Automated data processing in discrete nuclear spectrometry

V. Gadzhokov

Institute of Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia
Fiz. Elem. Chastits At. Yadra **11**, 1474–1518 (November–December 1980)

The review is devoted to automated computer processing of data obtained in measurements of discrete nuclear spectra. Such data are processed in two stages, the first being carried out interactively on a graphical display, while the second step, which is fully automated, solves the succession of monotypic nonlinear problems generated in the first step. The algorithms of both stages are considered in detail. The requirements that the algorithms must satisfy in order to be universally applicable to the given class of problems are discussed. The construction of such algorithms on the basis of regularized iteration processes is also discussed.

PACS numbers: 29.80.Cp

INTRODUCTION

Automation of spectrometric investigations involves in the first place the use of computers to accumulate, transmit, sort, and store experimental data and also to control the spectrometric experiment. These laborious processes can be formalized and described with comparative ease by means of the languages available for modern computers. It is natural to begin automated investigations with them. In the present paper, we consider effective ways of approaching the problem of automated mathematical processing of data obtained in measurements of the discrete spectra of nuclear radiation. In other words, we shall be concerned with a natural extension of the process of automating spectrometric investigations.

The present review is based mainly on experience accumulated in the Department of Nuclear Spectroscopy and Radiochemistry at the Laboratory of Nuclear Problems at Dubna from 1969 on. The author is very grateful to the head of the Department and the director of the Laboratory of Nuclear Problems for providing him with the possibility of collaborating with this outstanding group of investigators for a number of years. Of course, all comments in the review that might appear contentious or erroneous are solely the responsibility of the author and not of the group.

1. FORMULATION OF THE PROBLEM AND METHODS OF ITS SOLUTION

1.1. *Preliminary comments.* By the spectrum we shall mean the distribution of some physical quantity (for example, the intensity U of radiation) with respect to some parameter (for example, with respect to the energy E of the radiation). The function $U(E)$ describing the density of this distribution is usually called the spectral function.

The task of nuclear spectroscopy is fairly clear from its name. The following stages can be distinguished[1,2,3]: detection of the radiation, analog-to-digital conversion of the signals, accumulation of information, preliminary processing of the data, calculation of the basic parameters of the spectrum, calibration with respect to the efficiency and the energy, and interpretation of the results (construction of decay schemes, search for suitable nuclear models, etc.).

The development and improvements in the technical means of detection of radiation and the recording of signals has led to a situation in which for many years already it has been impossible to imagine a serious spectrometric investigation without the use of computers.[4,5] As regards the amount of computational work, the most laborious stages are the preliminary processing of the data and the calculation of the parameters of the spectrum. It is on these stages that we shall concentrate our attention in what follows. Without in any way belittling the intellectual and technical difficulties present in the other stages of nuclear spectroscopy, it can be asserted that it is in these two stages that the investigator encounters the greatest mathematical difficulties.

These two stages are characterized by the following features:

a) a large amount of information is fed in;
b) it is expedient to involve the human intellect in the preliminary processing;
c) the main mathematical problem to be solved is nonlinear;
d) a large number of calculations must be made;
e) a comparatively small amount of information is extracted.

In considering these features, we shall assume that the preceding stages have been completed, i.e., the measurements have ended and the accumulated information is stored in the computer memory (or in machine-readable form).

1.2. *The basic problem of spectrometry.* Any real measuring apparatus has a specific response function $K(E, e)$, and it is not the spectral function $U(E)$ but the instrumental spectrum $A(E)$ which is obtained at its output:

$$A(E)=\int_a^b K(E, e)\,U(e)\,de, \tag{1}$$

where a and b are the lower and upper limits of the sensitivity region of the spectrometer.

Equation (1) is a Fredholm integral equation of the first kind (see, for example, Refs. 1, 4, and 6–8). The solution of this equation, i.e., the finding of the unknown function $U(e)$ for given $A(E)$ and $K(E, e)$, is the basic problem of spectrometry. We note some fea-

 0090-4759/80/060584-20$02.80

tures of this problem:

a) it is improperly posed[1];

b) the spectrometer response function $K(E, e)$ is by no means always known with sufficient degree of accuracy, and serious difficulties are encountered in an attempt at specifying it analytically or measuring it directly;

c) the instrumental spectrum $A(E)$ is by nature a continuous function, and in the process of analog-to-digital conversion it is subjected to a "quantization" with step equal to the conversion step. Mathematically, this is equivalent to integrating the true instrumental spectrum within each conversion step. Because of features a) and b), one resorts to solution of the basic problem of spectrometry in the form (1) only when it is impossible to simplify it and reduce it to another, easier problem as, for example, in the case of bremsstrahlung or moderated neutrons, whose spectra are certainly continuous.[6,9] The feature c) is very often simply ignored. This is very understandable, since the spectrometric problem is sufficiently complicated in its own right[10] and its solution by traditional methods frequently leads to insuperable difficulties, so that one attempts to avoid additional complications associated with allowance for integration.

1.3. *The case of a discrete spectrum.* The spectrum is assumed to be discrete when it consists of individual, clearly separated lines. This means that the natural width Γ of the lines is much less than the distance ΔE between them. Ignoring the natural width, such a spectrum is conveniently represented in the form of a linear combination of a finite number of delta functions:

$$U(e)=\sum_{i=1}^{k} A_i\delta(e-e_i), \tag{2}$$

where e_i are the characteristic energies of the spectral lines, A_i are the intensities (amplitudes) of the corresponding lines, and k is the number of lines in the spectrum.

Substituting (2) in (1), we obtain

$$A(E)=\sum_{i=1}^{k} A_iK_i(E), \tag{3}$$

where we have adopted the notation $K_i(E)=K(E, e_i)$. It can be seen that when the investigated spectrum is discrete, the response function decomposes into k independent components, and the integration reduces to summation over the components with the weights A_i.

Equation (3) expresses the density of the distribution of signals at the output of the spectrometric circuitry before the analog-to-digital converter. Following the established terminology, we shall call the conversion step the channel width of the converter. In practice one usually employs (or, at least, attempts to) converters with constant channel width, so that we shall restrict ourselves to this case, assuming for simplicity that the channel width is unity. We shall assume that all the channels are numbered, the numbers being equal to the values of the upper threshold $q=1, 2, \ldots, Q$. Obviously, the total number of channels is Q. Since the converter reacts in the same way to all signals that arrive between the lower and upper thresholds of the given channel q, the total number Y of signals recorded in this channel is obtained by integrating the distribution (3) within the thresholds:

$$Y_q=\sum_{i=1}^{k} A_i\int_{q-1}^{q} K_i(E)\,dE, \quad q=1, 2, \ldots, Q. \tag{4}$$

The relations (4) constitute a system of nonlinear equations for the unknown parameters of the spectrum, which occur linearly on the right-hand sides of the equations (the line intensities occur explicitly and linearly, while the remaining parameters occur implicitly and nonlinearly under the integral). It is important that the left-hand sides of (4) are measured directly. Generally, these are the numbers of pulses measured in the channels of the analyzer, but the left-hand sides may also be powers of the photographic density on a plate used as a detector in a magnetic spectrometer for conversion electrons,[11] the counting rates of the detector in an α spectrometer at different currents of the magnet, etc. In this exposition, we shall usually be thinking of only the numbers of pulses in the channels of an analyzer, but the conclusions remain valid for other discrete spectra.

The set of numbers Y_q constitutes the data that are accumulated by means of the spectrometer when its detector is exposed to a mixture of k monochromatic lines of different intensities. The complete image of the required spectrum $U(e)$ is hidden in these data. The aim of the solution of the system (4) is to reconstruct this image, i.e., to find the individual discrete lines in the experimentally measured spectrum. It is clear that the solution of this problem depends on the resolution of the spectrometer (the half-width of the instrumental line). We note that, as a rule, the resolution of a given instrument depends weakly on the energy of the radiation.

Since the resolution of the spectrometers used is always worse than one would wish, the aim of the solution of the system (4) can be formulated differently as the reliable extraction of the greatest possible amount of useful information hidden in the experimental data. We shall show in what follows that if suitable algorithms are employed this aim can be well achieved even for lines separated by less than the resolution. In other words, by a more accurate solution of the system (4) one can "improve" the spectrometer.

Note that when the channel width is appreciably less than the half-width of an individual line (which is the case, for example, in scintillation spectrometers), the corrections for integration of the instrumental spectrum within an individual channel become negligibly small, and the comment made earlier ceases to hold.

1.4. *Generation of a succession of monotypic nonlinear problems.* The circumstance that all real spectrometers have a finite resolution leads, in conjunction with the physical fact of nonuniformity in the distribution of the lines over the spectrum, to very important consequences.[12-14]

First, the images of the true lines $\{Y_q\}$ in the mea-

sured spectrum are grouped naturally into clusters, in which the distances between the neighboring lines are comparable with the half-width h of an individual line, i.e., if the distances exceed the width, then by not more than a few times (for example, by 2–3 times in the case of a Gaussian profile of an individual line). In contrast, the distance between neighboring clusters is greater than the maximal distance between neighboring lines within a cluster.

Second, the contribution of an individual line to the number of recorded signals Y_q is extended to only the channel numbers q corresponding to the given cluster of lines. The contribution of the lines of a given cluster to the channels of the neighboring clusters is negligibly small compared with both the recorded numbers of signals and the unavoidable errors of measurement.

Third, since lines of a given cluster are close to each other, and the resolution of the spectrometer depends weakly on the parameter with respect to which the spectrum is represented, it can be assumed to a good accuracy that the resolution is the same for all lines of the cluster (and, of course, changes slightly on the transition to the neighboring clusters).

Suppose a cluster of lines lies completely between the channels with numbers from q_{in} to q_e. We shall call the ordered set

$$\{Y_q\}_{q=q_{in}, q_{in}+1, \ldots, q_e} \tag{5}$$

a section of the spectrum. By the definition of a section and the consequences just listed, the equations of the system (4) which correspond to the set (5) are independent of the remaining equations of the system and can be solved separately. Therefore, if the complete spectrum is divided suitably into sections, the system of nonlinear equations (4) decomposes into a finite number of independent nonlinear systems of lower dimension, these corresponding to the different sections. In this case, we can speak of a partial decomposition of the original spectrum, or we can say that the system (4) generates a succession of monotypic nonlinear problems[15] involving processing of sections of the type (5). If we can construct a universal algorithm[15] for solving the problems in this succession, the problem of automating the processing of discrete spectra will be to a considerable extent solved.

The reduction of the original problem to a succession of problems of reduced dimension considerably facilitates data processing on a computer, since it reduces the requirements on the operative memory of the computer, and also the total number of necessary elementary operations. This is very important, since without such reduction the processing of spectra by minicomputers is not possible at all.

For example, the γ-ray spectra of isotopes in the region of deformed nuclei may consist of several hundred individual lines. One such spectrum, measured with a spectrometer with a semiconductor detector and recorded in 4096 channels, usually breaks up into around 50–100 sections and therefore generates a succession of the same number of monotypic nonlinear processing problems. In a limiting case, an individual section may contain only one line. The average length of a section is 30–50 channels, which is considerably shorter than the length of the original spectrum.

The decomposition of the spectrum into sections is by no means a formal problem; moreover, its solution is not, as a rule, unique. To make the decomposition correctly, one must be able to recognize single and overlapping lines and clusters of them; in complicated cases, one also requires imagination, experience, and intuition. Since we have not yet learnt to program these attributes of the mind sufficiently well, participation of the human intellect in the decomposition stage is particularly important. If a sensible separation of the sections from the original spectrum is made, the final values of the calculated spectral parameters hardly depend on the method of decomposition (if, of course, the method remains sensible). Therefore, the problem of processing the spectrum still has a unique solution, although the decomposition is not unique. The algorithm that is employed is of great importance for achieving uniqueness of the solution. As we shall see later, a good algorithm for processing the sections can also indicate a bad decomposition. It is also helpful to invoke additional physical considerations; for example, a sharp drop in the value of the half-width of the lines of a given section from the general smooth dependence unambiguously indicates trouble.

