CLM-P 179

United Kingdom Atomic Energy Authority
RESEARCH GROUP
Preprint

ESTIMATION OF A NON-NEGATIVE FUNCTION

F. M. LARKIN

Culham Laboratory
Abingdon Berkshire

1968

UNCLASSIFIED
(Approved for Publication)

CLM - P 179

ESTIMATION OF A NON-NEGATIVE FUNCTION

by

F.M. LARKIN

(Submitted for Publication in B.I.T. (Tidskrift Informations Behandling)

ABSTRACT

It is pointed out that the problem of estimating a function from the values of a finite number of functionals is often posed, at least partially, as an aesthetic problem rather than as a mathematical one.

A particular problem of intepretation of experimental results, arising in the study of radiation source distribution in a plasma, is formulated mathematically in two different ways. Although the two formulations are different in character, they both rely on the idea of imposing a relative likelihood distribution on an appropriate function space; this is particularly fruitful when observational errors are taken into account.

U.K.A.E.A. Research Group
Culham Laboratory,
Abingdon,
Berks.
August 1968.

C O N T E N T S

APPENDICES

1. INTRODUCTION

A problem which occurs very frequently in the experimental sciences is that of estimating a function from given values of a finite number of discrete observations. Thus, for example, one might assume the existence of a functional relation $f(x)$ and attempt to estimate it from observations on the quantities $\{f(x_j) ; j = 1,2, \ldots m\}$. The work "estimate" is used here instead of the word "determine" since, firstly, the observations may be subject to experimental error and, secondly, even if the observations were known with complete accuracy, there may not be enough of them fix a unique function $f(x)$.

There is an important distinction between the two cases where

(a) the form of $f(x)$ is known, apart from the values of a finite number of assignable parameters; and

(b) the form of $f(x)$ is unknown, although its existence may be assumed.

In case (a) it would be possible in principle to fix $f(x)$ uniquely and precisely by making a sufficient (finite) number of exact observations. In most real experiments, of course, the observations are inexact, when it is customary to regard an observation not as a single number, but as a sample from the probability distribution of a number. By this means it is often possible to estimate the most probable values of the assignable parameters which could have resulted in the observations. In this context the technique of "least squares fitting" is so well known as to be used almost automatically!

We shall not consider case (a) in this paper, but will examine case (b) with a view to devising a workable mathematical formulation for the problem. For the purpose of illustration we shall consider a specific experimental situation, described in the next section, which requires the estimation of a non-negative function.

The concept of a relative likelihood in function space is introduce in section 5, where a particular form is developed. This form is used in section 6 in order to construct the first formulation of the estimation problem. Since this first formulation results in a solution with singular first derivative, in section 7 a criterion of smoothness is introduced which is used in the second formulation of the estimation problem, and it is shown how the smoothness criterion relates to an alternative relative likelihood functional, analagous to that used in Wiener integration. In section 8 a generalisation of Wiener integration is introduced in order to provide a basis for the computation of the means and variances used in section 9 for statistical improvement of observations subject to experimental error.

2. THE PARTICULAR EXAMPLE

A problem arising in the study of radiating cylinders of plasma, both astronomical and in the laboratory, is that of estimating a radial source density distribution from a limited number of integrated, edge-on measurements. Fig. 1(a) illustrates the situation diagramatically. A total of m measurements are made, the j^{th} reading being the integrated value of all radiation originating from the shaded region, bounded by the values x_j and x_{j+1} and the circumference of the plasma cylinder. Thus, assuming a radial distribution $f(r)$ for the source density of radiation, the observations are really attempts to determine the quantities

$$\varphi_j = 2\int_{x_j}^{x_{j+1}} dx \, . \int_{x}^{R} \frac{f(r) \, . \, rdr}{\sqrt{r^2 - x^2}} \; ; \; j = 1,2, \dots m, \qquad \dots (1)$$

where $f(r)$, being a physical source density is non-negative on $[0, R]$.

More generally, we could consider the quantities

$$\varphi_j = 2\int_0^R q_j(x) \, . \, dx \int_x^R \frac{f(r) \, . \, rdr}{\sqrt{r^2 - x^2}} \, ,$$

where the functions $\{q_j(x) \, ; \, j = 1,2, \ldots m\}$ are characteristic of of the measuring technique,

i.e.

where

$$\left.\begin{aligned} \varphi_j &= \int_0^R Q_j(r) \, . \, f(r) \, . \, rdr \\ Q_j(r) &= 2\int_0^r \frac{q_j(x) \, . \, dx}{\sqrt{r^2 - x^2}} \end{aligned}\right\} \quad j = 1,2 \ldots m. \qquad \ldots (2)$$

In the special case described by equations (1) it is easily verified that, for $j = 1,2, \ldots m$,

$$Q_j(r) = \begin{cases} 0 & ; \; 0 \leqslant r \leqslant x_j \\ 2 \, . \, \mathrm{Cos}^{-1}\left(\dfrac{x_j}{r}\right) & ; \; x_j \leqslant r \leqslant x_{j+1} \\ 2 \, . \left\{\mathrm{Cos}^{-1}\left(\dfrac{x_j}{r}\right) - \mathrm{Cos}^{-1}\left(\dfrac{x_{j+1}}{r}\right)\right\} & ; \; x_{j+1} \leqslant r \leqslant R \end{cases} \qquad \ldots (3)$$

Futhermore, as $m \to \infty$ and the points $\{ x_j \, , \, j = 1,2, \ldots m \}$ become more and more closely spaced, let

$$\frac{\varphi_j}{(x_{j+1} - x_j)} \to \psi(x_j) \quad , \quad \text{say,} \quad 0 \leqslant x_j < R, \qquad \ldots (4)$$

defining a function $\psi(x)$, which is actually an Abel transform of $f(r)$. Thus, we have

$$\psi(x) = 2\int_x^R \frac{f(r) \, . \, rdr}{\sqrt{r^2 - x^2}} \qquad \ldots (5)$$

which may be inverted by means of the formula (e.g. Titchmarsh, 1948)

$$f(r) = -\frac{1}{\pi}\int_r^R \frac{d[\psi(x)]}{\sqrt{x^2 - r^2}} \qquad \ldots (6)$$

provided we make the assumption

$$f(R) = 0 \qquad \ldots (7)$$

Since we assume that the source density vanishes identically for $r > R$, condition (7) simply expresses the requirement that $f(r)$ be continuous at $r = R$.

3. THE CONSISTENCY CONDITIONS

It is important to note that the given observational values $\{\varphi_j ; j = 1,2, \ldots m\}$ must be self-consistent. Using the functions defined in equations (3), and remembering that $f(r)$ is strictly a function of radius, it is not difficult to see from Fig. 1b that, if φ_m is given a non-zero value, a constraint is imposed upon the values which $\{\varphi_j ; j = 1,2, \ldots, m-1\}$ may attain. Specifically, we must have

$$\varphi_j \geqslant \int_{x_m}^{x_{m+1}} Q_j f \, . \, r dr \quad ; \quad j = 1,2, \ldots, m-1 \qquad \ldots (8)$$

given, of course, that

$$\int_{x_m}^{x_{m+1}} Q_m f \, . \, r dr = \varphi_m \, .$$

Similarly, as each succeeding value φ_j is prescribed, in order of decreasing j, further constraints are imposed upon φ-values corresponding to suffices of smaller order.

