JOURNAL OF APPROXIMATION THEORY 27, 355–371 (1979)

Probabilistic Estimation of Poles or Zeros of Functions

F. M. LARKIN

Department of Computing and Information Science Queen's University, Kingston, Ontario, Canada

Communicated by Oved Shisha

Received July 24, 1978

A Maximum Likelihood method is presented for the purpose of estimating the positions of poles or zeros of certain classes of functions of a complex variable from information on ordinate values sampled at arbitrary abscissae. General and asymptotic properties of the method are discussed and some numerical examples given.

1. INTRODUCTION

One of the classic problems of Numerical Analysis is the following:

Given $\{z_j\,;\, j = 1, 2,..., n\}$, a set of distinct, complex abscissae, and corresponding ordinate values $\{g_j\,;\, j = 1, 2,..., n\}$ of some function $g(\cdot)$, estimate a value z_0, where $g(z_0) = 0$. Slightly more generally, the given information may also include derivative values of $g(\cdot)$ at the given abscissae. Since a great variety of functions, with arbitrarily placed zeros or none, could have given rise to the observed information, it is clear that severe restrictions must be imposed on $g(\cdot)$ if there is to be any hope at all of estimating z_0. Nevertheless, many ad hoc approaches, of which the Secant Rule and its limiting form Newton's Method are perhaps the best known, often prove very useful in practice — although it is well-known that this is not always the case!

The question then arises: For a given class of functions $\{g(\cdot)\}$ and a given class of observations, can one formulate a criterion for comparative assessment of different techniques for estimating z_0? If, for example, the Secant Rule is not optimal according to such a criterion, how can a better estimate be constructed?

Prompted by the success of the techniques of linear, optimal approximation (Sard [12] and Davis [3]) and their reformulation within a probabilistic context (Larkin [6]), we address the question of the previous paragraph by relating the class $\{g(\cdot)\}$ to a Hilbert space H of analytic functions endowed with a weak Gaussian distribution which induces a finite dimensional probability distribution on every finite set of bounded, linear functionals

0021-9045/79/120355–17$02.00/0

on H. In the following section we shall see that this permits the construction of a prior, multivariate distribution on the reciprocals of observed ordinate values $\{1/g(z_j); j = 1, 2,..., n\}$ at any chosen set of distinct abscissae $\{z_j; j = 1, 2,..., n\}$. The quantity z_0 will appear as an unknown parameter in this distribution, and can be estimated by any reasonable statistical technique when the aforementioned ordinate values are known.

In this paper we tentatively answer the question posed by regarding as "optimal" the Maximum Likelihood estimate of z_0 (Fisher [4], or any standard statistics text, such as Cramér [2]), since in some circumstances it is known to be asymptotically efficient. The choice of the Maximum Likelihood method as a standard of comparison may be open to debate since, unlike the linear case, the issue is complicated by the possibility of bias in the estimate. However, analytic results near the limit where the quantities $\{|z_0 - z_j|; j = 1, 2,..., n\}$ are all small, as well as practical numerical experience, seem encouraging enough to suggest that the approach deserves further study.

In order to put the probabilistic approach into perspective, perhaps the following discussion is in order:

Any rule for estimating the value of a bounded linear functional, which is based on a finite number of observations on other bounded linear functionals (e.g. ordinate evaluations), can, in general, produce a result with an arbitrarily large error. It is inescapable that extra information of a non-linear nature be available, or assumed, if even the relative error in the estimate is to be bounded. Similar, though less sweeping, conclusions apply when estimating the value of a non-linear functional.

According to the notions of Optimal Approximation, this necessary, non-linear information is provided by an assumed bound on a norm or semi-norm of the function being observed (Davis [3]; Sard [12]). In the case of a norm bound, this is roughly equivalent to the assumption that, with probability 1, the function lies within a hypersphere of known radius centred on the origin in the function space in question. Thus, the effect of the non-linear information is to provide an a priori localization of the observed function.

By contrast, according to the probabilistic approach used here and elsewhere (Larkin [6], [7], [8], [9]), the assumption of a weak Gaussian distribution, with an unspecified variance, imposes an a priori localization of a different kind—one that is less severe in that the norm (or semi-norm) of the observed function need not be unequivocally limited, but more severe in that the shape of the distribution is more or less fixed. Thus, the probabilistic approach is similar in intent, but different in form, to the approach of Optimal Approximation. Advantages of the probabilistic approach are that, as well as agreeing identically with Optimal Approximation in the rules it provides for the linear estimation problem, it also provides practically useful, probabilistic error estimates in the linear case and a fresh approach to non-linear problems, such as the one considered in this paper.