Thus, if the spectrum has been measured and sensibly divided into sections, to solve the succession of monotypic nonlinear problems we must:

a) construct a suitable mathematical model of the response function $K_i(E)$ of the spectrometer. The model must clearly depend on at least two parameters—the position p_i of the line center and its half-width h_i (since the line is usually smeared, one usually takes the full width at half maximum; for brevity, we shall use the expression "half-width");

b) augment the mathematical model by terms which take into account the presence of background signals in Y_q;

c) substitute the explicit form of the chosen model into the relations (4);

d) for each section of the spectrum, i.e., for each of the succession of problems, specify initial approximations for the required unknowns, including the parameters of the background terms;

e) solve the individual problems of the succession, applying to each of them a suitable iteration procedure chosen in accordance with the universal algorithm for processing the given succession;

f) for each of the problems in the succession calculate the propagated errors of the required parameters at the point of the solution that are due to the inaccuracy of the initial data as well as the incompleteness of the adopted model.

1.5. *Construction of a mathematical model of the spectrum.* The physical processes responsible for the detection of radiation by the detector have a compli-

cated and undoubtedly statistical nature. This is true in all cases of measurement of discrete nuclear spectra. When models of definite spectra are constructed, the nature of the interaction between the radiation and the detector should be taken into account as far as possible. To be definite, we shall take the example of γ spectra measured by semiconductor detectors, the case which is encountered most frequently. The arguments which follow apply to a single monochromatic γ line.

The instrumental line. The interaction of γ rays with matter has been fairly well studied. In the material of the sensitive volume of the detector, a γ ray can undergo one of the following three processes: scattering, the Compton effect being the most important among the various kinds of scattering; total absorption with transfer of the energy to one of the bound electrons (photoelectric effect); and formation of electron–positron pairs with loss of energy equal to the rest energy of the produced particles (obviously, this is a process with a threshold). Combinations are also possible such as, for example, the formation of a pair with the subsequent total absorption of the remaining γ ray, etc.[16] At the energies in which we are interested ($\lesssim$10 MeV), production of heavier pairs does not occur.

Each process of interaction of γ rays with the matter, like every possible combination of the processes, leads to a characteristic distribution of the signals at the spectrometer output. The sum of these distributions is the response function of the spectrometer for γ radiation of the given energy or, as one says, the single-line profile. Because of the multiplicity of the physical processes, this profile is fairly complicated even in the case of a monochromatic detected spectrum. Satisfactory analytic expressions describing this profile do not exist, the attempts to find such expressions having so far led to severe difficulties. Let us attempt to establish the extent to which the efforts expended on overcoming the difficulties are justified.

The aim of the processing of the spectrum is the reliable determination of the intensities and energies of the individual lines. All the important information about these physical quantities is concentrated in the total-absorption peak, whose area and position are, respectively, equal to the intensity and energy up to a relatively simple (but not necessarily linear) calibration transformation. It is also expedient to include in the processing the peaks of simple and double emissions, since they may be helpful in identifying the lines of complex spectra consisting of a set of monochromatic components. With regard to the Compton part of the spectrum, in which we can also include the backward scattering peaks, the contributions of the individual lines are so mixed up in it that its processing appears, at least at the present state of our possibilities, hardly possible and perhaps unnecessary.

Thus, because the individual components of the instrumental spectrum of an isolated γ line in a semiconductor spectrometer are not equally informative, we introduce the following additional definitions and conditions.

1. Only well-defined peaks in the distribution of the signals will be regarded as lines of the instrumental spectrum. Physically, such a line corresponds to a total-absorption peak or, less often, peaks of simple or double emission. The profiles of isolated instrumental lines are similar (only the retarded annihilation peak is broadened).

2. The instrumental lines will be identified, i.e., an unambiguous connection between peaks and particular γ lines of the investigated spectrum will be established, after processing of all sections of the spectrum.

3. All the remaining components of the instrumental spectrum will be included with the background (together with the true physical background).

4. The resolution of the spectrometer will be determined relative to the concept of the instrumental line introduced above.

With these stipulations, all the conclusions drawn earlier can be applied to the γ spectra obtained from semiconductor detectors. Note that this is characteristic of only the specific spectrometers considered. For example, the same conclusions apply to magnetic spectrometers for internal conversion electrons without the introduction of additional conditions.

Analytic description of the instrumental line. The concept of the instrumental line having been clearly defined, we can now review descriptions of its profile.

a) If the spread of signals from the detector is due solely to random (statistical) factors, a normal distribution of these signals can be expected:

$$K(E)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\exp\left[-\left(\frac{E-p}{\sigma\sqrt{2}}\right)^2\right], \tag{6}$$

where the center of the distribution p is a measure of the energy of the absorbed γ ray, the variance σ^2 is related linearly to the half-width,

$$h=2\sqrt{2\ln 2}\,\sigma, \tag{7}$$

and the intensity is measured by the area of the instrumental line, which is simply equal to the amplitude A introduced earlier (in this section, we omit the line-number index i).

The distribution (6) is the oldest, most physical, and most frequently used model, and it entered semiconductor spectroscopy from scintillation spectroscopy. It has been used by so many authors that we should be hard put to say who was the first.

b) For many detectors, the assumption a) that the spread of the signals is due to purely statistical factors is an oversimplification.[17,18] The sequence of physical processes by means of which the incident γ ray is transformed by the detector into a signal of definite amplitude is very complicated,[18] and attempts at theoretical calculation of the distribution of the signals (see, for example, Ref. 19) have not yet found practical application in the processing of spectrometric data. There are two main reasons for this. First, the theoretical expressions which are derived are themselves fairly complicated and inconvenient; second, they also

contain unknown parameters and even unknown functions such as, for example, the efficiency of carrier collection. Even when these two problems are absent, the actual calculation of the signal distribution is made under inadmissible simplifications, and we are advanced no further.

Under these conditions, it is natural to avoid both the simplicity of the distribution (6) and the inconvenience of the theoretical calculations. This can be done either by a tabulated specification of the single-line profile or by introducing into (6) additional additive or multiplicative terms. Such an approach may be termed empirio-phenomenological.

c) Tabulated specification of the single-line profile[20-22] is the most empirical approach to the construction of a mathematical model of the spectrum. If we are to implement it, we must have at our disposal a sufficiently dense set of pure reference lines over the entire investigated energy range. This condition, and also the greater demands on the operative memory for storing the tabulated tables somewhat detract from the obvious advantages of this method.

Correction terms to (6) are usually selected on the assumption that the Gaussian distribution plays the leading part but its "tails" are distorted for a number of reasons (incomplete carrier collection, imperfections of the electronics, etc.). Many different ways have been proposed for analytic description of asymmetry of the instrumental line (see, for example, Refs. 23–26). The main methods are reviewed in Refs. 27–29, in which a comparative evaluation of the different methods is made. This involved the measurement of single total-absorption lines with statistics of more than 10^6 pulses and a background of the order of several hundred events per channel. The gain and channel width were chosen to make the half-width of the total-absorption peak equal to about 30–40 channels. Under these conditions, the corrections for integration of the signal distribution within one channel of the analyzer become insignificant. The χ^2 test was used to compare the different asymmetry models for optimal values of the parameters, and the following conclusions were obtained: The central part of the total absorption peak is described well by the Gaussian distribution; asymmetry is manifested only in the tails of the pure Gaussian and at low counting rates is appreciable only to the left of the peak; the asymmetry can be most accurately described in the form of two additive terms, which are called the step and tail functions; the total distribution function of the signals near the maximum of the total-absorption peak can be represented in the form

$$F(i)=G(i)+S(i)+D(i)+B(i), \tag{8}$$

where i is the channel number, $B(i)$ is a linear background, the Gaussian is given by

$$G(i)=H_G\exp[-(i-i_0)^2/2\sigma^2], \tag{9}$$

the step function by

$$S_i=(1/2)\,H_S[1-\operatorname{erf}((i_0-i)/\sigma\sqrt{2})], \tag{10}$$

the tail function by

$$D(i)=\begin{cases} H_D\exp[(i-c)/\beta], & i\leqslant i_1; \\ H_D\exp[-(i-c)^2/2\delta^2], & i\geqslant i_1, \end{cases} \tag{11}$$

and i_1 is the point at which the two parts of $D(i)$ are joined:

$$i_1=c-2\delta^2/\beta. \tag{12}$$

We see that five additional parameters must be introduced to give a complete description of the asymmetry. For their accurate determination, it was necessary to stretch the investigated spectrum considerably and adopt a special procedure to calculate the initial approximations for the parameters. At the same time, the main physical information is still contained in the parameters of the Gaussian. For this reason, such a complete treatment of the profile of a single total-absorption line can be recommended only in the case of particularly accurate measurements. Perhaps in the future it will be possible to use it as well for routine spectroscopic investigations, in which one must at present accept half-widths of the order of ten channels.

d) In our work on computer processing of discrete spectra we adopted at an early stage the symmetric Gaussian (6) to represent the single-line profile (Refs. 4, 10, and 30–35). The advantages of this model are obvious: a clear and unambiguous physical interpretation of the parameters is combined with simplicity, which is of the greatest importance for program implementation, especially with minicomputers. Its initially obvious disadvantage in failing to match exactly the real profile of the instrumental lines proved to be short-lived. The technological developments in the production of semiconductor radiation detectors and also analog and digital electronics led to widespread use of spectrometers with instrumental distribution of the signals only slightly different from a normal distribution. In this sense, one can say that universality was added to the advantages of the model (6).

Many investigations made using the model of a symmetric Gaussian demonstrated convincingly that it can be successfully used in physics experiments (see, for example, the reviews on the Dubna YaSNAPP program,[36,37] and the conclusions drawn in the preprints of Refs. 38–43 on methodology). It was shown in Ref. 44 that the model could be used in precision γ-ray spectroscopy.

The construction of a mathematical model of the spectrum in a given fixed section is completed by the addition of terms which take into account the background. Following the proved tradition, we shall represent the background by a polynomial in the channel number. In this representation, it is not necessary to take into account the integration of the background over one channel of the converter, since the definite integral of a polynomial of given degree within each channel leads to a polynomial of the same degree.[1)] We shall not restrict ourselves to representing the background by a straight line, since this does not allow us to take into account the possibility of Compton peaks from lines with higher energy coming within the investigated section. In general, the coefficients of the background

[1)] Indeed,

$$\int_{q-1}^{q} x^n\,dx=q^n-\frac{n}{2}\,q^{n-1}+\ldots+(-1)^n.$$

polynomial as well as its degree are unknown. The coefficients of the polynomial are found in the process of solving the nonlinear problem for the given section together with the remaining unknown parameters. The degree of the polynomial can be determined in accordance with known statistical criteria[45] by processing the same section with different fixed degrees. In fact, the need for such a detailed approach to the background arises comparatively seldom. It is much more often the case that the physicist, specifying the initial approximations, correctly guesses the appropriate degree on the basis of previous experience. For the majority of sections, a linear background is indeed the most suitable; in approximately 10% of the cases, one must approximate the background by polynomials of second and third degree; cases of higher degree are extremely rare and should be regarded with suspicion, since the behavior of polynomials with such degree is rather free (and certainly does not correspond to the physics of the phenomena under consideration), which may lead to a distortion in the values of the parameters of the spectrum. In the first place, this applies to low-intensity lines.

Formulation of the nonlinear problem for processing a section of the spectrum. Substituting the explicit form (6) into the equations of the system (4) corresponding to the section (5), and making some simple manipulations, we obtain the final form of the system of nonlinear equations which must be solved in order to process the data for the section:

$$Y_q=\frac{1}{2\sqrt{\pi}}\sum_{i=1}^{k} S_i\left[J\left(\frac{q-p_i}{\sigma\sqrt{2}}\right)-J\left(\frac{q-p_i-1}{\sigma\sqrt{2}}\right)\right]+\sum_{j=0}^{l} a_j q^j, \quad (13)$$

where

$$q=q_{in},\ q_{in}+1,\ \ldots,\ q_e \quad (14)$$

and we have adopted the following notation for the error function:

$$J(x)=\int_{-x}^{x}\exp(-u^2)\,du. \quad (15)$$

The other symbols have the following meaning: σ^2 is the single-line variance, S_i is the area of peak i, p_i is the position of the center of peak i, k is the number of lines in the section, l is the degree of the polynomial, and a_j are its coefficients. The number of equations in (13) is obviously

$$m=q_e-q_{in}+1, \quad (16)$$

the number of unknowns is

$$n=2k+l+2, \quad (17)$$

and

$$m\geqslant n. \quad (18)$$

The relations (13)–(18) give the explicit form[12-14,46] of one of the monotypic nonlinear problems in the succession generated in the processing of discrete nuclear spectra.

