \right]^{1/2} \ . \tag{33}$$

So the required vector y_i is

$$y_i^{(0)} = \frac{\sum (\alpha^{-1})_{ij} a_j}{\left[\sum (\alpha^{-1})_{ij} a_i a_j \right]^{1/2}} \tag{34}$$

and so the corresponding optimal functional $\langle \cdot, Y^{(0)*} \rangle$ is defined by the function

$$y^{(0)}(\phi)=\sum y_i^{(0)}\mathscr{P}(z_i,e^{i\phi})[\sigma(\phi)]^{-1} . \qquad (35)$$

Substitution from Eq. (34) into Eq. (29) gives the value of δ_0:

$$\delta_0=\left[\sum(\alpha^{-1})_{ij}a_ia_j\right]^{1/2} . \qquad (36)$$

The fact that we already have an explicit form [Eqs. (34) and (35)] for $y^{(0)}(\phi)$ allows us to determine also the optimal function $X^{(0)}(z)$: indeed, if the extremum is realized with $M^{(0)}$ and $y^{(0)}$ we have

$$\delta_0=||X^{(1)}-M^{(0)}||=\langle X^{(1)},Y^{(0)*}\rangle$$
$$=\langle X^{(1)}-M^{(0)},Y^{(0)*}\rangle , \qquad (37)$$

where the last step follows from Eq. (16b). But [setting $x^0(\phi)\equiv x^1(\phi)-m^0(\phi)$], from Schwarz's inequality we have the result

$$\langle x^0,y^0\rangle<||x^0||\times||y^0||\equiv||x^0|| \qquad (38)$$

(since $||y^0||\equiv 1$) unless $x^{(0)}(\phi)=ky^{(0)}(\phi)$, when equality occurs. Hence Eq. (37) tells us that the vectors $x^{(0)}(\phi)$ and $y^{(0)}(\phi)$ should be "aligned"; further, since [again Eq. (37)] $||x^{(0)}||=\delta_0$ and $||y^{(0)}||=1$, the constant k equals δ_0 and hence

$$x^{(1)}(\phi)-m^{(0)}(\phi)\equiv x^{(0)}(\phi)=\delta_0 y^{(0)}(\phi) . \qquad (39)$$

So the optimal function $x^{(0)}(\phi)$ can be written entirely in terms of known entities:

$$x^{(0)}(\phi)=\sum_{i,j}(\alpha^{-1})_{ij}a_j\mathscr{P}(z_i,e^{i\phi})[\sigma(\phi)]^{-1} . \qquad (40)$$

Finally, the corresponding complex function $X^{(0)}(z)$ is

$$X^{(0)}(z)=\sum_{i,j}(\alpha^{-1})_{ij}a_j\frac{1}{2\pi}\int_0^{2\pi}\left[\frac{e^{i\phi}+z}{e^{i\phi}-z}\right]\mathscr{P}(z_i,e^{i\phi})[\sigma(\phi)]^{-1}d\phi . \qquad (41)$$

We can immediately verify from Eq. (41) using Eq. (23), that, as must be the case, $X^{(0)}(z_i)$ does indeed have the value a_i.

The Neumann boundary condition—Problem B

In problem B we want to minimize the following integral:

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{\partial}{\partial\phi}\mathrm{Im}X(e^{i\phi})\right|^2\sigma(\phi)d\phi . \qquad (42)$$

Now the Cauchy-Riemann relations imply that

$$\frac{1}{r}\frac{\partial X_{\mathrm{Im}}}{\partial\phi}=\frac{\partial X_{\mathrm{Re}}}{\partial r}=\mathrm{Re}\left[\frac{dX}{dz}\frac{\partial z}{\partial r}\right]\equiv\frac{1}{r}\mathrm{Re}[X'(z)z] \qquad (43)$$

so that the minimization problem may be written as

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{\partial X_{\mathrm{Re}}(re^{i\phi})}{\partial r}\right|^2\sigma(\phi)d\phi\,_{|r=1}\to\text{least} . \qquad (44)$$

In order to proceed we need to be able to construct a complex function $X(z)$ which is holomorphic in the unit disk when the radial derivative of the real part $\partial X_{\mathrm{Re}}/\partial r$ is specified on the unit circle. This Neumann-type problem may be solved in terms of a Green's function analogous to the Poisson kernel of Eq. (19b). The Neumann kernel is derived in Appendix B of Ref. 12.[18] If we introduce the following notation,

$$\left.\frac{\partial F_{\mathrm{Re}}(re^{i\phi})}{\partial r}\right|_{r=1}=f_{,r}(\phi) ,$$

using $f_{,r}(\phi)$ to denote the radial derivative of the function $F_{\mathrm{Re}}(z)$ at the point $z=e^{i\phi}$, the required result takes the form [cf. Eqs. (B8)–(B10) of Ref. 12]:

$$F_{\mathrm{Re}}(z)=F_{\mathrm{Re}}(z_0)+\frac{1}{2\pi}\int_0^{2\pi}f_{,r}(\phi)\mathscr{N}(z_0;z,e^{i\phi})d\phi , \qquad (45)$$

where

$$\mathscr{N}(z_0;z,e^{i\phi})=-2\ln\left|\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right|$$
$$=-2\mathrm{Re}\left[\ln\left[\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right]\right] . \qquad (46)$$

$F(z)$ is given by

$$F(z)=F(z_0)-\frac{1}{\pi}\int_0^{2\pi}f_{,r}(\phi)\ln\left[\frac{e^{i\phi}-z}{e^{i\phi}-z_0}\right]d\phi\ . \tag{47}$$

As would be expected because of the nature of the Neumann condition a subtraction (arbitrary constant) is required.