From one point of view, the method of Maximum Likelihood (or any similar statistical principle) is not an estimation rule in the sense that a numerical analyst understands, for example, a quadrature rule or a root-finding rule, but rather is an intuitively attractive, and widely applicable, principle by means of which such rules can be constructed. Thus, the method of Maximum Likelihood permits a statistician to construct a rule for estimating some unknown quantity, in terms of other, measurable quantities; he would then expect to apply this estimation rule in many different circumstances, in much the same way as a numerical analyst might use Simpson's Rule. Furthermore, just as the numerical analyst is interested in, say, the truncation error of a quadrature rule and its convergence properties as the number of observations tends to ∞, so the statistician must assess similar properties of his estimation rule, albeit dealing in terms of stochastic variables.

If desired, an estimation rule derived using the method of Maximum Likelihood can subsequently be divorced from its origins and treated on an entirely ad hoc basis. Indeed, this is the approach taken in this paper with the root-finding estimator derived in section 3, below. To someone unwilling to presuppose the weak Gaussian distribution necessary for the probabilistic treatment one can then offer the rule to be assessed on the basis of whatever information he is willing to assume.

However, the main point of this paper is not the production of yet another competitor to Newton's method, and a rather complicated one at that, but rather to support the thesis that the probabilistic alternative to Optimal Approximation provides a unifying conceptual framework within which such traditionally disparate problems as numerical quadrature and root-finding can be treated on a similar basis. Indeed, investigations now in progress, arising in connection with Remark 5 in Section 5, below, suggest that s may be just as useful as $\hat{z}_0$ as an estimator of z_0, and perhaps more convenient.

Numerical analysts and statisticians are both in the business of estimating parameter values from incomplete information. The two disciplines have separately developed their own approaches to formalizing strangely similar problems and their own solution techniques; the author believes they have much to offer each other.

2. Problem Formulation

Let H denote a Hilbert space of functions analytic within some simple domain D in the complex plane. Since we shall be concerned primarily with information expressed in the form of ordinate values, we shall suppose that the linear functional of ordinate evaluation at $z \in D$ is bounded for all such points. This means that H must possess a reproducing kernel function (Aronszajn [1]), denoted by $K(\cdot, \bar{\cdot})$, with the defining properties

$$K(\cdot, \bar{z}) \in H; \ \forall z \in D$$
$$(h, K(\cdot, \bar{z})) = h(z); \ \forall h \in H \text{ and } \forall z \in D.$$

Here the bar signifies "complex conjugate".

Now consider the class F of functions $f(\cdot)$ of the form

$$f(z) = \frac{h(z)}{z - z_0}; \ \forall z \in D,$$

for any $h \in H$ and for some fixed, but unknown $z_0 \in \mathbb{C}$. Given some fixed $f \in F$, we are permitted to sample its ordinate values (and possibly also other bounded, linear functionals) at abscissae $\{z_j \in D; j = 1, 2, 3, \ldots\}$ and wish to devise a reasonable procedure for estimating the value of z_0 from the observations $\{f(z_j); j = 1, 2, 3, \ldots\}$, in spite of the fact that the information available will generally be insufficient to determine f completely. Of course, the problem of estimating the pole z_0 of $f(\cdot)$ is identical to that of estimating the zero of the function $g(\cdot)$ mentioned in the introduction if we identify

$$g(z) = \frac{1}{f(z)}; \qquad \forall z \in D.$$

Following the approach developed in previous papers (Larkin [6]–[9]), we endow H with a canonical, weak Gaussian distribution (Gross [5]) thus inducing a multivariate, normal distribution jointly on any finite number of bounded, linear functional values. For example (Larkin [6]), if the n-th order vector $\mathbf{h}$ is defined by

$$h_j = h(z_j); j = 1, 2, \ldots, n$$

and the n-th order Gram matrix K is defined by

$$K_{jk} = K(z_j, \bar{z}_k); j, k = 1, 2, \ldots, n,$$

the probability density function of $\mathbf{h} \in \mathbb{C}^n$ is given by

$$p(\mathbf{h}) = \left(\frac{\lambda}{\pi}\right)^n | K |^{-1} \exp(-\lambda \mathbf{h}' K^{-1} \mathbf{h}), \tag{1}$$

where λ is the (positive, real) dispersion parameter of the canonical, weak distribution on H. Here the prime signifies "Hermitean transpose."