1.6. *Features of the succession of problems generated by discrete spectroscopy.* We have already pointed out some general properties of the problems generated by spectroscopy. Let us now consider how these properties are manifested in the succession of problems to which the initial problem (1) for a discrete spectrum has been reduced.

a) The nonlinearity of each problem in the succession is due to the nonlinear dependence of the model on the parameters. Therefore, one must have recourse to an iteration procedure which begins with certain initial approximations to the unknowns. The initial approximations are selected individually for the separate problems.

b) The problem is overdetermined, the number of equations usually exceeding the number of unknowns. On account of the overdetermination, one does not solve the original system (13) on each section but rather its corresponding averaged problem, the derivative (Jacobi matrix) of the right-hand side of (13) being taken as averaging operator.[47] This operation is analogous to the transition from the conditional equations to the normal equations in the case of the least-squares method.

c) The problems are approximate because of the unavoidable spread of the measured ΔY_q. In the absence of systematic errors, the spread is assumed to be purely statistical. The different degrees of accuracy of Eqs. (13) are taken into account by the introduction of a diagonal matrix of statistical weights W, the elements of the matrix being inversely proportional to the squares of the spreads.

d) Ill conditioning of the linear steps in the solution procedure is a common phenomenon in successions of spectroscopic problems. It may be due to any one of the reasons given below, and also combinations of them.

1. The presence in the section of relatively very weak lines. In this case, the matrix of the linear step (see below) contains columns whose relative magnitude is much less than that of the others, and the numerical inversion of the matrix is difficult.

2. The presence of lines of approximately equal intensity very close to each other. In this case, the matrix of the linear step contains pairs of almost linearly dependent columns.[12] The numerical inversion of such a matrix again presents difficulties.

The inaccuracy of the initial data, although itself not capable of leading to ill conditioning, aggravates the effect of the two factors mentioned above. The universal algorithm for solving the problems must meet the requirement of being able to recognize cases of ill conditioning and must be able to adopt appropriate measures to ensure stability of the iteration process.

e) A certain number of complete problems for finding hidden features can in principle occur in the succession of problems,[34] i.e., there may be one or several sections in the succession which are so complicated that the intuition of the physicist operator by itself is inadequate to determine the number of lines k in the section. In such cases, the algorithm must satisfy the very stringent requirement of being able to determine objectively the number of lines.

Before we consider the construction of the universal algorithm for solving the succession of nonlinear problems, we consider some related questions.

1.7. *Iteration procedures for solving the systems of nonlinear equations.* Among the existing iteration procedures (see, for example, Ref. 48) that can be used to solve the problems (13)–(18), we shall consider only the ones that are actually used in the practical processing of discrete spectroscopy data. We introduce the notation

$$Y=\operatorname{col}\{Y_q\}_{q=q_{in}, q_{in}+1, \ldots, q_e},\ Y\in R^m; \tag{19}$$

$$x=\operatorname{col}(\sigma,\ S_1,\ p_1,\ \ldots,\ S_k,\ p_k,\ a_0,\ \ldots,\ a_l),\ x\in R^n; \tag{20}$$

$F\in C^1$ is the nonlinear operator of the right-hand side of (13), which maps the n-dimensional Euclidean space R^n onto R^m. Then (13) can be written in the form

$$Fx-Y=0, \tag{21}$$

or, with the notation

$$Fx-Y=fx, \tag{22}$$

we have

$$fx=0. \tag{23}$$

We also denote the derivative of the operator f by

$$f'(x)=[\partial f_\mu/\partial x_\nu]=[\partial F_\mu/\partial x_\nu], \tag{24}$$

where

$$\mu=1,\ 2,\ \ldots,\ m \tag{25}$$

and

$$\nu=1,\ 2,\ \ldots,\ n. \tag{26}$$

To take into account the approximate nature of the original equations (13), we introduce a diagonal matrix of weights W, whose nonvanishing elements are equal to

$$w_{\mu\mu}=1/(\Delta Y_\mu)^2. \tag{27}$$

The averaged problem now takes the form

$$\overline{f'(x)}Wfx=0, \tag{28}$$

where the bar over the operator denotes the transpose. Its approximate solution is usually taken to be the vector $\tilde{x}$ for which the functional

$$\theta=\sqrt{\overline{fx}Wfx/(m-n)} \tag{29}$$

is minimal. This functional is the weighted mean quadratic deviation of the left-hand side of (23) from zero. Therefore, a solution satisfying the criterion of a minimum of θ is also a solution in the least-squares sense. The presence of the normalization with respect to the degrees of freedom $m-n$ transforms the square of θ into the well-known[45] statistical χ^2 criterion.

Employed methods of solution. We shall assume that all the considered iteration processes of solution of (28) begin with some initial approximation to the vector of unknowns x^0. The specification of the initial approximations is considered later (see Sec. 2).

a) The most widely used method of solution of the posed problem is the Gauss–Newton method:

$$x^0:\ x^{t+1}=x^t-[V(x^t)]^{-1}\overline{f'(x^t)}Wfx^t, \tag{30}$$

in which we have the matrix of the linear step

$$V(x^t)=\overline{f'(x^t)}Wf'(x^t), \tag{31}$$

and the superscript t is the number of the iteration. The second term in (30) is called the vector of the linear step of the variation of the unknowns.

The advantage of the method is the rapid convergence (not worse than linear) near the point of the solution, and the main disadvantage is the small radius of convergence. This method is widely used to process discrete spectra,[20,49-51] although its successful use depends to a large extent on felicitous specification of x^0.[50] We note that *a priori* information such as bounds on the range of variation of the parameters and so forth[26] generally improves the convergence but does not guarantee it. Because of the features of the problems discussed earlier, the Gauss–Newton method frequently diverges and in principle is not suitable for solving the complete problem of finding hidden features.

b) The desire to achieve convergence of the iteration process even when (30) diverges led to a number of modifications of the Gauss–Newton method.

1. Damping of the linear step of the process, which is achieved by multiplying the vector of the linear step by a positive constant smaller than unity (sometimes by a diagonal matrix of damping coefficients that are different for the different components of the vector). This modification was used, for example, in Ref. 52; it cannot be called universal and, moreover, it greatly reduces the rate of convergence of the process.

A particular kind of damping by multiplicative regularization was used in Ref. 21. Besides the disadvantages mentioned above, multiplicative regularization is also disadvantageous in that it changes to the greatest extent precisely the elements of the matrix of the linear step V that least need it.

2. The matrix of the linear step V is kept positive definite, which makes it possible to carry out the iteration even in the case of an ill-conditioned step. More frequently used than others is the Levenberg–Marquardt method, which is a kind of additive regularization of the process, the additive term being found by means of empirical rules. Such an algorithm was realized, for example, in Refs. 53 and 54.

c) The difficulties encountered in implementation of the Gauss–Newton method prompted some authors to use the so-called quasi-Newtonian methods or variable-metric methods. In these methods, the matrix of the linear step V is not calculated and instead the inverse matrix V^{-1} is constructed directly from some initial approximation, which is improved in each iteration. One of these methods was realized in the programs of Refs. 17 and 55. The disadvantages of the variable-metric method are the slower convergence with greater number of unknowns and their basic unsuitability for solving a complete problem for finding hidden features.

d) Besides these methods, use is also made of the method of statistical adjustment.[56-59] It is a variant of classical direct search by trial and error which uses a number of auxiliary measures aimed at accelerating the progress of the wandering vector x toward the solution $\tilde{x}$ to the problem. The estimate of the errors of the required parameters in this method is of course also purely statistical.

In the statistical adjustment of the parameters of the processed spectrum we minimize the functional (29) directly without recourse to its derivatives. The main advantage of such a method is its speed, even when implemented on small and medium computers; this makes it very helpful for control purposes during an experiment. However, the method encounters characteristic difficulties in the processing of sections containing a higher number of overlapping lines or a combination of lines which differ strongly in intensity. Therefore, this method cannot be used to construct the required universal algorithm for solving the succession of monotypic nonlinear problems of discrete spectroscopy.

Regularized iteration processes. The content of the previous section illustrates fairly fully the characteristic difficulties that must be overcome in the creation of a universal algorithm for solving the succession of monotypic nonlinear problems generated by discrete spectroscopy. Similar difficulties are inherent in other problems.

For a large class of improperly posed problems, a method of regularizing the construction of approximate solutions has been developed by A. N. Tikhonov's school (see, for example, Ref. 1). In various papers (Refs. 9, 15, 47, and 60–63), Aleksandrov developed and justified regularized iteration processes (R processes) for solving systems of nonlinear equations. Among the R processes, a process of the type ERP with exponentially decreasing regularizer is particularly suitable for the solution of our discrete spectroscopy problems. It is realized in accordance with the scheme[15,61]

$$x^0:\ x^{t+1}=x^t-[V(x^t)+\alpha^t I]^{-1}\overline{f'(x^t)}\,Wfx^t, \tag{32}$$

where

$$\alpha^t=\alpha^0\exp(-rt), \tag{33}$$

and α^0 and r are positive constants; I is the unit matrix.

To avoid the premature termination of the R process when a first (possibly, local) minimum of the functional (29) is attained, the criterion for leaving the minimum is not a minimum of θ but a minimum of the parameter[14,61]

$$\Theta_s^t=\sum_{i=1}^{s}\rho_i\theta^{t-i+1},\quad t\geqslant s-1, \tag{34}$$

where

$$\theta^t=\theta(x^t), \tag{35}$$

and ρ_i are positive weight coefficients whose sum is normalized to unity:

$$\sum_{i=1}^{s}\rho_i=1. \tag{36}$$

We shall call θ_s^t the length discrepancy or, more precisely, the discrepancy of the length in s iterations. It is obvious that the weight coefficients can be chosen in such a way that for monotonic nonincrease of θ^t the criteria of minima of (29) and (34) are satisfied in the same iteration.

The propagated errors of the required parameters are estimated by the same method[15,45] for the R process (32) as in the case of the process (30). The purely statistical errors are calculated in accordance with

$$\varepsilon_\nu^{st}=\sqrt{[V(\tilde{x})^{-1}]_{\nu\nu}}, \tag{37}$$

and the more real "total" errors, which also take into account the inaccuracy of the adopted model, in accordance with

$$\varepsilon_\nu^{tot}=\varepsilon_\nu^{st}\theta(\tilde{x}). \tag{38}$$

In the solution of a complete problem for finding hidden features, the matrix $V(\tilde{x})$ is certainly degenerate and the estimate (37) cannot be made. In such cases, it can be replaced by a similar estimate, in which one uses instead of the operator $V(\tilde{x})$ the regularized operator $V(\tilde{x})+\tilde{\alpha}I$ ($\tilde{\alpha}$ is the value of the regularizer in the iteration in which the solution $\tilde{x}$ is attained). This leads to distorted values of ε_ν^{st} for at least the parameters of "spurious" lines. Since the strategy for solving the complete problem foresees a repeated processing of the section with corrected number of required lines, the distorted error estimates and are corrected subsequently.

The R process is capable of overcoming all the difficulties listed above. Its only disadvantage is a slight decrease in the rate of convergence compared with the Gauss–Newton method, espeically in the initial iterations, for which the regularizer is still large and acts as a damper. According to our estimates, the R process leads on the average to a solution in 7–10 iterations, and the slower convergence is completely offset by its unique positive features. Naturally, the R process must occupy a central position in the universal algorithm for solving the problems we are considering.