Self-consistent sets of the $\{\varphi_j\}$ are simply characterised by the following result:-

Theorem: A necessary and sufficient condition for the given values $\{\varphi_j \; ; \; j = 1,2, \ldots m\}$ to be consistent with the existence of a non-negative, integrable function $f(r)$ is that the quantities $\{\alpha_k \; ; \; k = 1,2, \ldots m\}$, satisfying the linearly independent equations

$$\sum_{k=j}^{m} \alpha_k \cdot Q_j(x_{k+1}) = \varphi_j \quad ; \quad j = 1,2, \ldots m, \qquad \ldots (9)$$

should all be non-negative.

Motivation for this result is provided by considering the sufficiency of the conditions of the theorem. We merely point out that the function

$$f^*(r) = \lim_{\varepsilon \to 0+} \cdot \sum_{k=1}^{m} \alpha_k \cdot \delta(r - x_{k+1} + \varepsilon) r^{-1} \qquad \ldots (10)$$

where the $\{\alpha_k\}$ are given by equations (9) and $\delta(x)$ is the Dirac δ-function, satisfies all the requisite conditions. Thus, the set of non-negative, integrable radial functions consistent with the given $\{\varphi_j\}$ is non-empty.

The proof of necessity is by induction. Note first that $Q_j(r)$ is monotonic increasing over the interval (x_j, x_{j+1}) and monotonic decreasing for $r > x_{j+1}$. Suppose that $f(r)$ is non-negative and define α_m by means of the equation

$$\varphi_m = \int_{x_m}^{x_{m+1}} Q_m f \cdot r dr = Q_m(x_{m+1}) \cdot \alpha_m , \qquad \ldots (11)$$

so that α_m is obviously non-negative.

Now define α_{m-1} by

$$\varphi_{m-1} = \int_{x_{m-1}}^{x_{m+1}} Q_{m-1} f \cdot r dr = Q_{m-1}(x_m) \cdot \alpha_{m-1} + Q_{m-1}(x_{m+1})\alpha_m, \quad \ldots (12)$$

whence, by substitution from (11), we find that

$$\int_{x_{m-1}}^{x_m} Q_{m-1} f \cdot r dr + Q_{m-1}(x_{m+1}) \int_{x_m}^{x_{m+1}} \left[\frac{Q_{m-1}(r)}{Q_{m-1}(x_{m+1})} - \frac{Q_m(r)}{Q_m(x_{m+1})}\right] \cdot fr\, dr$$

$$= Q_{m-1}(x_m) \cdot \alpha_{m-1} \quad \ldots (13)$$

By the monotonicity conditions and the non-negativity of $f(r)$, we now see that α_{m-1} must be non-negative. Proceeding radially inward, this argument may be repeated to demonstrate that all the quantities $\{\alpha_k ; k = 1,2, \ldots m\}$ must be non-negative if $f(r)$ is non-negative, thus completing the proof of necessity of the conditions of the theorem.

4. SOME FUNDAMENTAL CONSIDERATIONS

If the function $\psi(x)$ were known, the required function $f(r)$ could be found from equation (6). Several authors (Bracewell, 1956; Bockasten, 1961; Barr, 1962: Gorenflo and Kovetz, 1966) have described procedures which involve constructing a differentiable function $\psi^*(x)$ by fitting it somehow to 'smoothed' values of the observations $\{\varphi_j ; j = 1,2,\ldots m\}$. Birkeland and Oss (1967) have used a technique similar to that of Gorenflo and Kovetz, while allowing the plasma cylinder to absorb some of its emitted radiation. A disadvantage of this approach, pointed out by Gorenflo and Kovetz, is that equation (6) represents a process of ½-order differentiation, which may greatly amplify any errors present in the fitted function $\psi^*(x)$.

However, a more serious defect of this constructive approach stems from the fact that the problem: "given the exact values $\{\varphi_j \ ; \ j = 1,2,...m\}$, determine $\varphi(x)$, and $f(r)$", is not well posed. Clearly, there are very many conjugate pairs of functions $\varphi(x)$ and $f(r)$ which could have resulted in the observations $\{\varphi_j\}$, and the mere presentation of a method for constructing one particular pair begs several important questions of a mathematical nature. A constructive process may produce an answer, but one also needs to be clear about what mathematical problem is being solved!

Golomb and Weinberger (1958) have considered the question of how much one may legitimately infer about a function, considered as an element of a normed, linear space, from the values of a finite number of functionals. In particular, they showed that if the observations are on linear functionals only, extra information, in the form of a bound on a non-linear functional, must be supplied in order to restrict the required function to a bounded region in the space. In our case, the class of permissible functions $\{f(r)\}$ does not naturally form a linear space, so we need to re-examine the question of what extra information is required in order to make the problem well-posed. In each of sections 6 and 7 below, a unique, acceptable solution is ensured by insisting that a non-linear functional be minimised, not merely bounded.

The situation may be summarised informally, as follows:-

Physical intuition serves to convince us that $f(r)$, being a radiation source density distribution, is integrable and non-negative over the range $[0, R]$; the class of functions satisfying these conditions we shall denote by $\tilde{F}$. There is no compelling reason to bar any $f \in \tilde{F}$ satisfying equations (2) from being a candidate for our solution. However, many functions occurring in physical situations are known to be "smooth" in some loosely defined sense so, as a separate exercise, we restrict the class of allowable functions to those which are "smooth" in a sense defined precisely in section 7.

One of the objects of this paper is to emphasise the fact that the problem of estimating $f(r)$ from the given observations $\{\varphi_j \; ; \; j = 1,2, \ldots m\}$ may be made well-posed in many different ways. The purely constructive approach is rejected because of the temptation to use it as a substitute for proper formulation of a mathematical problem. However, the possibilities are indicated below by formulating a well-posed mathematical problem in two quite different ways, neither of which requires the use of the Abel inversion formula (6).

5. A RELATIVE LIKELIHOOD FOR FUNCTIONS

Although we have noted that a great variety of functions could have given rise to the set of observed quantities, there is still a natural inclination to regard some of these as more or less likely than others. For example, in Fig.2, if the marked points represent a set of observations on some function, one feels intuitively that curve 1 is much more likely to represent the required function than is curve 2, although either could have resulted in the observed ordinates. We shall attempt to formalise this natural preconception by associating with every function $f \in \tilde{F}$ a number which can be thought of as the relative likelihood of f. We shall then be in a position to answer the question : "of all $f \in \widetilde{F}$, which is the one most likely to have given rise to the observed quantities?"

There are many ways in which a relative likelihood function $\rho\{f\}$ could be defined, but in order to introduce the idea as naturally as possible we shall make use of an approach which is basic to statistical mechanics.

Regard the circle, radius R, as being dissected into n concentric, annular regions, all having the same area Δ. Now suppose that N particles are cast onto the circle, in such a way that the positional probability distributions of the particles are independently uniform over the area of the circle. Physically, these particles may be thought of as the radiating atoms in a unit length of the plasma cylinder mentioned in section 2; thus N is a very large, but finite, number.