Since we cannot directly observe the values $\{h_j\,; j = 1, 2,..., n\}$ we introduce the vector $\mathbf{f}$ of observable quantities defined by

$$f_j = \frac{h_j}{z_j - z_0}\,; \qquad j = 1, 2,..., n.$$

Thus, writing Z for the diagonal matrix whose j-th element is z_j, we have

$$\mathbf{h} = (Z - z_0 I)\mathbf{f}$$

so the Jacobian of the transformation, bearing in mind that $2n$ real variables are involved, is given by

$$\frac{\partial \mathbf{h}}{\partial \mathbf{f}} = |\det(Z - z_0 I)|^2 = \prod_{j=1}^{n} |z_j - z_0|^2$$

Hence, the probability density function of $\mathbf{f} \in \mathbb{C}^n$, given λ and z_0, may be written as

$$p(\mathbf{f} \mid \lambda, z_0) = \left(\frac{\lambda}{\pi}\right)^n \cdot |K|^{-1} \cdot \prod_{j=1}^{n} |z_j - z_0|^2$$
$$\cdot \exp(-\lambda \mathbf{f}'(Z' - \bar{z}_0 I)\, K^{-1} \cdot (Z - z_0 I)\,\mathbf{f}).$$

3. Formal Solution

The method of Maximum Likelihood consists in equating to zero the logarithmic derivatives of $p(\mathbf{f} \mid \lambda, z_0)$ with respect to λ and z_0 (actually, in the complex case, the real and imaginary parts of z_0), to obtain the following equations for $\hat{\lambda}$ and $\hat{z}_0$, the estimates of λ and z_0:

$$\hat{\lambda} = \frac{n}{\mathbf{f}'(Z' - \hat{\bar{z}}_0 I)\, K^{-1}(Z - \hat{z}_0 I)\,\mathbf{f}} \tag{3}$$

and

$$\frac{1}{n}\sum_{j=1}^{n} \frac{1}{z_j - \hat{z}_0} = \frac{\mathbf{f}'(Z' - \hat{\bar{z}}_0 I)\, K^{-1}\mathbf{f}}{\mathbf{f}'(Z' - \hat{\bar{z}}_0 I)\, K^{-1}(Z - \hat{z}_0 I)\,\mathbf{f}}\,. \tag{4}$$

Thus, the problem of finding a zero of the rather general function $g(\cdot)$ has been replaced by that of finding a root of the very special equation (4), constructed from observations on $g(\cdot)$. This, of course, is also typical of the more traditional methods.

A certain simplification can be achieved, presuming $t^2 \neq |s|^2$, if we now introduce quantities

$$s = \frac{\mathbf{f}'K^{-1}Z\mathbf{f}}{\mathbf{f}'K^{-1}\mathbf{f}}, \tag{5}$$

$$t^2 = \frac{\mathbf{f}'Z'K^{-1}Z\mathbf{f}}{\mathbf{f}'K^{-1}\mathbf{f}}, \tag{6}$$

$$w_0 = \frac{\hat{z}_0 - s}{(t^2 - |s|^2)^{1/2}}, \tag{7}$$

and

$$w_j = \frac{z_j - s}{(t^2 - |s|^2)^{1/2}}, \qquad j = 1, 2, \ldots, n, \tag{8}$$

permitting equation (4) to be expressed in the form

$$\frac{\bar{w}_0}{1 + |w_0|^2} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{w_0 - w_j}. \tag{9}$$

This is always possible if $\mathbf{f} \neq \mathbf{0}$ and the abscissae $\{z_j\,;\, j = 1, 2, \ldots, n\}$ are distinct since, by the Schwarz inequality, $t^2 = |s|^2$ only if $Z\mathbf{f}$ is proportional to $\mathbf{f}$, which is absurd. In the following, we shall invariably make these assumptions.

Although it is not immediately obvious that (4) has a solution, since both $\hat{z}_0$ and $\bar{\hat{z}}_0$ are involved, conditions sufficient for the existence of at least one root are given below. Furthermore, one of these roots will correspond to the global maximum of the function of z

$$p(z \mid \mathbf{f}) \overset{\text{def}}{=} \text{constant} \times \frac{\prod_{j=1}^{n} |z - z_j|^2}{[\mathbf{f}'(Z' - \bar{z}I)\, K^{-1}(Z - zI)\,\mathbf{f}]^n}. \tag{10}$$

In practice, one might perform the following sequence of operations in order to estimate z_0:

(i) Sample $f(\cdot)$ at chosen points $\{z_j\,;\, j = 1, 2, \ldots, n\}$

(ii) Evaluate s, t^2 and $\{w_j\,;\, j = 1, 2, \ldots, n\}$ from (5), (6) and (8).

(iii) Solve (9) for w_0, by an iterative method such as that described in section 5 below.

(iv) Invert (7) to compute

$$\hat{z}_0 = s + w_0 \cdot (t^2 - |s|^2)^{1/2}$$

(v) Terminate if $\hat{z}_0$ is regarded as being sufficiently accurate.

(vi) Otherwise, set $z_{n+1} = \hat{z}_0$, compute $f(z_{n+1})$, increment n by 1 and repeat steps (ii), (iii), (iv), (v) and (vi).