A regularized iteration process with exponentially decreasing regularizer was implemented for the first time in the program KATOK.[4,10,33] Later, an improved variant of the program was written in FORTRAN[12-14] and the program KOLOVOK was developed.[46] It uses the autoregularization iteration process ARP, which is realized in a program packet with the general designation REGN-СФМРIL.[15]

1.8. *Condition of the iteration steps and renormalization of the operators in the spaces R^n and R^m.* We have already noted that the linear steps in the iteration process used to solve the individual problems in a succession are as a rule ill conditioned. A quantitative measure of the condition of the n-dimensional linear problem

$$Ax=b$$

is the condition number[64] of the matrix A:

$$\operatorname{cond}A=\|A\|\cdot\|A^{-1}\|, \tag{39}$$

where $\|A\|$ is some norm of the matrix A. For Euclidean norms, which we shall use in the present paper,

$$\operatorname{cond}A=g_1/g_n, \tag{40}$$

where g_1 and g_n are the largest and smallest of the singular numbers of the matrix. It is well known[64] that the condition number gives an upper bound on the transfer of relative errors of the right-hand side of the initial system to the solution vector (for accurate matrix),

$$\|dx\|/\|x\|\leqslant\operatorname{cond}A\,\|db\|/\|b\|, \tag{41}$$

and also for the transfer of relative errors of the matrix of the problem to the solution vector (for accurate free term):

$$\| dx \| / \| x \| \leqslant \operatorname{cond} A \| dA \| / \| A \|. \tag{42}$$

Since numerical calculations on a computer always involve at least rounding errors[65] (and in practice there is always a set of additional sources of errors), it follows from the estimates (41) and (42) that it is worth attempting to lower the condition number of the problems to be solved by an appropriate transformation. This transformation cannot be orthogonal, since the condition number is invariant under orthogonal transformations in a Euclidean metric.

The simplest transformations of this type are represented by the class of nonorthogonal nondegenerate diagonal matrices. They are widely used in computational practice and are usually called scaling transformations. We prefer to use the expression "renormalization," since these transformations essentially change the norms of the transformed vectors and the corresponding norms of the operators, which are consistent with them. A good choice of the renormalization coefficients can significantly lower the condition number of the matrices which participate in the process, and this is the justification for using these transformations.

For example, for an arbitrary nondegenerate real 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{43}$$

one can express the singular numbers explicitly as the positive square roots of the eigenvalues of the positive-definite symmetric matrix $A\bar{A}$, from which we obtain

$$\operatorname{cond} A = \sqrt{\frac{Sp(A\bar{A}) + \sqrt{Sp^2(A\bar{A}) - 4 \det^2(A\bar{A})}}{Sp(A\bar{A}) - \sqrt{Sp^2(A\bar{A}) - 4 \det^2(A\bar{A})}}}, \tag{44}$$

where

$$\operatorname{Sp}(A\bar{A}) = a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2 \tag{45}$$

and

$$\det(A\bar{A}) = (a_{11}a_{22} - a_{12}a_{21})^2. \tag{46}$$

It is readily shown by means of the Cauchy inequality that the difference $\operatorname{Sp}^2(A\bar{A}) - 4 \det^2(A\bar{A})$ is always positive, i.e., the condition number is real and positive. Multiplication of the matrix A from the right by a diagonal renormalization matrix of the type

$$D = \begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix} \tag{47}$$

changes its trace and determinant:

$$\operatorname{Sp}(A'\bar{A'}) = a_{11}^2 + a_{21}^2 + k^2 (a_{12}^2 + a_{22}^2); \tag{48}$$

$$\det(A'\bar{A'}) = k^2 (a_{11}a_{22} - a_{12}a_{21})^2, \tag{49}$$

where

$$A' = AD. \tag{50}$$

Substituting (48) and (49) in (44), we can obtain the explicit dependence of the condition number on the renormalization coefficient (it is readily seen that the condition number depends only on the ratio of the diagonal elements of the renormalization matrix, so that the form (47) does not restrict the generality). Investigation of this dependence shows that for

$$k = k_0 = \sqrt{\frac{a_{11}^2 + a_{21}^2}{a_{12}^2 + a_{22}^2}} \tag{51}$$

there is a minimum equal to

$$\operatorname{cond} A'(k_0) = \sqrt{(1 + |\cos Z|)/(1 - |\cos Z|)}, \tag{52}$$

where

$$|\cos Z| = \sqrt{(a_{11}a_{12} + a_{21}a_{22})^2 / (a_{11}^2 + a_{21}^2)(a_{12}^2 + a_{22}^2)}, \tag{53}$$

and the angle Z can be regarded as the angle between the column vectors of the original matrix A in the basis before the renormalization transformation. Therefore, it is only when these vectors are orthogonal (for example, for $a_{12} = a_{21} = 0$ or for $a_{11} = a_{22} = 0$) that one can make the condition number equal to unity. In all other cases, the change in the condition number as a result of the transformation (50) is bounded below by the expression (52).

One can regard k_0 as a measure of the equilibration of the columns of the matrix A; then the transformation (50) with $k = k_0$ transforms the original matrix A into a matrix A' with columns equilibrated to the maximal extent possible for the given values of the elements of A. In the same way, one can derive expressions analogous to (51)–(53) for the case of renormalization and equilibration with respect to the rows, which occurs when A' is multiplied from the left by another diagonal matrix D'. These expressions are in fact identical to (51)–(53) with the elements a_{12} and a_{21} interchanged.

We give a numerical example. Suppose we have the matrix

$$A = \begin{pmatrix} 1.01\text{E}00 & 2.59\text{E}-3 \\ 5.09\text{E}-2 & 1.34\text{E}-4 \end{pmatrix}.$$

Its determinant is $\det A = 3.509\text{E}-6$, and the condition number is $\operatorname{cond} A = 3.28104\text{E}+5$ (to six significant figures). The optimal renormalization of A is obtained in accordance with (51) for $k_0^r = 3.89935\text{E}+2$, and the matrix then becomes

$$A' = \begin{pmatrix} 1.01\text{E}00 & 1.00993\text{E}00 \\ 5.09\text{E}-2 & 5.22513\text{E}-2 \end{pmatrix}$$

with condition number $\operatorname{cond} A' = 1.49485\text{E}+3$. Similarly, making an optimal renormalization from the left for $k_0^l = 1.95805\text{E}+1$, we obtain

$$A'' = \begin{pmatrix} 1.01000\text{E}00 & 1.00993\text{E}00 \\ 9.96650\text{E}-1 & 1.02311\text{E}00 \end{pmatrix},$$

for which $\operatorname{cond} A'' = 1.52284\text{E}+2$, i.e., a number which is more than three orders of magnitude smaller than the condition number of the matrix A.

If we work with a decimal computer and wish to avoid additional rounding errors associated with the renormalization, we must set $k_{10}^{r,l} = 10\text{INT}(\log k_0^{r,l} + 0.5)$, which in our case gives 1000 on the right and 10 on the left, respectively. After renormalization, the matrix A takes the form

$$A_{10}^{\sim} = \begin{pmatrix} 1.01\text{E}00 & 2.59\text{E}00 \\ 5.09\text{E}-1 & 1.34\text{E}00 \end{pmatrix}$$

and its condition number falls to 279, which is somewhat worse than the minimal value 152.

In the case of binary computers, for the same pur-

pose one must take $k_2^{r,l} = 2\mathrm{INT}(\log_2 k_0^{r,l} + 0.5)$. For the matrix A, these coefficients are 512 on the right and 16 on the left, the matrix itself becomes

$$A_2^* = \begin{pmatrix} 1.01000\mathrm{E}00 & 1.32608\mathrm{E}00 \\ 8.14400\mathrm{E}-1 & 1.09773\mathrm{E}00 \end{pmatrix},$$

and the condition number falls to 1.61647E+2, which is very close to the minimum. This is readily understood, since the integral powers of 2 form a denser mesh than the integral powers of ten, and as a result "binary" renormalization is almost optimal.

Thus, from our consideration of 2×2 real nondegenerate matrices, we can draw the following conclusions.

1. It is possible to achieve a change in the condition number of this class of matrices by means of a diagonal nonorthogonal renormalization transformation applied from the right and the left, i.e., with respect to the columns and with respect to the rows.

2. The possibility of reducing the condition number by means of a renormalization is associated with an equilibration with respect to the columns and rows of the matrix to be transformed.

3. If the matrix to be transformed is not equilibrated with respect to the columns and with respect to the rows, optimal renormalization from the right and the left leads to a minimum of the condition number of the transformed matrix.

4. The numerical value of the minimum admits a simple geometrical interpretation.

5. In the practical solution of actual problems on a computer, one can carry out a near-optimal renormalization, modifying only the orders of the elements of the matrices without introducing additional rounding errors. Because of the exact invertibility, such renormalization may be very helpful.

So far as we know, these conclusions have not been proved in the general case of n-dimensional square matrices. Nevertheless, the simplicity and geometrical naturalness of the conclusions suggest that such a proof is in principle possible, and the accumulated experience of the practical use of these conclusions (see Sec. 2) is an additional argument in favor of their validity.

Before we turn to a discussion of the features of the renormalization of nonlinear overdetermined problems of the type (23), let us consider which of the sources of ill conditioning of the linear steps (see Sec. 1.6) can be avoided by means of a renormalization transformation. It is clear that if two individual lines of the spectrum are close together, so that a pair (or pairs) of columns of the matrix of the linear step is (or are) nearly linearly dependent, the multidimensional analog of the angle Z will be small and, thus, not affected by the renormalization. However, differences in the amplitudes of the lines of the spectrum can be overcome by introducing a corresponding unit of measurement for each line. Therefore, renormalization of our problems will enable us to process the sections of the spectrum as if all the lines in them had approximately the same intensity.

This is a very important simplification. In fact, if one does not employ such a renormalization, the unregularized method of Newton is virtually useless under real conditions.

We now return to Eq. (23). We set

$$x = Cz, \tag{54}$$

where the renormalization matrix has the form

$$C = \mathrm{diag}\,(c_1,\ c_2,\ \ldots,\ c_n). \tag{55}$$

If we denote by G the nonlinear operator which carries $z \in R^n$ into $Gz \in R^m$, we obtain in place of (23) the new equation

$$Gz = 0, \tag{56}$$

and, as is readily seen, for its derivative we have

$$G'(z) = [\partial G_\mu / \partial z_\nu] = [(\partial f_\mu / \partial x_\nu)(\partial x_\nu / \partial z_\nu)] = f'(x)\,C. \tag{57}$$

All the expressions and estimates (19)–(38) remain valid if we substitute in them z, Gz, and $G'(z)$ instead of x, fx, and $f'(x)$, respectively. Note that the transformation (54) led to a renormalization of the rectangular matrix $f'(x)$ from the right, i.e., with respect to the columns in the space R^n, whereas the elements of the space R^m (the vector Y and the null vectors in Eqs. (23) and (56), as well as the vectors fx and Gz) remained unchanged. Therefore, the discrepancy θ is invariant under the transformation that we have made of the vector of unknowns.

In principle, one could also carry out a renormalization with respect to the components of the vectors in the space R^m. However, the actual form of Eq. (23) suggests that these components are of the same order of magnitude, and the use of weight matrices conceals still further the differences between them. This means that for the type of problems generated in the case we are considering a renormalization in R^m is not necessary. Equilibration of the matrix $G'(z)$ with respect to the columns leads to a linear-step matrix that is equilibrated with respect to the columns and with respect to the rows; this is by virtue of the very method of construction of $V(z)$ from $G'(z)$ [see (31)].

The transformation (54) also admits an interesting interpretation in terms of the metrics of the dual spaces R^n and R^m. Indeed, the differential of the discrepancy can be represented in the form

$$d\theta = \overline{Wfx} f'(x)\,dx/(m-n)\,\theta(x) = \overline{WGz} G'(z)\,dz/(m-n)\,\theta(z). \tag{58}$$

The denominators of the fractions on the left and right in the identity (58) are identical, and the numerators can be regarded as the scalar products of pairs of dual vectors with metric tensors $f'(x)$ and $G'(x)$, respectively. The stronger is the difference between the elements of the metric tensors, the greater is the degree of anisotropy of the corresponding pairs of dual spaces. The renormalization reduces the anisotropy and thus facilitates the iteration process, making it take place in a less "oblate" space. One can also say that it makes the anisotropy comparable with the accuracy with which numbers are represented in the computer.

We still have to establish how often a renormalization should be made when we solve a given nonlinear problem. Should it be done once at the beginning of the iter-

ation process, at every iteration, or in some other way? It is comparatively easy to find the answer to this question.

Modern computers usually work with an accuracy equivalent to six or seven significant figures in the decimal system. Keeping a certain number of places for calculations with reasonable rounding errors, one can evidently accept a linear-step matrix whose elements differ by three or four decimal orders. This enables us to construct the matrix of the derivative (the Jacobi matrix) from elements with maximal differences not exceeding 1.5–2 decimal orders. If we make a first renormalization before the beginning of the iteration process with respect to the values of the components of the vector of the initial approximations, and if these components are separated from the components of the solution by not more than two decimal orders, the first renormalization can well be the only one for the given problem. Otherwise it is desirable, and sometimes necessary, to make a second renormalization, if the vector of the unknowns has departed sufficiently far from the initial approximations.

