

# сообщения объединенного института Ядерных псследований дубна 

# 4.840 b <br> $G-12$ 

V.Gadjokov

PROCESSING
OF DISCRETE NUCLEAR SPECTRA
ON SMALL COMPUTERS.
A.Mathematical Considerations

# E10-12352 

## V.Gadjokov

PROCESSING:
OF DISCRETE NUCLEAR SPECTRA ON SMALL COMPITERS.
A.Mathematical Considerations


Обработка дискретных ядерных спектров на малых ЭВМ. А. Математические соображения
Эта работа - первая в серии из трех, посвященных детальному описанию алгоритма программы КАТøК, которая после девятилеьней эксплуатации на ЭВМ Минск-2 была усовершенствована и написана заново на ФОРТРАНе.

В работе показано возникновение решаемой задачи из основной задачи спектроскопии и рассмотрены ее основные особенности: нелинейность, приближенность, плохая обусловленность, переопределенность и массовость. Описаны математические средства преодоления трудностей, возникающих из-за этих особенностей.

Работа выполнена в Лаборатории ядерных проблем оияи.

Сообшение Объединенного института ядерных исследовании. Дубна 1878
Gadjokov V.
E10-12352
Processing of Discrete Nuclear Spectra on Small Computers. A. Mathematical Considerations
This paper is the first in a series of three dedicated to the detailed description of the KATØK-F algorithm. This code has recently been revised and re-written in FORTRAN after being run on Minsk-2 for nine years.

The paper shows how the problem to solve appears from the basic spectroscopy problem and discusses its main features: non-linearity, ill-condition, over-determination and also its approximate and mass character. The mathematical means of over-coming the difficulties due to these features are described.

The investigation has been performed at the Laboratory of Nuclear Problems, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1979

## 1. INTRODUCTION

Processing of spectroscopic data requires in general the solving of a first-kind Fredholm's integral equation $1 /$.. In the case of discrete spectra, supposing that a reasonably plausible mathematical model of the single line can be put forward, this is easily reduced to overdetermined nonlinear simultaneous equations (see, e.g.,/2/).These, in turn, break into a series of similar simultaneous equations of lesser dimensions which can be solved separately. Such a further reduction is not a purely mathematical procedure. Its physical foundation is at least twofold: the non-uniform distribution of lines over the spectrum which leads to line clustering, and the finite resolution of actual spectrometers. One is, therefore, justified to state that the original problem of processing discrete spectra generates a stream (a multitude) of nonlinear problems which are independent but closely resemble each other. Each individual problem consists of a certain number of overdetermined nonlinear simultaneous equation's and corresponds to a section of the spectrum containing one cluster of spectral lines.

A FøRTRAN-IV computer code, KAT $\emptyset K-F$, especially designed for automated processing of streams generated is described below. Automation of the mathematical procedures used is meant here; indeed, from operators's point of view all the problems of a given stream are as a single (although large) one.

The approach just outlined is neither completely new nor quite unknorn. In particular, an earlier and simpler machine-language version of the KATøK program is being used since $1970^{\circ \%}$. Up to date over 100000 spectrum sections have been processed, and divergence of the iteration procedure implemented was "practically never" /4,5/ observed.

The present paper aims at diving full description of the KATØK algorithm in its latest form. Despite extensive usage, details of this algorithm were never published in the past (except for in ref. $/ 2 /$ whose distribution is very limited).

## 2. MATHEMATICAL FORMULATION OF PROBLEMS TO SOLVE

Let us consider a spectrum section which consists of consecutive channels numbered respectively $q_{i n}, q_{i n}+1, \ldots, q_{\text {end }}$ with

$$
\begin{equation*}
m=q_{\text {end }}-q_{i n}+1 \tag{2.1}
\end{equation*}
$$

and let the number of pulses registered in the $q^{\text {th }}$ channel be $Y_{q}$. Suppose that this section contains a cluster of peaks with symmetric Gaussian shape on a background continuum represented by a polynomial of $\ell^{\text {th }}$ degree. Suppose also that the section considered is short enough to assume an equal Full-Width-atHalf-Maximum (FWHM) h for all the $k$ peaks present. Then, taking into account the integration of the $k$ pulse distributions within the channels of the analyser, and setting for convenience a unit channel width, the following $m$ simultaneous equations may be derived $2,3 /$ :

$$
\begin{equation*}
Y_{q}=\frac{1}{2 \sqrt{\pi}} \sum_{i=1}^{k} S_{i}\left[J\left(T_{q i}\right)-J\left(P_{q i}\right)\right]+\sum_{j=0}^{\ell} a_{j} q^{j} \tag{2.2}
\end{equation*}
$$

where $S_{i}$ is the area of the $i$-th peak; $J(y)$ is the Gaussial error integral

$$
\begin{equation*}
\mathrm{J}(\mathrm{y})=\int_{-\mathrm{y}}^{\mathrm{y}} \mathrm{e}^{-\mathrm{u}^{2}} d \mathrm{u} \tag{2.3}
\end{equation*}
$$