4. General Properties of the Method

The numerical values of the $\{w_j\ ;\ j = 1, 2, \ldots, n\}$ provide information about the existence and location of $\hat{z}_0$, as follows.

Making the substitutions defined in (5), (6), (7), and (8) into $p(\hat{z}_0 \mid \mathfrak{f})$, we see that (9) is the equation for a stationary point of the function of w

$$S(w) \overset{\text{def}}{=} \frac{\prod_{j=1}^{n} \mid w - w_j \mid^2}{(1 + \mid w \mid^2)^n}\,; \qquad \forall w \in \mathbb{C}. \tag{11}$$

THEOREM 1. *$S(\cdot)$ has an upper extremum in the finite part of the complex plane if there exists $\tilde{w}$ such that*

$$S(\tilde{w}) > 1 \tag{12}$$

Proof. Observe that, as a function of the real and imaginary parts of w, $S(w)$ is not constant and is bounded above, by M say. Also

$$\lim_{|w| \to \infty} S(w) = 1 \tag{13}$$

so, if such a $\tilde{w}$ exists, $M > 1$ and $\exists R$ such that $S(v) < (M + 1)/2$, $v\colon \mid v \mid > R$. However, the bounded, continuous function $S(\cdot)$ must attain an upper extremum on the compact set $\{w_r \mid w \mid = R\}$, so the proposition is proved.

COROLLARY 1. *Since $S(w)$ is continuously differentiable with respect to the real and imaginary parts of w, any upper extremum of S will be a maximum. Thus, condition* (12) *is sufficient for the existence of a root of equation* (9).

COROLLARY 2. *By considering the special case $w = 0$ we see that the condition*

$$\prod_{j=1}^{n} \mid w_j \mid > 1$$

is sufficient for (9) *to have a root at a maximum of $S(\cdot)$.*

THEOREM 2. *$S(\cdot)$ has a maximum if*

$$\sum_{j=1}^{n} w_j \neq 0. \tag{14}$$

Proof. Observe that in this case

$$S(w) \sim 1 - 2\,\mathrm{Re}\left\{\frac{1}{w} \sum_{j=1}^{n} w_j\right\}$$

for large $|w|$ and, by suitable choice of $\arg(w)$, this can be made to exceed 1. The required result then follows by appealing to Theorem 1.

Remark 1. It is not hard to construct examples, necessarily violating condition (14), in which

$$S(w) < 1, \qquad \forall \text{ finite } w$$

e.g. $n = 4$, $w_1 = w_2 = 1/2$, $w_3 = w_4 = -1/2$; in this case equation (9) has no finite solution.

Remark 2. If $S(\cdot)$ has a maximum, so does $p(\cdot, \mathbf{f})$.

In practical computation, of particular interest is the situation where the $\{w_j ; j = 1, 2,..., n\}$ are all large, $t^2 - |s|^2$ being small, resulting in a small value of $|w_0|$, a small value of $|\sigma|$ and a correspondingly accurate estimate $\hat{z}_0$ (see Theorem 5 below). The next result provides an easily checked condition for this.

THEOREM 3. *If a is real and positive, and*

$$|w_j| > \frac{1 + (1 + a^2)^{1/2}}{a}; \qquad j = 1, 2,..., n$$

then $S(\cdot)$ has a maximum within a complex disk of radius a, centred at the origin.

Proof. Consider that, if $|w| = a$,

$$\frac{|w_j - w|^2}{|w_j|^2(1 + |w|^2)} \leqslant \frac{(1 + a/|w_j|)^2}{1 + a^2} < 1; \qquad j = 1, 2,..., n.$$

Hence

$$\sup_{|w|=a} \prod_{j=1}^{n} \left\{\frac{|w_j - w|^2}{1 + |w|^2}\right\} < \prod_{j=1}^{n} |w_j|^2$$

i.e.

$$\sup_{|w|=a} S(w) < S(0)$$

which, by an argument similar to that used in Theorem 1, implies the required result.

5. Asymptotic Properties

We now consider the situation where all of the abscissae $\{z_j ; j = 1, 2,..., n\}$ are very close to z_0, and derive appropriate error estimates.