2. UNIVERSAL ALGORITHM FOR SOLVING THE SUCCESSIONS OF NONLINEAR PROBLEMS GENERATED BY DISCRETE SPECTROSCOPY

We recall that when the time comes for the universal algorithm to be applied we assume that the measurements have been completed and that the accumulated and sorted information is in a form suitable for direct use in the computer, i.e., the spectrum or spectra are stored in the operative or peripheral memories.

2.1. *Preliminary processing of raw experimental data.* The preliminary processing is none other than the necessary preparation for the calculation of the required parameters of the spectrum. Attempts to automate this stage have been made by many people; a good review of the results achieved and the main directions in the development of the automatic search for peaks and division of the spectrum into sections can be found in Ref. 66. In our view, none of the proposed solutions can be regarded as satisfactory for the purposes of investigative spectroscopy, although many of them are perfectly suitable for applied problems (for example, in activation analysis), in which the nature of the measured spectra is well studied or the conditions of measurement (statistics, mixture of isotopes, etc.) vary in relatively small ranges. For example, methods of automatic search for peaks may work well or badly, depending on an externally specified sensitivity parameter. If the sensitivity threshold is too low, nonexistent parasitic lines may be "found," but if the threshold is too high, low-intensity lines may be lost. For any given spectrum, the optimal value of the threshold can be found empirically, but we have only a vague idea of the way in which the threshold depends on the conditions of measurement. In fact, the existing situation is perfectly natural. The peaks are usually sought with respect to a formal parameter, whose calculation moreover is done improperly (for example, by numerical differentiation of the spectrum), and from this one cannot expect much. It can be asserted that the problem of automatic searching for peaks is analogous to the problem of shape recognition, and this is a fairly complicated problem still far from solution.

Therefore, at this stage we believe it is expedient to give up complete automation of the preliminary processing of the discrete spectra and concentrate attention on a semiautomatic variant in which the human mind can work in conjunction with the computer by means of a display. It seems to us that such an approach corresponds best to the present level of our capabilities and makes it possible to include in the preliminary processing human (and as yet unprogrammable) qualities such as imagination, intuition, the ability to recognize shapes, learning based on accumulated experience, and so forth.

The majority of mathematical operations in such a preliminary processing of the spectra are elementary and can be described without recourse to formulas. It can usually be implemented on a minicomputer, although a powerful computer in a multiprogram regime may be used (in practice, this is seldom). Our treatment is based on the assumption that each individual problem in the succession can be regarded as a problem for finding hidden features.[34,67] This is not a limitation; moreover, spectroscopic problems are typical representatives of this class, and a feature is here to be understood as each of the lines (peaks) of the spectrum being processed.

In its most general features, the algorithm for the preliminary processing in the chosen variant includes the following stages:

1) the spectrum is displayed (fully or in part);

2) a section containing an isolated cluster of lines is separated from the spectrum;

3) for the chosen section, one specifies the number of required features, the degree of the background polynomial, and the initial approximations for the unknown;

4) from the initial data for the section and the data of the preceding operation, a standard array containing the results of the preliminary processing of the section is constructed;

5) the standard array is put into the peripheral memory of the computer (or transmitted to a different computer) for the further and final processing;

6) if all the sections of the spectrum have been exhausted, the preliminary processing is ended; if not, the above sequence is repeated.

Some of the above operations are elementary and hardly require further discussion. We shall therefore consider only two of them, namely, the specification of the initial approximations and the structure of the standard array. We shall also discuss the language of intercourse with the computer during the preliminary processing.[67]

a) *The specification of the initial approximations* by means of a point display reduces to the indication of some characteristic points for each required feature.

Points in the set that form the graphical image are indicated or distinguished by the operator, who is guided by his experience, intuition, etc. In this case, the initial approximations can be expressed as elementary functions of the coordinates of the characteristic points.

The number of characteristic points needed to calculate the initial approximations of one required feature is called the multiplicity of the characteristic points. Note that the multiplicity is not necessarily equal to the number of parameters which determine the given type of feature. If discrete spectra are processed by means of the Gaussian model (6) of the single-line profile, each feature is determined by three parameters: the amplitude (area) of the line, its position, and half-width. But for the calculation of their initial approximations, it is sufficient to know the coordinates of two characteristic points: one at the maximum of the putative peak and one to the right or left of the maximum where, in the opinion of the operator, the peak "ends" (i.e., disappears into the background). In other words, the multiplicity of the characteristic points is equal to two.

Suppose that the coordinates of the maximum and minimum of the i-th peak "seen" by the operator on the display screen (we use the quotation marks, since the operator may note or merely suspect the existence of a line where it cannot be seen in the literal sense of the word) are, respectively, $(q_i^{\max}, Y_i^{\max})$ and $(q_i^{\min}, Y_i^{\min})$. Then as initial approximations one can take

$$h_i^0 = |q_i^{\max} - q_i^{\min}|; \tag{59}$$

$$S_i^0 = |q_i^{\max} - q_i^{\min}||V_i^{\max} - V_i^{\min}|; \tag{60}$$

$$p_i^0 = q_i^{\max}. \tag{61}$$

Further, since we assume (see Sec. 1.4) that the half-width is the same for all lines of a given section (strictly speaking, almost the same), it is natural to assume

$$h^0 = \frac{1}{k}\sum_{i=1}^{k} h_i^0, \tag{62}$$

and then the variance of the Gaussian curve can be found from (7).

It remains to find the initial approximations of the coefficients of the background polynomial. For this, it is sufficient to tell the computer what degree of the polynomial the operator regards as suitable; no additional characteristic points are specified. One can then set

$$a_0^0 = \min\{V_q\}_{q=q_{in}, q_{in+1}, \ldots, q_e}; \tag{63}$$

$$a_1^0 = a_2^0 = \ldots = a_k^0 = 0. \tag{64}$$

The determination of the vector of initial approximations in a section containing three peaks of different intensity on a background in the form of a polynomial of second degree is shown in Fig. 1. From the coordinates indicated by the operator for the characteristic points,

min.1(551, 1213), max.1(557, 1369),

min.2(560, 1341), max.2(565, 1616),

min.3(577, 749), max.3(570, 2316),

we obtain in accordance with Eqs. (59)–(64) the components of the vector of the initial approximations,

$$x^0 = \text{col}(2.55,\ 936,\ 557,\ 1375,\ 565,\ 10969,\ 570,\ 600,\ 0,\ 0),$$

where the order of the parameters is the same as in (20). The exact solution to the problem of this artificial example is known,

$$(2.2,\ 800,\ 557.5,\ 2400,\ 564.5,\ 7200,\ 570.5,\ 1209,\ -8,\ -1),$$

and comparing the two vectors we see that some components of the initial-approximation vector are fairly far from the corresponding true values, although in order of magnitude they are close to them. The last two coefficients of the background polynomial are an exception, which violates the single-renormalization condition for the problem (see Sec. 1.8); however, this violation is only apparent, since the renormalization with respect to the background variables can be made independently of their initial approximations.[12-14,46] Note that the origin of the background polynomial in the example is shifted by nine channels to the right of the start of the section, i.e., it is at channel 559.

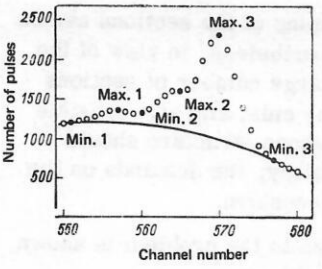

FIG. 1. Determination of the initial approximation by means of a point display (artificial example). The maxima and minima of three lines are indicated by the operator (black points), and the background is the continuous curve, which is represented by a polynomial of second degree. The initial values of the components of the vector of initial approximations are given in the text.

The specification of the vector of initial approximations on the display is a simple but rather crude operation. Nevertheless, having at our disposal a powerful algorithm for solving the systems of nonlinear equations, we can still find the point of the solution under these conditions. It is hardly worthwhile attempting to make the initial approximations more accurate by a more complicated preliminary processing [by using extra characteristic points, making Eqs. (59)–(64) more accurate, etc.]. The effort expended in this manner is not justified by the corresponding increased accuracy of the unknowns. In addition, algorithms based on regularized iteration processes are capable of handling problems in a succession at an accuracy of the initial approximations as described here.

b) *The structure of the standard array*, which includes both initial data of the section (i.e., the coordinates of the points of the spectrum) and the results of the preliminary processing (the coordinates of groups of characteristic points of corresponding multiplicity, the chosen degree of the background polynomial, etc.), is a question of convention. Any format which suits the

program for the final processing of the sections can be regarded as acceptable. Nevertheless, in view of the size of the problem and the large number of sections which must be processed, one must aim at a sensible organization of the format, whose structure should combine simplicity and flexibility; the demands on the memory should also not be excessive.

One of the possible solutions to the problem is shown in Table I. This format[67] combines

1) bytes, which are the shortest addressable memory units in the majority of modern computers;

2) a variable length, which makes it possible to store economically successive standard arrays in the peripheral memory of the computer;

3) the possibility of checking the transmitted information in accordance with the byte cyclic sum (and, if the working code is chosen appropriately, also on the basis of the parity of each byte);

4) independence of the working code;

5) the possibility of transferring both integer numbers and floating-point numbers.

An even more economic structure can be used if one dispenses with specifying all the abscissas in the standard array and includes in it only the initial channel number in the section and adopts a constant abscissa step for one channel (in this case, one loses the possibility of omitting the channels which are maverick and do not fit in the spectrum for instrumental reasons).

Obviously, the total length of the standard array in the format of Table I can be expressed in bytes as follows:

$$L = 81 + (m + Hk)(v + y). \tag{65}$$

For example, for the section in Fig. 1 for $v = 3$ and $y = 4$ (ASII code), we obtain $81 + (31 + 2 \times 3) \times (3 + 4) = 340$ bytes.

Special cases of such a format were realized with a Minsk-2 computer[68] (because of the particular architecture of the computer, a constant array length of 4008 words was used), and also with the minicomputers HP-21MX, IZOT-310, and EC-1010 (Refs. 12–14 and 46) (with variable length using the ASCII code).

TABLE I.

Numbers of bytes	Content
1—2	Total length L of standard array expressed in bytes in the form of a binary integer
3—66	Array identifier:
67—68	m: length of section in channels
69—70	n: number of required parameters
71—72	k: number of required features
73	l: degree of background polynomial
74	H: multiplicity of characteristic points
75	v: length of one abscissa in bytes
76	y: length of one ordinate in bytes
77—80	Reserve (for service information)
81—80+mv	m successive abscissas, each with length of v bytes
81+mv—80+$m(v+y)$	m ordinates in the same order as the abscissas, each with length of y bytes
81+$m(v+y)$—80 +$(m+Hk)(v+y)$	k groups of $H(v+y)$ bytes each; the first v bytes in the group represent the abscissa, and the following y bytes the ordinate of the corresponding characteristic point
81+$(m+Hk)(v+y)$	Control byte which makes up the cyclic sum of the array to 377_8 or 11111111_2

The sequence of standard arrays, which is the output information of the preliminary-processing algorithm, and the corresponding division of the original spectrum into sections are fed to the input of the program which implements the algorithm for solving the succession of nonlinear problems.

c) *The language of intercourse with the computer* in the preliminary processing must be as convenient as possible for the operator and also maximally adapted to the type of problem to be solved. It follows from this that it must be simple. The commands of the language can be divided into two groups.

I. The necessary commands:

1) illumination of the points of the spectrum (several or one, according to choice);

2) displacement of the image to the left or to the right through the spectrum;

3) indication of the start of the section;

4) indication of the end of the section;

5) indication of characteristic points of the necessary multiplicity for one required feature;

6) specification of the degree of the background polynomial;

7) construction and readout of the standard array of the data for the section (the array is constructed as the preceding commands are executed, and the final command merely implements the readout).

II. Additional commands giving greater convenience to the operator:

1) smoothing of a maverick point using the ordinates of its neighbors;

2) extinguishing of additional illumination;

3) calling of the following section of the spectrum to the screen (if the spectrum is put into the operative memory section by section, this command must be put in group I);

4) command for shifting the markers (i.e., the points with additional illumination);

5) commands for changing the scale of the Y axis from linear to logarithmic and vice versa;

6) omission of a maverick point (as an alternative to smoothing);

7) clearing of the system;

8) setting the multiplicities of the characteristic points;

9) input of spectrum identifier (with automatic numbering of the sections in accordance with their readout order), etc.

