SOVIET PHYSICS

USPEKHI

A Translation of Uspekhi Fizicheskikh Nauk

Vol. 13, No. 6, pp. 681-840 (Russian Original Vol. 102, Nos. 3 and 4) May-June 1971

519.24/.27

THE USE OF MATHEMATICAL-STATISTICS METHODS IN THE SOLUTION OF INCORRECTLY POSED PROBLEMS

V. F. TURCHIN, V. P. KOZLOV, and M. S. MALKEVICH

Institute of Atmospheric Physics and Institute of Applied Mathematics, U.S.S.R. Academy of Sciences

Usp. Fiz. Nauk **102**, 345–386 (November, 1970)

INTRODUCTION

THE term "incorrectly posed problems" now includes a wide variety of problems of different sorts and different origins, many of which lie outside the scope of the present review. Deferring a detailed explanation of the mathematical meaning of the term to Chapter 1, we note that the expression "incorrectly posed problems" is not to be taken too literally, although these problems, in contrast with the "correctly posed" ones, are very sensitive to the exact formulation, and therefore are often in fact found to be incorrectly formulated. One class of incorrectly posed problems, a correct formulation of which is of much value for their clear physical content, is the so-called incorrectly posed inverse problems of mathematical physics. According to an established terminology,* direct problems of mathematical physics are problems oriented along a cause-effect sequence—i.e., problems of finding out the consequences of given causes: the determination of fields in time and space from given sources, the calculation of the reaction of a device to a known signal, and so on.

In this sense the inverse problems are those associated with the reversal of the chain of causally related effects, i.e., problems of finding the unknown causes of known consequences: the determination of the characteristics of the sources of a field from the values of the field at certain points or in certain regions of space, the reconstruction of the input signal from the reaction at the output of a device, and so on. Inverse problems usually arise as problems of the interpretation of some sort of observations.

Properly speaking, any problem of measuring certain characteristics of a physical object is an inverse problem in the sense of this definition. There is, however, a tendency to use the term "inverse problem" to denote rather complicated problems of interpretation, concerning either the simultaneous and interdependent measurement of many parameters of a physical object, or else cases in which the number of parameters is indefinitely large (as, for example, when the state of an object is described by some function of the coordinates). Strictly speaking, only problems of the second sort can be called incorrectly posed, and they have received special attention; but when the number of parameters is large problems of the first kind can also display characteristic features of incorrectly posed problems, which, as we shall see later, are due to an indefiniteness of information. The methods of solution under conditions of such uncertainty are essentially the same for problems of both types.

Although it cannot be asserted that all "functional" inverse problems are clearly incorrectly posed in the mathematical sense, still in most cases they are incorrectly posed. They include the inverse problem of potential theory[2, 3] as it occurs in the interpretation of gravimetric observations in geological exploration, the inverse problem of heat conduction,[4] a class of inverse problems of radiative transfer[6-9, 12-18, 46-52] associated with the "probing" of media in terms of the optical characteristics of the emerging radiation, a class of "instrumental" inverse problems arising from the attempt to eliminate the influence of the measuring devices in optical and x-ray spectroscopy, instrumental optics, and radio astronomy,[5, 10, 11, 19-30, 34, 35] and so on. Even this far from complete list shows that incorrectly posed problems are deeply rooted in physics, and also, what is much more important, that the further development of the theory and technique of modern experimentation is impossible without a clear understanding

*This terminology evidently is due to A. N. Tikhonov.

of the mathematical nature of these inverse problems.

An important source of increased interest in incorrectly posed problems of measurement is due to the development of computational techniques, in particular electronic computers, which make possible the handling of large volumes of numerical material. There is often a tendency to exaggerate the possibilities of such processing, resulting in excess optimism regarding methods of measurement based on the solution of incorrectly posed inversed problems.* On the other hand, difficulties, sometimes only apparent ones,† in the solution of inverse problems can hinder the development and use of extremely useful methods of measurement. Therefore the exposition, and some degree of generalization, of presently available experience in the solution of incorrectly posed inverse problems seems very timely.

An important feature of problems of measurement, which are essentially all we shall be considering in the present review, is the stochastic nature of the quantities observed in an actual experiment. In the simplest cases one speaks of "random disturbances" or "noise" in the measuring apparatus, distorting the "useful signal." In more complicated cases this signal itself is random, as for example in measuring the intensity of a flux of quanta.

This stochastic property is an inevitable feature of every actual experiment, and naturally must appear explicitly in the formulation of the inverse problem. Statistical approaches and methods of solution of incorrectly posed inverse problems are therefore a direct consequence of the stochastic character of the experiment.

The purpose of this review is to elucidate recently developed methods for the solution of incorrectly posed inverse problems which depend essentially on mathematical statistics and information theory. Before proceeding to this task, we shall discuss in more detail the concept of "incorrectness," and shall give a cursory survey of the usual (nonstatistical) methods of solution of incorrectly posed problems, and of the difficulties that arise in them.

I. INCORRECTLY POSED INVERSE PROBLEMS AND DIFFICULTIES IN THEIR SOLUTION

The "Incorrect" Aspect of an Inverse Problem

A typical inverse problem, and one often encountered in mathematical physics, is the solution of the Fredholm integral equation of the first kind:

$$\int_a^b K(x, y)\,\varphi(x)\,dx = f(y), \qquad c \leqslant y \leqslant d, \tag{1}$$

where K(x, y), the kernel of the equation, determines the operator $\hat{K}$ of the direct problem, which converts the unknown function $\varphi(x)$, describing the state of the object of the measurement, into some other function f(y) which is accessible to direct registration and can therefore be regarded as known. Sometimes Eq. (1) takes the special form of the "convolution"

$$\int_{-\infty}^{+\infty} K(y-x)\,\varphi(x)\,dx = f(y), \tag{2}$$

if the kernel depends only on the difference of the arguments. We shall consider both Eq. (1) and Eq. (2), the latter because of the simplicity of the results relating to it.

The solution of Eq. (1) may not exist at all, or may not exist for every function f(y) on the right side. For example,* if

$$K(y, x) = y + x, \tag{3}$$

the right member f(y) must be a linear function of y. If a solution does exist, it need not be unique. In the case of our example, if $f(y) = f_0 + f_1 y$, any function $\varphi(x)$ that satisfies the two conditions,

$$\int_a^b \varphi(x)\,dx = f_1, \quad \int_a^b x\varphi(x)\,dx = f_0,$$

is a solution of Eq. (1) with the kernel (3).

In problems arising in physics we are usually sure of the existence of the function $\varphi(x)$ that appears in the integrand in the left member of Eq. (1). Its uniqueness can also often be guaranteed.†

However, even if the solution exists and is unique, Eq. (2) or Eq. (2) can have a specific peculiarity which makes the problem an incorrectly posed one. This peculiarity is the "smoothing" action of the kernel. We shall explain this with an example.

Let the kernel K(y, x) be a continuous function of its second argument. Let us consider as solutions two functions $\varphi(x)$: any function $\varphi_1(x)$ and the function $\varphi_2(x) + C \sin 2\pi nx/(b-a)$, where n is a sufficiently large integer. It is clear without further calculation that for an arbitrarily large value of C we can choose a value of n so large that the difference of the corresponding right members $f_1 = \hat{K}\varphi_1$ and $f_2 = \hat{K}\varphi_2$ (for a given value of the argument y) will be less in absolute value than any previously given (arbitrarily small) number ϵ, i.e., a kernel of the type in question "smooths out" even a very intense, but high-frequency component to an extremely small level (the smaller, the higher its frequency). If we knew the right member f(y) exactly, there would be no great harm in this, the uniqueness of the solution being guaranteed. But the presence of disturbances accompanying the registration of the function f(y) alters the situation catastrophically. In fact, suppose that under the experimental conditions we can check the agreement of the registered function f*(y) with the exact transform $f(y) = \hat{K}\varphi$ of the unknown function $\varphi(x)$ only to within an error e:

$$\max_{c \leqslant y \leqslant d} |f^*(y) - f(y)| \leqslant e. \tag{4}$$

It is easy to see that if the kernel of Eq. (1) has the smoothing action we have described, then we can always find two functions $\varphi_1(x)$ and $\varphi_2(x) = \varphi_1(x) + C \sin 2\pi nx/(b-a)$ whose transforms $f_1 = \hat{K}\varphi_1$ and $f_2 = \hat{K}\varphi_2$ both satisfy the inequality (4) with the same function f*(y). Accordingly, there are at least two dif-

*This is evidently the case with some inverse problems of satellite meteorology (cf., e.g., [15]).

†For example, the lack of an anlytic solution of an integral equation of the form (1).

*This example is taken from [38].

†For example, for the inverse problem of potential theory, [39] the inverse problem of spectroscopy and instrumental optics, [40] and a number of other problems.

ferent functions that satisfy Eq. (1) with the right member f*(y) taken from experiment to within the error e, and speaking more exactly there is an infinite set of such functions, among which there are specimens differing from each other by as much as we please. It can therefore be seen that having undertaken to solve Eq. (1) with the approximate right member f*(y) we have no sort of chance to get the true solution $\varphi(x)$ that characterizes the actual state of the object of the investigation, and we shall almost certainly arrive at some false solution containing indefinitely large rapidly oscillating components.

It is in this situation that the "incorrectness" of the problem (1) actually lies. The exact definition[39, 41, 42, 56] only enables us to dispense with the special form of the condition (4) and allows use of more general measures of the deviation.

The estimate of the difference between functions by the quantity (4), which is characteristic of the classical analysis, is inconvenient for applications, since the registered function f*(y) is usually a realization of some random process and subject only to probabilistic restrictions, so that we either can indicate its maximum deviation from the exact value "with a large safety factor" or else cannot indicate it at all. Therefore it is more suitable to use a statistical approach to the definition of correctness,[43] which means studying the statistical properties of the problem (1) in their dependence on the characteristics of the random process of which the function f*(y) registered in the experiment is a realization.

The peculiarities of the statistical approach may be conveniently demonstrated with an example from instrumental spectroscopy—the reduction of the spectrum recorded by a device with a finite resolution to the "ideal instrument" (with infinitely high resolution).[22] The corresponding inverse problem can be formulated in the form of Eq. (2),* supplemented by the assumption that the registered function f*(y) is the sum of the exact right member f(y) and a stationary random process $\delta(y)$ (the noise) with zero mean and correlation function $\Delta(\eta) = \langle \delta(y)\,\delta(y+\eta)\rangle$. The formal solution of Eq. (2) is easily obtained by means of the Fourier transformation

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{ipx}\frac{\tilde{f}(p)}{\tilde{K}(p)}\,dp, \qquad (5)$$

where the sign ~ denotes the Fourier transform of the corresponding function. Let us find the dispersion of the function $\varphi(x)$ when instead of f we put $f^* = f + \delta$ in the right member of Eq. (5). As Rautian has shown,[22] the expression for this is

$$D(\varphi) = \langle\varphi^2\rangle - \langle\varphi\rangle^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{G(p)}{|\tilde{K}(p)|^2}\,dp, \qquad (6)$$

where $G(p) = \tilde{\Delta}(p)$ is the power spectrum of the process $\delta(y)$. We now note that the Fourier transform $\tilde{K}(p)$ of the apparatus function (the transfer function) has the property that

$$|\tilde{K}(p)| \to 0 \quad \text{for} \quad |p| \to \infty. \qquad (7)$$

Therefore in order that the dispersion of the solution of the problem (2) remain finite the power spectrum G(p) of the noise must fall off sufficiently rapidly for $|p| \to \infty$. This imposes severe restrictions on the class of processes $\delta(y)$ that are admissible as noise. In practice these conditions are never satisfied, since the noise always contains a "white noise" component and consequently for $|p| \to \infty$ the spectrum G(p) approaches a finite limit. Therefore the dispersion of the solution as found from (6) is infinite, and consequently it is impossible to substitute the experimentally found function f*(p) for f(p) in Eq. (5). The source of this difficulty is obvious: the high-frequency components of f*(y), which arise from the presence of noise and which must not be present in the true function f(y), are divided in Eq. (5) by small eigenvalues k(p) and produce large (and for $|p| \to \infty$ infinitely large) oscillations in the solution.

It is easy to see that the condition (7) is equivalent to the classically defined incorrectness of the problem (2). It is useful to examine what the situation must be for the problem (2) to be a correctly posed one. For this the condition (7) must be replaced by its opposite: $|\tilde{K}(p)| > \epsilon > 0$ for $|p| \to \infty$; and the finiteness of the dispersion $D(\varphi)$ is then assured for a very much broader class of processes $\delta(y)$, it being sufficient that the dispersion of the process $\delta(y)$ itself be finite. According to the statistical approach[43] this is indeed the difference between incorrectly and correctly posed problems; the solution of a correctly posed problem is statistically stable with respect to a wider class of random processes than the solution of an incorrectly posed problem. In particular this is the explanation of the fact that methods for the solution of correctly posed problems can be developed without special consideration of statistical noise—the noise can always be taken into account at the final stage of the calculation as a perturbation of the exact solution.* This sort of method is no good for incorrectly posed problems. The point is that actual noise, as a rule, is not "admissible" for them, and the dispersion of the exact solution is in actual fact infinite, i.e., a formally exact solution of the type of (5) simply has no meaning.

The Algebraization of Incorrectly Posed Problems and the Question of Determinacy

In the numerical solution of an integral equation of the type (1) one always reduces it in one way or another to a system of linear algebraic equations

$$\sum_{i=1}^{n} K_{ji}\varphi_i = f_j, \qquad j = 1, 2, \ldots, n, \qquad (8)$$

where φ_i and φ_j are linear functionals of the functions $\varphi(x)$ and f(y), namely either their values at certain support points or the coefficients in their expansions in terms of a system of orthogonal functions.

It might seem that the reduction of Eq. (1) to a system of algebraic equations (8) exhausts the problem of solving it—we need only take the order of approximation n large enough to obtain any desired accuracy. When, however, we remember that the original problem (1) is incorrectly posed, we can expect that some obstacle will arise in the way of this procedure. The obstacle is

*In this case the kernel of Eq. (2) is known by the special name of the "apparatus function."

*An example of a correctly posed problem is the solution of the Fredholm equation of the second kind, which differs from Eq. (1) by having the unknown function present in the left member as an additional term.

the poor determinacy of the system of equations (8), i.e., the extraordinarily strong dependence of the solutions on variations of the inhomogeneous term, and also on errors in the coefficients in (8) and on computing errors.

As Fadeev[44, 45] has shown, the determinacy of the system (8) is closely connected with the set of eigenvalues μ_K of the matrix K*K [K is the matrix of the system (8), and K* is the transposed matrix].

The determinacy (stability) decreases with increase of the ratio μ_{max}/μ_{min}. The solution "shifts about" in the directions of the eigenvectors ψ_K of the matrix K*K that correspond to the smaller eigenvalues μ_K. More exactly, the sensitivity of the projection of the solution vector φ in the direction of ψ_K to variations of the components of the vector **f** and of the elements of the matrix K is proportional to μ_K^{-1}. Therefore, in particular, the demands on computational accuracy increase rapidly with the ratio μ_{max}/μ_{min} even for an exactly known vector **f**.

The algebraization of an incorrectly posed problem always (for a sufficiently large order of approximation n) gives a system of equations with poor determinacy, for it can be shown that if the original problem is incorrectly posed then μ_{max}/μ_{min} when $n \to \infty$. Consequently, by choosing the order of approximation large enough one can make the system of equations (8) arbitrarily poorly determinate.

The Approximate Solution of Incorrectly Posed Inverse Problems

Approaches to the solution of inverse problems have changed with the realization and changing understanding of their "incorrect" nature. In this section we list some of these approaches, adhering as far as possible to the chronological order.