The number of ways in which N_1 particles can fall into the first region (containing the origin), while N_2 fall into the second, etc., is

$$\frac{N!}{N_1!\ N_2!\ \ldots\ N_n!}$$

so the probability of the complexion $\underline{N} = (N_1, N_2, N_3, \ldots N_n)$ is given by

$$P(\underline{N}) = \frac{N!}{n^N \cdot \prod_{j=1}^{n} N_j!} \qquad \ldots (14)$$

Assuning that

$$N_j \gg 1 \quad ; \quad j = 1, 2, \ldots n \qquad \ldots (15)$$

and using Stirling's formula for the factorial, we find that equation (14) may be written

$$\log\left\{P(\underline{N})\right\} \sim N \cdot \log(N) - \sum_{j=1}^{n} N_j \cdot \log(N_j) - N \cdot \log(n) \quad \ldots (15)$$

We now define a histogram function $f_{\underline{N}}(r)$, associated with the complexion $\underline{N}$, by means of the equations

$$f_{\underline{N}}(r) = \frac{cN_j}{\Delta} = f_j \quad , \quad \text{say} \quad , \quad \begin{cases} r_{j-1} \leqslant r < r_j \\ j = 1, 2, \ldots n \end{cases} \qquad \ldots (17)$$

where c is constant for a given value of N, and r_{j-1} and r_j mark the inner and outer boundaries, respectively, of the j th annular region. Also, since we shall require $f_{\underline{N}}(r)$ to be integrable over $[0,R]$, note that equation (17) implies

$$c \sum_{j=1}^{n} N_j = cN = \Delta \sum_{j=1}^{n} f_j = 2\pi \int_0^R f_{\underline{N}} \cdot rdr = k \quad , \quad \text{say}, \ \ldots (18)$$

thence, equation (16) may be written in the form

$$\log \left\{ P(\underline{N}) \right\} \sim - \frac{N}{k} \cdot \Delta \sum_{j=1}^{n} f_j \log\left(\frac{Af_j}{k}\right) = - \frac{2\pi N}{k} \cdot \int_0^R f_{\underline{N}} \cdot \log \left(\frac{A}{k} \cdot f_{\underline{N}} \right) \cdot rdr$$

$$\ldots (19)$$

where

$$A = \pi R^2 = n\Delta . \qquad \ldots (20)$$

Thus we can assign a relative likelihood to $f_{\underline{N}}(r)$ by means of the relation

$$P\left\{ f_{\underline{N}}(r) \right\} = P \left\{ \underline{N} \right\} = \exp\left\{ - \frac{2\pi N}{k} \int_0^R f_{\underline{N}} \cdot \log \left(\frac{A}{k} f_{\underline{N}} \right) . rdr \right\} \ \ldots (21)$$

Although the above construction of functions $\left\{ f_{\underline{N}}(r) \right\}$, and their associated relative likelihoods, does not exhaust the whole of the function space $\tilde{F}$, by taking n and N large enough subject to

$$1 \ll n \ll N < \infty \qquad \ldots (22)$$

it does enable us to construct an $f_{\underline{N}}(r)$ whose mean deviation from any given non-negative, Riemann integrable function on $[0,R]$ is as small as desired.

6. THE FIRST FORMULATION

We are now in a position to cast the original, loosely worded requirement into mathematical form. As our estimate of the function which could reasonably have resulted in the given observations, we shall chose that function $f(r)$ which minimises the functional $\int_0^R f \cdot \log(f) \cdot rdr$ subject to the conditions of equations (2) and the condition that

$$f(r) \geqslant 0 \quad ; \quad 0 \leqslant r \leqslant R \qquad \ldots (23)$$

(Note that the factor $\frac{A}{k}$ in equation (21) only contributes a constant factor to the relative likelihood of $f_{\underline{N}}(r)$).

Posed in this form we have a non-linear programming problem in the function space $\tilde{F}$. However, the situation simplifies considerably if we make the additional assumption that the given observations $\left\{\varphi_j \;;\; j = 1,2 \ldots m\right\}$ are such as to support a strictly positive solution $f(r)$; in that case it can be determined by a standard variational technique, as follows:-

Using Langrange's method of undetermined multipliers, we minimise the functional

$$S(f) = \int_0^R f \cdot \log(f) \cdot rdr + \sum_{j=1}^{m} \varepsilon_j \left\{ \int_0^R Q_j f \cdot rdr - \varphi_j \right\} \qquad \ldots (24)$$

The Euler-Lagrange equation for this problem, which does not involve derivatives of f, is simply

$$1 + \log(f) + \sum_{j=1}^{m} \varepsilon_j \, Q_j = 0$$

e.g.

$$f(r) = \exp\left\{- 1 - \sum_{j=1}^{m} \varepsilon_j \cdot Q_j(r)\right\} \qquad \ldots (25)$$

where the constants $\left\{\varepsilon_j \;;\; j = 1, 2, \ldots m\right\}$ are determined by the auxiliary conditions

$$\varphi_j = \frac{1}{e}\int_0^R Q_j \cdot e^{-\sum_{k=1}^{m} \varepsilon_k Q_k} \cdot rdr \quad ; \quad j = 1,2, \ldots m \qquad \ldots (26)$$

It is worth noting that the variation could have been performed subject to the normalising condition

$$2\pi \int_0^R f \cdot rdr = B \quad , \quad \text{a constant,} \qquad \ldots (27)$$

thereby introducing an extra multiplying constant into the right-hand-side of equation (24). However, condition (27) is superfluous in the case of current interest, since the functions $\left\{Q_j \quad ; \quad j = 1, 2, \ldots m\right\}$ defined by equations (3) already imply that

$$2\pi \int_0^R fdr = 2\sum_{j=1}^{m} \varphi_j \qquad \ldots (28)$$

Fig. 3 illustrates in graphical form the result of using the above method on a particular set of observations, taking the special form given in equations (3) for the functions $\left\{Q_j \quad ; \quad j = 1, 2, \ldots m\right\}$. Numerical values of the input quantities for this example, and for the following ones, are given in Appendix III.

7. THE SECOND FORMULATION

It is at least partly an aesthetic problem to ask for the most likely smooth function $f(r)$ which could have resulted in the given observations $\left\{\varphi_j \; ; \; j = 1, 2, \ldots m\right\}$. Our next task, therefore, will be to formulate mathematically what could reasonably be meant by 'smoothness'.

In a particular experimental situation there might be some a priori indication of how the smoothness of functions might be compared. However, in the absence of such guidance let us regard a constant as being the smoothest kind of function, and if a function deviates more and more from being constant we shall regard it as being less and less smooth. Thus, a reasonable measure of the lack of smoothness of the function $h(r)$, over the circle of radius R, is the functional

$$\xi\left\{h\right\} = 2\pi \int_0^R \left(\frac{dh}{dr}\right)^2 . \, r dr \qquad \ldots (29)$$

We might immediately attempt to find the smoothest function, in the sense just outlined, which could have resulted in the observations $\left\{\varphi_j \; ; \; j = 1, 2, \ldots m\right\}$, thus minimising the error amplification noted in section 4. However, a technical point arises here, because there is no reason to restrict $h(r)$ in equation (29) to being non-negative, although it will have to be piecewise differentiable. The space $\tilde{F}$ is no longer necessarily a good source of candidates for the solution to our problem. A natural way out of this difficulty is to express the required, non-negative function $f(r)$ in the form

$$f(r) = \left\{h(r)\right\}^2 \qquad \ldots (30)$$

where h is allowed to range over the Hilbert space $\tilde{H}$ of real-valued, piecewise differentiable functions on $[0,R]$ which satisfy

$$h(R) = 0, \qquad \ldots (31)$$

and have an inner product defined by

$$(g,h) = \int_A (\nabla g) \cdot (\nabla h) \cdot dA = 2\pi \int_0^R g' h'_{\circ} \, r \; dr \qquad \ldots (32)$$

and norm given by

$$|| \, h \, ||^2 = (h,h). \qquad \ldots (33)$$

Formally, then, we are seeking a function $h \in \tilde{H}$, which minimises the functional $\xi\left\{h\right\}$ subject to the constraints

$$\varphi_j = \int_0^R Q_j \circ h^2 \cdot r dr \quad ; \quad j = 1, 2, \ldots m \qquad \ldots (34)$$

Using the method of undetermined multipliers, the Euler-Lagrange equation for this problem is

$$\frac{1}{r} \cdot \frac{d}{dr}\left(r \frac{dh}{dr}\right) + h \cdot \sum_{j=1}^{m} \varepsilon_j Q_j = 0 , \qquad \ldots (35)$$

with boundary conditions

$$h'(0) = 0 = h(R) , \qquad \ldots (36)$$

where the constants $\left\{\varepsilon_j \quad ; \quad j = 1, 2, \ldots m\right\}$ are determined by equations (34).