A more detailed description of the commands of the language is given in Ref. 67.

An individual command can be made in a diagrammatic manner (for example, by means of a light pencil

pointed onto the images of the light points on the screen) or by a program command decoder (using the control panel). To prevent the execution of commands made by the operator which violate the correct sequence of the preliminary processing of the section, it is expedient to introduce a square "syntax matrix" of the language. The number of rows and columns of the matrix is equal to the number of commands of the language, the rows corresponding to the last command given and the columns to the command that follows it. The elements of the matrix are equal to unity if the corresponding sequence of commands is permitted, and zero if the sequence is not. In this way, the computer can check the correctness of the given commands; of course, this does not guarantee error-free preliminary processing, but it does completely eliminate purely technical errors.

2.2. *Solution of the individual nonlinear problems in an automated regime.* We now consider the central and most important part of the universal algorithm—the part which solves the individual nonlinear problems of the succession in an automated regime. The input information in this stage is the succession of standard arrays, and the output information is the values of the required parameters of the spectrum (20) and their propagated errors (38). The general strategy of the algorithm is as follows[12,33]:

1) the Gauss–Newton method is applied to the successive problems;

2) if a divergence of the Gauss–Newton method is found, the corresponding problem is solved by the regularized iteration process;

3) if a divergence is found here too, the problem is declared to be unsolvable and the standard array of the next problem is called.

Obviously, the frequency of occurrence of the last situation can serve as a reciprocal measure of the universality of the algorithm.

General description of the algorithm. We list the main stages of the algorithm.

A1. Introduce the number of problems in the succession to be processed.

A2. Introduce one standard array of data; transform it to a form convenient for calculation; calculate the components of the vector of initial approximations x^0; carry out automatic renormalization on the basis of the values of the components of x^0.

A3. Solve the renormalized nonlinear system of equations by the Gauss–Newton method (30); after each iteration, test for convergence of the iteration process and satisfaction of the criteria for ending it; if there is no convergence, go to A4; if the process is advancing stably and the criteria for ending it are satisfied, go to A6.

A4. Beginning with the same vector of initial approximations, solve the renormalized nonlinear system of equations by the regularized iteration process (32); after each iteration, test for convergence of the process and satisfaction of the criteria for ending it; if there is no convergence, go to A5; if the process is advancing stably and the criteria for ending it are satisfied, go to A6.

A5. Declare the particular problem to be unsolvable by means of the employed algorithm; remove the input data of the problem; go to A7.

A6. Calculate the propagated errors of the required unknowns; by a transformation which inverts the renormalization, go over to the physically meaningful values of the unknowns and their propagated errors; extract these values.

A7. Test whether all problems of the given succession have been processed; if unprocessed sections remain, go to A2; if all the problems have been exhausted, declare the processing of the succession to be complete; the processing of the following succession (if there is one) can be begun by starting the algorithm again at A1.

Preparation for solution. Stage A1 is so elementary that we can turn directly to stage A2, which prepares for what follows. Here, various points must be made.

First, the standard array contains integer data, and for a start these must be transformed to a format with floating point. If an external exchange code was used, it must be converted to the internal working code of the particular computer.

Second, using data on the dimension of the particular problem at the start of the standard array (see Table I), it is necessary to extract the coordinates of the characteristic points and use (59)–(64) to calculate the components of the vector x^0, which must be arranged in order in accordance with (20).

Third, it is necessary to consider the renormalization, for which we must get a good understanding of the dependence of the model (13) on the unknown parameters. Therefore, we consider (13) and, taking

$$P_i=(q-p_i-1)/\sigma\sqrt{2}; \tag{66}$$

$$T_i=(q-p_i)/\sigma\sqrt{2}, \tag{67}$$

write

$$Y_q=\frac{1}{2\sqrt{\pi}}\sum_{i=1}^{k}S_i\,[J(T_i)-J(P_i)]+\sum_{j=0}^{l}a_jq^j. \tag{68}$$

Then, differentiating with respect to the parameters, we obtain in explicit form the components of the derivative of the operator:

$$\frac{\partial f}{\partial\sigma}=\frac{1}{\sigma\sqrt{\pi}}\sum_{i=1}^{k}S_i[P_i\exp(-P_i^2)-T_i\exp(-T_i^2)]; \tag{69}$$

$$\partial f/\partial S_i=(1/2\sqrt{\pi})\,[J(T_i)-J(P_i)]; \tag{70}$$

$$\partial f/\partial p_i=\frac{S_i}{\sigma\sqrt{2\pi}}[\exp(-P_i^2)-\exp(-T_i^2)]; \tag{71}$$

$$\partial f/\partial a_j=q^j. \tag{72}$$

Investigation of the dependences (68)–(71) shows that, as one would expect, the contribution of an individual line is significant only over a restricted interval to the left and to the right of the position of the maximum; this applies not only to the instrumental spectrum (68) itself but also to all its derivatives (except the param-

eters of the background, which we shall consider later). The magnitude of this "interval of significance" does not exceed the parameter σ by more than 3–3.5 times; outside this interval, Eqs. (68)–(71) give negligibly small values. In addition, it can be noted that all the terms in the equations, except the areas S_i, are near unity. Therefore, a renormalization with respect to the areas is required. A renormalization with respect to the remaining parameters is not necessary, and as the coefficient of area renormalization one can take the area itself or an integral power of the base (10, 2, or other) close to it.

We demonstrate this for the artificial example of a section in channels 546–554 containing one line with an intensity of 10 000 pulses; the half-width of the line is 2.355 channels, and there is no background. The center of the distribution of the signals before the transformation is at the mark corresponding to channel 549.5. The data for the section are given in Table II (note the shift of the maximum due to the integration within one channel).

The Jacobi matrix in this case has dimension 9×3 and is represented by the last three columns of Table II. From it, we construct the matrix of the linear step,

$$V(x) = \begin{pmatrix} 1.729\text{E}+7 & -1.250\text{E}+3 & 5.000\text{E}-6 \\ -1.250\text{E}+3 & 2.712\text{E}-1 & -6.000\text{E}-10 \\ 5.000\text{E}-6 & -6.000\text{E}-10 & 1.251\text{E}+7 \end{pmatrix},$$

in which the columns and rows are not equilibrated by up to four decimal orders, and the eigenvalues are, respectively, 1.729E+7, 1.2510E+7, and 1.8085E−1. By virtue of its method of construction, the matrix V is positive definite and its singular numbers are equal to the eigenvalues. Therefore, its condition number is 9.5604E+7 (note that two pairs of elements of V are nonvanishing only because of rounding errors). The inversion of such a matrix requires some precautionary measures even when one is working with a decimal computer with 12-place mantissa, which was used to calculate the example.

If we renormalize in accordance with (54), taking

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10^4 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we obtain the new linear-step matrix

$$V(z) = \begin{pmatrix} 1.729\text{E}+7 & -1.250\text{E}+7 & 5.000\text{E}-6 \\ -1.250\text{E}+7 & 2.712\text{E}+7 & -6.000\text{E}-6 \\ 5.000\text{E}-6 & -6.000\text{E}-6 & 1.251\text{E}+7 \end{pmatrix}$$

TABLE II.

q	Fx	$\partial f/\partial\sigma$	$\partial f/\partial S$	$\partial f/\partial p$
546	2.37E00	2.98E+1	2.37E−4	−8.57E00
547	5.98E+1	4.08E+2	5.98E−3	−1.67E+2
548	6.06E+2	1.50E+3	6.06E−2	−1.12E+3
549	2.42E+3	−1.82E+2	2.42E−1	−2.23E+3
550	3.83E+3	−3.52E+3	3.83E−1	−1.60E−9*
551	2.42E+3	−1.82E+2	2.42E−1	2.23E+3
552	6.06E+2	1.50E+3	6.06E−2	1.12E+3
553	5.98E+1	4.08E+2	5.98E−3	1.67E+2
554	2.37E00	2.98E+1	2.37E−4	8.57E00

*Obviously because of rounding errors; the exact value is zero.

with eigenvalues 3.5718E+7, 1.2437E+7, and 8.8046E+6; its condition number has been reduced to 4.0567E00. It can be seen that even in the simplest cases one can achieve a significant decrease in the condition number of the linear-step matrix of the problem by a simple renormalization. For more complicated sections containing lines of different intensities, the gain with respect to the condition number can exceed what has just been demonstrated by seven decimal orders.

With regard to the renormalization with respect to the coefficients of the background polynomial, for the majority of which the initial approximations are zero, we can point out the following two features:

1) to avoid excessively high values when the channel numbers are squared, cubed, etc., it is more convenient to place the coordinate origin at a point within the section (the transfer applies only to the background and does not affect the parameters of the spectrum);

2) after the transfer, one can calculate the coefficients of the renormalization with respect to the background variables using the mean value of the corresponding degree over the section rather than the initial approximations for the coefficients.

Thus, the algorithm of the automatic renormalization must contain the following steps.

B1. Extract the corresponding area of the line from the x^0 components.

B2. Separate the characteristic from the mantissa.

B3. Ascribe to the mantissa a constant lowest characteristic (for example, 1) and take the new number as the corresponding component of the vector z^0.

B4. Ascribe the characteristic to the component of x^0 to the unit mantissa (to ensure that the product of the two obtained numbers is equal to the original component of the vector x^0) and retain the obtained integral power of the working base as element of the matrix C.

B5. Test whether all the lines of the section have been exhausted; if not, return to B1; if so, go to B6.

B6. Calculate the mean values of the degrees of the polynominal with shifted origin and select the integral powers of the base nearest to them.

B7. Construct the diagonal renormalization matrix C; place on its diagonals units in accordance with the integral powers of the base obtained in B4 and B6, observing the order of the components in accordance with (20).

B8. Proceed to the iteration process (see A3).

In the example we have given (see Table II and the explanations to it) we have for simplicity ignored the matrix of statistical weights. In reality, the individual equations of the nonlinear system to be solved have different accuracy and, therefore, different weights. However, this fact does not alter the renormalization principle.

Criteria of convergence and termination of the process. When the preparation for calculation has been

completed, we can proceed to the solution of the problem for the section by means of the classical Gauss–Newton method. We shall use it in the pure form (30) without recourse to modifications.

As the criterion for convergence of the iteration process we shall take monotonic nonincrease of the discrepancy (29) from iteration to iteration. If in a given iteration the discrepancy is greater than in the preceding iteration, it is necessary to abandon solution by Newton's method and apply the regularized iteration process (32). Somewhat later we shall see that an increase of the discrepancy is not the only sign of deviation from normal progress of the process.

From the set of criteria for terminating the iteration process in processing programs,[12-14,33] four have been used.

1. The completion of a fairly large number of iterations, which is the same for all successions of problems.

2. A discrepancy is obtained which is lower than a certain constant, this being the same for all problems of a given succession and for all successions.

3. Vanishing of the increments of all unknowns to within the corresponding propagated errors.

4. A vector is obtained on the left-hand side of (23) with all its components not larger in modulus than a certain constant whose value is characteristic for the particular problem being solved and is calculated automatically (uniform approximation criterion).

The criteria for convergence and termination are tested after the completion of each linear iteration step. Before the start of the first iteration, it is necessary to calculate the discrepancy $\theta(x^0)$ of the initial approximations; this is needed to test for convergence and, in addition, is used in the calculation of the initial value of the regularizer (in the cases when it is not possible to find a solution by Newton's method; see below).

As the constant of uniform approximation one can take

$$\Delta_0 = \sqrt{a_0^0}, \tag{73}$$

where a_0^0 is taken from (63).

If Newton's process is found to diverge, it is natural to attribute this to degeneracy or ill conditioning of the linear-step matrix V of the problem. These two concepts are basically different from the point of view of a rigorous mathematical approach but become indistinguishable and can go over into each other in actual computer calculations. The limited accuracy with which numbers can be represented and its consequences (accumulation of rounding errors, possible overflow, and so forth) smear the sharp boundary between degeneracy and ill conditioning. If we return to the unrenormalized matrix V of the example in Table II, we see that if we work with five significant decimal figures it appears to be degenerate, whereas in reality it is ill conditioned.