If we are to proceed in analogy with the Dirichlet case, the next step should be to try to define a norm for $F(z)$ by means of the boundary function $f_{,r}(\phi)$:

$$||F(z)||=\left\{\frac{1}{2\pi}\int_0^{2\pi}[f_{,r}(\phi)]^2\sigma(\phi)d\phi\right\}^{1/2}\ . \tag{48}$$

The difficulty which immediately faces us is that because of the subtraction in Eq. (47), the right-hand side of Eq. (48) does not define a valid norm for $F(z)$ since the latter might be nonzero even if the right-hand side of Eq. (48) is zero.

This difficulty can be circumvented by restricting ourselves to the space[15] $\{F(z)\}$ of $F(z)$ vanishing at $z=z_0$. For convenience we choose one of the points z_i at which the values a_i of $F(z)$ are prescribed as the subtraction point z_0; to be specific we take $z_0\equiv z_1$. The optimization condition, Eq. (44), is not altered if we replace the initial function $X(z)$ by $X(z)-a_1$, that is, if we replace the set of values $a_1,\ldots,a_n$ by $0, a_2-a_1\ldots,a_n-a_1$. Proceeding otherwise as before, the functions $M(z)$ will be required, as usual, to be zero at all the points $z_1,\ldots z_n$. This restriction on the space $\{F(z)\}$ allows us to define a unique function $F(z)$ associated with each real radial derivative function $f_{,r}(\phi)$:

$$F(z)=-\frac{1}{\pi}\int_0^{2\pi}f_{,r}(\phi)\ln\left[\frac{e^{i\phi}-z}{e^{i\phi}-z_1}\right]d\phi \tag{49}$$

and the integral from the right-hand side of Eq. (48) is now indeed a norm for $F(z)$.

As before, we start, now with a function $X^{(2)}(z)$ defined to be holomorphic and to take the values 0, $a_2-a_1,\ldots a_n-a_1$ at the points z_i. Specifically, we choose

$$X^{(2)}(z)\equiv X^{(1)}(z)-a_1\ , \tag{50}$$

where $X^{(1)}(z)$ is defined in Eq. (11). We then have to solve the infimum problem

$$\delta_0\equiv\inf_M||X^{(2)}-M||\ , \tag{51}$$

where the norm is defined as in Eq. (48). The infimum is with respect to the set of functions $M(z)$ defined in Eq. (14). The function $M(z)$ which gives the least value of δ_0 will be denoted by $M^{(0)}(z)$ and the corresponding $X(z)$ by $X^{(0)}(z)$:

$$X^{(0)}(z)=X^{(2)}(z)-M^{(0)}(z)\ . \tag{52}$$

As before we use the duality theorem to replace the infimum problem by a supremum one:

$$\delta_0\equiv\inf_M||X^{(2)}-M||=\sup_{Y^*}\langle X^{(2)},Y^*\rangle\ , \tag{53}$$

where the supremum is with respect to the set of functionals $\langle\cdot,Y^*\rangle$ defined by Eqs. (16a) to (16c). In this case the set of functionals satisfying Eqs. (16a) to (16c) turns out to be (see Ref. 12) the set defined by

$$y(\phi)=\sum_{i=2}^{n}y_i\mathcal{N}(z_1;z_i,e^{i\phi})[\sigma(\phi)]^{-1}\ , \tag{54}$$

where the constants y_i are real and take all possible values subject to the normalization condition (16a). The summation is from 2 to n since

$$\mathcal{N}(z_1;z_1,e^{i\phi})=0\ . \tag{55}$$

The extremum problem expressed by Eq. (53) may now be written as

$$\begin{aligned}\delta_0&=\sup\sum_{i=2}^{n}y_i\frac{1}{2\pi}\int_0^{2\pi}\mathcal{N}(z_1;z_i,e^{i\phi})x^{(2)}{}_{,r}(\phi)d\phi\\&=\sup\sum_{i=2}^{n}y_i(a_i-a_1)\ .\end{aligned} \tag{56}$$

The coefficients y_i must satisfy the normalization condition

$$\sum_{i,j=2}^{n}\alpha_{ij}y_iy_j=1\ , \tag{57}$$

where

$$\alpha_{ij}=\frac{1}{2\pi}\int_0^{2\pi}\mathcal{N}(z_1;z_i,e^{i\phi})\mathcal{N}(z_1;z_j,e^{i\phi})[\sigma(\phi)]^{-1}d\phi\ . \tag{58}$$

This integral is evaluated in Appendix C of Ref. 12 where it is shown that [for the definition of $S(z)$ see Eq. (24)]:

$$\alpha_{ij}=\left[\ln\left|\frac{z_i}{z_1}\right|\ln\left|\frac{z_j}{z_1}\right|-\pi^2\theta(-z_1)\theta(z_i)\theta(z_j)\right]S(0)+A_{ij}+A_{ji}\ , \tag{59}$$

with A_{ij} defined by

$$A_{ij}=-\mathrm{P}\int_{z_1}^{z_i}\left\{\ln\left|\frac{1-z'z_j}{1-z'z_1}\right|+\ln\left|\frac{z'-z_j}{z'-z_1}\right|\right\}\frac{S(z')}{z'}dz'\ . \tag{60}$$