$\mathrm{T}_{\mathrm{qi}}$ and $\mathrm{P}_{\mathrm{qi}}$ are respectively

$$
\begin{equation*}
\mathrm{T}_{\mathrm{q} i}=\frac{\mathbf{q - \mathbf { p } _ { \mathbf { i } }}}{\sigma \sqrt{2}} \tag{2.4}
\end{equation*}
$$

and

$$
\begin{equation*}
\mathbf{P}_{\mathrm{qi}}=\frac{\mathbf{q}-\mathbf{p}_{\mathbf{i}}-\mathbf{1}}{\sigma \sqrt{2}} \tag{2.5}
\end{equation*}
$$

$p_{i}$ is the maximum position of the $i$-th peak; $\sigma$ is connected with the FWHM by the relation

$$
\begin{equation*}
\mathrm{h}=2 \sqrt{2 \ln 2} \sigma \tag{2.6}
\end{equation*}
$$

$\mathbf{a}_{j}$ is the $j$-th coefficient of the background polynomial. In (2.2) the $m$ quantities $Y_{q}$ are given while the peak parameters $\sigma, p_{1}, S_{1}, \ldots, p_{k}, S_{k} \quad$ as well as the background coefficients $a_{0}, a_{1}, \ldots, a_{\ell}$ are unknown. One can easily express the number of unknown $n$ as

$$
\begin{equation*}
\mathbf{n}=2 k+\ell+2 \tag{2.7}
\end{equation*}
$$

Usually

$$
\begin{equation*}
\mathrm{n}<\mathrm{m}, \tag{2.8}
\end{equation*}
$$

i.e. the non-linear system (2.2) is averdetermined. It is also approximate, since the statistical quantities $Y_{q}$ are measured with finite accuracy $\Delta Y_{q}$ which is estimated by means of Poisson's distribution as

$$
\begin{equation*}
\Delta Y_{q}=\sqrt{Y_{q}} \tag{2.9}
\end{equation*}
$$

Another substantial feature of this system is that it is very often ill-conditioned which leads to divergent iteration processes applied for its solving. In the following sections we shall explain in more detail how the ill-condi tioning is manifested and how it can be overcome. And, last but not least, it should be borne in mind that we have to deal not with a single system (2.2) but rather with many such systems of various dimensions (i.e., with different values of $m, k$ and $l$ ) which represent the stream to be processed. The number of non-linear problems of the type (2.2) in a stream will be denoted as $M$.

So far we have considered the number of peaks $k$ in a spectrum section to be known. When processing streams generated by actual spectroscopic measurements, however, a limited number of sections may occur where $k$ cannot be guessed a priori. In such cases the system (2. 2 ) becomes a full problem of the analysis of latent regularities ${ }^{16 /}$. It is highly desirable that the algorithm used to solve (2.2) be able to process such full problems as well as ill-conditioned systems. If, in addition, this algorithm possesses an enlarged convergence domain, we shall term it universal. A universal algorithm is the only means of solving the multitude of (2.2) systems authomatically, i.e., as a single large problem. In our judgement, the
$K A T \varnothing K-F$ algorithm has all the prerequisits of a universal one. It consists of two distinct parts which will be dealt with in section 3 (preliminary processing) and 4 (actual stream-line solution of problems (2.2)) respectively.

## 3. PRELIMINARY PROCESSING OF RAW SPECTRA

The preliminary processing of raw experimental data is carried out in a man-machine dialogue by means of CRT point-display. As such a procedure is beyond the scope of this paper it will only be summarized here for the sake of completeness. The reader is referred for further details to refs. 7,8\%.