We first note that incorrectly posed inverse problems can in a certain sense be "solved," and indeed very successfully, not only without the use of the mathematical methods especially developed for the purpose in very recent times, but often without any suspicion of the incorrectness and the problems associated with it. The point is to extract physically reliable information about particular characteristics of the exact solution φ (or, in physical terms, the state of the object of the investigation) without seeking the solution itself; in such cases the form of the exact solution is either known (except for parameters which are to be determined) or else is unimportant. We shall give an example. Our confidence in the validity of one of the most important physical theories—quantum mechanics—is to a large degree based on the interpretation of data from the experimental spectroscopy of atoms and molecules, i.e., essentially on the solution of the inverse problem (2). But does it follow from the mathematical "incorrectness" of the problem (2) that quantum mechanics is incorrect as a physical theory? Although it is not hard to guess the answer, it may be useful to go through the argument.* The point is that the experimental foundation of classical quantum mechanics was the aggregate of experimental data on the positions and intensities of the lines in the emission spectra of atoms and simple molecules. However, as is well known [in the general case it follows from the form of Eq. (2)], the intensity of an isolated spectral line can be found from the instrumental spectrum f independently of the form of the true spectrum φ (the line shape) and even of the kernel of Eq. (2) [about this one needs to know only the normalization factor $\tilde{K}(0)$, the sensitivity of the device]. The same is true, with a further assumption that the line shape and the kernel $\tilde{K}$ are symmetric, of the determination of the frequency of a spectral line (the frequency of the transition). Furthermore, an analysis shows (see, e.g., [22]) that both of these characteristics are determined with a finite error under natural assumptions about the character of the noise, so that the incorrectness of the problem (2) does not affect the reliability of the results, if one uses ordinary precautions which are well known from practical work with spectroscopic devices. Other, by no means so trivial, examples of this sort of situation are provided by the spectroscopy of light-scattering substances. Here considerable analytic effort[51, 52] is often required to select an experimental arrangement in which the results of the direct measurements will on one hand permit the determination of the optical parameters of the substance in question, and on the other hand be independent of the exact form (often unknown) of the structure of the light-scattering medium, which, together with the measuring device, determines the kernel of an in general incorrectly posed inverse problem relating to the absorption spectrum of the substance.*

A direct possibility for evading the difficulties caused by the incorrect formulation of the problem is sometimes provided by a suitable parametrization of the solution φ, based on the concrete physical nature of the problem. For example, in our example of an isolated spectral line we may know (if the spectrum is taken with the substance in the gaseous phase) the line shape—assumed in the Doppler or the Lorentz form, depending on the experimental conditions. Then from the functional equation (2) we can derive the algebraic equation for the only unknown parameter, the line width,† and this equation can be solved, for example, by the method of least squares. In the general case such arguments lead to a system of simultaneous equations, which can be solved in the usual way (provided it is not underdetermined). This kind of procedure is often used in processing experimental data.

As for more general methods of solution, essentially the first attempt to solve the incorrectly posed problem (2) was made by Lord Rayleigh in 1871, when he proposed an iteration method for correcting slit distortions in spectroscopy. Since then such methods have been proposed repeatedly (see the review [22]). As Rautian has shown,[22] they are all equivalent to some iteration method for solving Eq. (2). The point is that in spec-

*This argument of course does not exhaust the question of the foundation of quantum mechanics itself.

*The question of the permissible uncertainty of the kernel of Eq. (1) is an important and still unsolved problem. [51,52] All existing methods for solving Eq. (1) require an exact knowledge of the kernel (with regard to the sensitivity of the solution to errors in the kernel see below).

†The intensity and the position, as indicated above, can be determined independently.

troscopy the kernel of Eq. (2) is essentially a very narrow pulse; i.e., the corresponding operator $\hat{K}$, when applied, to a sufficiently smooth function, does not differ much from the unit operator $\hat{I}$ (with appropriate normalization). Therefore for the solution of (2) we can try to use the formal expansion of the inverse operator of the problem in Neumann series:

$$\hat{K}^{-1} = \sum_{n=0}^{\infty} (\hat{I} - \hat{K})^n, \tag{8a}$$

i.e., to find the corrections to the zeroth approximation $\varphi_0 = f$ by successive applications of the operator $\hat{I} - \hat{K}$. The initial corrections are indeed relatively small, if we start with a sufficiently smooth function f, but later they increase rapidly, showing more and more rapid oscillations, this growth being due to the presence of noise in the right member of (2).* Therefore in the application of iteration algorithms the research worker must himself decide when to break off the iteration process, being guided by some sort of ideas about the genuine or noise origin of the details of the solution that appear after each new iteration.

Another general method for solving a problem of the type of Eqs. (1), (2) is simply to solve the corresponding algebraic system (8). If there has been good luck with the algebraization, one can sometimes get an acceptable solution by working to a small order of approximation n, at which the instability does not yet show up. The algebraization is most often done by expanding $\varphi(x)$ in terms of a system of orthogonal functions (it is desirable that they have a physical meaning) and keeping the first n terms of the expansion. This method, along with the iteration method, was widely applied in practice† in the time when owing to technical difficulties of computation it was indeed not possible to use high orders of approximation, so that the very fact that a problem of the type (1) was incorrectly posed could remain unknown to those doing the calculations. But with the development of computing techniques, when it became possible to deal with systems of large dimensionalities (or high-order iterations), the incorrectness began to show up in the form of "shifting about" of the solution with increases of the order of approximation. The original procedure in this case was to adjust the degree of approximation according to the character of the solution obtained, stopping at an approximation in which the "shifting about" (as estimated by comparing the solution with the form expected for the unknown function) was still not too large. It is convenient to examine this approach with the example of Eq. (2), for which there is a simple possibility of controlling the degree of approximation by using the formal solution (5).

Confining ourselves to the reconstruction of the signal spectrum $\tilde{\varphi}(p)$ in a finite frequency band $|p| \leq p_0$, we rewrite (5) in the form

$$\varphi^*(x) = \int_{-p_0}^{+p_0} e^{ipx} \frac{\tilde{f}(p)}{\tilde{K}(p)}\, dp, \tag{5a}$$

which gives an approximate solution of Eq. (2). The dispersion of the approximate solution φ^* is then determined by a "cut off" form of the integral (6) and remains finite if p_0 is not too large. By changing p_0 one can change the relation between the "accuracy of the approximation" of the unknown function and the size of the error; here an increase of the "accuracy" (owing to an increase of p_0) leads automatically to an increase of the error.[22]

Various considerations have been used to choose the cut-off frequency p_0: exclusion of zeroes of the transfer function $\tilde{K}(p)$,[21] the expected form of the unknown function,[22, 34] or the form of the registered function.[25]*

There is one other, somewhat more general way of adjusting the degree of approximation in the use of the solution (5). This is to multiply the integrand in (5) by some function $\tilde{g}(p)$, which generally speaking is arbitrary, and which falls off rapidly enough at $|p| \to \infty$ so that the corresponding integral (6) will remain finite:

$$\varphi^*(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{g}(p) \frac{\tilde{f}(p)}{\tilde{K}(p)} e^{ipx}\, dp. \tag{5b}$$

We note that (5b) goes over into (5a) for a special choice of the factor $\tilde{g}$. It is not hard to see that the use of (5b) is equivalent to looking for a "smoothed version" of the true solution,

$$\varphi^*(x) = \int_{-\infty}^{+\infty} g(x - x')\, \varphi(x')\, dx',$$

where g(x) is obtained from $\tilde{g}(p)$ by the inverse Fourier transformation. The degree of this "smoothing" can be adjusted by changing the parameters of the function $\tilde{g}(p)$ and even the very form of this function. But such an extreme freedom in acting on the solution of Eq. (2) must already make us cautious: it is clear that by a subjective choice of the "smoothing factor," and in general of the method for approximate solution, it is extremely easy to distort the result. That this is so is shown by examples of ambiguous interpretations of radioastronomical observations, given in [21].

The effort to dispense with arbitrary factors and to develop general methods for the solution of incorrectly posed problems, along with mathematical researches on the nature of their incorrectness, has led in the last decade to new approaches and methods for the solution of incorrectly posed problems.

The following considerations are of fundamental importance for these approaches. An incorrectly posed problem can be regarded as effectively not fully defined. In fact, with the classical concept of incorrectness, a solution of the problem (1) can be any function φ that satisfies the condition

$$r_e(f^*, \hat{K}\varphi) \leqslant e, \tag{9}$$

where $r_e(f^*, f)$ is a measure of the deviation of the registered function f* from the exact right member, which depends on the conditions of the experiment. Since among these functions there are "bad" ones, the prob-

*The series (8a) converges only if the right member f coincides with the exact image $\hat{K}\varphi$ of the true solution.

†It is neither possible nor necessary to list the papers on this matter.

The paper [25] describes a program of numerical Fourier transformation in which one excludes from the integral (5) those frequency ranges in which the spectrum of the realization f(p) seems to be controlled by the noise alone [$|f^*(p)| < kG(p)$, where k is a safety factor].

lem (9) so as to get a unique solution as close as possible to the true one.

Generally speaking one can complete the definition of the problem (9) in various ways. However, any method of completing the definition must be based on some sort of ideas about the nature of the desired solution, or in other words, on a priori information about the solution. Different methods for solving incorrectly posed problems, including the statistical methods to be considered here, differ explicitly or implicitly in the form of the a priori information that is used.

The first work in this direction was a paper by Phillips,[53] in which it was suggested that from the set of functions that satisfy the condition (9) one should choose the "smoothest" function, or more exactly the function that minimizes the norm of the derivative,

$$\int_a^b \left(\frac{d\varphi}{dx}\right)^2 dx = \min. \tag{10}$$

One can usually show that the desired minimum is attained on the boundary of the region defined by the inequality (9); consequently we can replace (9) by the equation

$$r_e(f^*, \hat{K}\varphi) = e. \tag{9a}$$

We now have a problem of a conditional extremum. Solving it by the Lagrange method, we get the equation

$$r_e(f^*, \hat{K}\varphi) + \alpha \int_a^b \left(\frac{d\varphi}{dx}\right)^2 dx = \min, \tag{11}$$

where α is an undetermined multiplier, which can be determined from (9a). For a Euclidean metric the condition (11) leads after algebraization to a system of linear equations. In applying this method in practice, Phillips noted that when the parameter α is determined from (9a) the function obtained is excessively smoothed. One can, however, choose a value of α (in each concrete case) such that the solution of (11) gives a solution much closer to the true one. Accordingly, in the Phillips method the parameter α is in actual fact undetermined.

For the algebraized system (8) Twomey[54] considered a somewhat more general form of the condition (10):

$$\Omega(\varphi) = \sum_{i,j}^{n} H_{ij}\varphi_i\varphi_j = \min, \tag{8a}$$

where the matrix H is positive definite. Otherwise Twomey's approach is the same as that of Phillips. The condition (11) takes the form

$$K^*K\varphi + \alpha H\varphi = K^*\mathbf{f}, \tag{11a}$$

and the Lagrangian multiplier is determined from the relation

$$|K\varphi - f| = e.$$

Another way of completing the definition of the problem (9) (see the bibliography in [42]: mathematical questions relating to this are also investigated there) is in a certain sense complementary to that just described. We shall examine it for the example of Eq. (8). We give the name of a quasisolution (the terminology is from [42]) of the system (8) to a vector φ which satisfies the two conditions:

$$\Omega(\varphi) = \sum_{i,j} H_{ij}\varphi_i\varphi_j \leqslant C, \tag{12}$$

$$|K\varphi - f| \to \min, \tag{13}$$

where the matrix H is positive definite, as before. If it is known that the minimum is reached on the boundary of the region defined by the inequality (12), the problem of Eqs. (12) and (13) reduces to Eq. (11a), but the Lagrangian multiplier will be determined from the relation (12), which is to be used for this with the equals sign.

The condition (12) singles out a certain bounded closed region as the set of admissible solutions, and is equivalent to the presence of a priori information that the required solution belongs to a given closed manifold. In [55], weaker restrictions are used, which are taken from the physical meaning of the solution of problem (1) as a particle-size distribution (see [46])—namely, this density is not negative. In this case the condition (12) is replaced by

$$\sum_{i=1}^{n} H_{ij}\varphi_i \geqslant 0 \quad \text{for all } j, \tag{12a}$$

and the problem is solved by the methods of linear programming.

A. N. Tikhonov[56-58] has introduced the concept of regularization of the solution of an incorrectly posed problem. This is taken to mean the construction of a family of correctly posed problems depending on a regularization parameter α, which has the property that for $\alpha \to 0$ and when the errors of the right member also simultaneously approach zero the solution of the correctly posed problem approaches the true solution of the incorrectly posed problem. Equation (11) was postulated and studied by Tikhonov, independently of Phillips, as a regularized equation with the regularization parameter α. In its practical application the Tikhonov method is identical with the Phillips method, perhaps with the difference that the indefiniteness of the parameter α follows naturally from the very idea of regularization, whereas in Phillips' approach it is a bit of an embarrassment, to a certain extent discrediting the method (on this see below). It must be remarked, by the way, that in a number of papers elaborating Tikhonov's idea, algorithms have been proposed for the determination of α, but they can scarcely be regarded as having an real foundation, since the source or even the exact form of the a priori information remains an open question.

As has already been pointed out, these methods for the solution of incorrectly posed problems start from the assumption that experiment allows the setting of an exact upper limit on the error $|f^*(y) - f(y)|$. A consistent use of this assumption leads in practice to excessively smoothed solutions; this is evidently due to the fact that the actual (random) error is usually smaller than its maximum value. Moreover, the assumption does not correspond to the nature of an actual experiment, and does not permit an accurate estimate of the error of the reconstruction. Therefore it is more natural to consider the problem of solving an incorrectly posed problem by taking into account the statistical nature of the experimental errors and by using other types of a priori information, including statistical information.