This extremum problem is completely analogous to that described by Eqs. (29) and (22) except that it is now $n-1$ dimensional. In Fig. 3 the ellipsoid is now in an $(n-1)$-dimensional space and the n-vector c_i is replaced by the $(n-1)$-dimensional vector (a_i-a_1), $i=2$ to n. The extremum calculation, using a Lagrange multiplier to take account of Eq. (57), yields the value of δ_0 and the functional $\langle\cdot,Y^{(0)*}\rangle$ which gives the supremum

$$\delta_0=\sup\langle X^{(2)},Y^*\rangle=\langle X^{(2)},Y^{(0)*}\rangle\ . \tag{61}$$

The results obtained are

$$\delta_0=\left[\sum_{i,j=2}^{n}(\alpha^{-1})_{ij}(a_i-a_1)(a_j-a_1)\right]^{1/2}, \tag{62}$$

and $\langle\cdot,Y^{(0)*}\rangle$ is defined by

$$y^{(0)}(\phi)=\sum_{i=2}^{n}y_i^{(0)}\mathscr{N}(z_1;z_i,e^{i\phi})[\sigma(\phi)]^{-1}\ , \tag{63}$$

where the coefficients $y_i^{(0)}$ have the values

$$y_i^{(0)}=\frac{\sum_{j=2}^{n}(\alpha^{-1})_{ij}(a_j-a_1)}{\left[\sum_{i,j=2}^{n}(\alpha^{-1})_{ij}(a_i-a_1)(a_j-a_1)\right]^{1/2}}\ . \tag{64}$$

The required function $X^{(0)}(z)\equiv X^{(2)}(z)-M^{(0)}(z)$ may now be determined. As in the Dirichlet boundary condition case we observe that

$$\delta_0=||X^{(2)}-M^{(0)}||=\langle X^{(2)},Y^{(0)*}\rangle$$
$$=\langle X^{(0)},Y^{(0)*}\rangle\ , \tag{65}$$

where the last step follows from Eq. (16b). Now the equality between norm and functional

$$||X^{(0)}||=\langle X^{(0)},Y^{(0)*}\rangle=\delta_0\ , \tag{66}$$

together with the fact that the functional $\langle\cdot,Y^{(0)*}\rangle$ has unit norm, implies, by the Schwarz inequality, that

$$x^{(0)}{}_{,r}(\phi)=\delta_0 y^{(0)}(\phi)\ . \tag{67}$$

So we have the following result:

$$x^{(0)}{}_{,r}(\phi)=\sum_{i,j=2}^{n}(\alpha^{-1})_{ij}(a_j-a_1)\mathscr{N}(z_1;z_i,e^{i\phi})[\sigma(\phi)]^{-1}\ . \tag{68}$$

The corresponding complex function $X^{(0)}(z)$ is

$$X^{(0)}(z)=\sum_{i,j=2}^{n}(\alpha^{-1})_{ij}(a_j-a_1)\frac{1}{2\pi}\int_0^{2\pi}\left\{-2\ln\left[\frac{e^{i\phi}-z}{e^{i\phi}-z_1}\right]\mathscr{N}(z_1;z_ie^{i\phi})[\sigma(\phi)]^{-1}\right\}d\phi\ . \tag{69}$$

IV. INCORPORATION OF ERRORS IN THE DATA

In this section we extend the results already obtained so as to take account of the errors which will normally be associated with the data. The numerical data input will usually be imprecise whether it is experimental or theoretical in origin. The data set $\{a_i\}$ is now replaced by $\{a_i,\epsilon_i\}$. If an attempted fit to the data gives values $\tilde{a}_i$ at the data points, then the goodness of fit is measured by the usual χ^2 function, defined in this case by[19]

$$\chi^2=\sum_{i=1}^{n}\left[\frac{\tilde{a}_i-a_i}{\epsilon_i}\right]^2\ . \tag{70}$$

We have shown how to construct that function which, while constrained to assume values $\tilde{a}_i$ at the

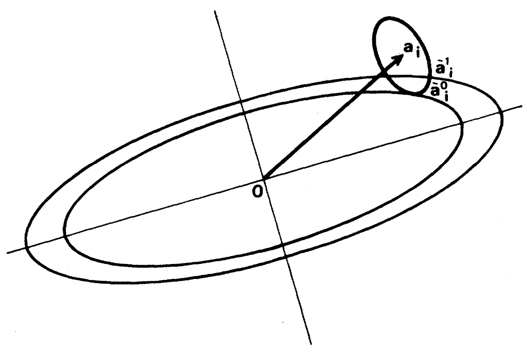

FIG. 4. The larger ellipsoids, with center at the origin, are the surfaces of constant $\delta_0{}^2=\sum(\alpha^{-1})_{ij}\tilde{a}_i\tilde{a}_j$ (these ellipsoids are dual to those of Fig. 3). The smaller ellipsoid, with center at a_i is the surface of constant $\chi^2=1$; clearly, $\tilde{a}_i^0$ gives the least possible value of δ_0 for $\chi^2=1$.

data points z_i, gives a minimum value δ_0 for the norm $||X||\equiv||X^{(1)}-M||$. In the case of the Dirichlet-type boundary condition [problem A, with the norm defined by Eq. (8)], δ_0 is given by Eq. (36) and for the Neumann problem [problem B, norm defined by Eq. (48)], δ_0 is given by Eq. (62).