The procedure of preliminary processing includes the following four steps
3.1. Sectioning, i.e., breaking the experimental spectrum into separate sections, each of them containing an isolated cluster of spectral lines which eventually may overlap. Sections are later processed independently of each other since they give raise to independent (decoupled) non-linear simultaneous equations. These are said to represent a stream of similar non-linear problems.
3.2. Pointing out a pair of characteristic points for each spectral peak observed or suspected. One of these points should indicate the peak maximum and the other its basis, where it practically becomes undistinguishable from the background. No specific order of the characteristic points is prescribed. and the basic point can lie either on the left or on the right to the maximum one. The co-ordinates of the characteristic points are later used to calculate the initial guesses of peak parameters
3.3. Selecting the degree of polynomial to fit the background on the section. Since Compton-scattering pulses from lines of higher energies are also included into the background, this may in general be non-monotonous and, accordingly, require a higherdegree polynomial. Most often a linear presentation of the background is considered sufficiently accurate, however some sections require second, third and even higher degrees.
3.4. Forming a standard formatted data set which represents the input to the stage of final processing.

Among the above listed steps, 3.4 is the only one which can be termed formal; the preceding three steps are for the time being extremely difficult (if not impossible) to formalize, and their successful completion depends essentially on the operator's past experience, intuition and recognition capability.

As to the format structure of the standard data set, it is clearly a question of convention. The one accepted by the KAT $К K-F$ code is presented in Table 1 each row of which corresponds to a record in ASCII-code.
4. ITERATION PROCESSES AND AUXILIARY THECHNIQUES FOR SOLVING THE PROBLEMS POSED

### 4.1 Iteration Processes

Let us introduce the notations

$$
\begin{align*}
& x=\operatorname{col}\left(h, p_{1}, S_{1}, \ldots, p_{k}, S_{k}, a_{0}, a_{1}, \ldots, a_{\ell}\right) \in R^{n}  \tag{4.1.1}\\
& Y=\operatorname{col}\left(Y_{q_{i n}}, Y_{q_{i n}+1}, \ldots, Y_{q_{\text {end }}}\right) \in R^{m} \tag{4.1.2}
\end{align*}
$$

and rewrite (2.2) in the form

$$
\begin{equation*}
Y=F x \tag{4.1.3}
\end{equation*}
$$

where $F \in C^{1}$ is the non-linear operator of the right-hand side of (2.2). Moreover, if we put

$$
\begin{equation*}
\mathrm{f} \mathbf{x}=\mathrm{F} \mathbf{x}-\mathrm{Y} ; \quad \mathrm{f} \in \mathrm{C}^{\mathbf{1}} \tag{4.1.4}
\end{equation*}
$$

we come to the operator equation

$$
\begin{equation*}
\mathrm{f} x=0 \tag{4.1.5}
\end{equation*}
$$

which is equivalent to (2.2). As the various components of (4.1.5) have different accuracy, we introduce the square $(\mathrm{m} \times \mathrm{m}) \quad$ weight matrix $W$

$$
\begin{equation*}
\mathrm{W}=\operatorname{diag}\left(\mathrm{w}_{1}, \mathrm{w}_{2}, \ldots, \mathrm{w}_{\mathrm{m}}\right) \tag{4.1.6}
\end{equation*}
$$

where

$$
\begin{equation*}
\mathbf{w}_{\mathrm{i}}=1 /\left(\Delta Y_{i}\right)^{2} . \tag{4.1.7}
\end{equation*}
$$

Following Aleksandrov 9 / we shall be using the regularized iteration process ( $R$-process) for finding the solution of (4.1.5)

$$
\begin{equation*}
x^{0}: x^{t+1}=x^{t}-\left[V\left(x^{t}\right)+a^{t} I\right]^{-1} \overline{f^{\prime}\left(x^{t}\right) W f x^{t},} \tag{4.1.8}
\end{equation*}
$$

КАТФК-F $\frac{\text { Table } 1}{\text { Input Format }}$

| Record <br> format Record content | Record <br> length <br> in sym- <br> bols | Notes |
| :--- | :--- | :--- | :--- |

where $x^{0}$ is the vector of initial guesses and the superscript in general is the iteration number;
$f^{\prime}\left(x^{t}\right)$ is the first derivative (Jacobi matrix) of $f$ at point $x^{t}$;
$V\left(x^{t}\right)$ is the iteration-step matrix, i.e.

$$
\begin{equation*}
V\left(x^{t}\right)=f^{\prime}\left(x^{t}\right) W f^{\prime}\left(x^{t}\right) \tag{4.1.9}
\end{equation*}
$$

the overscore indicates a transposed matrix;
I is a unit matrix of rank $n$;
$a^{\text {t }}$ is a regularizing scalar

$$
\begin{equation*}
a^{\mathrm{t}}=a^{0} \mathrm{e}^{-\Gamma \mathrm{t}}+a_{\infty} \tag{4.1.10}
\end{equation*}
$$

$r$ and $\alpha_{\infty}$ are experimentally chosen constants, while
$a^{0}$ depends on the vicinity of $x^{0}$ to the solution point $\bar{x}$.
Note, that in the case of $\alpha^{t}=0(4.1 .8)$ is identical to the classical Gauss-Newton iteration process. A third variant may be obtained by setting $a^{t}=$ const 9 !