In this connection, it is helpful to consider the steps of Newton's process at which ill conditioning may be detected, which can happen despite formal satisfaction of the criterion of a nonincreasing discrepancy. Note that by means of the renormalization we ruled out the possibility of overflow, and we therefore now look for additional criteria for the appearance of degeneracy and ill conditioning, ignoring the steps in the algorithm in which overflow could be expected. There exist at least two such criteria which can be followed without modifications of the process (30).

C1. The execution of the linear step of the iteration requires inversion of the matrix V. When V is inverted, one can also calculate its determinant, and as soon as it takes a value equal to the computer zero V must be regarded as degenerate in the particular computational situation.

C2. As we have already noted, the matrix of the linear step is positive definite by construction. The diagonal elements of the matrix V itself, as well as of its inverse V^{-1}, must be positive. If V is sufficiently ill conditioned for the actual accuracy with which the calculations are performed, negative diagonal elements of V^{-1} cannot be ruled out. Therefore, it is necessary to test the signs of the diagonal elements of V^{-1} and, if a minus is found, regard V as insufficiently well conditioned for the normal progress of Newton's iteration process.[12-14]

In either case it is necessary to terminate the solution by the Gauss–Newton method even if the discrepancy does not increase and go over to the regularized iteration process.

Regularized processes. The ERP process is the most powerful tool at our disposal. It can be used either in the form (32)–(36) or with the following modifications.

First, the value of the regularizer can be bounded below by the introduction of an additional term:

$$\alpha^t = \alpha^0(\theta^0)\exp(-rt) + \alpha_\infty . \tag{74}$$

Second, the choice of the initial value of the regularizer can be made to depend on the discrepancy of the vector of the initial approximations:

$$\alpha^0(\theta^0) = \begin{cases} \exp(r^0\theta^0 + r), & \theta^0 < \theta^0_{\max}; \\ \exp(r^0\theta^0_{\max} + r), & \theta^0 \geqslant \theta^0_{\max}, \end{cases} \tag{75}$$

where r, α_∞, r^0, and $\theta^0_{\max}$ are positive constants chosen by means of a computational experiment and are the same for all problems of a given succession as well as for all processed successions.

The lower bound on the regularizer is introduced in order to prevent the ERP process becoming the Gauss–Newton process through the regularizer's taking the value of the computer zero (since if we have arrived at this stage, the given problem cannot be solved without regularization). The reason for the dependence (75) is to begin the solution of the problems with relatively good initial approximations at smaller values of the regularizer. It is only when the vector of the initial approximations is so far from the point of the solution that the discrepancy θ^0 exceeds the constant $\theta^0_{\max}$ that (75) goes over into (33).

Third, it is desirable to foresee the possibility of automatic joining of the ERP process to the R_ε process, in which the value of the regularizer does not change from iteration to iteration. This can be done if the ERP process is continued until the achievement of a minimum of the length discrepancy (34), after which it is necessary to go back one iteration and begin from it the R_ε process to seek a new and lower minimum of the length discrepancy (34). Thus, using the R_ε process, one can improve the solution found by the ERP process.

The criterion of convergence of the ERP process is modified during the iterations:

1) if the number of iterations is not yet sufficient for the calculation of the first length discrepancy (and note that then the value of the regularizer is still fairly large), all three criteria of the previous section remain in force;

2) if it is possible to calculate the length discrepancy, the criterion of monotonic nonincrease of θ^t ceases to hold, and the process is regarded as divergent if situations C1 and C2 are encountered.

If a divergence is found, we assume that the concrete problem is unsolvable by the universal algorithm and go over to the special regime for removing its initial data.

The criterion for terminating the regularized iteration process remains a minimum of the length discrepancy defined by (34). At the same time, the criteria 2, 3, and 4 cease to hold, and they should not be tested for. The testing for criterion 1 can be retained to avoid going into a cycle, but termination on the basis of this criterion must be regarded as an indication of possible difficulties in the particular problem being solved.

Estimate of the errors and extraction of the information. A special regime for extracting the data of a section for which the problem cannot be solved by means of the universal algorithm is foreseen in stage A5. It could be called an impotence regime, but it is also a regime of organized transition to the next problem without breakdown of the fully automated processing of the succession of problems. If the initial data are properly prepared, this regime is not used.[5]

The statistical propagated errors of the required parameters are estimated after each linear iteration step. If a particular problem is solved by the Gauss–Newton method, such an estimate is necessary to test in accordance with criterion 3; at the same time, the two additional criteria C1 and C2 for normal progress of the calculations are also tested. However, if the solution is sought by the regularized iteration process, criterion 3 is eliminated, and the occurrence of C1 and C2 indicates that it is necessary to go over to the special extraction regime A5.

Once the solution to the problem has been found, it is necessary to make the final estimate of the propagated errors and extract the results. The algorithm for this procedure is as follows.

D1. Calculate the matrix which is the inverse of the unregularized linear-step matrix at the point of the solution and test for the presence of situations C1 and C2. If at least one is present and:

1) the solution is found by Newton's method, then proceed to A4;

2) the solution is found by the regularized iteration process, then proceed to D2.

But if the situations C1 and C2 are not found, then proceed to D3.

D2. Calculate the matrix which is the inverse of the regularized linear-step matrix at the point of the solution and test for the presence of situations C1 and C2. If at least one of them is found and:

1) the solution is found by the regularized iteration process, then proceed to A5;

2) the estimate is made in the special extraction regime A5, then take the absolute value of the negative diagonal elements and, staying in the same regime, proceed to D3.

But if the situations C1 and C2 are not found, then proceed to D3.

D3. Calculate in accordance with (37) the statistical propagated errors and in accordance with (38) the total propagated errors.

D4. Extract the data for the solved problem, the discrepancy at the point of solution, the calculated values of the required parameters and their total propagated errors, and also information about the method used to find the solution.

D5. Proceed to A7 (this step does not relate to the algorithm for calculating the errors and extracting the results, but, since it is elementary and follows directly after the extraction, we mention it here so as not to get involved with it in more detail).

2.3. *Solution of the complete problem for finding hidden features by means of the universal algorithm.* The details of the universal algorithm for processing the data of discrete spectroscopy have been described earlier (Refs. 4, 11–14, 33, and 67). The algorithm consists of two parts, of which the first is implemented in an interaction regime between the operator and the computer and generates a succession of monotypic nonlinear problems; the second part is fully automated and solves all the problems of the succession without the intervention of the operator. An important element of the second part of the universal algorithm is the regularized iteration process with exponentially decreasing regularizer.

The reader may have noted that at all stages of the universal algorithm we are dealing with a certain fixed number of required features, which, in the case of discrete spectroscopy, are the lines of the spectrum to be processed that lie within the given section. The number of these lines is specified by the operator during the preliminary processing. But as we have already noted, there are sections in real spectra which are so complicated that the operator is unable to determine

"by eye" the number of required lines. Such sections generate complete problems for finding hidden features.[34]

In solving these problems, we shall use a characteristic property of regularized iteration processes of Newton's type[15] to which attention was already drawn in Ref. 34, namely, the attempt to use the regularized iteration process to find a number of features that is greater than the true number for the given concrete problem leads to a pseudosolution, in which there are either features with zero or negligibly small amplitudes or pairs of features with different amplitudes but whose remaining parameters are equal or very nearly equal. In real computational situations, combinations of these two cases can also be encountered.

Thus, regularized iteration processes of Newton's type have the property of sifting spurious features in definitely degenerate problems.

In the solution of the successions of problems generated by discrete spectroscopy, equal or nearly equal parameters indicate coincident or nearly coincident positions of two lines in a given section. This occurs because, in accordance with the adopted model, the half-widths of all the lines in the section are assumed to be equal. Proximity of positions must, of course, be estimated on the basis of the resolution of the spectrometer that was employed. For example, if the single-line half-width is 5–6 channels, the calculated positions of two peaks separated by less than 0.25–0.3 channels may be regarded as completely coincident.

On the basis of this property of regularized iteration processes, we can indicate two possible ways for solving a complete problem by means of the universal algorithm. For brevity, we shall refer to a section of the spectrum that generates such a problem as a "complete section."

a) Processing of a complete section in a semiautomatic regime.

a1. In the preliminary processing stage, the complete section is treated like all the others, i.e., the operator indicates the number of lines that he notes.

a2. After this, the complete section is processed together with all the others in the succession.

a3. The results of the "final" processing of the complete section are examined specially with a view to finding features with negligibly small amplitudes or with very small separations of the positions; in addition, its discrepancy at the point of the solution and the obtained value of the line half-width are compared with the corresponding values in the neighboring (normal) sections. If the results of such examination are satisfactory, the problem is regarded as solved. Otherwise, the preliminary processing of the complete section is repeated with specification of a new—larger or smaller—number of features, the section is included in a new succession (see a2 above), etc., until a satisfactory solution to the problem has been found.

b) Processing of a complete section in an automated regime.

b1. The preliminary processing of the complete section is repeated several times with different numbers of features indicated by the operator; these begin with a value that is definitely below the expected value and end with one that is definitely above it. Accordingly, the complete section generates not one but several monotypic nonlinear problems. All these problems become members of the succession on an equal footing.

b2. The generated succession of problems is processed in accordance with the second part of the universal algorithm in the automated regime.

b3. The series of solutions obtained for the complete section is analyzed separately from the remaining results of the processing and the same criteria as in a3 above are used to select the one which agrees best with the results of the processing of the neighboring (normal) sections.

Method a) requires less computer time but greater human involvement than method b), to which we therefore give preference.

Note that the use of additional physical information about the neighboring sections actually becomes necessary only when there is a very poor accuracy of the measured data (in the case of inadequate statistics). Spurious features of "ordinary" complete sections can be sifted only by means of the regularized iteration process. In fact, there is no need to bother about its inclusion in the processing—the universal algorithm calls it automatically when necessary.[33] Some of the physical estimates of the significance of the results of the processing are also amenable to automated program analysis.[13,14,46]

CONCLUSIONS

Both parts of the universal algorithm were realized for the first time in 1969–1970 on a Minsk-2 computer at the Department of Nuclear Spectroscopy and Radiochemistry at the Laboratory of Nuclear Problems at the JINR, Dubna, and they have been used since then regularly in research studies (Refs. 10, 11, 33, 69, and 70).

In recent years, improved variants of the programs have been developed for the minicomputers HP-2116 and 21MX, IZOT-310, and EC-1010[12-14]; the program KOLOVOK[46] has also been developed for the computer EC-1040. Intensive work is in progress on the improvement of the programs for the initial processing of the spectra and also more complete automation of this processing. The reliability of the estimates of the propagated errors of the parameters of the spectrum have been investigated specially[71] by the method of comparative calculation. The conclusions of this work indicate that the estimates of the accuracy of the algorithm can be trusted and satisfy the requirements of high-precision spectroscopy.[44]

The reliability of operation of the algorithm is recognized to be exceptionally high.[3-5] In particular, the regime DUMP, which corresponds to the impossibility of processing a given section, is observed only for definitely erroneous input data. The universal algorithm

and programs developed on the basis of it have been used to process a huge amount of experimental material and important physical information has been obtained (Refs. 35–44 and 71–97). The algorithm itself has been considered from various points of view in Refs. 5, 66, 71, and 98–101. It is assumed that the strategy of the approach to the processing of spectroscopic data will remain essentially the same when the research program YaSNAPP-2 is completed; the new technique and new program developments are largely oriented toward the interactive regime, which must lead to more active participation of the physicists in the processing of the data and, ultimately, to more physical estimates of the required parameters of the spectrum.

[1]A. N. Tikhonov and V. Ya. Arsenin, Metody resheniya nekorrektnykh zadach (Methods of Solution of Improperly Posed Problems), Nauka, Moscow (1974).

[2]P. N. Zaikin, Sistemy polnoĭ matematicheskoĭ obrabotki rezul'tatov spektrometricheskikh êksperimentov (Systems of Complete Mathematical Processing of the Results of Spectroscopy Experiments), Doctoral Dissertation, State University, Moscow (1977).

[3]V. Gadzhokov, Avtomatizatsiya obrabotki dannykh diskretnoĭ spektrometrii yadernykh izlucheniĭ (Automated Processing of Data of Discrete Nuclear Spectroscopy), Doctoral Dissertation, Institute of Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia (1978).