We shall first describe the procedure for the Dirichlet problem—the extension to the Neumann case is relatively straightforward. Equation (36) may be written as

$$\delta_0{}^2=\sum_{i,j=1}^{n}(\alpha^{-1})_{ij}\tilde{a}_i\tilde{a}_j\ , \tag{71}$$

where the values $\tilde{a}_i$ are ascribed to the function $X(z)$ at the data points. By varying $\{\tilde{a}_i\}$ we change $\delta_0{}^2$ and also χ^2. One could at this stage simply apply a relative weighting to χ^2 and $\delta_0{}^2$, and carry out a simultaneous optimization. However we prefer to proceed as follows. We specify a value of χ^2, e.g., $\chi^2=1$, and then look for that vector $\tilde{a}_i^0$ which gives χ^2 the value 1 and subject to that leads to a minimum value of δ_0.

The geometrical description of this procedure is very simple. We see at once that, since both χ^2 and $\delta_0{}^2$ are positive definite, Eq. (70) for a constant value of χ^2 and Eq. (71) for a constant value of $\delta_0{}^2$ each represent an ellipsoid in the n-dimensional space $\{\tilde{a}_i\}$. This is illustrated in Fig. 4. The ellipsoids centered at the origin are the surfaces of constant δ_0. Through any point $\tilde{a}_i^1$ lying on the surface $\chi^2=1$ there passes an ellipsoid of constant δ_0. It is clear that the smallest such ellipsoid, and hence the one giving the smallest value of δ_0, is that which touches the ellipsoid $\chi^2=1$ externally. The vector $\tilde{a}_i^0$ to the point of contact is the value of the data vector $\tilde{a}_i$ which fits the data with $\chi^2=1$ and which, subject to this, gives the least value of δ_0.

To solve this problem analytically and to obtain the vector $\tilde{a}_i^{(0)}$ we use Lagrange multipliers. We write

$$F=\sum_{i,j=1}^{n}(\alpha^{-1})_{ij}\tilde{a}_i\tilde{a}_j+\lambda\left[\sum_{i=1}^{n}\left[\frac{\tilde{a}_i-a_i}{\epsilon_i}\right]^2-1\right]. \tag{72}$$

Then

$$\frac{\partial F}{\partial\tilde{a}_i}=2\sum_{j=1}^{n}(\alpha^{-1})_{ij}\tilde{a}_j+2\lambda\frac{(\tilde{a}_i-a_i)}{\epsilon_i{}^2}=0\ . \tag{73}$$

We may solve Eq. (73) for $\tilde{a}_i$ and the result, which is the vector $\tilde{a}_i^0$ giving the smallest value of δ_0 subject to $\chi^2=1$, is

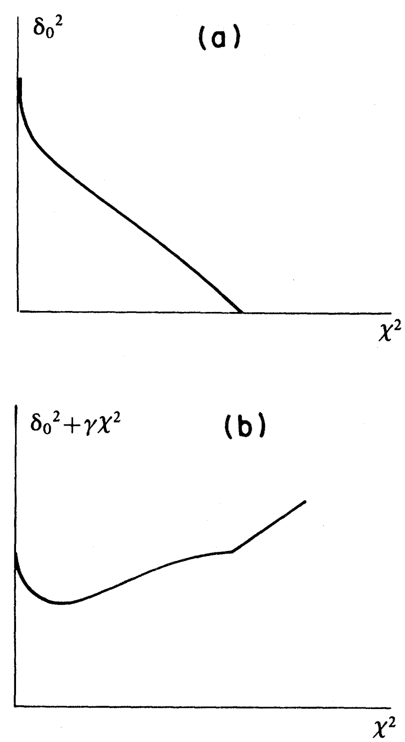

FIG. 5. (a) $\delta_0{}^2$ plotted as a function of χ^2. (b) $\delta_0{}^2+\gamma\chi^2$ plotted as a function of χ^2. Combining $\delta_0{}^2$ with χ^2 in this way is in the spirit of Cutkosky's modified χ^2 test.

$$\tilde{a}_i^0 = \sum_{j=1}^{n} (\mathscr{A}^{-1})_{ij} a_j \ , \tag{74}$$

where

$$\mathscr{A}_{ij} \equiv \delta_{ij} + \frac{\epsilon_i^{\,2}}{\lambda} (\alpha^{-1})_{ij} \ . \tag{75}$$

We substitute this into the equation $\chi^2 = 1$ to obtain the following equation for λ:

$$\sum_{i=1}^{n} \left[\frac{1}{\epsilon_i} \sum_{j=1}^{n} (\mathscr{A}^{-1} - I)_{ij} a_j \right]^2 = 1 \ , \tag{76}$$

where the value of $\mathscr{A}^{-1}$, which is a function of λ, is inserted from Eq. (75). Equation (76) has to be solved numerically for λ. Given λ, Eqs. (75) and (76) yield the vector $\tilde{a}_i^0$ which may then be substituted in place of $\tilde{a}_i$ in Eq. (71) to give δ_0. Of the real values of λ one clearly chooses that which gives the least value of δ_0.

Having thus determined the optimal $\chi^2 = 1$ vector $\tilde{a}_i^0$ we may use Eqs. (40) and (41), with $\tilde{a}_i^0$ in place of a_i, to obtain the corresponding optimal functions $x^{(0)}(\phi)$ and $X^{(0)}(z)$. It is clear that χ^2 can be given any value we choose, not necessarily 1, in the above calculation. So we can calculate $\delta_0^{\,2}$ as a function of χ^2 [Fig. 5(a)]. If we wish to give a relative weighting to $\delta_0^{\,2}$ and χ^2 we can define an overall optimum coming from that value of χ^2 for which $\delta_0^{\,2} + \gamma\chi^2$ has a minimum, where γ is a positive real constant determining the relative weighting and where $\delta_0^{\,2}$ is a function of χ^2 [Fig. 5(b)]. One would normally want to check that the minimum occurred for a sufficiently small value of χ^2 that the fit to the data could be regarded as acceptable. If this is not the case one would take the value of δ_0 obtained for the largest value of χ^2 which is acceptable.