Since (4.1.5) is both approximate and overdetermined, we shall be looking for its approximate solution $\vec{x}$ which minimizes the normalized solution defect

$$
\begin{equation*}
\theta^{\mathrm{t}}=\frac{\mathrm{fx}^{\mathrm{t}} \mathrm{~W} \mathbf{f x ^ { \mathrm { t } }}}{\mathrm{~m}-\mathrm{n}} \tag{4.1.11}
\end{equation*}
$$

which can be seen to represent the weighted Euclidean norm of fx . Therefore, $\overrightarrow{\mathrm{x}}$ is also a solution in the sense of the least squares.

The criterion of minimizing (4.1.11) can only be applied if (4.1.8) converges with a null regularizer, i.e. in the case of pure Gauss-Newton process. If regularization becomes necessary, (4.1.11) will be replaced by the behavioristic "long solution-defect" 1 /

$$
\begin{equation*}
\Theta_{s}^{t}=\sum_{i=1}^{s} v_{i} \theta^{t-i+1} \tag{4.1.12}
\end{equation*}
$$

whose minimum will be sought for. The long solution-defect represents the gliding weighted average of $\theta^{t}$ over the last $s$ iterations, $v_{i}$ being the respective weights which satisfy the relation
$\sum_{i=1}^{s} v_{i}=1$.

The above-mentioned "vicinity" of $\mathrm{x}^{0}$ to $\tilde{x}$ is evaluated in terms of $\theta^{0}$; accordingly, $a^{0}$ is determined as

$$
a_{0}= \begin{cases}\mathrm{Ae}^{\mathrm{B} \theta_{\max }^{0}} & \text { for } \theta^{0} \geq \theta_{\max }^{0}  \tag{4.1.14}\\ \mathrm{Ae}^{\mathrm{B} \theta^{0}} & \text { for } \theta^{0}<\theta_{\max }^{0}\end{cases}
$$

where $A, B$ and $\theta_{\text {max }}^{0}$ are experimentally chosen constants. Once chosen at the stage of program debugging and set-up, such constants become fixed internal parameters which the common user is totally unaware of.

Once the solution point $\overrightarrow{\mathbf{x}}$ has been found, the i -th component $\mathrm{e}_{\mathrm{i}}^{\text {tot }}$ of the corresponding vector of total inherited errors $e^{\text {tot }} \in R^{\mathrm{n}}$ is expressed ${ }^{11 / /}$ as

$$
\begin{equation*}
\mathrm{e}_{\mathrm{i}}^{\mathrm{tot}}=\theta(\tilde{\mathrm{x}}) \sqrt{\left[\mathrm{V}^{-1}(\widetilde{\mathrm{x}})\right]_{\mathrm{ii}}} \tag{4.1.15}
\end{equation*}
$$

In the case of full problems of the analysis of latent regularities $V(\vec{x})$ degenerates and inversion becomes impossible. Nevertheless, an estimation of $e^{\text {tot }}$ is still according to ${ }^{19 /}$

$$
\begin{equation*}
\mathrm{e}_{\mathrm{i}}^{\text {tot }} \approx \theta(\overrightarrow{\mathrm{x}}) \sqrt{ } \sqrt{[\mathrm{V}(\overrightarrow{\mathrm{x}})+\vec{a} \mathrm{I}]_{\mathrm{ii}}^{-1}}, \tag{4.1.16}
\end{equation*}
$$

where $\vec{a}$ is the value of regularizer at the $i-t h$ iteration when $\tilde{x}$ has been reached. Then, repeating the processing of the same section with reduced number of peaks $k$ ', a nondegenerate solution $\vec{x}^{\prime}$ may be obtained which renders the more accurate error estimation (4.1.15) possible.