II. THE INFORMATION OBTAINABLE FROM EXPERIMENT, AND THE SOLUTION OF THE INVERSE PROBLEM

Resolving Power and Informational Metric

It is natural to begin an analysis of the "informing possibilities" of an experiment with the simplest problem in which these possibilities appear at all, namely with the problem of distinguishing two closely spaced states of the object being measured. The formulation of this sort of problem in instrumental optics and spectroscopy in order to estimate the limiting possibilities of optical devices—the problem of resolving power—stems from Rayleigh (for a survey of further papers see [22]). The basis of Rayleigh's approach is the idea of comparing (by means of a given optical or spectroscopic device) two standard "objects"—for example, in the case of a spectroscope, a spectral doublet and a single line. The corresponding statistical approach is the formulation of the problem in terms of the theory of statistical decisions.[31]*

Let $\varphi_1(x)$ and $\varphi_2(x)$ be two functions describing fixed states of an object of measurement, and let the object be in one of these states. As the result of an experiment on the object we get an observed function $f^*(y) = f(y) + \delta(y)$, where $f(y)$ is the function associated with the true state $\varphi(x)$ in Eq. (1), and $\delta(y)$ is a random function (noise). Following [31], we shall regard $\delta(y)$ as a normal stationary random process with zero mean and power spectrum $G(p)$. The problem is, by observing $f^*(y)$, to decide in which of the states, φ_1 or φ_2, the object actually is. We shall characterize the distinguishability of the states by the maximum probability P_r of correct decisions under the condition that the states being compared are a priori equally probable and that the optimal decision procedure is used. The quantity P_r takes values between $1/2$ and 1; the lower limit corresponds to absolutely indistinguishable states, and the upper to absolutely distinguishable states. It can be shown that P_r depends on the data of the problem [the kernel $K(y, x)$, the functions φ_1 and φ_2, and the statistical noise] only through a quantity ρ, which indeed directly characterizes the degree of distinguishability of the states:

$$P_r = 1/2 + 1/2\Phi(\rho/2), \tag{14}$$

$$\rho^2 \equiv \|\varphi_1 - \varphi_2\|^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{|\widetilde{\hat{K}(\varphi_1 - \varphi_2)}|^2}{G(p)}\,dp, \tag{15}$$

where $\Phi(x)$ is the probability integral, and the sign $\sim$ denotes the Fourier transform of the expression whose squared absolute value is indicated. The expression (15) has been derived on the assumption that near the edges of the range of observation (c, d) the function $\hat{K}\varphi$ goes to zero for arbitrary φ and that the correlation range of the noise is much smaller than the interval (c, d).[36] The quantity ρ defined by Eq. (15) as a measure of the distinguishability of the states φ_1 and φ_2, by means of the "instrument" described by the operator $\hat{K}$ and the noise power spectrum $G(p)$, has extremely important properties. First, the case $\rho(\varphi_1 - \varphi_2) = 0$ corresponds to complete indistinguishability (equivalence) of the states in the given experiment. The second limiting case $\rho = \infty$ would correspond to "infinitely large" (absolute) distinguishability, which ordinarily does not occur in practice. It is convenient to take $\rho = 1$ as a threshold value defining the limit of distinguishability.* It is natural to call the quantity ρ the informational distance in the state space of the object under study. In fact, it can be shown that ρ has all of the properties of a geometrical distance—nonnegativity, symmetry with respect to the states being compared, and validity of the "triangle inequality": $\rho(\varphi_1 - \varphi_2) \leq \rho(\varphi_1 - \varphi) + \rho(\varphi - \varphi_2)$. The distance ρ characterizes the information contained in the experimental results about the states φ_1 and φ_2 from the point of view of their distinguishability.†

The Uncertainty of the Solution of the Inverse Problem

The results of the preceding section show that for each "true'solution" φ_0 of the problem (1), characterizing the actual state of the object of measurement, there exist infinitely many functions corresponding to states that do not differ from the actual one in the results of the real (statistical) experiment—namely states that satisfy the inequality

$$\|\varphi - \varphi_0\| \leqslant 1. \tag{16}$$

In other words, the functions φ fill a sphere of unit radius in the function space of the solutions of the problem (1), which it is natural to call the "sphere of uncertainty." In fact, any one of the functions φ satisfying the inequality (6) can be accepted as a solution of the inverse problem (1), since the experiment does not give enough information to distinguish this function from any other one that also satisfies the inequality (16).‡ In the sense of the metric (15) the sphere (16) characterizes the maximum accuracy which can be attained in solving the problem (1). Therefore, in particular, every attempt to "refine" the solution further by purely mathematical means, without bringing in additional information, is analogous to an attempt to devise an "informational perpetual motion" which produces information out of nothing.

However, the practical value of a method for solving Eq. (1) that leads to an arbitrary function in the sphere (16) would be small: if the original problem is incorrectly stated this sphere contains for the most part functions φ to which there do not correspond any sort of physical states of the object of measurement. To single out the "physical part" of the sphere of uncertainty it is necessary to indicate some kind of features that distinguish the "physical" solutions from the "nonphysical" ones, or, as one usually says, we need a pri-

*In the theory of optical devices an analogous approach has been used independently in [32] and [33].

*Setting $\rho = 1$ in (14), we find that in this case the probability of a correct identification of the state φ is 13.7 percent larger than the probability of guessing correctly (50 percent), and that it falls off rapidly if ρ is decreased.

†For a quantitative approach to the informational volume of a manifold in signal space see [62].

‡Strictly speaking, to make this a completely precise statement, one must replace 1 by ½ in the right member of (16), which changes nothing in principle. For what follows it is more convenient to keep the value unity.

ori information about the possible states of the object of measurement or, what is the same thing, about the functions that are admissible as solutions of Eq. (1). Leaving the more detailed discussion of questions concerning the use of a priori information in the solution of inverse problems to the next section, let us examine the case in which we have no a priori information of any kind.[37]

It is clear, of course, that in this case it makes no sense to solve Eq. (1) directly by any method. Sometimes, however, what is of interest is not so much the actual solution of Eq. (1), but rather a knowledge of certain functionals of this solution, for example the value of the total intensity of a spectral line without regard to its shape. A way to analyze such problems by the use of the metric (15) has been proposed in [37] for application to a class of linear functionals of the desired solution.

Following [37], let us confine ourselves to the case of the difference kernel (2), for which all of the results can be written especially simply. In this case Eq. (15) can be rewritten in the form

$$\rho^2 = \| \varphi_1 - \varphi_2 \|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{| \tilde{K}(p) |^2 | \tilde{\varphi}_1(p) - \tilde{\varphi}_2(p) |^2}{\tilde{G}(p)} dp, \qquad (15a)$$

where, as before, $\tilde{K}(p)$ is the Fourier transform of the kernel K(x). Let us estimate the limiting accuracy of the determination of the linear functional

$$A = A_\varphi = \int_{-\infty}^{+\infty} a(x) \varphi(x) dx, \qquad (17)$$

where a(x) is a weight function defining the functional. Let us locate the sphere of uncertainty at the origin of coordinates by setting $\varphi_0(x) \equiv 0$ in the condition (16). Substituting the functions $\varphi(x)$ contained in this sphere in Eq. (17), we get values of the quantity A which, according to the definition of the sphere of uncertainty, are practically to be identified with the value $A_{\varphi_0} = 0$ corresponding to the center of the sphere. In other words, if we denote by $\|A\|$ the maximum value of A on the sphere (16), then all values of A satisfying the inequality $A \leq \|A\|$ cannot be distinguished from zero by means of the given apparatus, i.e., $\|A\|$ is the error in the determination of A from the experimental data.* Since the distance (15a) depends only on the difference of the functions to be compared, the sphere (16) is displaced as a whole when its center is changed. From this (and from the linearity of the functional A) it follows that the error is independent of the true spectrum.

It can further be shown that the quantity $\|A\|$ can be calculated from the formula

$$\| A \|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{| \tilde{a}(p) |^2 G(p)}{| \tilde{K}(p) |^2} dp, \qquad (18)$$

where $\tilde{a}(p)$ is the Fourier transform of the weight function of the functional. The quantity $\|A\|^2$ in fact is the dispersion of the error caused in the calculation of the functional A by the noise in the experiment.

We must call attention to the fact that the right member of Eq. (18) remains finite only for a rather narrow class of functionals. If the integral (18) diverges for some functional, this means that it is impossible to determine this functional in the given experiment without additional information about the unknown function $\varphi(x)$. The structure of the expression (10) shows that the class of admissible functionals ($\|A\| < \infty$) consists of functionals with weight functions a(x) that are essentially smoother than the kernel of Eq. (2), provided that the noise power spectrum G(p) does not fall off pathologically rapidly for $|p| \to \infty$. In particular, values of the function φ at specified points are certainly not admissible functionals.

III. RIGID A PRIORI RESTRICTIONS

For greater clarity we shall consider the algebraized version of the inverse problem, Eq. (8), assuming that the order n is sufficiently large to assure the required accuracy of approximation.

To impose rigid a priori restrictions on the unknown function $\varphi(x)$ (or, in the algebraized version, on the vector $\boldsymbol{\varphi}$ in n-dimensional vector space R^n) means to single out some region D in the space R^n and look for solutions only in this region, declaring that all solutions outside the region D are meaningless, "unphysical." It is easy to see that the question of the informational uncertainty of the problem (8) reduces to the question of the relative positions of the region D and the sphere of uncertainty (16), or, in other words, to a question about the geometry and dimensions of the region D in the metric (15).

The system (8) must be regarded as informationally undefined if the sphere (16) with center at any point $\varphi_0 \in D$ contains points not belonging to the region D. In fact, in those directions in which the sphere (16) "extends beyond" the region D the experiment does not make it possible to improve the accuracy of a state parameter in comparison with the a priori range. Moreover, a formal solution of the system (8), for example by the method of least squares, will give an error of the coordinate in such a direction which is beyond this range by a large factor. This means, essentially, that the experimental data do not contain any information about the corresponding parameter of state. The presence of such parameters "unguaranteed" by information among the unknowns of the system (8) is the meaning of the definition given above. The concept of an informationally incompletely undefined system can be more conveniently expressed in terms of the geometrical characteristics of the region D in the metric (8). Namely, we shall call the region D degenerate in one or more directions if the maximum diameter of the region D in these directions does not exceed unity. Then, obviously: 1) degeneracy of the region D is equivalent to incompleteness of definition of the system (8), and 2) the system (8) (corresponding to the experiment) does not allow us to determine the parameters corresponding to the directions of the degeneracy.

The geometrical formulation shows particularly clearly that the property of incomplete definition does not depend on the choice of the basis for the algebraization of the original problem (1),* i.e., on the choice of the coordinates in the space R^n. This property is due to the experiment itself and the region D.

*The characteristics of the apparatus and of the noise enter through the definition of the metric, Eq. (15a).

*Provided that the basis allows us to describe all points of the region D, i.e., is of large enough dimensionality.

Let us now turn to the problem of determining the "physical" solution of the system (8). It is clear that to do this we must first eliminate the degenerate directions, since they are the source of the "unphysical" nature of solutions. Let R^r be the maximal subspace of the space of solutions R^n that does not contain any directions of degeneracy of the region D—the allowed subspace. Then it is reasonable to take as the solution of the poorly determined system ($r < n$) the projection of the unknown vector φ on the subspace R^r.

For the case in which the region D is given by its second-order central moments, the inertia tensor C,

$$C^{ik} = \langle (\varphi^i - \langle \varphi^i \rangle)(\varphi^k - \langle \varphi^k \rangle) \rangle,$$

a method is proposed in [63] for successive "exhausting" of the principal directions, which leads to the actual construction of the allowed subspace of the problem. From a computational point of view the method reduces to the problem of the eigenvalues of the operator GC:

$$\sum_{i,k=1}^{n} g_{ik} C^{il} u^k = \lambda u^l, \tag{19}$$

where the g_{ik} are the elements of the Fisher informational matrix.[64] Let $\lambda_1 > \lambda_2 > \ldots \lambda_n$ be the eigenvalues of the problem (19), and let $u_1, u_2, \ldots u_n$ be the corresponding eigenvectors. It can be shown that each eigenvalue λ_k is numerically equal to the ratio of the mean square (the dispersion) of the a priori allowed variation of the component of the vector φ in the direction of u_k to the dispersion of the error in determining this component from the system (8). It is clear that for $\lambda_k < 1$ the corresponding direction u_k must be regarded as degenerate. Accordingly, the allowed subspace R^r must contain only directions u_k for which $\lambda_k > 1$, and all their linear combinations. In other words, the first eigenvectors $u_1, u_2, \ldots, u_k$ form the basis of the allowed subspace R^r, and the dimensionality of this subspace is equal to the number of eigenvalues λ_k that are larger than unity. The desired solution of the problem (1) can then be expressed in the form

$$\varphi^*(x) = \varphi_0(x) + \sum_{k=1}^{r} \varphi'^k \eta'_k(x), \tag{20}$$

where the basis $\{\eta'_k\}$ is constructed from the basis $\{\eta_i\}$ used in obtaining the system (8) by means of the eigenvectors:

$$\eta'_k(x) = \sum_{i=1}^{n} u_k^i \eta_i(x),$$

and $\varphi(x) = \sum_{i=1}^{n} \langle \varphi^i \rangle \eta_i(x)$ is the solution corresponding to the center of the region D. The explicit solution of the problem is given by formulas expressing the coefficients of the expansion (20) in terms of measured values

$$\varphi'^k = \sum_{i=1}^{n} \sum_{j=1}^{m} f_j k_{ji} u_k^i \Big/ \sum_{i,l=1}^{n} \sum_{j=1}^{m} k_{ji} k_{jl} u_k^i u_k^l. \tag{21}$$

The solution obtained from Eqs. (2) and (21) satisfies, as far as is possible, the requirements which it is natural to ask that a solution of an incorrectly posed problem with rigid a priori restrictions satisfy. The operation which makes $\mathbf{f} \to \varphi$ "almost" preserves the information about the difference between any vectors of R^n, since if the vectors to be compared lie in R^r the distance ρ between them is not diminished; if, on the other hand, they have components in the complementary subspace R exp $(n - r)$, then the decrease is of the order of $(\lambda_{r+1})^{1/2} < 1$—i.e., states are identified with each other which are practically indistinguishable in the given experiment. Moreover, the solution φ^*, a random quantity, almost certainly belongs to the "physical" region D, since its mathematical expectation, taken over the noise distribution, coincides with the actual solution, which belongs to D by definition, and the noise acting in nondegenerate directions cannot carry the solution significantly beyond the limits set by the a priori range.

IV. PROBABILISTIC RESTRICTIONS AND THE METHOD OF STATISTICAL REGULARIZATION

The method of statistical regularization was developed in [65-68]. Its main feature is that the a priori information about the unknown function is introduced by giving a probability distribution. This leads to the replacement of the exact solution of the equation by an approximate, "regularized" solution. There are various ways of prescribing an a priori probability distribution, and therefore there are also various versions of the method of statistical regularization; these will be briefly described. First, however, we shall state the theoretical propositions on which this method is based. Although as compared with the introduction of rigid restrictions on the unknown function the introduction of probabilistic restrictions in many cases requires more a priori information, it nevertheless has the following advantages.

First, the problem of solving an incorrectly posed equation occurs in applications as the problem of processing experimental data, for which the introduction of probabilistic concepts is inescapable, because the error in the right member is of a random nature and can be characterized only in a probabilistic way. Therefore the probabilistic way of giving a priori information leads to the use of a single kind of apparatus and is more natural. This will be seen especially clearly when the question of the error of the constructed solution is dealt with.

Second, the probabilistic method allows more complete use of previous experience, by including it in the a priori distribution.

Third, when there is no such experience, the probabilistic method allows the formulation of extremely weak assumptions about the unknown function, which in principle are not expressible in set-theory language (ascription to a "laminar ensemble," discussed below).

As in the preceding section, we assume that our equation has already been algebraized, i.e., reduced to a system of equations

$$\sum_{i=1}^{n} k_{ji} \varphi_i = f_j, \qquad j = 1, 2, \ldots, m. \tag{22}$$

In contrast with the equations (8), we here allow the number of equations m to be unequal to the number of unknowns n. The presence or absence of an exact solution of Eq. (22) will play no part here. The quantities φ_i can be any linear functionals of $\varphi(x)$, but for the method of statistical regularization it is most natural to take as the φ_i the values of $\varphi(x)$ at certain refer-

ence points, and this will be assumed in what follows for convenience in the exposition. We shall often refer to the vectors φ and $\mathbf{f}$ in the spaces R^n and R^m simply as the functions φ and f.

To explain the essence of the method of statistical regularization, it is necessary to formulate our problem as a problem of mathematical statistics and to introduce the appropriate concepts.