The Neumann problem with errors

Predictably, the Neumann problem is somewhat more involved, the complication arising from the need to subtract away one of the data points. Equation (62) gives[20]

$$\delta_0^{\,2} = \sum_{i,j=2}^{n} (\alpha^{-1})_{ij} (\tilde{a}_i - \tilde{a}_1)(\tilde{a}_j - \tilde{a}_1) \ . \tag{77}$$

If we write the condition $\chi^2 = \text{const}$ in the form

$$\sum_{i=2}^{n} \left[\frac{(\tilde{a}_i - \tilde{a}_1) - (a_i - \tilde{a}_1)}{\epsilon_i} \right]^2 = \text{const} - \left[\frac{\tilde{a}_1 - a_1}{\epsilon_1} \right]^2 , \tag{78}$$

then for each value of $\tilde{a}_1$ which we may choose, the surfaces of constant $\delta_0^{\,2}$ and those of constant χ^2 are $(n-1)$-dimensional ellipsoids in the space of the vectors $\{\tilde{a}_i - \tilde{a}_1\}$ and hence we have an $(n-1)$-dimensional problem completely analogous to that already described for the Dirichlet case. So for each value of $\tilde{a}_1$ we find an optimal $(n-1)$-dimensional vector $\{\tilde{a}_2, \tilde{a}_3, \ldots, \tilde{a}_n\}$ giving, for that particular value of $\tilde{a}_1$ and for $\chi^2 = \text{const}$, the minimum value of $\delta_0^{\,2}$. The next step is to select that value of $\tilde{a}_1$ for which the $\delta_0^{\,2}$ thus obtained is an overall minimum. Combining this value of $\tilde{a}_1$ with the corresponding $(n-1)$-vector $\{\tilde{a}_2, \ldots, \tilde{a}_n\}$ yields the optimum n-vector $\{\tilde{a}_1^{\,0}, \ldots, \tilde{a}_n^{\,0}\}$.

Lagrange multipliers can be used as before. We write, in this case,

$$F = \sum_{i,j=2}^{n} (\alpha^{-1})_{ij} (\tilde{a}_i - \tilde{a}_1)(\tilde{a}_j - \tilde{a}_1) + \lambda \left[\sum_{i=1}^{n} \left[\frac{\tilde{a}_i - a_i}{\epsilon_i} \right]^2 - \text{const} \right] . \tag{79}$$

Then

$$\frac{\partial F}{\partial \tilde{a}_1} = -2 \sum_{i,j=2}^{n} (\alpha^{-1})_{ij} (\tilde{a}_j - \tilde{a}_1) \ 2\lambda \frac{(\tilde{a}_1 - a_1)}{\epsilon_1^{\,2}} = 0 \tag{80}$$

and, for $i \neq 1$

$$\frac{\partial F}{\partial \tilde{a}_i} = 2 \sum_{j=2}^{n} (\alpha^{-1})_{ij} (\tilde{a}_j - \tilde{a}_1) \ 2\lambda \frac{(\tilde{a}_i - a_i)}{\epsilon_i^{\,2}} = 0 \ . \tag{81}$$

The solution to Eq. (81) is

$$(\tilde{a}_i^0 - \tilde{a}_1^{\,0}) = \sum_{i=2}^{n} (\mathscr{A}^{-1})_{ij} (a_j - \tilde{a}_1^{\,0}) \ , \tag{82}$$

where the matrix $\mathscr{A}$ is defined as

$$\mathscr{A}_{ij} \equiv \delta_{ij} + \frac{\epsilon_i^{\,2}}{\lambda} (\alpha^{-1})_{ij} \quad (i,j = 2 \text{ to } n) \ . \tag{83}$$

Note that although this equation has the same form as Eq. (75), in this case the matrix $\mathscr{A}_{ij}$ is $(n-1)\times(n-1)$, with i,j running from 2 to n, and the matrix α_{ij} is the Neumann α matrix. The value $\tilde{a}_1{}^0$ is obtained by substituting Eq. (82) into Eq. (80), giving the result

$$\tilde{a}_1{}^0=\frac{\sum_{i,j=2}^{n}(\mathscr{B}^{-1})_{ij}a_j+\frac{\lambda}{\epsilon_1{}^2}a_1}{\sum_{i,j=2}^{n}(\mathscr{B}^{-1})_{ij}+\frac{\lambda}{\epsilon_1{}^2}}\ , \qquad (84)$$

where the $(n-1)\times(n-1)$ matrix $\mathscr{B}$ is defined as

$$\mathscr{B}_{ij}=\alpha_{ij}+\frac{\epsilon_i{}^2}{\lambda}\delta_{ij},\quad (i,j=2 \text{ to } n) \qquad (85)$$

or equivalently by

$$\mathscr{B}=\mathscr{A}\alpha\ . \qquad (86)$$

Equation (84) may now be used to substitute for $\tilde{a}_1{}^0$ in Eq. (82) which will then give $\tilde{a}_i^0$ in terms of $a_1,\ldots a_n$, and λ. Finally, these results for the vector $\{\tilde{a}_i^0\}$ are substituted into the equation $\chi^2=$const to obtain an equation which may be solved numerically for λ. One can then proceed precisely as in the Dirichlet case.[21]

V. THE DISCREPANCY METHOD

In the preceding sections we have shown how to construct a holomorphic function which fits the data, with errors, to within a required value of χ^2 and which is optimal in terms of the boundedness or smoothness condition, defined by Eqs. (1) or (2), which we choose to impose with a suitably chosen weight function σ, on the boundary of the holomorphy domain. In this section we show how we can use these results to test whether the data support a hypothesis of a certain type of structure on the boundary for the physical amplitude (Green's function, etc.), $A(z)$. The structure which one would expect and for which the data is to be tested, is expressed through the discontinuity function along the cuts and typically is related to the positions and widths of resonances. Resonances are described by nearby poles on the second Riemann sheet: we have to ask whether the data do or do not support the existence of such resonances and if they do, what are the preferred values of the resonance parameters (the pole positions and residues).