### 4.2 Calculation of the Components of Vector $\mathrm{x}^{0}$

The initial guesses of all the unknows are needed for starting up the iterations according to (4.1.8) and for computing the regularizer initial value $a^{0}$ (in case a regularization turns out indispensible). The various components of the initial guess vector $\mathrm{x}^{0}$ are calculated as follows:

$$
\begin{equation*}
h^{0}=\frac{1}{k} \sum_{i=1}^{k}\left|q_{1}^{(i)}-q_{2}^{(i)}\right| \tag{4.2.1}
\end{equation*}
$$

$$
\left.\begin{array}{c}
\mathrm{p}_{\mathrm{i}}^{0}=\left\{\begin{array}{cccc}
\mathrm{q}_{1}^{(\mathrm{i})} & \text { if } & \mathrm{Y}_{\mathrm{q}_{1}}^{(\mathrm{i})}>\mathrm{Y}_{\mathrm{q}_{2}}^{(\mathrm{i})} \\
\mathrm{q}_{2}^{(\mathrm{i})} & \text { if } & \mathrm{Y}_{\mathrm{q}_{1}^{(i)}}^{(\mathrm{i})}<\mathrm{Y}_{\mathrm{q}_{2}}^{(\mathrm{i})}
\end{array}\right\} \\
\mathrm{i}=1,2, \ldots, \mathrm{k}  \tag{4.2.3}\\
\left.\mathrm{~S}_{\mathrm{i}}^{0}=\left|\mathrm{q}_{1}^{(\mathrm{i})}-\mathrm{q}_{2}^{(\mathrm{i})}\right| \cdot\left|\mathrm{Y}_{\mathrm{q}_{1}}^{(\mathrm{i})}-\mathrm{Y}_{\mathrm{q}_{2}}^{(\mathrm{i})}\right|\right\} \\
\mathrm{i}=1,2, \ldots, \mathrm{k}
\end{array}\right\}
$$

In case $\mathrm{Y}_{\mathrm{q}_{1}}^{(\mathrm{i})}=\mathrm{Y}_{\mathrm{q}_{2}}^{(\mathrm{i})} \quad$ for a certain value of i , the input is considered irregular and processing of the corresponding section is skipped.

$$
\left.\begin{array}{l}
\mathrm{a}_{0}^{0}=\min \left\{\mathrm{Y}_{\mathrm{q}}\right\} \\
\mathrm{q}=\mathrm{q}_{\mathrm{in}}, \mathrm{q}_{\mathrm{in}}+1, \ldots, \mathrm{q}_{\mathrm{end}} \tag{4.2.5}
\end{array}\right\}
$$

The geometric foundation of formulae (4.2.1)-(4.2.5) is illustrated by Figure 1 on which a simple one-peak section is presented.

### 4.3 Condition of Matrix $V(x)$ and Scaling in $R^{n}$

As already mentioned in section 2, a well-conditioned problem of the type (2.2) is rather an exception to the rule. In the lack of an overall condition-number definition for non-linear problems we can evaluate these numbers in discrete points

$$
\begin{equation*}
x^{0}, x^{1}, \ldots, x^{t}, \ldots \in R^{n} \tag{4.3.1}
\end{equation*}
$$

by means of the condition numbers of the respectively iteration spet matrices $V\left(x^{0}\right), V\left(x^{1}\right)$, etc. To achieve this we accept the definition 10 /

$$
\begin{equation*}
\text { cond } V=\|V\| \cdot\left\|V^{-1}\right\| \tag{4.3.2}
\end{equation*}
$$

where weighted Euclidean norms conjugated with (4.1.11) are meant. In single-precision calculations on standard computers with four-byte length of floating point variables, a matrix $V$ may be considered ill-conditioned in $V$, $-10^{5} \div 10^{6}$ and above. Numerical inversion of such matricies


Fig. 1.
A one-peak spectrum section with characteristic points (deliberately inaccurate here, as it happens in actual processing) shown as dark circles. The shaded triangle represents the initial guess of peak area $S_{1}^{0}$.
is usually not accurate enough to allow for stable and convergent iteration process. Hence, means to lower cond $V\left(x^{t}\right)$ should be sought for.