Define

$$\sigma = s - z_0 = \frac{\mathbf{f}'K^{-1}(Z - z_0I)\,\mathbf{f}}{\mathbf{f}'K^{-1}\mathbf{f}} = \frac{\mathbf{f}'K^{-1}\mathbf{h}}{\mathbf{f}'K^{-1}\mathbf{f}} \tag{14}$$

and

$$\tau^2 = \frac{\mathbf{f}'(Z' - \bar{z}_0I)\,K^{-1}(Z - z_0I)\,\mathbf{f}}{\mathbf{f}'K^{-1}\mathbf{f}} = \frac{\mathbf{h}'K^{-1}\mathbf{h}}{\mathbf{f}'K^{-1}\mathbf{f}} \tag{15}$$

Of course, $\mathbf{h}'K^{-1}\mathbf{h}$ is the squared norm of $\hat{h}$, say, that member of H with smallest norm subject to the constraints

$$\hat{h}(z_j) = h_j\,; \qquad j = 1, 2,..., n.$$

LEMMA 1. *The quantity $t^2 - |\, s\,|^2$ is invariant under a uniform translation of the $\{z_j\,; j = 1, 2,..., n\}$ so, in particular,*

$$0 \neq t^2 - |\, s\,|^2 = \tau^2 - |\,\sigma\,|^2 = \frac{\mathbf{f}'(Z' - \bar{s}I)\,K^{-1}(Z - sI)\,\mathbf{f}}{\mathbf{f}'K^{-1}\mathbf{f}}$$

Furthermore

$$|\,\sigma\,| \leqslant \tau.$$

The proof of this is merely a matter of algebraic manipulation, and use of the Schwarz inequality.

Remark 3. A promising aspect of this method is that as n increases the $\{|\, w_j\,|; j = 1, 2,..., n\}$ will often become very large since

$$(\tau^2 - |\,\sigma\,|^2)^{1/2} < \tau = \frac{\|\,\hat{h}\,\|}{\|\hat{f}\|} \leqslant \frac{\|\,h\,\|}{\|\hat{f}\|}$$

where $\hat{f}$ is that member of H with smallest norm such that

$$\hat{f}(z_j) = f_j\,; j = 1, 2,..., n,$$

and $\|\hat{f}\|$ should increase rapidly with n since $f(\cdot)$ is not a member of H. Thus, from Theorem 3, $|\, w_0\,|$ should decrease rapidly with n even if the abscissae $\{z_j\,; j = 1, 2,..., n\}$ are poor estimates of z_0.

Introducing the notation $\gamma[u_1u_2 \cdots u_r]$ to denote the $(r - 1)$-th. divided difference of a function $\gamma(\cdot)$, based on the abscissae $\{u_j\,; j = 1, 2,..., r\}$, we shall also need the following result.

LEMMA 2.

$$f[z_1z_2 \cdots z_r] = h[z_0z_1z_2 \cdots z_r] - h(z_0) \cdot \prod_{j=1}^{r} (z_0 - z_j)^{-1}.$$

Proof. It is well known (eg. Milne-Thomson [11]) that

$$\mathbf{f}[z_1 z_2 \cdots z_r] = \sum_{j=1}^{r} f_j \prod_{k=1}^{r}{}' (z_j - z_k)^{-1},$$

where the prime indicates omission of the j-th. (zero) factor. But

$$f_j = \frac{h_j}{z_j - z_0}, \qquad j = 1, 2, \ldots, n$$

so, writing h_0 for $h(z_0)$,

$$\begin{aligned} f[z_1 z_2 \cdots z_r] &= \sum_{j=1}^{r} h_j \prod_{k=0}^{r}{}' (z_j - z_k)^{-1} \\ &= \sum_{j=0}^{r} h_j \prod_{k=0}^{r} (z_j - z_k)^{-1} - h_0 \prod_{k=1}^{r} (z_0 - z_k)^{-1}, \end{aligned}$$

as required.

We now introduce some quantities required for the subsequent statement of the asymptotic results.

Recall that the operation of constructing the divided differences $\{f[z_1 z_2 \ldots z_r];\ r = 1, 2, \ldots, n\}$ from the ordinate values $\{f_j\ ;\ j = 1, 2, \ldots, n\}$ is linear, and also nonsingular if the abscissae $\{z_j\ ; j = 1, 2, \ldots, n\}$ are distinct. Denote the matrix of this operation by D and define vectors $\boldsymbol{\delta}\mathbf{f}$ and $\boldsymbol{\delta}\mathbf{h}$, by

$$(\delta f)_j = f[z_1 z_2 \ldots z_j],$$

and

$$(\delta h)_j = h[z_1 z_2 \ldots z_j], j = 1, 2, \ldots, n.$$

Writing $\mathbf{p}$ for the vector whose r-th element is given by

$$p_r = \prod_{j=1}^{r} (z_0 - z_j)^{-1}; \qquad r = 1, 2, \ldots, n \tag{16}$$

we see from Lemma 2 that

$$\boldsymbol{\delta}\mathbf{f} = \boldsymbol{\delta}\mathbf{h} - h_0\mathbf{p}.$$