The procedure is as follows. We construct a trial function $T_\kappa(z)$ which will usually be given by a pair of conjugate second-sheet poles. We may take, for example,

$$T_\kappa(E)=\frac{\kappa_1+i\kappa_2}{\sqrt{E}-(\kappa_3+i\kappa_4)}-\frac{\kappa_1-i\kappa_2}{\sqrt{E}-(-\kappa_3+i\kappa_4)}$$
$$(\kappa_1,\ldots,\kappa_4 \text{ real}\ ,\ \ \kappa_4=\text{ negative})\ , \qquad (87)$$

where the square root is defined to have a nonnegative imaginary part in the complex plane cut along the positive real semiaxis. The data a_i for $A(z)$ are now replaced by new values obtained by subtracting from the values a_i the values of this trial function when evaluated at the appropriate data points z_i. The modified data $a_i^{(D)}$ obtained in this way are now submitted to the analyses of Secs. III and IV. A function $D_\kappa^0(z)$ is obtained via Eq. (41) or Eq. (69) which fits the adjusted data values $a_i^{(D)}$ and which is optimal in terms of the stabilizing condition imposed on the boundary. $D_\kappa(z)\equiv A(z)-T_\kappa(z)$ is called the discrepancy function.[13] If we combine the trial function $T_\kappa(z)$ with the optimal discrepancy $D_\kappa^0(z)$, then the function $A_\kappa^0(z)\equiv T_\kappa(z)+D_\kappa^0(z)$ represents the fit to the original data a_i which has the resonance structure of the trial function $T_\kappa(z)$ but otherwise possesses the minimum possible structure due to the requirement that the norm δ of $D_\kappa^0(z)$ should be minimal. The Neumann condition [Eq. (2)] is the more appropriate one for imposing this minimum structure requirement on $D_\kappa(z)$:

$$\inf_{(D_\kappa)}\delta(\kappa)\equiv\inf_{(D_\kappa)}\left\{\frac{1}{2\pi}\int_0^{2\pi}\left[\frac{d}{d\phi}\operatorname{Im}D_\kappa(e^{i\phi})\right]^2\sigma(\phi)d\phi\right\}^{1/2}=\delta_0(\kappa)\ . \qquad (6')$$

Expressing $\tilde{a}_i^{0(D)}$ in terms of the modified $a_i^{(D)}=a_i-T_\kappa(z_i)$ and the errors ϵ_i by means of Eqs. (82)–(86), we obtain for the optimal function $A_\kappa^0(z)$

$$A_\kappa^0(z)=T_\kappa(z)-\sum_{i,j=2}^{n}(\alpha^{-1})_{ij}(\tilde{a}_j^{0(D)}-\tilde{a}_1^{0(D)})\frac{1}{\pi}\int_0^{2\pi}\ln\left[\frac{e^{i\phi}-z}{e^{i\phi}-z_1}\right]\mathscr{N}(z_1;z_i,e^{i\phi})\sigma(\phi)^{-1}d\phi\ , \qquad (88a)$$

while the norm of the corresponding optimal discrepancy $D_\kappa^0(z)$ is

$$\delta_0(\kappa)=\left[\sum_{i,j=2}^{n}(\alpha^{-1})_{ij}(\tilde{a}_i^{0(D)}-\tilde{a}_1^{0(D)})(\tilde{a}_j^{0(D)}-\tilde{a}_1^{0(D)})\right]^{1/2} . \tag{88b}$$

Since $D_\kappa^0(z)$ is constructed from the modified data it depends on the choice of the trial function $T_\kappa(z)$, so that δ_0 will be a function of the parameters κ_1, κ_2, κ_3, and κ_4 of $T_\kappa(z)$. If for some specific $\kappa_i=\kappa_i^0$ the function $T_{\kappa^0}(z)$ adequately represents the structure which the data require, then $D_{\kappa^0}(z)$ can be relatively free of structure and the corresponding δ_0 should be small. However if the trial pole is substantially different from the pole sought by the original data, then the effect of subtracting $T_\kappa(z)$ will be to increase rather than decrease the structure demanded by the data: in this case $D_\kappa(z)$ will necessarily possess significant structure on the boundary—more structure in fact than would have been required by the original data—so that $\delta_0(\kappa)$ will be large, larger than the value obtained from the unmodified data.