The numerical solution of this generalised eigenvalue problem is dealt with in Appendix I. The result of using this second formulation of the problem of estimating a radiation source distribution from exact observations is shown in Fig.4; this is directly comparable with Fig.3, which resulted from the first formulation.

8. FUNCTIONAL INTEGRATION

Of course, many measures of lack of smoothness, other than that defined in equation (29), could be used. However $\xi(h)$ is of particular interest to physicists because of its analagy with Wiener integration. In its original form (Wiener, 1924) Wiener integration enables one to compute expectation values of functionals over the space of functions continuous on $[0,1]$ and vanishing at the lower limit. One definition of the expectation value of $F\left\{h\right\}$, if it exists, is

$$\tilde{E}\left\{F\right\} = \int F\left\{h(r)\right\} d_W h = \lim_{\substack{S\to\infty \\ \operatorname{Max}_j(\Delta r_j)\to 0}} \cdot B_S \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F\left\{h^*(r)\right\} e^{-\lambda \int_0^1 \left(\frac{dh^*}{dr}\right)^2 dr} \cdot dx_1 dx_2 \ldots dx_S \qquad \ldots (37)$$

where $h^*(r)$ is a piecewise linear, interpolatory approximation to $h(r)$ over the range $[0,1]$

i.e.
$$\left.\begin{aligned} & h^*(r_j) = h(r_j) = x_j \quad ; \quad j = 1, 2, \ldots s , \\ & 0 = r_0 < r_1 < r_2 < \ldots < r_{s-1} < r_s = 1 \end{aligned}\right\} \quad \ldots (38)$$

and B_s is an appropriate normalising factor. Notice that the exponential factor in equation (37) plays the role of the relative likelihood of $h^*(r)$.

Our situation is a radial variant of that considered by Wiener, with functions vanishing at R instead of at 0. Later on we make use of an expectation definition analagous with (37) in order to make statistical improvements to inexact observational data. However, for the relatively simple situation when the given observations are exact it is sufficient to note that, if

$$\exp\left\{-\lambda\int_0^R\left(\frac{dh}{dr}\right)^2 \cdot r dr\right\}$$

is regarded as the relative likelihood of $h \in \tilde{H}$, a maximum likelihood estimate of $h(r)$, subject to conditions (2) results in the smoothest function satisfying these conditions.

It is worth pointing out that the general inner product defined in equation (32) is not limited to the case of radial dependence of the functions in $\tilde{H}$. Clearly, the development may be extended to the situation where $\tilde{H}$ comprises the set of suitably differentiable functions of two (or possibly more) variables, which vanish on some convenient boundary enclosing the region of interest. This more general formulation would, in principle, allow one to estimate the 'most likely' function of several variables consistent with given observed functional values.

Preparing the way for computation of the means and variables necessary for section 9, we now introduce a generalisation of Wiener integration.

Let $\left\{g_j\right\}$ be a complete sequence of distinct elements in $\widetilde{H}$, and $F\left\{h\right\}$ a functional whose domain includes $\widetilde{H}$.

Let
$$E_n\left\{F\right\} = \frac{\int^{(n)} F\left\{h_n\right\} \cdot e^{-\lambda \, || \, h_n \, ||^2} \cdot dx_1 \, dx_2 \ldots dx_n}{\int^{(n)} e^{-\lambda \, || \, h_n \, ||^2} \cdot dx_1 \, dx_2 \ldots dx_n} \qquad \ldots (39)$$

where each of the n-folded integrals ranges over $(-\infty, \ \infty)$, and h_n is that element of $\widetilde{H}$ with least norm which satisfies the conditions

$$\tau_j = (g_j \, , \, h_n) \quad ; \quad j = 1,2,\ldots n. \qquad \ldots (40)$$

Then, we define the expectation value of the functional $F\left\{h\right\}$ to be

$$\widetilde{E}\left\{F\right\} = \lim_{n\to\infty} \cdot E_n\left\{F\right\} \qquad \ldots (41)$$

if the limit exists. It may be shown (Larkin 1968) that if this limit exists it is independent of the particular complete sequence $\left\{g_j\right\}$ used in its construction. In the present example we have

$$\widetilde{E}\left\{\psi_j\right\} = \widetilde{E}\left\{ \int_0^R Q_j(r) \cdot h^2 \, (r) \cdot rdr\right\} = \int_0^R Q_j \, (r) \cdot r \cdot \widetilde{E}\left\{h^2(r)\right\} \cdot dr \qquad \ldots (42)$$

Now notice that $\widetilde{E}\left\{h(r)\right\}$ <u>for a specific value of r</u>, depends only upon one single, bounded, linear functional $h(r)$; thus the multiple integrals in equation (39) reduce to single integrals if we include the representer of $h(r)$ in the sequence $\left\{g_j\right\}$.

The function $h_1(t)$ which has smallest norm while satisfying

$$h_1(r) = h(r) \qquad \ldots (43)$$

may be found by a standard variational technique; it turns out to be of the form

$$h_1(t) = \begin{cases} h(r) & ; \quad 0 \leqslant t \leqslant r \\ \dfrac{h(r) \cdot \log (\frac{t}{R})}{\log(\frac{r}{R})} & ; \quad r \leqslant t \leqslant R , \end{cases} \qquad \ldots (44)$$

so that

$$|| h_1 ||^2 = \frac{2\pi h^2 (r)}{\log(^R/r)} . \qquad \ldots (45)$$

Hence,

$$\widetilde{E} \left\{ h^2 (r) \right\} = \frac{\int_{-\infty}^{\infty} h^2 e^{- \frac{2\pi h^2 \lambda}{\log (^R/r)}} \cdot dh}{\int_{-\infty}^{\infty} e^{- \frac{2\pi h^2 \lambda}{\log (^R/r)}} \cdot dh} , \qquad \ldots (46)$$

i.e.

$$\widetilde{E} \left\{ h^2 (r) \right\} = \frac{\log (^R/r)}{4\pi \lambda} , \qquad \ldots (47)$$

$$\widetilde{E} \left\{ \psi_j \right\} = \frac{1}{4\pi \lambda} \cdot \int_0^R \log (^R/r) \cdot Q_j (r) \cdot rdr. \qquad \ldots (48)$$

The elements H_{jk} of the variance matrix H may be found in a similar fashion. In order not to burden the text the derivation is given in Appendix II, and here we simply quote the result

$$H_{jk} = \widetilde{E} \left\{ \psi_j \psi_k \right\} - \widetilde{E} \left\{ \psi_j \right\} \cdot \widetilde{E} \left\{ \psi_k \right\} = \frac{1}{8\pi^2 \lambda^2} \cdot \int_0^R \left[\log (\frac{R}{r}) \right]^2 \cdot \left[Q_j(r) \cdot S_k(r) + Q_k(r) \cdot S_j(r) \right] \cdot rdr$$

$$j, k = 1, 2, \ldots m \qquad \ldots (49)$$

where

$$S_j (r) = \int_0^r Q_j (t) \cdot tdt \quad ; \quad j = 1, 2, \ldots m. \qquad \ldots (50)$$