Hence

$$\mathbf{f}'K^{-1}\mathbf{h} = (\boldsymbol{\delta}\mathbf{h}' - \bar{h}_0\mathbf{p}')(DKD')^{-1}\boldsymbol{\delta}\mathbf{h} \tag{17}$$

and

$$\mathbf{f}'K^{-1}\mathbf{f} = (\boldsymbol{\delta}\mathbf{h}' - \bar{h}_0\mathbf{p}')(DKD')^{-1}(\boldsymbol{\delta}\mathbf{h} - h_0\mathbf{p}).$$

Let A denote the matrix defined by

$$(A^{-1})_{jk} = \frac{1}{(j-1)!\,(k-1)!} \cdot \frac{\partial^{j+k-2} K(z_0, \bar{z}_0)}{\partial z_0^{j-1}\, \partial \bar{z}_0^{k-1}}; \qquad j, k = 1, 2, \ldots, n,$$

where $\partial/\partial z_0$ denotes formal differentiation with respect to z_0, keeping $\bar{z}_0$ constant, and vice versa.

Let $\mathbf{dh}$ denote the n-th order vector defined by

$$dh_j = \frac{1}{(j-1)!} \cdot \frac{d^{j-1} h(z_0)}{dz_0^{j-1}}; \qquad j = 1, 2, \ldots, n,$$

write $\mathbf{a} = A\,\mathbf{dh}$ and define

$$b_n = -\frac{a_n}{A_{nn} h_0}; \qquad d_n = \frac{\| \hat{h} \|}{A_{nn}^{1/2} \mid h_0 \mid}$$

and

$$c_n = \frac{[\| \hat{h} \|^2 - \mid a_n \mid^2 / A_{nn}]^{1/2}}{A_{nn}^{1/2} \mid h_0 \mid}.$$

Observe also that, in the limit as all of the $\{z_j ; j = 1, 2, \ldots, n\}$ approach z_0,

$$\boldsymbol{\delta}\mathbf{h} \to \mathbf{dh} \tag{19}$$

and

$$DKD' \to A^{-1}, \tag{20}$$

by a standard property of divided differences.

Assuming now, and for the rest of this section, that

$$\mid z_j - z_0 \mid \ll 1, \qquad j = 1, 2, \ldots, n$$

we obtain the following results.

LEMMA 3.

$$\sigma \sim b_n \prod_{j=1}^{n} (z_j - z_0)$$

$$\tau \sim d_n \prod_{j=1}^{n} \mid z_j - z_0 \mid$$

$$(\tau^2 - \mid \sigma \mid^2)^{1/2} \sim c_n \prod_{j=1}^{n} \mid z_j - z_0 \mid$$

and

$$w_k \sim \frac{(z_k - z_0)}{c_n} \cdot \prod_{j=1}^{n} | z_j - z_0 |^{-1}; \qquad k = 1, 2, \ldots, n. \tag{21}$$

Proof. Considering the orders of magnitude of the quantities involved in (17) and (18) and recalling (19) and (20), we see that

$$\mathbf{f}'K^{-1}\mathbf{h} \sim -\bar{h}_0 \cdot \mathbf{p}'A\, \mathbf{dh} \sim -a_n\bar{h}_0 \cdot \prod_{j=1}^{n} (\bar{z}_j - \bar{z}_0)^{-1}$$

and

$$\mathbf{f}'K^{-1}\mathbf{f} \sim | h_0 |^2 \cdot \mathbf{p}'A\mathbf{p} \sim A_{nn} | h_0 |^2 \cdot \prod_{j=1}^{n} | z_j - z_0 |^{-2}.$$

Then, from Lemma 1 and the definitions of σ, τ, and $\{w_k ; k = 1, 2, \ldots, n\}$ the required results follow by simple substitution.

We are now in a position to prove the main results.

THEOREM 4. *There exists a w_0 satisfying equation* (9) *and maximizing $S(\cdot)$ such that*

$$\bar{w}_0 \sim -c_n \left[\frac{1}{n} \sum_{j=1}^{n} \frac{1}{z_j - z_0}\right] \cdot \prod_{j=1}^{n} | z_j - z_0 | .$$

Proof. From Lemma 3 we see that all of the quantities $\{| w_k |; k = 1, 2, \ldots, n\}$ can be made arbitrarily large by choosing the abscissae $\{z_j ; j = 1, 2, \ldots, n\}$ sufficiently close to z_0. Let

$$b = \min_k \left[c_n^{-1} \cdot \prod_{j \neq k} | z_j - z_0 |^{-1}\right] \stackrel{\text{def}}{=} \frac{1 + (1 + a^2)^{1/2}}{a}, \qquad \text{say.}$$

Then $| w_k | > b; k = 1, 2, \ldots, n$ and, from Theorem 3, there exists w_0 satisfying (9) such that

$$| w_0 | < a = \frac{2b}{b^2 - 1} \ll 1$$

Hence we can approximate (9) by

$$\bar{w}_0 \simeq -\frac{1}{n} \sum_{j=1}^{n} \frac{1}{w_j},$$

leading immediately to the required result, using (21).