So it is possible to test the hypothesis that a particular type of trial function, for specific values of its parameters which are to be determined, should provide a structure which meets the demands of the data, as follows: δ_0 is constructed as a function of the parameters κ of the trial function $T_\kappa(z)$. If within this parameter space $\delta_0(\kappa)$ has a significant minimum (at which it would be expected to have a value much less than the value of δ_0 for the unmodified data a_i) then we can say that for the values of the parameters κ_i^0 giving the minimum of δ_0, the structure described by $T_{\kappa^0}(z)$ is favored by the data. If, on the other hand, δ_0 does not have a significant minimum then the data cannot be said, on the basis of this analysis, to support the resonance hypothesis. This negative result should not automatically be regarded as a reason to reject the hypothesis, as it will often be the case that the data are not sufficiently accurate to support a resonance hypothesis. However, even in this case, the present method should provide the strongest possible statement which the data can support, in the sense that there is no other mathematical procedure which can give a better result. This is so since, if it happens that for a whole range of values of the κ's the functional $\delta_0(\kappa)$ does not differ significantly from its minimum value, then all the corresponding functions $T_\kappa(z)+D_\kappa^0(z)$ will be analytic, will have little structure outside the resonance region, and will all have acceptable values at $z=z_i$ (i.e., fitting the data a_i within the given value of χ^2) although the shape and position of the resonance $T_\kappa(z)$ may vary considerably.

A standard minimization program can be used to determine the minimum of δ_0 if one (or more) exists. In practice it will often be useful to evaluate δ_0 at various points over a lattice so that one can visualize its dependence on the parameters κ.

An important element in this discrepancy analysis is the choice of the weight σ. First, σ may serve as a sieve to reject functions of an unwanted type. For example, σ can be chosen so that an undesirable threshold behavior would give a large (divergent) contribution to δ_0 and so be rejected. Asymptotic behavior can be controlled in the same way. A second particularly useful facility provided by σ is to emphasize particular limited segments of the resonance region. One would often expect many resonances to contribute but, provided that there is a reasonable separation between these resonances, σ may be defined to single out a limited energy range within which a single resonance may be identified.

VI. CONCLUSION

In this paper we were concerned with the construction of analytic-continuation procedures which would be optimal with respect to some functional boundary conditions. As was pointed out in the Introduction it is essential to introduce stabilizing conditions since, due to the imprecise nature of the input data and the finite number of data points, the analytic continuation is not only no longer unique but (in the absence of such a stabilizing condition) is also widely unstable, in the sense that the outputs may differ by arbitrarily large amounts even if the imprecision of the initial data is small or tends to zero.

Analytic continuation can be an important adjunct to a theoretical calculation when it is necessary to extend the results obtained by some procedure—such as a perturbation expansion—into a region where this procedure is not directly applicable. It has been a long-standing challenge to attempt to use calculated results for unphysical values of the energies to predict properties, such as resonances, at values of physical interest. This is not simply a matter of analytic continuation from one region to another: given that the perturbation approximation is itself an analytic function, a straightforward continuation will just yield the result which would be obtained if the perturbation calculation were applied directly within the resonance region where it is known not to be valid.

In order to get the correct physical answer from the (false) perturbative input, we have first to relax the uniqueness of the continuation procedure. This is possible, since our method allows errors to be incorporated. It is in fact necessary to do this, since the truncated perturbation input, even where that expansion is valid, is only an approximation to the actual function. But the incorporation of errors (even though these are small) also relaxes the uniqueness of the analytic continuation and allows the real function to enter the set of possible continuations.

Instead of basing our method on the requirements that mathematical properties of a somewhat remote kind should be satisfied (such as the special shape of the holomorphy domain of the amplitude in the complex plane of the coupling constant, which is required for Borel summability—and which is difficult if not impossible to prove), we have designed our continuation procedure to use only properties which might be easily verified by experiment, such as the separability of resonance peaks, where this does occur. Moreover, the stabilization conditions [Eqs. (1) and (2)] have been chosen to suit these particular physical requirements. In defining these conditions we retain considerable flexibility of application. First, there is the freedom of choosing a discrepancy function D_κ constructed specifically to test a hypothesis such as the likelihood of any set of resonance parameters; δ_0 gives a direct measure of this likelihood. Second, the weight σ may be constructed to perform various functions: It may be chosen so as to sieve out some unwanted functions; for example, the truncated perturbation expansion might have an unsuitable threshold behavior and functions with this property could be rejected if σ causes the integral in Eqs. (1) or (2) to diverge for them. Also the function σ may be chosen to single out different segments of the resonance region: this means that one can ask questions separately about individual resonances.

Summary of the main results

In Sec. III we show how to find that particular analytic function which assumes specified values at the data points and which best satisfies the constraint conditions. Our method for approaching this problem uses some of the functional analysis techniques described in Ref. 12; the results may then be expressed in a remarkably elegant and convenient closed form [see Eq. (88)]. A key feature of our approach is to recognize that the input data is given on a finite set of n points. By working directly from this data set we obtain results for the optimum function which can be set in an n-dimensional Euclidean space, where they have a simple geometric description.

We then proceed in Sec. IV to extend these results to take account of the errors associated with the data values, so that at this stage we can solve the following problem: We can find that function which fits the data, with errors, to a required χ^2 value and which is optimal in terms of the smoothness or boundedness conditions which we choose to impose on the boundary of the domain of holomorphy. The next step is to adapt these results to enable us to test whether the data support a hypothesis of a certain type of structure on the boundary. Typically this structure is associated with a pole on a second Riemann sheet which can manifest itself as a physical resonance. To solve this problem we introduce in Sec. V a so-called discrepancy function in the following way: From the original data we subtract the values of a trial function, defined in terms of some parameters, which is selected to represent the structure (such as that due to a second-sheet pole) which we are looking for. We then ask if, for some values of these parameters, the discrepancy function, in which the structure represented by the trial function has been subtracted away from the original function, satisfies particularly well the smoothness constraints on the boundary. A positive result would indicate that the data favor the structure described by the trial function for the observed values of the parameters.