9. STATISTICAL IMPROVEMENT OF THE OBSERVATIONS

Let us suppose that the measured quantities $\left\{\varphi_j,\ j = 1, 2, \ldots m\right\}$ result from the additive contamination of the true values $\left\{\psi_j\ ;\ j = 1, 2, \ldots m\right\}$ by noise components $\left\{\eta_j,\ j = 1, 2, \ldots m\right\}$. Thus

$$\varphi_j = \psi_j + \eta_j \quad ; \eta_j = 1, 2, \ldots m \qquad \ldots (51)$$

The terms 'true values' and 'noise' have merely a notational significance, since the n th order vectors $\underline{\psi}$ and $\underline{\eta}$ may be regarded mathematically as partitions of a compound stochastic variable vector of order 2 m. In many practical situations it is reasonable to assume that $\underline{\eta}$ has zero mean value and that $\underline{\psi}$ and $\underline{\eta}$ are uncorrelated; these assumptions are made in the following development.

Let us define the mean vector $\bar{\underline{\psi}}$ by

$$\bar{\psi}_j = \tilde{E}\left\{\psi_j\right\} \quad ; \quad j = 1, 2, \ldots m, \qquad \ldots (52)$$

and the variance matrices H and K by

$$H_{jk} = \tilde{E}\left\{\psi_j\psi_k\right\} - \tilde{E}\left\{\psi_j\right\}\cdot\tilde{E}\left\{\psi_k\right\} \qquad \ldots (53)$$

and

$$K_{jk} = \tilde{E}\left\{\eta_j\eta_k\right\}, \qquad \ldots (54)$$

where the expectation operator represents the process of averaging over the Hilbert space $\tilde{H}$, as described in the previous section.

It may be verified (e.g. Deutsch, 1965, p.67) that the expression

$$\hat{\underline{\psi}} = \bar{\underline{\psi}} + H(H + K)^{-1}\ (\underline{\varphi} - \bar{\underline{\psi}}) \qquad \ldots (55)$$

is an unbiased, minimum variance estimator of $\underline{\psi}$.

The most information an experimenter is usually prepared to give about the accuracy of his observations is an estimate of the 'error', by which he means the standard deviation of each observation from its preferred value. Thus, the matrix K will usually be strictly diagonal.

An example of the use of formula (55) for 'smoothing' inaccurate observations, prior to estimating the radial source distribution, is given in Appendix III and illustrated in Fig.5. Notice that, in contrast with the situation in which the observations are exact, the final estimate of the source distribution depends upon the parameter λ which, roughly speaking, measures the a priori dispersion of elements of $\tilde{H}$ about their mean. λ can be thought of as a 'smoothing parameter', since the measure of lack of smoothness $\xi\left\{h\right\}$, defined in equation (29), decreases as λ increases. In the numerical example λ was chosen so as to satisfy the condition

$$\sum_{j=1}^{m} \bar{\psi}_j = \sum_{j=1}^{m} \varphi_j \qquad \dots (56)$$

Another reasonable condition might be

$$\sum_{j=1}^{m} \hat{\psi}_j = \sum_{j=1}^{m} \varphi_j$$

which would necessitate iterative determination of λ.

10. CONCLUDING REMARKS

We have been concerned with the problem of estimating a function f(r), about which we are given a limited amount of experimental information. Although the results of experiments are analysed one at a time, we need to consider the class of all possible results, and be specific about the class $\tilde{C}$ of all functions $f(r) \in \tilde{C}$ which we can accept as conceivably giving rise to an observable result.

Our information about f(r) is of two kinds:-

(i) General properties, such as non-negativity, continuity and integrability of f(r) and/or its derivatives. This information, which may be genuinely given or arbitrarily imposed, limits a priori the class $\tilde{C}$ of allowable solutions.

(ii) Specific properties, in the form of a finite number of observations on f(r). These observations may be thought of as linear or non-linear functionals of f, which limit a posteriori the sub-class of $\tilde{C}$ to which f can belong.

In general, an infinite number of exact observations will be required in order to characterise a particular $f \in \tilde{C}$, an impossible requirement in a real, physical situation. Thus, unless more assumptions are made, we can say no more about f(r) than that it is a member of that sub-class of $\tilde{C}$ whose elements could have resulted in the given observations - a tautology which is hardly likely to appeal to an experimental physicist!

A way out of this unpalatable but incontrovertible dilemma is to express precisely any intuitive feeling we may have that certain members of $\tilde{C}$ are a priori more likely than others. This enables us to formulate a variational problem whose solution, roughly speaking, represents the a posteriori most likely member of $\tilde{C}$. Furthermore, it turns out that this approach extends very naturally to cover the situation in which the given observations are subject to experimental error.

The particular experimental situation discussed in the text has been chosen deliberately in order to illustrate the various ways in which a corresponding mathematical problem may be formulated out of the original aesthetic one. It is, of course, the privilege of the experimenter to decide, by looking at the final results if necessary, which of the many possible mathematical formulations he prefers to use.

11. ACKNOWLEDGEMENTS

The author is grateful to Mr. A. Sykes (U.K.A.E.A. Culham Laboratory) for his help with one of the numerical examples, and also to Dr. R. Gorenflo and Mrs. Y. Kovetz (Institut für Plasmphysik, Garching bei München) for bringing the example of section 2 to his attention.

12. REFERENCES

1. BARR, W.L., Method for Computing the radial distribution of emitters in a cylindrical source. J. Opt. Soc. Am. 52, 885-888 (1962).

2. BIRKELAND, J.W. and OSS, J.P. Spatial Resolution of Emission and Absorption Coefficients in Self-Absorbing Sources of Cylindrical Symmetry. A.R.L. 67-0279, Dec. 1967.

3. BRACEWELL, R.N., Strip integration in radio astronomy. Austral. J. Phys. 9, 198-217 (1956).

4. BOCKASTEN, K., Transformation of observed radiances into radial distribution of the emmission of a plasma J. Opt. Soc. Am. 51, 943-947 (1961).

5. DEUTSCH, R., Estimation Theory. Prentice Hall, Englewood Cliffs, N.J. (1965).

6. GOLOMB, M., and WEINBERGER, H.F., Optimal Approximation and Error Bounds, ONA 117-190, MR22, 12697 (1959).

7. GORENFLO, R., and KOVETZ, Y., Solution of an Abel type integral equation in the presence of noise by Quadratic Programming. Numerische Mathematik, 8, 392-406 (1966) (and earlier papers).