Strategies and Estimates

Suppose one measures m quantities f_j which, if there were no errors of measurements, would be connected with the n unknown quantities φ_i by the relations (22). Owing to the errors of measurement the quantities f_j found from the measurements are different from their ideal values given by Eq. (22). In order not to introduce new notations, we shall from now on use f_j to mean not the ideal, but the real values of the quantities f_j. Then Eq. (22) will be satisfied only approximately, and the exact equations will be of the form

$$\sum_{i=1}^{n} k_{ji}\varphi_i + \delta_j = f_j, \qquad j = 1, 2, \ldots, m,$$

where the δ_j are the errors in the measurements of the quantities f_j, which form a random vector $\boldsymbol{\delta}$ in the space R^m. We denote by $P_\delta(\boldsymbol{\delta})$ the probability density of this vector. Obviously the quantities f_j are also random variables and depend both on the true values φ_i and on the errors δ_j. The conditional probability density of the vector $\mathbf{f}$ for a given vector φ is

$$P(\mathbf{f} \mid \varphi) = P_\delta(\mathbf{f} - \hat{K}\varphi). \tag{23}$$

(We note that the function $P(\mathbf{f}|\varphi)$ of the m + n variables f_j and φ_i completely characterizes the experimental arrangement, i.e., both the connection between φ and $\mathbf{f}$ and the statistical properties of the errors of measurement.) We can now formulate the problem in the following way. The unknown quantities φ_i characterize a state of nature. Some random process gives us the quantities f_j, and the conditional probability $P(\mathbf{f}|\varphi)$ is known. From the quantities f_j, what can we say about the quantities φ_i? The answer to this sort of question is the fundamental problem of mathematical statistics.

The question "what can we say about the φ_i?" is insufficiently exact. When we try to make it more precise, the first formulation that presents itself is: How probable are various values of the quantities φ_i? In other words, what is the probability density $P(\varphi|\mathbf{f})$ of the vector φ under the condition that the measurements have given the result $\mathbf{f}$?

In order to answer this question, it is necessary to introduce some a priori probability density $P(\varphi)$—that is, to assume that the experiment of measuring $\mathbf{f}$ is one of a series of such experiments which are made with different states of nature φ, chosen randomly in accordance with a probability density $P(\varphi)$. Then the a posteriori probability can be determined from the Bayes formula:

$$P(\varphi \mid \mathbf{f}) = \frac{P(\varphi)\,P(\mathbf{f} \mid \varphi)}{\int P(\varphi)\,P(\mathbf{f} \mid \varphi)\,d\varphi}. \tag{24}$$

If no a priori probability is introduced the concept of the a posteriori probability $P(\varphi|\mathbf{f})$ loses all meaning.

If we do not demand such detailed information about the function as an a posteriori probability distribution for it, then we are not obliged to introduce an a priori probability distribution. However ("if not through the door, then through the window"), a priori distributions appear even in the most general formulation of the question, in which mathematical statistics is regarded as the theory of reaching decisions under conditions of incomplete knowledge of the state of nature. For what do we need any sort of information at all? Obviously, in order to make decisions about the performance of some action or other. If we knew the true state of nature φ, we would make a certain decision. But we do not know φ; we know only $\mathbf{f}$. Mathematical statistics must give us recommendations as to how to act in such cases; it must indicate a strategy for making decisions with unknown φ but known $\mathbf{f}$. In order to avoid discussing the specific peculiarities of each concrete problem, statistics must give this sort of recommendation: "Here is a vector φ^0 for you; act as if $\varphi = \varphi^0$." Such a value φ^0 is called an estimate of φ, and the algorithm for calculating φ^0 from $\mathbf{f}$ and its further use instead of the true φ is called estimative strategy.

By adopting a particular strategy, we may, depending on the case, get a better or a worse result; i.e., the estimate φ^0 may approximate more or less closely to the true φ. It is natural to raise the question of finding the strategy that will give the best result on the average. But in order to define the concept "on the average" it is necessary to introduce an a priori probability distribution for the unknown $P(\varphi)$. To calculate the Bayesian estimate of any quantity, in particular the estimate φ^0 of the vector φ, we must use the Bayes' formula (24) to calculate the a posteriori distribution $P(\varphi|\mathbf{f})$ and average the quantity in question over this distribution. In particular,

$$\varphi_i^0 = \int \varphi_i P(\varphi \mid \mathbf{f})\,d\varphi, \qquad i = 1, 2, \ldots, n. \tag{25}$$

Besides the Bayesian strategies, one also considers in mathematical statistics so-called minimax strategies, which minimize the "damage" from the replacement of the true quantities by their estimates in the least favorable case. For physical applications these strategies are less natural. Moreover, the Bayesian strategies as a whole have a sort of "completeness." We call a strategy reasonable if there is no strategy that is always better, i.e., better independently of the outcome of any random events. It can be shown that every reasonable strategy is the Bayesian strategy corresponding to some a priori probability distribution. Consequently, by studying the Bayesian strategies we study all reasonable strategies.

From the Bayes formula (24) we see that the a posteriori probability is the product of the a priori probability $P(\varphi)$ and the conditional probability $P(\mathbf{f}|\varphi)$ (the denominator does not depend on φ and is introduced for normalization). Therefore the a posteriori information about φ is the sum of the a priori information and that obtained from experience. If one of the factors is much more informative than the other, the a posteriori probability is close to the former factor and almost independent of the other. This case is very often encountered, the informative factor being, of course, the conditional probability $P(\mathbf{f}|\varphi)$—otherwise there would be no

reason to do the experiment. Suppose, for example, that $P(\mathbf{f}|\varphi)$ as a function of φ has a sharp maximum in some region V_0 of the space and goes rapidly to zero outside this region, while $P(\varphi)$ varies slowly in the region. Then the a posteriori probability $P(\varphi|\mathbf{f})$ is practically independent of the specific form of $P(\varphi)$, and therefore we can set $P(\varphi) = \text{const}$ in any region V_1 containing V_0.

So, every reasonable strategy for adopting solutions under indefinite conditions is a Bayesian strategy corresponding to some a priori probability distribution. When we do not have any a priori information, we adopt the solutions which seem natural to us in this situation, and possibly seem to be the best. In reality they are the best for a world in which all thinkable possibilities (described in a definite way) are equally probable. We shall see below that the usual direct solution of the equations (22) without the application of regularization methods, together with the usual estimate of the errors by the methods of the classical theory of errors, is the Bayesian strategy for the a priori probability density $P(\varphi) = \text{const}$. Accordingly, the use of a priori information is not an exclusive privilege of the regularization methods: in principle, it always occurs; the regularization methods are distinguished by the fact that they use nontrivial a priori information.

Before proceeding to the examination of various ways of introducing a priori information, we shall adopt a specific form of the probability density function of the errors. We make the usual, and as a rule fully justified, assumption that the errors δ_j for different j are independent and distributed according to the normal law with mathematical expectation zero. We denote the root-mean-square error in the measurement of the quantity δ_j by s_j. Then the conditional probability (23) takes the form

$$P(\mathbf{f}|\varphi) = \prod_{j=1}^{m} (2\pi s_j^2)^{-1/2} \exp\left\{-\frac{1}{2s_j^2}\left[f_j - \sum_{i=1}^{n} k_{ji}\varphi_i\right]^2\right\}. \quad (26)$$

This expression can be considerably simplified if we introduce quantities g_j, proportional to the f_j and measured with equal absolute accuracies,

$$g_j = \frac{s}{s_j} f_j,$$

and the corresponding matrix L with the matrix elements

$$L_{ji} = \frac{s}{s_j} k_{ji}.$$

We define the quantity s (the error of measurement of g_j) so that the transformation $\mathbf{f} \to \mathbf{g}$ is unitary:

$$s^m = \prod_{j=1}^{m} s_j.$$

Accordingly, s is the geometric mean of all the errors of measurement of the f_j. The equations (22) become

$$L\varphi = \mathbf{g}, \quad (27)$$

but as before we shall regard the quantities f_j as the fundamental physical variables, and treat the vector $\mathbf{g}$ as an auxiliary construction.

We can now write the conditional probability (26) in the form

$$P(\mathbf{f}|\varphi) = (2\pi s^2)^{-m/2} \exp\left\{-\frac{1}{2s^2}|\mathbf{g} - L\varphi|^2\right\}. \quad (28)$$

The Solution without Regularization

Let us set $P(\varphi) = 1/V$ in an extremely large volume V (which we shall let go to infinity later on, if this turns out to be possible) and $P(\varphi) = 0$ outside this volume. Then in the volume V the a posteriori probability $P(\varphi|\mathbf{f})$ will be equal up to a constant factor to the $P(\mathbf{f}|\varphi)$ given by Eq. (28).

With every probability distribution one can associate an infinite set, called a statistical ensemble, whose elements are obtained by independent random choices subject to the given distribution. The use of the concept of a statistical ensemble often assists the brevity and intuitiveness of our language by appealing to set-theoretical ideas. The a posteriori distribution $P(\varphi|\mathbf{f})$ for the case in which $P(\varphi) = \text{const}$ defines an ensemble of vectors (functions) φ which we shall call the complete ensemble of unregularized solutions and shall denote by $P_c(\varphi)$ (we identify ensembles by indicating the corresponding probability densities of φ). Thus,

$$P_c(\varphi) = c_1 \exp\left\{-\frac{1}{2s^2}|\mathbf{g} - L\varphi|^2\right\}. \quad (29)$$

The components φ_i^0 of the unregularized solution are the averages of the φ_i over this ensemble, and their errors are the square roots of the dispersions of the φ_i.

Expanding the square of the absolute value of the vector in (29) and including a factor not involving φ in the constant, we get

$$P_c(\varphi) = c_2 \exp\left\{-\frac{1}{2s^2}[(\varphi, L^*L\varphi) - 2(L^*\mathbf{g}, \varphi)]\right\}. \quad (30)$$

We can think of the ensemble $P_c(\varphi)$ as a set of points in the space R^n with density proportional to $P_c(\varphi)$. To elucidate the character of this set, we introduce new coordinate axes in the space R^n, namely the orthonormal system of eigenvectors ψ^k of the matrix L*L. Since L*L is a symmetric positive semidefinite matrix, all of the vectors ψ^k are real, and the corresponding eigenvalues, which we denote by λ_k^2, are nonnegative.

Let us denote by $\tilde{\varphi}_k$ the projection of φ on ψ^k. The quantities $\tilde{\varphi}_k$ can be taken as the new coordinates of the vector φ. In the new variables the density of the distribution of the complete ensemble of solutions is of the form

$$P_c(\varphi) = c_2 \exp\left\{-\sum_{k=1}^{n} \frac{1}{2s^2}(\lambda_k^2 \tilde{\varphi}_k^2 - 2\tilde{h}_k\tilde{\varphi}_k)\right\}, \quad (30a)$$

where $\tilde{h}_k$ is the projection of the vector $L^*\mathbf{g}$ on ψ^k.

Accordingly those quantities $\tilde{\varphi}_k$ for which $\lambda_k \neq 0$ are statistically independent and are distributed according to the normal law with mathematical expectations $\tilde{h}_k/\lambda_k^2$ and dispersions s^2/λ_k^2. Consequently, the region in which the points of the ensemble $P_c(\varphi)$ are mainly concentrated is of extent of order s/λ_k in the direction of ψ^k. Those quantities $\tilde{\varphi}_k$ for which $\lambda_k = 0$ do not actually occur in the expression (30a) (it is easy to show that if $\lambda_k^2 = 0$, then also $\tilde{h}_k = 0$). In a displacement along the vector ψ^k the function $P_c(\varphi)$ does not change in value. From this it can already be seen that if even one of the eigenvalues λ_k goes to zero, we cannot let $V \to \infty$, because in this case the integral $\int P_c(\varphi)d\varphi$ does not

exist. There is no information at all about the $\tilde{\varphi}_k$ in this direction, and we are obliged to impose some sort of a priori restrictions on this $\tilde{\varphi}_k$.

If all of the λ_k are different from zero [the necessary and sufficient condition for this is that $\det(L^*L)$ be different from zero], then we may assume that $V < \infty$. The Bayesian strategy for this case reduces to the method of least squares. The Bayesian estimate φ^0 is found from the equation

$$L^*L\varphi^0 = L^*g,$$

which for the case m = n is the same as the original equation (27), and the error of the determination σ_i at the point i can be found from the equation

$$\sigma_i^2 = s^2((L^*L)^{-1})_{ii}, \qquad i = 1, 2, \ldots, n. \tag{31}$$

For a quantity characterizing the error of the solution φ^0 as a whole we can calculate the mean square of the error σ taken over all i:

$$\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = \frac{s^2}{n}\operatorname{Sp}((L^*L)^{-1}) = \frac{s^2}{n}\sum_{k=1}^{n}\frac{1}{\lambda_k^2}$$

(we have made use of the invariance of the trace of a matrix under coordinate transformations).

Accordingly, a strong increase of the error of the unregularized solution occurs whenever any of the eigenvalues λ_k of the matrix L*L is extremely small, and it happens because of the uncertainty of the component along the corresponding ψ_k.

If the matrix K has even one pair of rows that are nearly equal up to a constant factor, or if more generally its rows are almost linearly dependent, or, qualitatively speaking, if they "resemble each other," then the matrix L will have this same property, and there will be small numbers among the eigenvalues of the matrix L*L. Then the unregularized solution is useless, since the error in it is many times larger than the reconstructed function itself. Precisely such matrices, with nearly linearly dependent rows, are obtained in the algebraization of an incorrectly posed problem. This can also happen, by the way, with a correctly posed original equation.

The Solution in an Ensemble Prescribed by Finite Selection

Let us suppose that the a priori ensemble of possible solutions is characterized by giving its N vectors φ^ν, which consequently can be regarded as obtained by a random selection subject to the a priori probability distribution $P(\varphi)$. This case can occur, for example, when one makes systematic direct measurements of the quantities φ_i under certain stable conditions in order to obtain statistical material which can be used for reliable indirect determination of the φ_i from measurements made under the same conditions of the quantities f_j, when no direct measurements are available.

If the selection is sufficiently representative, the regularized solution and its error can be obtained as the Bayesian solution and its error, the calculations being made by replacing (approximately) the averaging over the a priori distribution by averaging over the finite-selection ensemble.

Suppose some function $F(\varphi)$ of the quantities φ_i is given. The average of this function over the a posteriori distribution (24) can be put in the form

$$\langle F(\varphi)\rangle = \int F(\varphi)\,P(\varphi\,|\,f)\,d\varphi = \frac{\langle F(\varphi)\,P(f\,|\,\varphi)\rangle_{apr}}{\langle P(f\,|\,\varphi)\rangle_{apr}},$$

where $\langle G(\varphi)\rangle_{apr}$ denotes the average of the function $G(\varphi)$ over the a priori distribution $P(\varphi)$.

Let us replace the average over the ensemble by that over the finite selection:

$$\langle G(\varphi)\rangle_{apr} \approx \frac{1}{N}\sum_{\nu=1}^{N} G(\varphi^\nu).$$

Then for $\langle F(\varphi)\rangle$ we get

$$F(\varphi) = \sum_{\nu=1}^{N} w_\nu F(\varphi^\nu)\Big/\sum_{\nu=1}^{N} w_\nu, \tag{32}$$

where

$$w_\nu = \exp\left\{-\sum_{j=1}^{m}\frac{1}{2s_j^2}\left[f_j - \sum_{i=1}^{n} k_{ji}\varphi_i^\nu\right]^2\right\} \tag{33}$$

are the f-dependent weight factors for the various vectors φ^ν and are proportional to the conditional probabilities $P(f\,|\,\varphi^\nu)$.

When we set $F(\varphi) = \varphi_i$ and $F(\varphi) = (\varphi_i - \langle\varphi_i\rangle)^2$ we get the regularized solution and its error from Eqs. (32) and (33).

The Solution in an Ensemble Prescribed by the Correlation Matrix

Let us assume that the mathematical expectation of the vector φ is zero in the a priori ensemble $P(\varphi)$. This assumption will be understood to be made also in all the following discussions. (It is clear that it does not limit the generality of the treatment, since one can always make a parallel displacement.) We also assume that in some way or other we are provided with knowledge of the correlation matrix of the a priori ensemble:

$$C_{ij} = \langle\varphi_i\varphi_j\rangle_{apr} = \int \varphi_i\varphi_j P(\varphi)\,d\varphi. \tag{34}$$

In order to introduce into our determination of the a priori probability $P(\varphi)$ the least possible information beyond that contained in the relations (34), we choose from among all functions satisfying (34) the one that contains the minimum of information about φ, i.e., the one that minimizes the functional

$$I[P(\varphi)] = \int \ln P(\varphi)\,P(\varphi)\,d\varphi, \tag{35}$$

which gives (to within a constant factor) a quantitative measure of this information.