[4]V. Gadzhokov, in: ÉVM v êksperimental'noĭ fizike (Computers in Experimental Physics), Dubna (1970), pp. 155–157.

[5]V. M. Tsupko-Sitnikov, in: ÉVM v yadernykh issledovaniyakh (Computers in Nuclear Research), Dubna (1974), pp. 285–307.

[6]A. N. Tikhonov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **29**, 815–818 (1965).

[7]A. N. Tikhonov, in: Vychislitel'nye metody i programmirovanie (Computational Methods and Programming), No. 8, Izd. MGU (Moscow State University), Moscow (1967), pp. 3–33.

[8]D. W. Green, Nucl. Instrum. Methods **76**, 349–356 (1969).

[9]L. Aleksandrov and E. Khristov, Dokl. Bolg. Akad. Nauk **21**, 195–198 (1968).

[10]V. Gadzhokov, Preprint R10-6655 [in Russian], JINR, Dubna (1972).

[11]L. A. Vylova *et al.*, Preprint R10-6751 [in Russian], JINR, Dubna (1972).

[12]V. Gadzhokov (Gadjokov), Preprint E10-12352, JINR, Dubna (1979).

[13]V. Gadzhokov (Gadjokov), Preprint E10-12353, JINR, Dubna (1979).

[14]V. Gadzhokov (Gadjokov), Preprint E10-12354, JINR, Dubna (1979).

[15]L. Aleksandrov, Preprint B1-5-9969 [in Russian], JINR, Dubna (1976).

[16]K. Siegbahn (ed.), Alpha-, Beta-, and Gamma-Ray Spectroscopy, Vol. 1, North–Holland, Amsterdam (1965) [Russian translation published by Atomizdat, Moscow (1969)].

[17]J. T. Routti and S. G. Prussin, Nucl. Instrum. Methods **72**, 125–142 (1969).

[18]D. C. Camp, Preprint UCRL-50156, University of California (1967).

[19]R. Trammel and F. J. Walter, Nucl. Instrum. Methods **76**, 317–321 (1969).

[20]O. Dragoun and P. Jagn, Preprint 1967/4, Max-Planck-Institut für Kernphysik, Heidelberg (1967).

[21]G. Winter, Preprint ZfK-182, Zentralinstitut für Kernforschung, Rossendorf bei Dresden (1969).

[22]V. B. Zlokazov, Preprint E10-9059, JINR, Dubna (1975).

[23]J. S. Geider, R. L. Graham, and M. W. Johns, Paper presented at the 18th Conf. on Nuclear Spectroscopy and Structure of the Atomic Nuclei, Riga (1968).

[24]M. Stevens and J. A. Harvey, Report WASH-1071 of the AEC Nuclear Cross Section Advisory Group, presented at the Meeting in Argonne, Illinois (1966).

[25]I. Zvol'ski and V. I. Prikhod'ko, Preprint 6-3517 [in Russian], JINR, Dubna (1967).

[26]R. G. Helmer *et al.*, Nucl. Instrum. Methods **57**, 46–57 (1967).

[27]L. A. McNelles and J. L. Campbell, Nucl. Instrum. Methods **127**, 73–81 (1975).

[28]H. H. Jorch and J. L. Campbell, Nucl. Instrum. Methods **143**, 551–559 (1977).

[29]J. L. Campbell and H. H. Jorch, Nucl. Instrum. Methods **159**, 163–170 (1979).

[30]Kh. K. Kaburov *et al.*, Paper presented at the Symposium on Nuclear Spectroscopy and Nuclear Structure, Dubna (1969) [see also: Yadrena Energiya **5**, 102–107 (1977)].

[31]Ch. Stoyanov, L. Aleksandrov, and V. Gadzhokov, At. Energ. **29**, 203 (1970) [for English version, see: J. Radioanal. Chem. **10**, 75–81 (1972)].

[32]V. Gadzhokov, L. Aleksandrov, and Ch. Stoyanov, Soobshchenie (Communication), R10-5043 [in Russian], JINR, Dubna (1970).

[33]V. Gadzhokov, Preprint R10-5035 [in Russian], JINR, Dubna (1970) [for English version, see: Instrum. Exp. Tech. No. 5, 1338–1341 (1970)].

[34]L. Aleksandrov and V. Gadzhokov, Preprint R5-5294 [in Russian], JINR, Dubna (1970) [for English version, see: J. Radioanal. Chem. **9**, 279–292 (1971)].

[35]V. Gadzhokov *et al.*, Preprint R6-5434 [in Russian], JINR, Dubna (1970) [see also: Izv. Akad. Nauk SSSR, Ser. Fiz. **35**, 2255–2265 (1970)].

[36]R. Arl't *et al.*, Fiz. Elem. Chastits At. Yadra **5**, 843–891 (1974) [Sov. J. Part. Nucl. **5**, 339 (1964)].

[37]K. Ya. Gromov, Fiz. Elem. Chastits At. Yadra **6**, 971–1039 (1975) [Sov. J. Part. Nucl. **6**, 393 (1975)].

[38]Ts. Vylov *et al.*, Preprint R6-9071 [in Russian], JINR, Dubna (1975).

[39]Ts. Vylov *et al.*, Preprint R6-9072 [in Russian], JINR, Dubna (1975).

[40]Ts. Vylov *et al.*, Preprint R6-9073 [in Russian], JINR, Dubna (1975).

[41]Ts. Vylov *et al.*, Preprint R6-10414 [in Russian], JINR, Dubna (1977).

[42]Ts. Vylov *et al.*, Preprint R6-10415 [in Russian], JINR, Dubna (1977).

[43]Ts. Vylov *et al.*, Preprint R6-10416 [in Russian], JINR, Dubna (1977).[2)]

[44]Ts. Vylov, Preprint R6-10417 [in Russian], JINR, Dubna (1977).

[45]D. Hudson, Statistika dlya fizikov (Statistics for Physics), [Russian translation], Mir (1967).

[46]V. Gadzhokov *et al.*, Preprint R10-12724 [in Russian], JINR, Dubna (1979).

[47]L. Aleksandrov, Preprint R5-10366 [in Russian], JINR, Dubna (1977).

[48]J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, New York (1970) [Russian translation published by Mir, Moscow (1975)].

[49]P. Quittner and R. E. Wainerdi, At. Energ. Rev. **8**, 361–415 (1970).

[50]J. W. Tepel, Report PEL-207, Pelindaba (South Africa) (1970).

[51]É. Rupp, Preprint R10-5776 [in Russian], JINR, Dubna (1971).

[52]I. A. Slavić and S. P. Bingulac, Nucl. Instrum. Methods **84**, 261–268 (1970).

[53]L. Varnell and J. Trischuk, Nucl. Instrum. Methods **76**, 109–114 (1969).

[2)]The papers of Refs. 38–43 were also published in the book Prikladnaya yadernaya spektroskopiya [Applied Nuclear Spectroscopy), No. 6, Atomizdat, Moscow (1976)].

[54]H. P. Blok, J. C. de Lange, and J. W. Schotman, Nucl. Instrum. Methods **128**, 545–556 (1975).
[55]G. W. Phillips and K. W. Marlow, Nucl. Instrum. Methods **137**, 525–536 (1976).
[56]G. Éler *et al.*, Preprint R10-7364 [in Russian], JINR, Dubna (1973).
[57]G. Éler *et al.*, Preprint R10-7365 [in Russian], JINR, Dubna (1973).
[58]G. Éler *et al.*, Preprint R10-7366 [in Russian], JINR, Dubna (1973).
[59]G. Éler *et al.*, Preprint R10-7367 [in Russian], JINR, Dubna (1973).
[60]L. Aleksandrov, Preprint R5-5136 [in Russian], JINR, Dubna (1970).
[61]L. Aleksandrov, Preprint R5-5137 [in Russian], JINR, Dubna (1970).
[62]L. Aleksandrov, Preprint R5-5515 [in Russian], JINR, Dubna (1970).
[63]L. Aleksandrov, Preprint R5-5790 [in Russian], JINR, Dubna (1971).
[64]G. Forsythe and C. B. Moler, Compute Solution of Linear Algebraic Systems; Prentice–Hall, Englewood Cliffs, New York (1967).
[65]V. V. Voevodin, Rotaprint, Computational Center, State University, Moscow (1969).
[66]L. P. Kabina and I. A. Kondurov, Preprint, A. F. Ioffe Physicotechnical Institute, USSR Academy of Sciences, Leningrad (1972).
[67]V. Gadzhokov, Yadrena Energiya No. 9, 73 (1978).
[68]L. A. Vylova *et al.*, Preprint R10-7061 [in Russian], JINR, Dubna (1973).
[69]Z. V. Lysenko, I. Tomik, and V. R. Trubnikov, Preprint 10-3331 [in Russian], JINR, Dubna (1967).
[70]Z. Zaĭdel' *et al.*, Preprint 13-5044 [in Russian], JINR, Dubna (1970).
[71]V. S. Aleksandrov, Ts. Vylov, and V. I. Gudov, in: Programma i tezisy dokladov XXIV soveshchaniya po yadernoĭ spektroskopii i strukture atomnogo yadra, Khar'kov, 1974 (Program and Abstracts of 24th Symposium on Nuclear Spectroscopy and Nuclear Structure, Khar'kov, 1974), Nauka, Leningrad (1973), pp. 484–485.
[72]Ts. Vylov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **36**, 718–743 (1972).
[73]B. Amov *et al.*, Yad. Fiz. **16**, 881–889 (1972) [Sov. J. Nucl. Phys. **16**, 487 (1973)].
[74]Ts. Vylov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **36**, 2111–2117 (1972).
[75]Ts. Vylov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **36**, 2118–2123 (1972).
[76]V. P. Afanas'ev *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **37**, 25–31 (1973).
[77]Ts. Vylov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **37**, 43–47 (1973).
[78]V. S. Aleksandrov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **37**, 948–952 (1973).
[79]V. S. Buttsev *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **37**, 953–964 (1973).
[80]Ts. Vylov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **38**, 701–709 (1974).
[81]A. A. Aleksandrov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **38**, 2096–2102 (1974).
[82]A. A. Aleksandrov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **38**, 2487–2492 (1974).
[83]N. A. Bonch-Osmolovskaya *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **38**, 2487–2492 (1974).
[84]V. S. Aleksandrov, Izv. Akad. Nauk SSSR, Ser. Fiz. **39**, 468–477 (1975).
[85]Ts. Vylov *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **39**, 506–522 (1975).
[86]I. Adam *et al.*, Bolg. Fiz. Zh. **2**, 12–30 (1975).
[87]R. Arlt *et al.*, J. Inorg. Nucl. Chem. **34**, 3001–3005 (1972).
[88]G. Yu. Baĭer *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **36**, 782–785 (1972).
[89]V. S. Buttsev, K. Ya. Gromov, and V. G. Kalinnikov, Izv. Akad. Nauk SSSR, Ser. Fiz. **37**, 1024–1034 (1973).
[90]A. Latuszynski *et al.*, Nucl. Instrum. Methods **120**, 321–328 (1974).
[91]I. Adam *et al.*, Bolg. Fiz. Zh. **3**, 145–160 (1976).
[92]I. Adam *et al.*, Bolg. Fiz. Zh. **3**, 26–38 (1976).
[93]Ts. Vylov *et al.*, Preprint R6-9369 [in Russian], JINR, Dubna (1975).
[94]K. Zuber *et al.*, Preprint 6-8517 [in Russian], JINR, Dubna (1975).
[95]K. Zuber *et al.*, Preprint R6-8669 [in Russian], JINR, Dubna (1975).
[96]S. R. Avramov *et al.*, Preprint R6-8781 [in Russian], JINR, Dubna (1975).
[97]K. Ya. Gromov *et al.*, Preprint R13-10611 [in Russian], JINR, Dubna (1975).
[98]R. Arl't *et al.*, Prib. Tekh. Eksp. No. 6, 71–83 (1972).
[99]E. M. Lobanov, Yu. A. Levushkin, and S. P. Vlasyuga, At. Energ. **35**, 247–252 (1973).
[100]E. M. Lobanov, Yu. A. Levushkin, and S. P. Vlasyuga, Dokl. Akad. Nauk BSSR **17**, 805–807 (1973).
[101]V. Mikhaĭlov *et al.*, Yadrena Energiya No. 6, 65 (1977).

Translated by Julian B. Barbour