8. LARKIN, F.M., Gaussian Measure in Hilbert Space, and Applications in Numerical Analysis (in preparation).

9. TITCHMARSH, E.C., Introduction to the Theory of Fourier Integrals. O.U.P., N.Y. (1948).

10. WIENER, N., The Average Value of a Functional. Proc. London Math. Soc. 22, 454 (1924).

APPENDIX I : NUMERICAL TECHNIQUE FOR THE EIGENVALUE PROBLEM

We define the linear, self adjoint, positive definite operators $\{L_j\}$ by means of the relations

$$(h,L_jh) = \int_o^R Q_jh^2 rdr \quad ; \quad j = 1,2,\ldots m \qquad \ldots (1)$$

from which it follows that

$$L_jh = \frac{1}{2\pi} \cdot \int_r^R \frac{dt}{t} \cdot \int_o^t Q_j(s).h(s).sds \quad ; \quad j = 1,2,\ldots m \qquad \ldots (2)$$

i.e.

$$L_jh = \int_o^R K(r,s).Q_j(s).h(s).sds \quad ; \quad j = 1,2,\ldots m \qquad \ldots (4)$$

where

$$K(r,s) = \begin{cases} \frac{1}{2\pi}.\log \frac{R}{s} & ; \quad 0 \leqslant r \leqslant s \leqslant R \\ \frac{1}{2\pi}.\log \frac{R}{r} & ; \quad 0 \leqslant s \leqslant r \leqslant R \end{cases} \qquad \ldots (5)$$

Our problem now is to find h(r) such that (h,h) is minimised, subject to

$$(h,L_jh) = \varphi_j \quad ; \quad j = 1,2,\ldots m. \qquad \ldots (6)$$

The iteration adopted may be described as follows:

(i) Choose a starting function h_o

(ii) Compute normalising constants $\{\alpha_j\}$ such that

$$(\alpha_jh_o,L_j\alpha_jh_o) = \varphi_j$$

i.e.

$$\alpha_j = \left\{\frac{\varphi_j}{(h_o,L_jh_o)}\right\}^{\frac{1}{2}} \quad ; \quad j = 1,2,\ldots m. \qquad \ldots (7)$$

(iii) Construct h_1 by projecting the origin onto the linear manifold common to the supporting hyperplanes at the points $\left\{\alpha_j h_o; \quad j = 1,2,\ldots m\right\}$

(iv) Construct h_2, h_3,..etc., as required, by repeating operations (ii) and (iii) on the most recently constructed member of the sequence.

Algebraically, the iteration may be expressed as

$$h_{r+1} = \sum_{j=1}^{m} \beta_j L_j h_r \qquad \ldots (8)$$

where the vector $\underline{\beta}$ satisfies the equation

$$A\underline{\beta} = \underline{\Upsilon} \qquad \ldots (9)$$

and

$$A_{jk} = (L_j h_r, L_k h_r) \quad ; \quad \left.\begin{array}{l} j = 1,2,\ldots m \\ k = 1,2,\ldots m \end{array}\right\} \qquad \ldots (10)$$

$$\Upsilon_j = \left\{\varphi_j (h_r, L_j h_r)\right\}^{\frac{1}{2}} \quad ; \quad j = 1,2,\ldots m \qquad \ldots (11)$$

Assuming that the given values $\left\{\varphi_j \quad ; \quad j = 1,2,\ldots m\right\}$ are self-consistent, and that the iteration converges to h(r), it is easy to verify that h satisfies equations (6) and (h,h) has a local minimum subject to these conditions.

APPENDIX II : DERIVATION OF THE VARIANCE MATRIX

Consider

$$\tilde{E}\left\{\psi_j\psi_k\right\} = \tilde{E}\left\{\int_o^R\int_o^R h^2(r).h^2(s).Q_j(r).Q_k(s).rs.dr.ds\right\} \quad \ldots (1)$$

i.e. $$\tilde{E}\left\{\psi_j\psi_k\right\} = \int_o^R\int_o^R \tilde{E}\left\{h^2(r).h^2(s)\right\}.Q_j(r).Q_k(s).rs.dr\ ds \quad \ldots (2)$$

Suppose $$o < r < s < R \quad \ldots (3)$$

then the function $h_2\ (t) \in \tilde{H}$ with minimum norm, subject to the conditions

$$\left.\begin{array}{l} h_2(r) = h(r) \\ h_2(s) = h(s) \end{array}\right\} \quad \ldots (4)$$

is given by

$$h_2(t) = \begin{cases} h(r) & ;\ o \leqslant t \leqslant r \\ h(r) + [h(s) - h(r)]\dfrac{\log\left(\frac{t}{r}\right)}{\log\left(\frac{s}{r}\right)} & ;\ r \leqslant t \leqslant s \\ h(s).\dfrac{\log\left(\frac{t}{R}\right)}{\log\left(\frac{s}{R}\right)} & ;\ s \leqslant t \leqslant R \end{cases} \quad \ldots (5)$$

Hence $$\|h_2\|^2 = \frac{2\pi[h(s) - h(r)]^2}{\log\left(\frac{s}{r}\right)} + \frac{2\pi h^2(s)}{\log\left(\frac{R}{s}\right)} \quad \ldots (6)$$

Now, if we make the substitutions

$$\alpha = \frac{2\pi}{\log\left(\frac{s}{r}\right)} \quad ; \quad \beta = \frac{2\pi}{\log\left(\frac{R}{s}\right)} \quad \ldots (7)$$

it is clear from the definition that, for fixed r and s,

$$\tilde{E}\left\{h^2(r).h^2(s)\right\} = \gamma\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h^2(r).h^2(s).e^{-\lambda[\alpha(h(s)-h(r))^2+\beta.h^2(s)]}.dh(r).dh(s) \quad \ldots (8)$$

where $$\gamma^{-1} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\lambda[\alpha(h(s)-h(r))^2+\beta h^2(s)]}.dh(r).dh(s) \quad \ldots (9)$$

The integrals in equations (8) and (9) may be evaluated explicitly, yielding

$$\tilde{E}\left\{h^2(r).h^2(s)\right\} = \frac{1}{16\pi^2\lambda^2}\cdot\left\{\log\left(\frac{R}{r}\right)\cdot\log\left(\frac{R}{s}\right) + 2\left[\log\left(\frac{R}{s}\right)\right]^2\right\} \quad ; \quad o<r<s<R \quad \ldots(10)$$

Thus, for o<s<r<R

$$\tilde{E}\left\{h^2(r).h^2(s)\right\} = \frac{1}{16\pi^2\lambda^2}\cdot\left\{\log\left(\frac{R}{r}\right)\cdot\log\left(\frac{R}{s}\right) + 2\left[\log\left(\frac{R}{r}\right)\right]^2\right\} \quad \ldots(11)$$

These results may be substituted into equation (2), giving

$$\tilde{E}\left\{\psi_j\psi_k\right\} = \begin{cases} \frac{1}{16\pi^2\lambda^2}\cdot\int_0^R\int_0^R \log\left(\frac{R}{r}\right)\cdot\log\left(\frac{R}{s}\right)\cdot Q_j(r).Q_k(s).rs.dr\,ds \\ + \frac{1}{8\pi^2\lambda^2}\cdot\int_0^R Q_j(r)\cdot r\,dr\int_0^r\left[\log\left(\frac{R}{r}\right)\right]^2\cdot Q_k(s)\cdot sds \\ + \frac{1}{8\pi^2\lambda^2}\cdot\int_0^R Q_j(r)\cdot rdr\int_r^R\left[\log\left(\frac{R}{s}\right)\right]^2\cdot Q_k(s)\cdot sds \end{cases} \quad \ldots (7)$$

Referring now to equation (53) of the text we see that

$$\tilde{E}\left\{\psi_j\psi_k\right\} - \tilde{E}\left\{\psi_j\right\}\tilde{E}\left\{\psi_k\right\} = \frac{1}{8\pi^2\lambda^2} \cdot \int_0^R \left[\log\frac{R}{r}\right]^2 \cdot \left[Q_j(r).S_k(r)+Q_k(r).S_j(r)\right].rdr$$

$$j,k = 1,2,\ldots m, \qquad \ldots(8)$$

where

$$S_j(r) = \int_0^r Q_j(t).tdt \quad ; \quad j = 1,2,\ldots m, \qquad \ldots(9)$$

as required.