It can be shown that this distribution density is of the form

$$P(\varphi) = c_3\exp\left\{-\frac{1}{2}(\varphi, C^{-1}\varphi)\right\}. \tag{36}$$

The correlation matrix C is a symmetric positive-definite matrix. Consequently, there exists an orthonormal system of eigenvectors χ^l, $l = 1, \ldots, n$, of the matrix C with positive eigenvalues, which we denote by γ_i^2. (The case in which some $\gamma_1 = 0$ can be regarded as a limiting case.) Let us denote by $\dot{\varphi}_1$ the projection of the vector φ on the axis χ^1. In the coordinates $\dot{\varphi}_1$ the density distribution $P(\varphi)$ takes the form

$$P(\varphi) = c_3\exp\left\{-\sum_{i=1}^{n}(\dot{\varphi}_i^2/2\gamma_i^2)\right\}.$$

Consequently, the components $\dot{\varphi}_1$ of the expansion of

φ in terms of the eigenvectors of the correlation matrix are statistically independent and are normally distributed with root-mean-square values γ_l.

Using (36) and (28), we get from the Bayes formula (24) the a posteriori distribution

$$P(\varphi \mid f) = c_4 \exp\left\{-\frac{1}{2s^2}[(\varphi, (L^*L + s^2C^{-1})\varphi) - 2(L^*g, \varphi)]\right\}, \quad (37)$$

which is again a normal distribution. According to (25) the regularized solution $\varphi^C = \langle\varphi\rangle$ is the solution of the equation

$$(L^*L + s^2C^{-1})\varphi^C = L^*g. \quad (38)$$

The mean square error of this construction is

$$\sigma_i^2 = s^2((L^*L + s^2C^{-1})^{-1})_{ii}. \quad (39)$$

The method of solving the inverse problem by the formulas (38) and (39) has also been pronounced in [83] and [84] on the basis of similar arguments.

A statistical ensemble with a probability density which is the product of the probability densities of two other ensembles can be called the intersection of these ensembles. This intersection has much in common with the set-theory intersection. If the ensembles combined by multiplication are such that they are described by probability distributions of the "all or nothing" type, i.e., distributions which take a constant value in some region of space and are zero outside this region, then their intersection is also an ensemble of this type, and the region it occupies is the intersection of the regions occupied by the original ensembles. In our case the ensembles are described by normal distributions, which theoretically extend to infinity, but outside a certain region they decrease so rapidly that the situation is almost the same as for the ensembles just now described.

The a posteriori ensemble (37) is the intersection of the complete ensemble of solutions (29) and the a priori ensemble (36); consequently, in the set-theoretic interpretation it consists of those vectors which are members of both ensembles. Let us choose a definite value level, for example 0.99, and represent an ensemble by the region (or body) in the space R^n, bounded by a surface $P(\varphi) = \text{const}$, such that the integral of $P(\varphi)$ over this region is equal to 0.99. Owing to the fact that the distributions are normal, all of these regions will be bounded by ellipsoids (sometimes degenerate). The principal axes of the ellipsoid representing the complete ensemble of solutions are directed along the vectors ψ^k, and those of the ellipsoid corresponding to the a priori ensemble along the vectors χ^l. In general these axes do not coincide.

Let us investigate the picture of the intersection of the ensembles for an error of measurement s which approaches zero.

If all of the eigenvalues λ_k are different from zero, then for $s \to 0$ the complete ensemble of solutions contracts to a point corresponding to the exact solution of Eq. (8). Consequently, the a posteriori ensemble of regularized solutions also contracts to this same point [independently of the form of $P(\varphi)$, provided that $P(\varphi^a)$, where φ^a is the exact solution, does not become zero]. Accordingly, in this case the regularized solution approaches the exact solution, and its error goes to zero.

If one or more of the eigenvalues λ_k is (are) equal to zero, then for $s \to 0$ the complete ensemble of solutions approaches a hyperplane parallel to the corresponding vectors ψ^k. The intersection of this hyperplane with the ellipsoid of the a priori ensemble determines some degenerate elliptical region in the space R^n, which characterizes the a posteriori ensemble. The center of this region is the regularized solution, and its dimensions determine the error of the reconstruction, which accordingly approaches a finite and nonzero limit for $s \to 0$.

The study of the intersections of ensembles in the space R^n shows that the information content of an experiment depends on the relative orientation of the principal axes of the ellipsoids of the complete ensemble of solutions and the a priori ensemble. If the axes of the former ellipsoid that correspond to the largest values of λ_k are "remote" directions for the second ellipsoid, i.e., directions such that the distance from the center to the boundary in these directions is close to its largest value, then the experiment is very informative. If these axes are directed along "near" directions of the a priori ellipsoid, the experiment gives little information. In particular, we can imagine a case in which with a given a priori distribution the experiment gives no information at all. Suppose the γ_l are equal to zero for some group of indices l; the corresponding vectors χ^l define a subspace of the space R^n. Also let the matrix L*L be such that all of its eigenvectors ψ^k for which $\lambda_k \neq 0$ lie in this subspace. It is easy to see that in this case the a posteriori distribution is the same as the a priori distribution. From the experiment one can determine only those components of the vector φ which are a priori equal to zero.

The analysis of the regularized solution is very much simplified if we assume that the matrices L*L and C can be simultaneously brought to diagonal form (this is the case, for example, if the problem as a whole has translational symmetry). For definiteness we denote the common eigenvectors by ψ^k and number them so that the λ_k are nonincreasing. Then from (38) and (39) we find in the ψ representation

$$\tilde{\varphi}_k^C = \frac{h_k}{\lambda_k^2 + s^2/\gamma_k^2}, \quad (40)$$

$$\sigma_k^2 = \frac{s^2}{\lambda_k^2 + s^2/\gamma_k^2}. \quad (41)$$

We can connect the regularized solution (40) with the exact solution φ in the following was (if the exact solution exists):

$$\tilde{\varphi}_k^C = z_k\tilde{\varphi}_k, \quad (40a)$$

where

$$z_k = \frac{\lambda_k^2\gamma_k^2}{\lambda_k^2\gamma_k^2 + s^2} \quad (40b)$$

is a "cutoff factor" whose value depends on the ratio between $\lambda_k^2\gamma_k^2$, the mean square of the k-th component in the right member, and s^2, the mean square error of the measurement of this component. If $\lambda_k^2\gamma_k^2 \gg s^2$, i.e., if the useful signal is much larger than the noise, then $z_k \approx 1$ and the k-th component in the regularized solution is practically the same as in the unregularized (exact) solution. If $\lambda_k^2\gamma_k^2 \ll s^2$ (the signal is much smaller than the noise), then $z_k \ll 1$ and the k-th component in the solution, reflecting exclusively errors of measurement, is suppressed. Comparing (41) with (31) we see

that the σ_k^2 for the regularized solution is also in the ratio z_k to the σ_k^2 for the unregularized solution. For $\lambda_k^2\gamma_k^2 \gg s^2$ the error is practically the same as without regularization; for $\lambda_k^2\gamma_k^2 \ll s^2$ we find from (41) that $\sigma_k^2 \approx \gamma_k^2$; the uncertainty of the k-th component is equal to its uncertainty in the a priori ensemble. Thus the effect of the regularization is to replace the false information which the unregularized solution gives about k-th components with small λ_k with the information about these components which is contained in the a priori distribution.

The inequality $\lambda_k^2\gamma_k^2 \ll s^2$ is a condition whose satisfaction gives us grounds for thinking that λ_k is practically equal to zero and that the conditional probability $P(f|\varphi)$ is practically independent of $\bar{\varphi}_k$. There may be a large break in the spectrum of the matrix L*L, so that for some index k_0 the eigenvalue $\lambda_{k_0}^2$ is a hundred or a thousand times as large as the next eigenvalue $\lambda_{k_0+1}^2$. Then for some s a situation will occur such that for $k \leq k_0$ the cutoff factor z_k can be regarded as equal to unity, and for $k > k_0$, as equal to zero, and this situation will continue while s decreases by a large factor [as long as the $(k_0 + 1)$-st component does not become informative]. In the solution of the regularized equation this leads to the phenomenon, strange at first glance, that neither the regularized solution nor its error changes while the error s is being decreased by a large factor.

The Solution in an Ensemble of Bounded or Smooth Functions

Suppose that we know that the unknown function $\varphi(x)$ is "more or less smooth." We can make this information more precise by introducing some sort of functional characterizing the degree of smoothness of a function, for example the norm of its q-th derivative (we shall not specify the particular value of q):

$$\Omega[\varphi(x)] = \int \left[\frac{d^q\varphi(x)}{dx^q}\right]^2 dx, \tag{42}$$

and by fixing an expected approximate value of this functional:

$$\Omega[\varphi(x)] \approx \omega.$$

In the same way we can use information about the boundedness of the function $\varphi(x)$. For this it suffices to set q = 0 in Eq. (42). We can also form Ω as a linear combination of the norms of several derivatives. In all these cases after the algebraization is performed the a priori information will be available in the form

$$(\varphi, \Omega\varphi) = \sum_{i,j=1}^{n} \varphi_i\Omega_{ij}\varphi_j \approx \omega,$$

where Ω is a symmetric positive semidefinite matrix which is the finite-difference equivalent of the corresponding functional.

Obviously we must also introduce an a priori distribution $P(\varphi)$ such that

$$\int (\varphi, \Omega\varphi)\, P(\varphi)\, d\varphi = \omega. \tag{43}$$

In order to introduce as little arbitrariness as possible, we choose, as in the previous case, from among all $P(\varphi)$ satisfying the condition (43) the particular distribution $P(\varphi)$ with the minimum information (35). This function $P(\varphi)$ will be

$$P_\alpha(\varphi) = c_\alpha \exp\left\{-\frac{\alpha}{2}(\varphi, \Omega\varphi)\right\}, \tag{44}$$

where

$$\alpha = n/\omega. \tag{45}$$

Accordingly we have arrived at the case considered in the previous section, with the correlation matrix $C = (\alpha\Omega)^{-1}$. All of the results obtained previously remain valid. In particular, the regularized solution φ^α obeys the equation (38):

$$(L^*L + s^2\alpha\Omega)\,\varphi^\alpha = L^*g. \tag{46}$$

This equation, as already indicated, was first derived by Phillips.[53] Independently of Phillips it was postulated by Tikhonov[56] as the correctly posed equation approximately representing the incorrectly posed equation (27).

In solving the regularized equation (46) one must keep in mind that the matrix Ω may have eigenvalues equal to zero. If, for example, Ω describes the norm of the first derivative, then for a vector φ that represents a constant function, i.e., one that has all of its components φ_i equal, $(\varphi, \Omega\varphi) = 0$. For the matrix Ω that gives the norm of the second derivative there are two linearly independent vectors with this property, and so on. This means that the corresponding a priori distribution extends to infinity along these vectors with a nonzero value ("infinitely remote" directions). In this case the distribution (44) is to be understood as defined in some large volume V. If the infinitely remote directions of the a priori ensemble are also infinitely remote, or simply extremely remote, directions of the complete ensemble of solutions, then the regularized solution will, respectively, be nonexistent or have a very large error. In this case it is necessary to change Ω (for example, to include in the functional the norm of the function itself), in order to limit the probability density $P(\varphi)$ along the infinitely remote directions. As a rule, however, this is not necessary, because the functional Ω representing the norm of the derivative is used when information about the smoothness of the function is sufficient for the regularization, and consequently the solution is adequately stable against combination with a constant, a linear function, and so on. In the language of ensembles this means that the infinitely remote directions of the a priori ensemble are near directions of the complete ensemble of solutions. Therefore we may assume that the a priori distribution (44) is different from zero only inside some volume V, and after multiplying it by the complete ensemble of solutions we can let V go to infinity.

The Solution in the Narrowest Admissible Ensemble

Let us think again about probability density functions of the "all or none" type. If both the a priori ensemble and the complete ensemble of solutions are described by probability densities of this type, then it can happen that the corresponding regions in the space R^n have no common points, and consequently there does not exist any ensemble which is the intersection of these two ensembles. This situation indicates a contradiction between the information obtained from experiment and the

a priori information, and we must state that no solution of the problem [i.e., no solution of Eq. (8) with an indicated accuracy] in the given a priori ensemble exists.

If the ensembles are described by normal distributions, then their intersection always exists, but it is intuitively clear that if the central regions of the ensembles, where, say, 99 percent or 99.9 percent of all the points are concentrated, do not intersect, then the intersection is a sort of spurious effect and has no physical meaning. Therefore a criterion is necessary that will enable us to separate those cases in which the a posteriori ensemble is actually a solution of the problem from those in which it is a "game with the tails" of the normal distributions. It is natural to base this criterion (see [65]) on an estimate of the error with which the solution appearing in the a posteriori ensemble satisfies Eq. (8) or Eq. (27) (the latter is preferable, since in the complete ensemble of solutions the error g_i does not depend on i). The expression

$$\frac{1}{n}(g-L\varphi)^2$$

gives the average (over the component number i) of the square of the error with which Eq. (27) is satisfied. By averaging it over some ensemble, we obtain the mean square error for all the functions (vectors) of the ensemble. It is obvious that if for the averaging we take the complete ensemble of solutions (29) we get the quantity s^2 by definition:

$$s^2=\left\langle\frac{1}{n}(g-L\varphi)^2\right\rangle_c.$$

If, on the other hand, we average over the a posteriori ensemble, then, generally speaking, we get a different quantity, which we denote by s'^2:

$$s'^2=\left\langle\frac{1}{n}(g-L\varphi)^2\right\rangle_{apost}=\frac{1}{n}\int(g-L\varphi)^2P(\varphi\mid f)\,d\varphi. \qquad (47)$$

We can now formulate the criterion about which we were speaking earlier as the inequality $s'^2 \le s^2$. It has the following meaning. We get the a posteriori ensemble by selecting from the complete ensemble of solutions those functions φ that belong to the a priori ensemble. If as the result of this selection the mean square deviation of $L\varphi$ from g is not increased, i.e., $s'^2 \le s^2$, this means that belonging to the a priori ensemble does not hinder a function φ from being a solution of Eq. (27). In this case the a posteriori ensemble satisfies the requirements imposed on it and is the most complete solution of the problem. If, on the other hand, $s'^2 > s^2$, this means that for the function φ to belong to the a priori ensemble is (statistically) incompatible with the validity of Eq. (27) to the accuracy s. In this case the a priori ensemble consists of functions φ which (on the average) do not satisfy the original equation to the necessary accuracy; consequently, no solution of the problem exists in the given a priori ensemble.

If on the basis of previous experience the correlation matrix is well known to us, there will hardly be any need to subject the solution obtained from (38) to the test with the criterion $s'^2 \le s^2$. In fact, in this case we have no doubt that the true function φ belongs to the a priori ensemble we have chosen, and if this is so, then $s'^2 \approx s^2$ (see [65]). (We note that the inequality $s'^2 \le s^2$ is to be understood in the spirit of statistics as $s'^2 < s^2$ or $s'^2 \approx s^2$.) The same is also true of the solution in an ensemble of smooth (or bounded) functions, when we have sufficiently reliable knowledge of the quantity ω, and consequently also of the quantity α. If often happens, however, that there are no reliable data on the quantity ω. Then we may pose the problem of finding the largest α for which a solution still exists (see [66]). Let us consider the family of a priori ensembles (44) with arbitrary α and call an ensemble such that $s'^2 \le s^2$ an admissible ensemble. For $\alpha = 0$ the ensemble (44) covers all functions φ, giving them equal probabilities. With increasing α the ensemble narrows (there is no possibility here of giving this concept a rigorous definition, but it is intuitively obvious), keeping in its membership only smoother and smoother functions, and for $\alpha = \infty$ it degenerates into the ensemble which contains only the function which is identically zero.