THEOREM 5.

$$\hat{z}_0 - z_0 \sim b_n \cdot \prod_{j=1}^{n} (z_j - z_0) \sim \sigma \tag{22}$$

Proof. From (7), Lemmas 1 and 3 and Theorem 4 we see that

$$\hat{z}_0 - z_0 \sim -\frac{a_n}{A_{nn}h_0} \cdot \prod_{j=1}^{n} (z_j - z_0) - c_n{}^2 \left[\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\bar{z}_j - \bar{z}_0}\right] \cdot \prod_{j=1}^{n} | z_j - z_0 |^2.$$

However, the second of the two expressions on the right hand side of this relation is insignificant compared with the first, so the proposition is proved.

COROLLARY 3.

$$| \hat{z}_0 - z_0 | \leqslant d_n \cdot \prod_{j=1}^{n} | z_j - z_0 | .$$

Proof. If $\mathbf{u}_n$ is an n-th order vector with zero elements except for a 1 in the n-th position, we have

$$a_n = \mathbf{u}_n' A \, \mathbf{d}\mathbf{h}$$

so, by the Schwarz inequality,

$$| a_n | \leqslant (\mathbf{u}_n' A \mathbf{u}_n)^{1/2} \cdot (\mathbf{dh}' \, A \, \mathbf{dh})^{1/2} \sim A_{nn}^{1/2} \, \| \hat{h} \| ,$$

and

$$\frac{| b_n |}{d_n} \leqslant \frac{| a_n |}{A_{nn}^{1/2} \, \| \hat{h} \|} \sim 1.$$

Hence, taking absolute values of both sides of (22), the required result follows.

Remark 4. A_{nn} is the squared norm of that element $\hat{e}$ of H with smallest norm satisfying the constraints.

$$\left(\frac{d^j \hat{e}}{dz^j}\right)_{z_0} = 0; \qquad j = 0, 1, 2,..., n - 2$$

$$\left(\frac{d^{n-1} \hat{e}}{dz^{n-1}}\right)_{z_0} = 1.$$

Remark 5. Since

$$s - z_0 = \sigma \sim b_n \prod_{j=1}^{n} (z_j - z_0),$$

from Lemma 3, it may be advantageous to use s as an estimate of z_0 when the quantities $\{| z_j - z_0 |; j = 1, 2,..., n\}$ are all small. This avoids having to solve the implicit equation (4), at the expense of a small deviation from the Maximum Likelihood estimate. The effect of this is illustrated by comparing σ with $\hat{z}_0$ in Table 1, below. Note that care should be exercised when computing s since the matrix K may be ill-conditioned.

TABLE 1

Estimates of the Pole of e^z/z; $K(z, \bar{w}) = 1/(1 - z\bar{w})$; $n = 3$

z_1	0.5	0.5	0.05	0.00546
z_2	−0.4	0.05	0.005	−0.00987
z_3	−0.5	0.005	0.0005	−0.00123
σ	$7.4000.10^{-2}$	$-1.0653.10^{-5}$	$5.6574.10^{-8}$	$3.3450.10^{-8}$
$(\tau^2 - \lvert\sigma\rvert^2)^{1/2}$	$1.2845.10^{-1}$	$1.8787.10^{-4}$	$1.7917.10^{-7}$	$9.3548.10^{-8}$
w_1	3.3166	2661.5	$2.7907.10^5$	$5.8365.10^4$
w_2	−3.6903	266.20	$2.7907.10^4$	$-1.0551.10^5$
w_3	−4.4688	26.671	$2.7904.10^3$	$-1.3149.10^4$
w_0	$6.0324.10^{-2}$	$-1.3872.10^{-2}$	$-1.3260.10^{-4}$	$2.2799.10^{-5}$
$\hat{z}_0$	$8.1748.10^{-2}$	$-1.3259.10^{-5}$	$5.6550.10^{-8}$	$3.3453.10^{-8}$

6. A Numerical Example

Let us choose H to be the Szegö-Hilbert space of functions analytic within the unit disc and square integrable around its perimeter, with inner product

$$(h_1, h_2) = \frac{1}{2\pi} \oint_{|z|=1} h_1(z) \cdot \overline{h_2(z)} \cdot |dz|; \qquad \forall h_1, h_2 \in H$$

and natural norm. The functions of z $\{z^k;\ k = 0, 1, 2, \ldots\}$ form a complete ortho-normal basis for H and the reproducing kernel (e.g. Meschkowski [10]) is given by

$$K(z, \bar{w}) = (1 - z\bar{w})^{-1}$$

The function e^z is certainly a member of H, so it is legitimate to choose the example

$$f(z) = \frac{e^z}{z}; \qquad \forall z \in \mathbb{C}$$

and apply the foregoing method for the purpose of estimating the pole $z_0 = 0$.