APPENDIX III : NUMERICAL EXAMPLES

For all three of the numerical examples given here the input parameters are summarised in the following table:

TABLE 1

INPUT PARAMETERS

j	x_j	x_{j+1}	φ_j	σ_j
1	0·0000	0·2000	3·2000	1·0000
2	0·2000	0·4000	3·4000	1·0000
3	0·4000	0·6000	3·8000	1·0000
4	0·6000	0·8000	2·8000	1·0000
5	0·8000	1·0000	1·6000	1·0000

Here, $m = 5$ and $R = 1{\cdot}0$. The quantities $\left\{\sigma_j \quad ; \quad j = 1,2,\ldots 5\right\}$ represent given estimates of the mean deviations of the corresponding observations; thus, in this case the noise variance matrix K reduces to the unit matrix. Of course, the values of the $\left\{\sigma_j\right\}$ are only significant in the third example, whose results are illustrated in Fig. 5.

In Figs. 3,4 and 5, the given observations $\left\{\varphi_j \quad ; \quad j = 1,2,\ldots 5\right\}$ are shown as histograms, the area under each segment being equal to the corresponding observational value. The smoothed values, in the third case, are shown as a broken histogram in Fig. 5. In all three cases it is clear that $\psi(x)$, the Abel transform of the radial distribution function, does follow the appropriate histogram quite closely, although the first formulation gives results which may be less acceptable than those of the second formulation. Furthermore, comparison of Figs. 4 and 5 indicates that the double peak in the solution to the 'unsmoothed case' is probably spurious.

Table 2 lists the computed values of the quantities $\left\{ \varepsilon_j \; ; \; j = 1,2,..5 \right\}$ satisfying equations (26); Table 3 and Table 4, respectively, give the computed values of the mean vector $\bar{\underline{\psi}}$ and the variance matrix H, corresponding to a unit value of the smoothing parameter λ.

TABLE 2

CONSTANTS IN 1ST FORMULATION

j	ε_j
1	2·314
2	2·693
3	4·793
4	3·041
5	3·356

TABLE 3

SIGNAL MEAN VECTOR

j	$\bar{\underline{\psi}}_j$
1	0·4307E-02
2	0·2916E-02
3	0·1743E-02
4	0·8076E-03
5	0·1705E-03

TABLE 4

SIGNAL VARIANCE MATRIX

j \ k	1	2	3	4	5
1	0.78830519E-04	0.33055824E-04	0.11166399E-04	0.26559307E-05	0.21879508E-06
2	0.33055824E-04	0.23884418E-04	0.99185329E-05	0.24719909E-05	0.20754065E-06
3	0.11166399E-04	0.99185329E-05	0.65329588E-05	0.20502862E-05	0.18305404E-06
4	0.26559307E-05	0.24719909E-05	0.20502862E-05	0.11190839E-05	0.13764012E-06
5	0.21879508E-06	0.20754065E-06	0.18305404E-06	0.13764012E-06	0.39713986E-07

For computational purposes, the functions employed in the second formulation (cases 2 and 3) were represented by their values at 101 equi-spaced abscissae, spanning the range [0,1]; all quadratures were effected by the trapezoidal rule. The slight discrepancy, in Figs.3 and 4, between $\psi(x)$ and the first section of the histogram is a result of the discretization. Each of these cases required about 25 seconds of computer time (FORTRAN program, E.E.C. KDF9 machine).

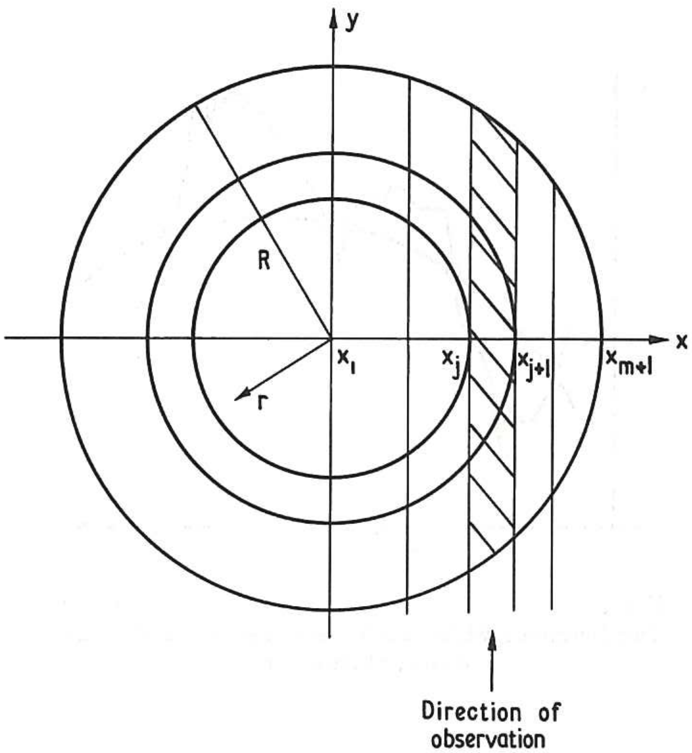

Fig. 1(a) Diagram of the experimental situation (CLM-P179)

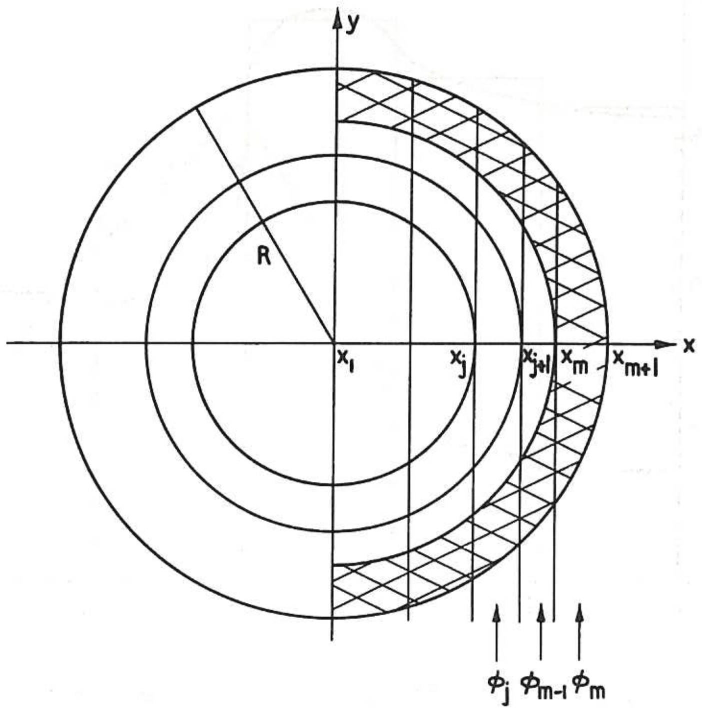

Fig. 1(b) (CLM-P179)
Illustration of the interdependence of the observations

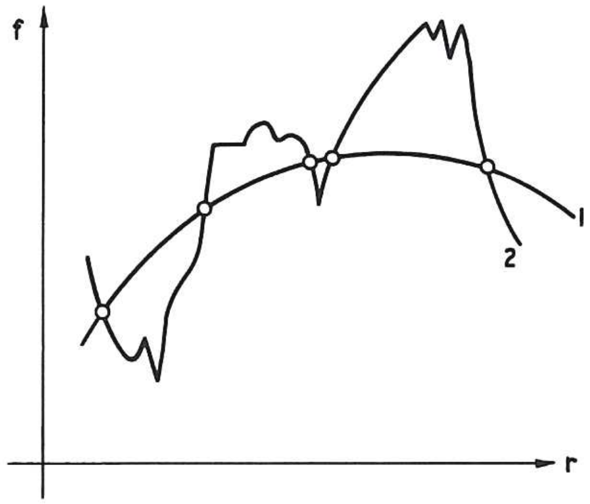

Fig. 2 (CLM-P 179)
Two functions which could have resulted in the same observed data points