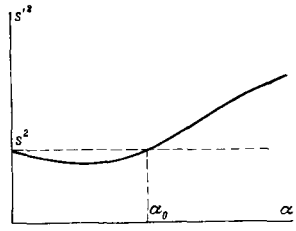

FIG. 1. The dependence of s'^2 on α.

In [66] the dependence of s'^2 on α is investigated for the translationally invariant case, and it is shown that it is of the form shown in Fig. 1. The point $\alpha = \alpha_0$ characterizes the narrowest admissible ensemble of smooth functions. It can be found by numerical calculation of the curve $s'^2(\alpha)$.

Thus if the parameter α, which characterizes the degree of smoothness of the unknown function, is not known, we can determine the value of α corresponding to the "smoothest" admissible ensemble, and find the solution in this ensemble. Accordingly, the solution found must in some sense be the smoothest, and therefore it is interesting to compare this approach with that of Phillips,[53] which also consists of finding the smoothest solution that approximately satisfies the original equation. According to the approach of Phillips the parameter α would have to be determined from the equation

$$\frac{1}{n}(g-L\varphi^\alpha)^2=s^2, \qquad (48)$$

where φ^α is the regularized solution. However, as has already been stated, the function obtained with this sort of algorithm turns out to be excessively smoothed, and better results are obtained if one chooses an α corresponding to a right member of (48) smaller than s^2. With the statistical approach the parameter α is determined from the condition

$$s'^2=\frac{1}{n}(g-L\varphi^\alpha)^2+\frac{1}{n}\langle(L(\varphi-\varphi^\alpha))^2\rangle=s^2. \qquad (49)$$

Equation (49) differs from (48) by the addition to the left member of a quantity depending on the dispersion in the a posteriori ensemble, and this is equivalent to using Eq. (28) with a smaller right member. Therefore, by using the condition (49) we get a less smoothed solution. It is shown in [66] that this algorithm for determining the parameter α leads to extremely satisfactory results.

The Solution in a Laminar Ensemble

A different method of regularization is proposed in [67]; it gives not the very smoothest (in any sense) solution, but the solution with the most probable degree of smoothness.

Suppose we know that the unknown function is smooth to some degree and could be obtained by a selection from some ensemble of smooth functions $P_\alpha(\varphi)$, but the parameter α characterizing this ensemble, and consequently the degree of smoothness of the unknown function, is not known to us. Does this mean that regularization is impossible and that the best we can do is to use the unregularized solution?

We have seen that the unregularized solution corresponds to the assumption $P(\varphi) = \text{const}$. But the uncertainty of the value of α by no means implies the assumption $P(\varphi) = \text{const}$, for this means complete absence of correlations between the values of the φ_i. At the same time, if φ represents a continuous physical function there must exist a correlation between values of φ_i with nearly equal values of i, although the degree of correlation (which depends on α) is not known to us and may be arbitrary. This sort of assumption about φ can be described by the a priori probability

$$P(\varphi) = \int_0^\infty P(\alpha)\, P_\alpha(\varphi)\, d\varphi, \tag{50}$$

where $P(\alpha)$ is an a priori probability for α, which we can regard as taking a constant value in any arbitrarily large range of positive α.

An ensemble with the probability density (50) can be called "laminar," the layers being ensembles of smooth functions with different values of α. The laminar ensemble (50) can be used to express the a priori information about φ in cases in which the physical nature of φ is known but the parameter which characterizes the main physical factor is not known. Suppose, for example, that φ represents the time dependence of the coordinate in the motion of a point mass m in a one-dimensional force field F. Then the function $\varphi(x)$ (x is the time) is a solution of Newton's equation, and the second derivative $\varphi''(x)$ is of the order of magnitude of F/m. We may know nothing about the quantities m and F in our physical system and nevertheless state that φ is subject to the a priori distribution (50), where $P_\alpha(\varphi)$ is given by Eq. (44) and the matrix Ω represents the norm of the second derivative of the function $\varphi(x)$ [q = 2 in Eq. (42)]. This information is sufficient for the determination of the regularized solution and its error.

Using (50) as the a priori distribution, we get by Bayes' formula (24) the a posteriori distribution

$$P(\varphi \mid \mathbf{f}) = \frac{\int P(\alpha)\, P_\alpha(\varphi)\, P(\mathbf{f} \mid \varphi)\, d\alpha}{\int P(\alpha)\, P_\alpha(\varphi)\, P(\mathbf{f} \mid \varphi)\, d\alpha\, d\varphi}. \tag{51}$$

The Bayesian estimate of any function $F(\varphi)$ is the average of $F(\varphi)$ over this distribution:

$$\langle F(\varphi)\rangle = \int P(\varphi \mid \mathbf{f})\, F(\varphi)\, d\varphi.$$

We now have the problem of expressing $\langle F(\varphi)\rangle$ in terms of the integral over α of the Bayesian estimate of this function in the case of a definite α:

$$\langle F(\varphi)\rangle_\alpha = \int P(\varphi \mid \mathbf{f}, \alpha)\, F(\varphi)\, d\varphi, \tag{52}$$

where $P(\varphi \mid \mathbf{f}, \alpha)$ is the a posteriori distribution for the a priori distribution $P_\alpha(\varphi)$ [it is obtained from (37) by replacing C^{-1} by $\alpha\Omega$].

Introducing the distributions

$$P(\mathbf{f} \mid \alpha) = \int P(\mathbf{f} \mid \varphi)\, P_\alpha(\varphi)\, d\varphi, \tag{53}$$

$$P(\alpha \mid \mathbf{f}) = \frac{P(\alpha)\, P(\mathbf{f} \mid \alpha)}{\int P(\alpha)\, P(\mathbf{f} \mid \alpha)\, d\alpha}, \tag{54}$$

we find

$$\langle F(\varphi)\rangle = \int \langle F(\varphi)\rangle_\alpha\, P(\alpha \mid \mathbf{f})\, d\alpha. \tag{55}$$

Accordingly, $\langle F(\varphi)\rangle$ is obtained by a further averaging of $\langle F(\varphi)\rangle_\alpha$ over α with the weight function $P(\alpha \mid \mathbf{f})$, i.e., the a posteriori probability of the smoothness parameter α for the known result f of the measurements.

Using the previous results for $\langle F(\varphi)\rangle_\alpha$, one can calculate $\langle F(\varphi)\rangle$ by numerical integration over α in Eq. (55). If, however, the measurement of f gives a sufficiently large amount of information about the parameter α, the function $P(\alpha \mid \mathbf{f})$ will have a sharp maximum at some value α_0. Then instead of the average over α we can use the formulas for a definite α, setting (a posteriori) $\alpha = \alpha_0$. The value α_0 can be found by numerical analysis of the curve of $P(\alpha \mid \mathbf{f})$.

The quantity α_0 is essentially a statistical estimate of the parameter α by the method of maximum likelihood. The boundaries between the various "layers" of smooth functions occurring in the ensemble (50) are not strictly fixed owing to the probabilistic statement of the entire problem. If, however, the function φ is given, then we can estimate the value of α characterizing the layer from which this function most probably comes by the methods of mathematical statistics. When we measure the function f in order to determine the function φ, the latter is of course not known to us. But we do know the probabilistic character of the process which produces f for a given φ. Accordingly we get the following two-step scheme for the determination of α:

$$\alpha \xrightarrow{P_\alpha(\varphi)} \varphi \xrightarrow{P(\mathbf{f} \mid \varphi)} \mathbf{f}$$

from which we can find the conditional probability of f for given α, as expressed by Eq. (53). If this function is known, the estimation of α for a given f is a classical problem of mathematical statistics.

The question arises: In what cases can we replace the averaging over α by the solution with the most probable α? To answer this question we must calculate the width in the α-scale of the peak of the function $P(\alpha \mid \mathbf{f})$, or in other words the error $\Delta\alpha$ of our determination of the parameter α. The expression for the error $\Delta\alpha$ has a very clear physical meaning in the case when the matrices Ω and L can be simultaneously diagonalized. For this case we get (see [68])

$$\Delta\alpha/\alpha_0 = \sqrt{2/n_{\text{eff}}},$$

where

$$n_{\text{eff}} = \sum_{k=1}^{n} z_k^2,$$

and z_k is the cutoff factor (40b). The quantity n_{eff} is the effective number of independent components of the

unknown function determined in the experiment. Those components for which $z_k \ll 1$ contain no information about α and do not contribute to reducing the value of $\Delta\alpha$.

The algorithm for the determination of the most probable α and the algorithm expounded in the preceding section for determining the α for the narrowest admissible ensemble lead in general to different values of α_0. It is shown in [65], however, that the α_0 that corresponds to the narrowest admissible ensemble can also be regarded as a well-grounded estimate of α. Therefore in those cases in which $\Delta\alpha$ is not too large the two methods lead to nearly equal results, which coincide for $\Delta\alpha \to 0$.

Some Results and Problems

Recent papers[66-68] have given descriptions of computer programs which realize the method of statistical regularization, and the results of some mathematical experiments which were carried out to test this method. The experiments were done in the following way. A kernel (matrix) K and a "true" function (vector) φ were chosen. The "true" right member of the equation was then determined by multiplying K by φ. To the vector $K\varphi$ was added a random-error vector with independent normally distributed components. The vector f so obtained to imitate the experimental results was used to reconstruct the vector φ by means of the programs in question. The result of the reconstruction was compared with the true vector φ, with the theoretically found error of the reconstruction (vector σ) taken into account.

The various programs used the methods of the narrowest admissible ensemble, of the most probable α, and of averaging over α. The function $P(\alpha|f)$ was also studied. In all of the experiments it was found that the reconstructed function coincides with the true function to an accuracy corresponding to the error indicated by the theory. As an illustration we give an example from [67] (Fig. 2). The kernel of the equation was of the difference form, corresponding to the resolution function shown in the figure. The "experimental" right member f is shown together with the error of measurement. The crosses give the result of the reconstruction with the choice of the most probable α (regime 2), and the histogram gives the result for averaging over α (regime 3).

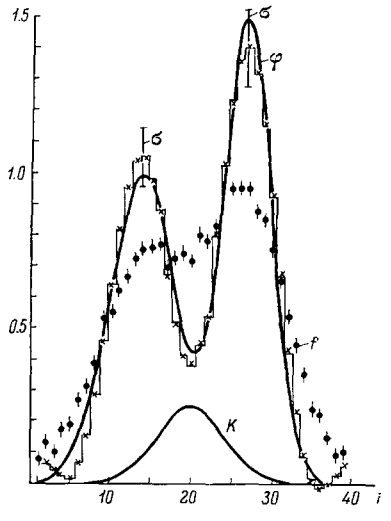

FIG. 2. Reconstruction of a function $\varphi(i)$ from the result of its convolution with a resolution curve K(i). σ is the error of the reconstruction.

The application to some geophysical problems of methods of solving incorrectly posed problems in the ensemble given by the correlation matrix is described in Chapter V of the present article.

The use of an a priori probability distribution in cases when all we know is that the unknown function is smooth, and its actual statistical characteristics are unknown, presents a serious and general problem. A test of whether the finally obtained regularized solution is genuine can be obtained with reasonably large probability by selection from an a priori ensemble. For this one can employ the usual statistical criteria of likelihood of hypotheses.

It would be desirable to make such a test both with an α given in advance and with a posteriori determination of α. In those cases in which the intersection of the a priori ensemble and the complete ensemble of solutions occurs in the "tails" of the Gaussian distributions the likelihood criterion obviously gives a negative answer. In the examples that were solved in [66-68] it could scarcely be doubted that the likelihood criterion must give a positive answer, since n_{eff} was small and the adjustment of the single parameter α could give a high enough level of likelihood. In more complicated cases, however, in which the experiment contains a large amount of information about φ, this may not be so. How does one proceed in such cases?

An answer presents itself: Introduce two or more parameters in the a priori ensemble and determine them a posteriori, as we did with the single parameter α. Then the a priori probability density will be of the form of a "many-dimensional laminar" ensemble:

$$P(\varphi) = \int P(\alpha_1, \ldots, \alpha_\nu) P_{\alpha_1, \ldots, \alpha_\nu}(\varphi)\, d\alpha_1 \ldots d\alpha_\nu.$$

Instead of at once determining the most probable set of parameters $\alpha_1, \ldots, \alpha_\nu$, we can propose the following procedure. We first give to all the parameters α_i except one certain definite initial values (chosen, say, on the basis of past experience). We find the most probable value of this parameter and determine its degree of likelihood. If it is satisfactory, then we look for the solution in the corresponding ensemble. If it is unsatisfactory, we increase the number of unknown parameters by one, and so on. The number of parameters that finally turn out to be estimated a posteriori depends on the informativeness of the experiment with respect to φ.

This formulation of the question gives rise to the following problem: what sort of set of parameters $\alpha_1, \ldots, \alpha_\nu$ should we use, and in what order should they be subjected to estimation? As the second parameter it is natural to suggest the order q of the derivative in Eq. (42). But what sort of initial value of q should we use?

To approach the solution of problems of this sort one must make a fundamental study of the statistical structure of the functions encountered in the various classes of physical problems. Such studies will be an addition to and a refinement of the a priori information

necessary for the satisfactory solution of the incorrectly posed, and essentially undetermined, problems of physics.

V. INVERSE PROBLEMS OF THE OPTICAL PROBING OF THE ATMOSPHERE

The method of statistical regularization considered in Chapter IV has been tested on problems of determining the vertical distributions (profiles) of atmospheric temperature and humidity from measurements of spectra of the Earth's own radiation. These problems are urgent at present in connection with the use of artificial satellites to probe the atmosphere by optical methods, i.e., to determine the temperature and humidity and other parameters of the lower layers of the atmosphere (0–50 km) at any point on the Earth. Since in problems of atmospheric physics the spatial and temporal variations of the temperature and humidity must be known to high accuracy (1 to 2 degrees of temperature over a range 200–320° K, and 0.1 to 0.2 g/kg of specific humidity over a range 0.5–20 g/kg), the securing of reliable solutions to the appropriate inverse problems is a practical necessity.

Since, first, these variations are of a random nature, i.e., the temperature profile $T(\zeta)$ and the humidity profile $q(\zeta)$ are random functions of the height,* and second, for the Earth's atmosphere there is a wealth of data on the statistical characteristics of the vertical structure of the fields $T(\zeta)$ and $q(\zeta)$, the application of the method of statistical regularization to the solution of these problems is quite natural.