Although the constants h_0, A_{nn}, a_n, c_n and d_n will not normally be required, the example has been chosen so that they are easy to evaluate, and so serve to confirm the asymptotic estimates derived earlier. Specifically, it may be verified that A is the unit matrix,

$$h_0 = 1,$$

$$dh_j = \frac{1}{(j-1)!}\,; \qquad j = 1, 2, \ldots, n$$

and

$$\hat{h}(z) \sim \sum_{j=0}^{n-1} \frac{z^j}{j!}\,,$$

so that

$$\| \hat{h} \|^2 \sim \sum_{j=1}^{n} [(j-1)!]^{-2} \quad .$$

Hence, the constants appearing in the asymptotic forms of σ, τ and $(\tau^2 - |\sigma|^2)^{1/2}$ are given by

$$b_n \sim -\frac{1}{(n-1)!}$$

$$d_n \sim \left\{ \sum_{j=1}^{n} [(j-1)!]^{-2} \right\}^{1/2}$$

$$c_n \sim \left\{ \sum_{j=1}^{n-1} [(j-1)!]^{-2} \right\}^{1/2}$$

and, specializing to the case $n = 3$, these become -0.5, 1.5 and $\sqrt{2} \simeq 1.414$, respectively. Table 1 shows the values of various significant quantities, resulting from different choices of z_1, z_2, and z_3. In each case equation (9) was solved by the iteration

$$v_0 = 0$$

$$v_{k+1} = \frac{1 + |v_k|^2}{n} \cdot \sum_{j=1}^{n} \frac{1}{\bar{v}_k - \bar{w}_j}\,; \qquad k = 0, 1, 2, \ldots,$$

terminating at the first m for which

$$| v_m - w_0 | < 10^{-8}.$$

Table 2 illustrates a similar computation for the function

$$f(z) = \frac{1}{e^z - 1}\,; \qquad \forall z \in \mathbb{C},$$

except that at each stage the list of abscissae is augmented by the previously computed estimate of $\hat{z}_0$.

TABLE 2

Iterations to the Pole of $1/(e^z - 1)$; $K(z, \bar{w}) = 1/(1 - z\bar{w})$

n	2	3	4	5
z_1	0.4	0.4	0.4	0.4
z_2	0.3	0.3	0.3	0.3
z_3		$8.4102.10^{-2}$	$8.4102.10^{-2}$	$8.4102.10^{-2}$
z_4			$4.7366.10^{-3}$	$4.7366.10^{-3}$
z_5				$7.1363.10^{-6}$
$\hat{z}_0$	$8.4102.10^{-2}$	$4.7366.10^{-3}$	$7.1363.10^{-6}$	$0(10^{-14})$
m	4	3	3	2

Acknowledgment

Thanks are due to Devendra Kumar, a graduate student in the Department of Computing and Information Science, Queen's University, for assistance with the computations.

References

1. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.
2. H. Cramér, "Mathematical Methods of Statistics," Princeton Univ. Press, Princeton, N. J., 1951.
3. P. J. Davis, "Interpolation and Approximation," Baisdell, New York/London, 1963.
4. R. A. Fisher, On the mathematical foundations of theoretical statistics, *Phil. Trans. Roy. Soc.* **222** (1921), 309.
5. L. Gross, Measurable functions on a Hilbert space, *Trans. Amer. Math. Soc.* **105** (1962), 372–390.
6. F. M. Larkin, Gaussian measure in Hilbert space, and applications in numerical analysis, *Rocky Mountain J. Math.* **2** (1972), 397–421.
7. F. M. Larkin, Probabilistic error estimates in spline interpolation and quadrature, *in* "Information Processing 74," pp. 606–609, North–Holland, Amsterdam, 1974.
8. F. M. Larkin, Some remarks on the estimation of quadratic functionals, *in* "Theory of Approximation with Applications" (A. G. Law and B. N. Sahney, Eds.), pp. 43–63, Academic Press, New York, 1976.
9. F. M. Larkin, A further optimal property of natural polynomial splines, *J. Approximation Theory* **22** (1978), 1–8.

10. H. MESCHKOWSKI, "Hilbertsche Räume mit Kernfunktion," Springer-Verlag, Berlin/Göttingen/Heidelberg, 1962.
11. L. M. MILNE-THOMSON, "The Calculus of Finite Differences," Macmillan Co., New York/London, 1965.
12. A. SARD, "Linear Approximation," Math. Surveys No. 9, Amer. Math. Soc., Providence, R. I., 1963.