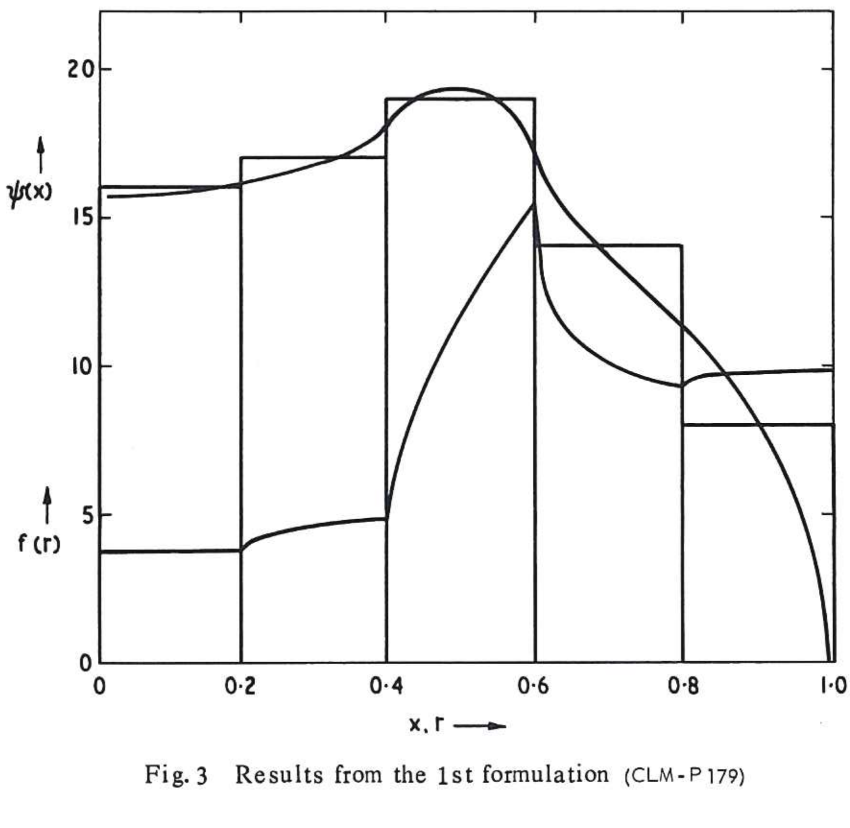

Fig. 3 Results from the 1st formulation (CLM-P 179)

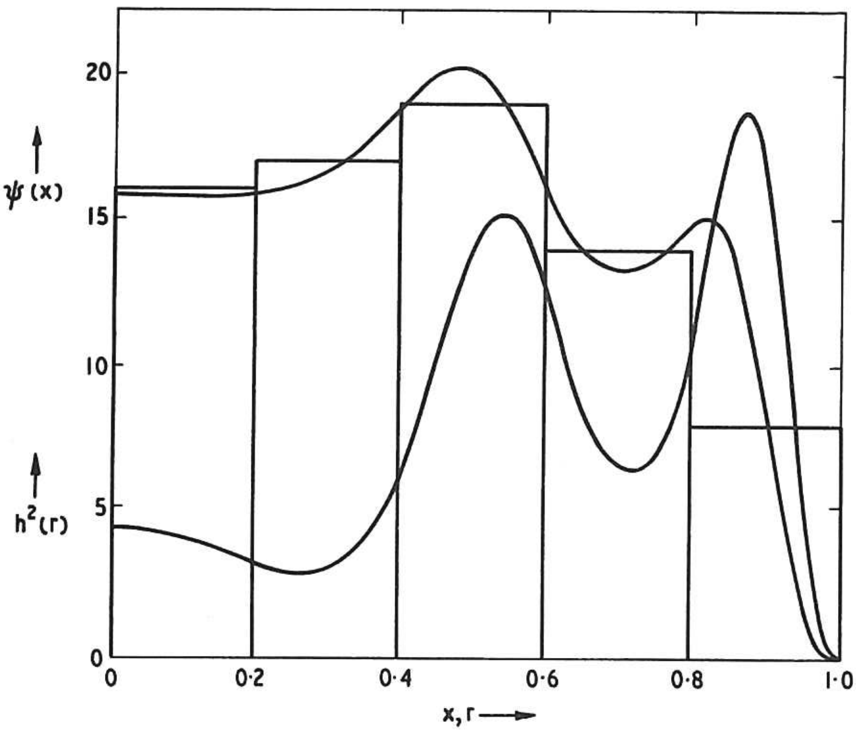

Fig. 4 (CLM-P 179)
Results from 2nd formulation, after 5 iterations; observations not smoothed

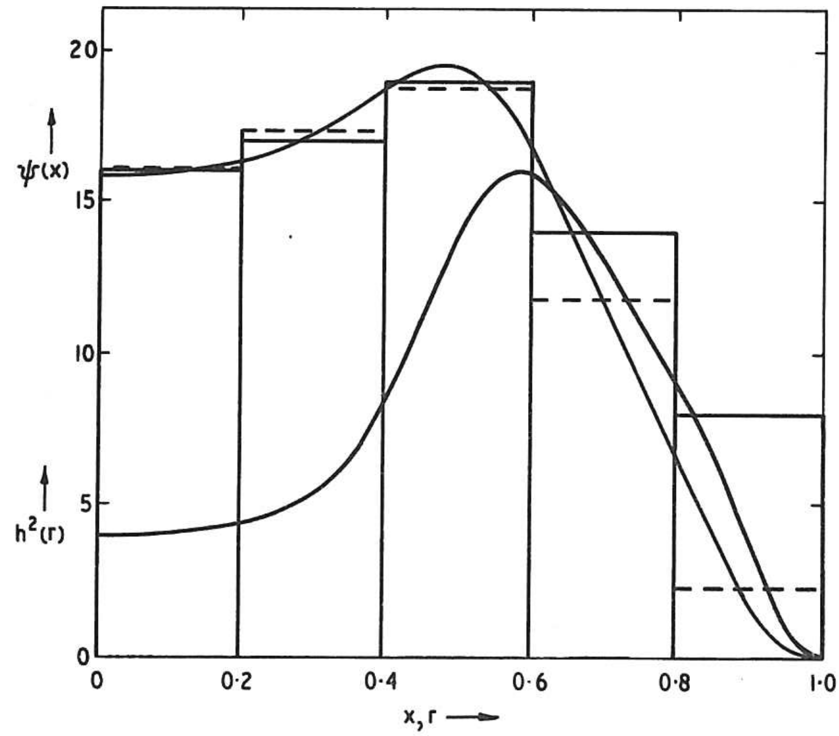

Fig. 5 (CLM-P 179)
Results from 2nd formulation after 4 iterations ; observations 'optimally' smoothed ($\lambda = 6{\cdot}72 \cdot 10^{-4}$)