The Determination of the Vertical Temperature Profile of the Atmosphere

The idea of determining the vertical temperature profile of the atmosphere from measurements of the Earth's own radiation in sufficiently narrow spectral intervals of the CO_2 absorption band near 15 μ with an artificial satellite was proposed in a paper by Kaplan.[6] The basis of this idea is the physical fact that the radiation in various sections of this band is generated in different layers of the atmosphere, and consequently is determined by the temperature of these layers. This correspondence is sufficiently unambiguous, since the relative concentration of CO_2 is constant and well known up to very large heights, and the absorption of water vapor and other atmospheric substances can be neglected.†

The relation between the spectral intensity I_ν of the radiation of frequency ν, as measured with a satellite, and the temperature $T(\zeta)$ of the atmosphere is described by the equation of transfer of thermal radiation. For the simplest case of an absolutely black terrestrial surface and measurements of I_ν in the direction of the local vertical this relation will be of the form

$$I_\nu[T(\zeta),\ q(\zeta)] = B_\nu[T(1)]\,P_\nu[w(1)] - \int_0^1 B_\nu[T(\zeta)]\frac{\partial P_\nu[w(\zeta)]}{\partial\zeta}\,d\zeta. \quad (56)$$

Here $B_\nu[T(\zeta)]$ is the Planck function, and $P_\nu[w(\zeta)]$ is the transmission function of the atmosphere, which characterizes the weakening of the radiation from a source by a column of atmosphere $(0, \zeta)$ of unit cross section, containing an effective mass of absorbing gas $w(\zeta) = \int_0^\zeta t^n q(t)\,dt$ (n is a constant which allows for the change of the halfwidth of the absorption lines of the gas in the atmosphere nonuniform in height). The function P_ν in (56) is taken to include the spectral sensitivity of the apparatus, and the frequency ν corresponds, for example, to the middle of the interval of spectral resolution (sometimes we shall take the index ν to mean simply the number of a section of the spectrum for which a measurement of I_ν has been made).

The relation (56), regarded as an integral equation for $T(\zeta)$, allows us to determine the vertical temperature profile if we know the emission intensity I_ν as a function of ν and the transmission function P_ν, which is the kernel of the equation. The apparent simplicity of Eq. (56) was the reason that in the first attempts to solve it, in papers by Wark and by Yamamoto,[69, 70] use was made of the formal reduction of (56) to systems of algebraic equations obtained either by approximation of the integral by a finite sum,[70] or by expanding the unknown function in a series of polynomials.[69] As was to be expected, from the solution of such systems without special precautions one could get values arbitrarily far from the true solution, and even physically quite meaningless (cf., e.g., [71]), and a refinement of the approximations, meaning an increase of the order of the algebraic systems, leads to increasing stability of the solution. This displays the incorrectness of the formulation of the inverse problem of optical probing, which is a direct consequence of the physical mechanism which determines the transfer of the intrinsic radiation in the stratified atmosphere. Owing to this mechanism there is a smoothing out of the variations of the radiation in the individual layers of the atmosphere, the degree of smoothing being characterized by the kernel of Eq. (56). For this reason some of the details of the unknown functions, such as the previously mentioned sharp changes of the temperature gradient, are either completely smoothed away and do not show up at all in the measured intensity I_ν, or else contribute an amount at the level of the random errors of the measurements of I_ν or of the errors of the calculations. In the inversion of equations of the type of Eq. (56) such errors are amplified, which leads to instability of the solution.

Equation (56) is nonlinear in the function $T(\zeta)$, but it can be linearized to an accuracy sufficient for practical purposes. By using the relation $\epsilon_\nu(T) = B_\nu(T)/B_0(T)$, where $B_0(T)$ is the Planck function for one of the N sections of the CO_2 band that are being used, and neglecting the dependence of ϵ_ν on T (which is permissible only for a very narrow spectral interval in the band), Yamamoto[69] derived the equation

$$\frac{I_\nu}{\epsilon_\nu} = B_0[T(1)]\,P_\nu[w(1)] - \int_0^1 B_0[T(\zeta)]\frac{\partial P_\nu[w(\zeta)]}{\partial\zeta}\,d\zeta, \quad (57)$$

which is linear in $B_0(T)$.

The solution of Eq. (57) has been performed by the

*As the variable characterizing the height we here follow the common practice of using the quantity ζ, the pressure of the atmosphere at the given height.

†Actually the absorption of water vapor and aerosol can make an appreciable contribution to the radiation in the wings of the CO_2 band. Besides this, the variation of the concentration of CO_2 existing in the atmosphere can be of some importance.

use of expansions of the unknown function in terms of Legendre and Chebyshev polynomials[69] or of trigonometric functions.[72] In [16-18], and recently also in [74,75], use has been made for this purpose of the natural orthogonal functions (vectors) which are the eigenfunctions (vectors) of the correlation function (matrix) and are obtained by statistical processing of the data from aerological probing of the atmosphere. In this connection it must be noted that the natural orthogonal functions also assure the optimal approximation to any random profile from the ensemble considered, i.e., they allow one to parametrize the solution of Eq. (56) with the smallest number of expansion coefficients (as compared with any other orthogonal basis). This, however, does not automatically guarantee that the corresponding inverse problem will be solved with the accuracy indicated by the errors of measurement of I_ν. The point is that in the formal application of such expansions one does not always have good matching of the number of terms used with the kernel of Eq. (56) (such a matching has been carried out, for example, in the previously mentioned papers [62, 63]).

The problem of determining vertical temperature profiles has been solved in [76, 77] by the Tikhonov method.[56-58].

A linearization of Eq. (56) which is convenient for the application of statistical methods of regularization can be obtained by representing $T(\zeta)$ as a sum

$$T(\zeta)=\overline{T}(\zeta)+T'(\zeta), \tag{58}$$

where $\overline{T}(\zeta)$ is the mean vertical temperature distribution for the given ensemble and $T'(\zeta)$ is the deviation from $\overline{T}$ ($T' \ll \overline{T}$). Substituting (58) in (56) and using only the linear terms in the expansion $B_\nu(T) = B_\nu(\overline{T}) + (\partial B_\nu(\overline{T})/\partial T)T' + \dots$, we get a linear equation for $T'(\zeta)$:

$$f_\nu = \frac{\partial B_\nu[\overline{T}(1)]}{\partial T}\tilde{P}_\nu(1) - \int_0^1 \frac{\partial B_\nu[\overline{T}(\zeta)]}{\partial T} T'(\zeta)\frac{\partial \tilde{P}_\nu(\zeta)}{\partial \zeta}\,d\zeta, \tag{59}$$

where $f_\nu = I_\nu - \overline{I}_\nu$, $\overline{I}_\nu = I_\nu[\overline{T}(\zeta)]$, $\tilde{P}_\nu(\zeta) = P_\nu[w(\zeta)]$.

Determinations of vertical temperature profiles by applying the statistical method of regularization to Eq. (59) were made in [78] from balloon measurements of the proper radiation I_ν, made with a many-channel spectrometer in five sections of the CO_2 band at 15 μ (677.5, 691, 703, and 709 cm^{-1}) and in the "transparency window" at 899 cm^{-1} [75] (resolution ~ 5 cm^{-1}), and also emission spectra measured by means of a Fourier spectrometer with a resolution of the order of 2 cm^{-1}.[79] The measurements of [75] and [79] were made in the same districts (Palestine, Texas and Sioux Falls, South Dakota) in summer. According to the authors of [75] and [79] the errors of the measurements did not exceed 1 percent.

The vertical temperature and humidity profiles were measured simultaneously with I_ν, which made it possible, on one hand, to check the methods of reconstructing $T(\zeta)$, and on the other hand, to take the actual transmissivity of the atmosphere into account in the kernel of Eq. (59). For the kernel of Eq. (59) use was made of the transmission functions calculated in [75] for the districts where I_ν was measured.

The determination of the variations of the vertical profiles $T(\zeta)$ was carried out in [78] with closed and unclosed schemes. In the closed scheme (which we shall call A_1) Eq. (56) was used to calculate I_ν and $\overline{I}_\nu$ from a known realization of the profile $T(\zeta)$ and the mean profile $\overline{T}(\zeta)$. Then a random error with mean-square deviation 1 percent was added to I_ν, after which $T(\zeta)$ was reconstructed by application of Eqs. (38) and (39) to Eq. (59). Comparisons of profiles $T(\zeta)$ reconstructed with scheme A_1 with the original profiles are shown in Fig. 3.

In evaluating these results we must keep in mind that we must judge the effectiveness of the statistical regularization method itself only in terms of the "closed scheme." When the "unclosed scheme" is used physical factors not taken into account (or incorrectly taken into account) in the original equation may be of importance. In fact, the mean-square error σ_T of the recon-

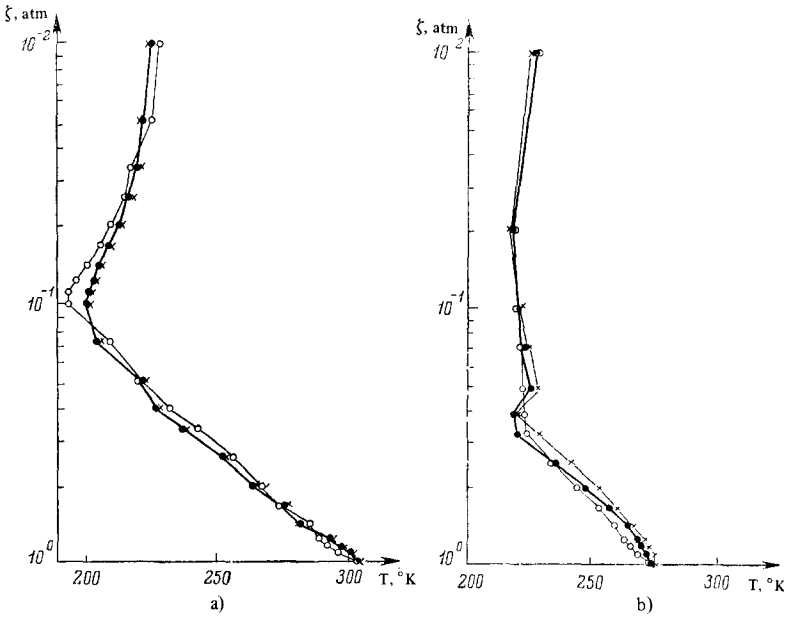

FIG. 3. Examples of the reconstruction of vertical temperature profiles by the statistical-regularization method O and X are the respective reconstructions by the closed and unclosed schemes from balloon measurements of I_ν in the CO_2 band near 15μ [75]: (a) over Palestine, Texas, and (b) over Sioux Falls, South Dakota; points O show the true profiles.

struction of $T(\zeta)$ with the scheme A_1 does not exceed 3° at the level most unfavorable for the method, that of the tropopause ($\zeta = 0.25$–0.3 atm, $z = 10$–12 km). On the other hand the actual errors in the reconstruction of $T(\zeta)$ by use of Eqs. (38) and (39) with direct measurements of I_ν (the unclosed scheme A_2) were somewhat larger, as can be seen from Fig. 3. The point is that with a given kernel and method of calculation the actual error of the original information f_ν is in most cases several times the error of the measurements of I_ν.

The question of the influence of errors in the kernel of Eq. (59) on the errors in determining the profile of temperature or some other atmospheric parameter is extremely important for the practical application of any method to problems of the optical probing of the atmosphere. Under actual conditions the transmission function P_ν in the CO_2 band 15 μ depends on a number of factors, for whose determination it is necessary to solve inverse problems on the basis of certain supplementary information. The most important of these factors is the absorption of thermal radiation by the aqueous aerosol, including water droplets in clouds and water vapor, which gives bands whose wings overlap the CO_2 band 15 μ.

No well-founded method for taking the aerosol absorption into account can as yet be proposed, since there has not been much study of the optical properties of aerosol and clouds in the infrared region of the spectrum. As for correcting for the water vapor, here it is necessary to solve an inverse problem to determine the vertical profile of its concentration, using data on the Earth's emission in the bands of H_2O vapor.

Determination of the Vertical Humidity Profile of the Atmosphere

The vertical profile of the relative water-vapor concentration $q(\zeta)$, or the profile $w(\zeta)$ of the effective mass of the vapor, can be determined from Eq. (56) if one has measured the radiation of the Earth in some water-vapor band, for example, in the band at 6.3 μ. In this case the kernel of Eq. (56) is the function $B_\nu[T(\zeta)]$, i.e., to determine $q(\zeta)$ or $w(\zeta)$ one must have the temperature profile of the atmosphere. Consequently it is appropriate to formulate the complex problem of the determination of $T(\zeta)$ and $q(\zeta)$ from simultaneous measurements of I_ν in bands of CO_2 and of H_2O. Then the profile $T(\zeta)$ determined from the radiation of the Earth in the CO_2 band near 15 μ without taking the water-vapor absorption into account, regarded as a first approximation, can be used to determine $q(\zeta)$ from measurements in the H_2O band near 6.3 μ, after which, if necessary, one can improve both solutions by taking into account the overlap of the CO_2 and water-vapor bands.

The physical principle on which the possibility of determining the humidity profile is based is analogous to that for determining the temperature profile. The atmospheric radiation in the central part of the band carries information about small concentrations of water vapor in the upper troposphere or the stratosphere. In the regions of smaller absorption and in the wings of the band I_ν gives information about the humidity in the lower layers of the atmosphere. The weights with which this information affects I_ν are determined by the vertical temperature profiles and the nature of the dependence of the transmission function on the humidity.

The problem of determining $q(\zeta)$ is an incorrectly posed problem, somewhat more complicated than that of determining $T(\zeta)$. The complication is primarily due to the essentially nonlinear dependence of the transmission function on $q(\zeta)$. Besides this, P_ν depends directly on $w(\zeta)$, and consequently with respect to $q(\zeta)$ there is a double smoothing effect in height. This increases the sensitivity of $q(\zeta)$ to the random errors of measurement and of the calculations, as has been illustrated in [18], where the profile of $q(\zeta)$ was determined by expansion of the deviation $q'(\zeta) = q(\zeta) - \bar{q}(\zeta)$ in terms of the eigenvectors of the correlation matrix of an ensemble of humidity profiles [$\bar{q}(\zeta)$ is the mean profile for this ensemble]. But this approach, like that proposed in [80] and [81], of representing the profiles $q(\zeta)$ in parabolic form $q_0\zeta^\kappa$, with unknown parameters q_0 and κ, does not assure a reliable solution of an incorrectly posed problem. Particularly large errors are obtained in a region of strong variation of the unknown function (for example, in the surface layer of the atmosphere).

The application of the method of statistical regularization to the equation

$$f_\nu = B_\nu[T(1)]\frac{\partial P_\nu[\bar{w}(1)]}{\partial w}w'(1) - \int_0^1 B_\nu[T(\zeta)]\frac{\partial}{\partial\zeta}\left\{\frac{\partial P_\nu[\bar{w}(\zeta)]}{\partial w}w'(\zeta)\right\}d\zeta, \tag{60}$$

obtained by linearization of (56) with respect to $w'(\zeta) = w(\zeta) - \bar{w}(\zeta)$, where

$$\bar{w}(\zeta) = w_0\int_0^\zeta t^n\bar{q}(t)\,dt, \quad w'(\zeta) = w_0\int_0^\zeta t^n q'(t)\,dt,$$

$$f_\nu = I_\nu - \bar{I}_\nu, \quad \bar{I}_\nu = I_\nu[T(\zeta), \bar{q}(\zeta)],$$

makes it possible to determine the vertical profile of $w(\zeta)$ directly.[82] As the a priori information one uses the correlation matrix constructed from the variations of the profiles $w'(\zeta)$. The profiles $q(\zeta)$ are calculated from the $w(\zeta)$ so reconstructed, by differentiating with respect to ζ. As before, the reconstruction of $w(\zeta)$ and $q(\zeta)$ is done with closed (A_1) and unclosed (A_2) schemes. The scheme A_1, which allows us to check the effectiveness of the method of inversion, begins with the calculation of the quantities I_ν and $\bar{I}_\nu$ from known profiles $T(\zeta)$, $q(\zeta)$, and $\bar{q}(\zeta)$. Then, on the assumption that $T(\zeta)$ and $\bar{q}(\zeta)$ are known, the profiles $w(\zeta)$ and $q(\zeta)$ are reconstructed from the calculated I_ν after a random error with given dispersion has been added to them. In the scheme A_2 the profiles $q(\zeta)$ and $w(\zeta)$ are reconstructed directly from measured values of I_ν, whose errors of measurement are characterized by a given dispersion.