In pursuing the above program a key quantity (number) is the functional δ_0 computed for the discrepancy function $D_\kappa(z)$ [Eq. (6)], which is defined to be the least norm, as defined in Eqs. (3) or (4) which is compatible with the analyticity of the amplitude, the data, and with a set of resonance parameters κ which enter δ_0 via the discrepancy function $D_\kappa(z)$. The two alternative norms of Eqs. (1) and (2) represent two different types of stabilizing conditions. The first of these (the Dirichlet case) is based on the L^2 norm of the real part of the function, integrated over the cuts, the second (the Neumann case) is related to the smoothness of the imaginary part on the cuts.

The reader who is interested in the mathematical aspects of the problem is referred to Ref. 12 and in particular to the Appendices in that paper where the main results from functional analysis needed to carry out these calculations are derived.

---

[1]See, for example, M. Gell-Mann, M. L. Goldberger, and W. E. Thirring, Phys. Rev. 95, 1612 (1954); M. L. Goldberger, *ibid.* 99, 979 (1955); R. Karplus and M. A. Ruderman, *ibid.* 98, 771 (1955); M. L. Goldberger, H. Miyazawa, and R. Oehme, *ibid.* 99, 988 (1955).

[2]For example, see G. Kallen, *Handbuch der Physik,* edited by S. Flugge (Springer, Berlin, 1958).

[3]M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. B147, 385 (1979); B147, 448 (1979); B147, 519 (1979).

[4]For example, see E. C. Poggio, H. R. Quinn, and S. Weinberg, Phys. Rev. D 13, 1958 (1976).

[5]J. S. Bell and R. A. Bertlmann, Nucl. Phys. B187, 285 (1981); B177, 218 (1981).

[6]The "distance" between the regions over which the continuation is made does not play an essential role: marked instability can appear in continuation even to nearby points.

[7]R. E. Cutkosky and B. B. Deo, Phys. Rev. Lett. 22, 1272 (1968); Phys. Rev. 174, 1859 (1968); R. E. Cutkosky, J. Math. Phys. 14, 1231 (1973).

[8]S. Ciulli, Nuovo Cimento 61A, 787 (1969); 62A, 301 (1969); S. Ciulli and J. Fischer, Nucl. Phys. B24, 537 (1970); S. Ciulli, C. Pomponiu, and I. S. Stefanescu, Phys. Rep. 17, 133 (1975).

[9]See also K. Miller, SIAM J. Appl. Math. 18, 346 (1970); K. Miller and G. A. Viano, J. Math. Phys. 14, 1037 (1973).

[10]E. Pietarinen, Nuovo Cimento 12A, 522 (1972); Nucl. Phys. B49, 315 (1972).

[11]For a scattering amplitude in the $\cos\theta$ plane the condition $X(\bar{z})=\bar{X}(z)$ is not satisfied. However for $-1<z<1$, $\mathrm{Re}X(z)$ $\mathrm{IM}X(z)$ define two separate functions $X_1(z)$ and $X_2(z)$, which are analytic in the $z\equiv\cos\theta$ complex plane with certain cuts and possesses the above property. The form factor, on the other hand, will satisfy this condition directly.

[12]S. Ciulli and T. D. Spearman, J. Math. Phys. (to be published).

[13]This idea was introduced and applied by J. Hamilton, P. Menotti, T. D. Spearman, and W. S. Woolcock, Nuovo Cimento 20, 519 (1961). For a more recent application, see C. B. Lang and A. MasParareda, Phys. Rev. D 19, 956 (1979).

[14]In this connection see the papers of I. Caprini, S. Ciulli, C. Pomponiu, and I. S. Stefanescu, Phys. Rev. D 5, 1658 (1972), and I. S. Stefanescu, Nucl. Phys. B56, 287 (1973); an excellent mathematical discussion can be found in I. S. Stefanescu, J. Math. Phys. 21, 175, (1980), where the L^1-norm problem in particular, is treated.

[15]See the discussion of this in Ref. 12.

[16]G. Auberson, L. Epele, G. Mahoux, and F. R. A. Simao, Ann. Inst. H. Poincaré XXII, 317 (1975); Nucl. Phys. B94, 311 (1975).

[17]S. Ciulli and T. D. Spearman (unpublished); see also Report No. DIAS-STP-82-14 (unpublished).

[18]Alternatively one may obtain this by using the Schwarz-Villat formula for $zX'(z)$, availing of Eq. (43).

[19]The errors may well be correlated. In particular, we would expect this to be the case for inputs obtained by theoretical calculations. Correlated errors give a nondiagonal χ^2 of the form

$$\chi^2=\sum_{i,j}^{n}\epsilon_{ij}(\tilde{a}_i-a_i)(\tilde{a}_j-a_j)\ .$$

It is straightforward to repeat the calculations of Sec. IV with this form for χ^2 instead of Eq. (70). We are grateful to Dr. J. S. Bell for a discussion of this point.

[20]The matrix α_{ij} which appears here is, of course, quite distinct from the matrix α_{ij} which arises in the Dirichlet problems and which features in Eqs. (72)–(76). The Dirichlet α matrix is $n\times n$ ($i,j=1$ to n) whereas the Neumann α matrix is $(n-1)\times(n-1)$ ($i,j=2$ to n). The explicit form of the two matrices is given in Eqs. (25) and (59), respectively.

[21]An alternative method for treating the Neumann case is to construct the envelope of this ellipsoid in Eq. (78), which is itself an ellipsoid. The problem then is simply to find the value of $\delta_0{}^2$ for which the ellipsoid of Eq. (77) touches this envelope.