The results of the determination of $w(\zeta)$ and $q(\zeta)$ by the method of statistical regularization, as obtained in [82] from the measurements of I_ν given in [79], are shown in Fig. 4. It is not hard to see (cf. Fig. 4, a) that the profile $\tilde{w}(\zeta)$ reconstructed by the closed scheme practically coincides with the original profile. With the unclosed scheme the agreement of $\tilde{w}(\zeta)$ and $w(\zeta)$ is also quite satisfactory, although the error in the reconstruction of $w(\zeta)$ in the layer next the ground may be as much as 10 percent. However, even these errors become substantial as soon as we proceed to the determination of the profiles $q(\zeta)$ by differentiation of $w(\zeta)$.

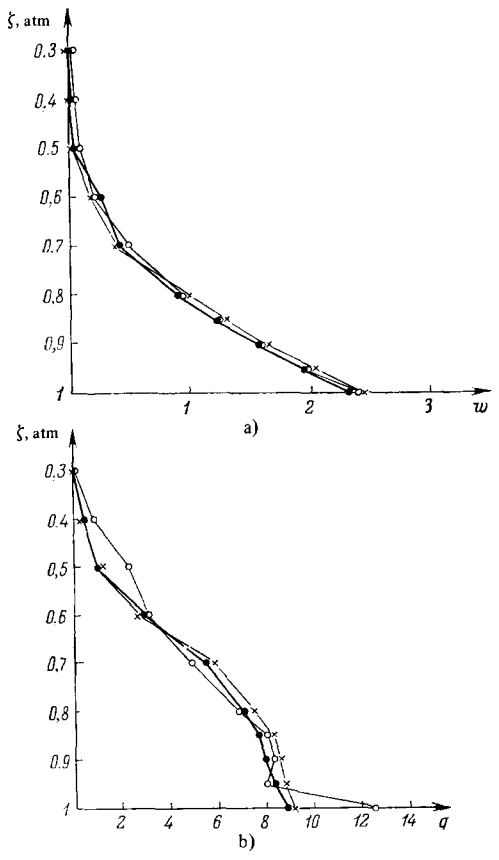

FIG. 4. Examples [82] of the reconstruction of vertical profiles of (a) the effective mass w(ζ) and (b) the concentration q(ζ) of water vapor from balloon measurements of I in the 6.3 band [79].

It is seen from Fig. 4, b that in the layer next the ground we cannot get an approximation to the original distribution q(ζ), which here has a large gradient, no matter which scheme is used, the closed or the unclosed.

We can, however, apply the method of statistical regularization for a profile q(ζ) represented as an expansion in terms of eigenvectors. The directly determined quantities are then the coefficients q_k in the expansion. The a priori information necessary for the determination of the q_k will be present in the diagonal matrix of the eigenvalues of the correlation matrix of the profiles q(ζ). It must be stated concerning this that for a satisfactory reconstruction of the profile q(ζ) in the surface layer of the air and in other regions of strong variation of the gradient of q(ζ) one needs more detailed a priori information in the ζ scale, which in turn requires the carrying out of balloon soundings of the atmosphere at a large number of levels.

CONCLUSION

The concept of incorrectly posed problems arose in mathematical physics, and therefore attempts to solve such problems were originally confined to the framework of the analytic approach. In applications, however, incorrectly posed problems usually arise as problems of the processing of experimental data, so that they began to be approached more and more often from the standpoint of mathematical statistics. With this approach it became clear that the difference between incorrectly and correctly posed problems is that satisfactory solution of the former depends on nontrivial additional information about the unknown function, independent of (a priori with respect to) the experiment on the basis of which the equation to be solved was set up. The a priori information can be more or less detailed and have as its source either general considerations arising from the physical nature of the problem (for example, smoothness of the unknown function) or concrete experimental data.

The explicit introduction of the a priori information makes it possible to determine the error of the solution rigorously. Unlike the purely analytic approach, which gives only an upper limit (usually put much too high) on the error, the statistical approach gives the mean-square error the physicist needs to interpret the results. It also makes it possible to estimate the information content of the experiment, i.e., the quantity and structure of the actual information independently of the procedure of solution. These estimates are the actual basis for the optimal planning of the experiment, including the choice of the scheme for the measurements and the optimization of the measuring apparatus.

In conclusion we note that statistical methods for the solution of incorrectly posed problems have demonstrated their effectiveness in a number of practically important applications and are now being more and more widely applied. Further progress of the statistical methods for solution of incorrectly posed problems will mainly consist of an increase in skill in introducing insufficient a priori information into the formulation of the problem and in development of the mathematical technique of solution, when the problem has been formulated and the kernels of the equations which describe the transformation of the unknown quantities into those that are measured are known with sufficient accuracy. In cases in which the kernels of the equations are subject to intrinsic errors (many problems of the processing of experimental data are cases of this sort), we have the important problem of insuring that the parameters used in the treatment are sufficiently representative. This problem, however, is inherent in all methods for the solution of incorrectly posed problems.

The writers regard it as their pleasant duty to express their gratitude to G. V. Rozenberg for his interest in this work and for helpful discussions.

[1] A. N. Tikhonov, Dokl. Akad. Nauk SSSR 39, 195 (1943).
[2] P. S. Novikov, Dokl. Akad. Nauk SSSR 18, 165 (1938).
[3] B. C. Bullard and R. J. B. Cooper, Proc. Roy. Soc. (London) A194, 332 (1948).
[4] F. John, Ann. mat. pura e appl. 4, 129 (1955).
[5] B. S. Tsybakov and V. P. Yakovlev, Izv. vuzov (radiophysika) 1, No. 5–6, 98 (1958).
[6] L. D. Kaplan, J. Opt. Soc. Amer. 49, 1004 (1959).
[7] J. A. Curcio, J. Opt. Soc. Amer. 51, 547 (1961).
[8] M. V. Maslennikov, Zh. vychisl. matem. i matem. fiz. 2, 1044 (1962).
[9] S. Twomey and G. T. Severynse, J. Atm. Sci. 20, 329 (1963).

[10] A Hora, Optik 19, 357 (1963).
[11] A. Hora, Optik 19, 409 (1963).
[12] M. P. Freeman and S. Katz, J. Opt. Soc. Amer. 53, 1172 (1963).
[13] P. Elder, T. Jerric, and J. M. Birkenland, Appl. Opt. 4, 589 (1965).
[14] R. N. Tourin and B. Kranow, Appl. Opt. 4, 237 (1965).
[15] M. T. Chahine, J. Opt. Soc. Am. 58, 1634 (1968).
[16] M. S. Malkevich, Some problems on interpretation of radiation measurements from satellites. Proc. XV Astron. Congress, v. II, Warszawa, 1965, p. 123.
[17] M. S. Malkevich and V. I. Tatarskiĭ, Kosm. issledovaniya 3, 444 (1965).
[18] M. S. Malkevich and V. I. Tatarskiĭ, in: Issled. kosm. prostranstva (Research in Cosmic Space), Moscow, "Nauka," 1965, p. 104.
[19] N. D. Nyuberg, Dokl. Akad. Nauk SSSR 4, 278 (1934).
[20] R. N. Bracewell and J. A. Roberts, Austral. J. Phys. 7, 615 (1954).
[21] R. N. Bracewell, Proc. IRE 46, 106 (1958).
[22] S. G. Rautian, Usp. Fiz. Nauk 66, 475 (1958) [Sov. Phys.-Usp. 1, 245 (1958)].
[23] H. J. London and W. L. Miranker, J. Math. Analysis and Appl. 2, 97 (1963).
[24] J. Comes and V. Nazel, J. Phys. et le Radium 22, 359 (1961).
[25] J. S. Rollett and L. A. Higgs, Proc. Phys. Soc. 79, 87 (1962).
[26] A. F. Bondarev, Optika i spektroskopiya 12, 510 (1962) [Optics and Spectroscopy 12, 282 (1962)].
[27] B. N. Grechushnikov, Optika i spektroskopiya 12, 135 (1962) [Optics and Spectroscopy 12, 70 (1962)].
[28] A. A. Dmitriev, Kosm. issledovaniya 1, 221 (1963).
[29] A. Girard, Communication presented at symposium on molecular structure and spectroscopy. Ohio State Univ., Columbus, 1963, Preprint.
[30] A. S. Kagan and V. M. Snovidov, Zh. Tekh. Fiz. 34, 759 (1964) [Sov. Phys.-Tech. Phys. 9, 579 (1964)].
[31] V. P. Kozlov, Optika i spektroskopiya 16, 501 (1964) [Optics and Spectroscopy 16, 271 (1964)].
[32] J. L. Harris, J. Opt. Soc. Am. 54, 606 (1964).
[33] C. W. Helstrom, The detection and resolution of optical signals, IEEE Trans. on Infrared Theory JT-10, 1964, page 275.
[34] G. G. Petrash, Trudy Fiz. Inst. Akad. Nauk SSSR 27, 3 (1964).
[35] R. G. Akopdzhanov and E. E. Vaĭnshteĭn, Optika i spektroskopiya 18, 495 (1964) [Optics and Spectroscopy 18, 278 (1964)].
[36] N. S. Shestov, Vydelenie opticheskikh signalov na fone sluchaĭnikh pomekh (Detection of Optical Signals in a Background of Random Noise), Moscow, "Soviet Radio," 1967.
[37] V. P. Kozlov, K voprosu ob optimal'noĭ reduktsii v teorii spektral'nykh priborov (The Problem of Optimal Reduction in the Theory of Spectral Devices), Proc. 16th All-Union Conf. on Spectroscopy (Moscow, 1965), Moscow, "Nauka", 1969, p. 73.
[38] G. Fox and E. T. Goodwin, Phil. Trans. Roy. Soc. (London) A245, 501 (1953).
[39] M. M. Lavrent'ev, O nekotorykh nekorrektnykh zadachakh matematicheskoĭ fiziki (Some Incorrectly Posed Problems of Mathematical Physics), Novosibirsk. Izd. Sib. Otd. Akad. Nauk SSSR, 1962.
[40] L. A. Khalfin, O Teoretiko-informatsionnom podkhode k teorii spektral'nykh priborov (Information-theory Approach to the Theory of Spectral Devices), Proc. All-Union Conf. on Prob. Theory and Math. Statistics (Erevan), Izd. Akad. Nauk Arm.SSR, 1960, p. 187.
[41] J. Hadamard, Sur les problemes aux derivees particelles et leur significations physiques, Bull. Princeton Univ. 13, 82 (1902).
[42] V. K. Ivanov, O nekorrektno postavlennykh zadachakh (On Incorrectly Posed Problems), Matem. Sb., nov. ser. 61 (103), 211 (1963).
[43] L. A. Khalfin and V. N. Sudakov, Dokl. Akad. Nauk SSSR 157, 1058 (1964).
[44] D. K. Fadeev, Trudy Mat. Inst. Akad. Nauk SSSR 53, 387 (1959).
[45] D. K. Fadeev and V. N. Fadeeva, Zh. vychisl. matem. i matem. fiz. 1, 412 (1961).
[46] K. S. Shifrin and A. Ya. Perel'man, Optika i spektroskopiya 15, 533, 667, 803 (1963) [Optics and Spectroscopy 15, 285, 362, 434 (1963)].
[47] K. S. Shifrin and A. Ya. Perel'man, Dokl. Akad. Nauk SSSR 158, 578 (1964) [Sov. Phys.-Dokl. 9, 809 (1965)].
[48] K. S. Shifrin and A. Ya. Perel'man, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 1, 964 (1965).
[49] K. S. Shifrin and I. B. Kolmakov, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 2, 851 (1966).
[50] K. S. Shifrin, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 2, 928 (1966).
[51] G. V. Rozenberg, Usp. Fiz. Nauk 91, 569 (1967) [Sov. Phys.-Usp. 10, 188 (1967)].
[52] G. V. Rozenberg, Usp. Fiz. Nauk 95, 159 (1968) [Sov. Phys.-Usp. 11, 353 (1968)].
[53] D. L. Phillips, J. Associat. Comput. Machin. 9, 84 (1962).
[54] S. Twomey, J. Associat. Comput. Machin. 10, 97 (1963).
[55] F. Fridrikh, in: Metody vychisleniĭ (Computational Methods), No. 4, Izd. Leningrad Univ., 1967, p. 102.
[56] A. N. Tikhonov, Dokl. Akad. Nauk SSSR 151, 501 (1963).
[57] A. N. Tikhonov, Dokl. Akad. Nauk SSSR 153, 49 (1963).
[58] A. N. Tikhonov and V. B. Glasko, Zh. Vychisl. matem. i matem. fiz. 4, 564 (1964).
[59] V. Ya. Arsenin and V. K. Ivanov, Zh. vychisl. matem. i matem. fiz. 8, 310 (1968).
[60] V. Ya. Arsenin and V. K. Ivanov, Dokl. Akad. Nauk SSSR 182, 9 (1968).
[61] V. K. Ivanov, Zh. vychisl matem. i matem. fiz. 6, 1089 (1966).
[62] V. P. Kozlov, Dokl. Akad. Nauk SSSR 166, 779 (1966).
[63] V. P. Kozlov, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 2, 137 (1966).
[64] S. Kul'bak, Teoriya informatsii i statistika (Information Theory and Statistics), Moscow, "Nauka", 1967.
[65] V. F. Turchin, Zh. vychisl. matem. i matem. fiz. 7, 1270 (1967).
[66] V. F. Turchin, Zh. vychisl. matem. i matem. fiz. 8, 230 (1968).
[67] V. F. Turchin and V. Z. Nozik, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana," 5, 29 (1969).

[68] V. F. Turchin and V. Z. Nozik, Preprint 138, Physics-Energetics Inst., 1969.
[69] G. Yamamoto, J. Meteorol. 18, 581 (1961).
[70] D. Q. Wark, J. Geophys. Res. 66, 77 (1961).
[71] M. S. Malkevich, Kosm. issledovaniya 2, 246 (1964).
[72] D. T. Hilleary, D. Q. Wark, and D. G. James, Nature 205, 489 (1965).
[73] M. S. Malkevich, V. P. Kozlov, and I. A. Gorchakova, Tellus 21, 389 (1969).
[74] D. Q. Wark and H. E. Fleming, Monthly Weath. Rev. 94, 351 (1966).
[75] D. Q. Wark, F. Saiedy, and D. Q. James, Monthly Weath. Rev. 95, 468 (1967).
[76] V. B. Glasko and Yu. M. Timofeev, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 4, 303 (1968).
[77] O. M. Pokrovskiĭ and Yu. M. Timofeev, Izv. Akad. Nauk SSSR, Ser. "Fiz. atm. i okeana" 5, 1082 (1969).
[78] V. F. Turchin, M. S. Malkevich, and I. A. Gorchakova, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 5, 449 (1969).
[79] L. W. Chaney, L. T. Loh, and M. T. Surh, A Fourier transform spectrometer for measurement of atmospheric thermal radiation, Techn. Rep. Univ. of Michigan, 05863-12-T, May, 1967.
[80] W. L. Smith, Monthly Weath. Rev. 95, 363 (1967).
[81] W. L. Smith, Monthly Weath. Rev. 96, 387 (1968).
[82] I. A. Gorchakova, M. S. Malkevich, and V. F. Turchin, Izv. Akad. Nauk SSSR, ser. "Fiz. atm. i okeana" 6, 565 (1970).
[83] O. N. Strand and E. R. Westwater, J. Associat. Comput. Machin. 15, 100 (1968).
[84] E. R. Westwater and O. N. Strand, J. Atm. Sci. 25, 750 (1968).

Translated by W. H. Furry