The search of such means may be facilated by an understanding of basic reasons which lead to ill-conditioned iteration-step matrices. There are two such basic reasons in the problems of type (2.2). Although in general an ill-conditioned matrix is not necessarily next to generation, in this particular case each of the two reasons individually, as well as their simultaneous appearance, lead to both high values of cond $V$ and nearly vanishing det $V$.
4.3.1. Suppose we process a section where some pair of peaks $i_{1}$ and $i_{2}$ are positioned very near to each other in terms of the FWHM h, i.e.

$$
\begin{equation*}
\left|\mathrm{p}_{\mathrm{i}_{1}}-\mathrm{p}_{\mathrm{i}_{2}}\right| \ll \mathrm{h} . \tag{4.3.1.1}
\end{equation*}
$$

Then, if we observe the same order of variables as in (4.1.1), it may be shown that

$$
\begin{equation*}
\left\|\mathrm{V}_{\mu_{1}}+K V_{\mu_{2}}\right\|=\epsilon_{1}\left(\mathrm{p}_{\mathrm{i}_{1}}-\mathrm{p}_{\mathrm{i}_{2}}\right) \tag{4.3.1.2}
\end{equation*}
$$

where $\mu_{1,2}=2 i_{1,2}, V_{\mu}$ denotes the $\mu^{\text {th }}$ column-vector of $V$ (or also, row rector, as V is symmetric) and $\epsilon_{1}>0$ tends to zero when $\mathbf{p}_{\mathbf{i}_{1}} \rightarrow \mathbf{p}_{\mathbf{i}_{2}}$. If the norm in (4.3.1.2) were null, it might be said that a linear dependence between the two column-(row-) vectors existed. In analogy, we may call (4.3.1.2) an equation of " $\epsilon$-quasilinear dependence" of $\mathrm{V}_{\mu_{1}}$ and $\mathrm{V}_{\mu_{2}}$. Hence, when $\mathrm{p}_{\mathrm{i}_{1}}$ is close to $\mathrm{p}_{\mathrm{i}_{2}}$ in terms of $h$, the iteration step matrix tends to degeneration and, accordingly, its condition number grows high.

When applying the iteration formula (4.1.8) with a nonzero regularizer $a^{\text {t }}$ we effectively destroy the " $\epsilon$-quasilinear dependence" as $a^{\text {t }}$ is added to different elements of $V_{\mu_{1}}$ and $V_{\mu_{2}}$ (to those on the main diagonal of $V$ ). Therefore, regularization is a good remedy against illconditioned iteration steps caused by peaks close to each other. This is of particular importance in the case of full problems of the analysis of hidden regularities /6/ when $k$ is deliberately set higher than the value expected, and at the solution point coinciding peaks are expected to appear, i.e. a degeneration at the solution point is present a priori.
4.3.2. Suppose now that the same two peaks $\mathrm{i}_{1}$ and $\mathrm{i}_{2}$ do not obey (4.3.1.1) but rather their intensities (areas) $S_{i_{1}}$ and $S_{i_{2}}$ differ strongly, e.g.

$$
\begin{equation*}
\mathrm{S}_{\mathrm{i}_{1}} \ll \mathrm{~S}_{\mathrm{i}_{2}} \tag{4.3.2.1}
\end{equation*}
$$

Again, it can be shown that

$$
\begin{equation*}
\left\|V_{\mu_{1}}\right\|=\epsilon\left(-\frac{S_{i_{1}}}{S_{i_{2}}}\right) \cdot\left\|V_{\mu_{2}}\right\| \tag{4.3.2.2}
\end{equation*}
$$

where $\mu_{1,2}=2 i_{1,2}+1$ and $\epsilon_{2}$ tends to zero when $\frac{\mathrm{S}_{\mathrm{i}_{1}}}{\mathrm{~S}_{\mathrm{i}_{2}}} \rightarrow 0$. This fenomenon may be termed " $\epsilon$-quasidegeneration" of the iteration step matrix $V$ whose condition number raises sharply under these circumstances. One can easily see that regularization destroys the $\epsilon$-quasidegeneration too. However, the value of $a^{t}$ may not suffice, since it depends on the overall precision of vector $x^{0}$ and, in addition, it falls exponentially with the iteration number. A remedy which acts independently of the iteration number and, therefore, does not fade out when improving the solution $\mathbf{x}^{\mathbf{t}}$, should affect directly the magnitude of areas involved. It may consist of introducing individual units for all the peak intensities so that they all have the same order of magnitude; that is nothing suitable scaling.

The type of scaling we are going to use is represented by the class of square real diagonal non-orthogonal matrices

$$
\begin{equation*}
\mathrm{C}=\operatorname{diag}\left(\mathrm{c}_{\mathrm{h}}, \mathrm{c}_{\mathrm{p}_{1}}, \mathrm{c}_{\mathrm{s}_{1}}, \ldots, \mathrm{c}_{\mathrm{p}_{\mathrm{k}}}, \mathrm{c}_{\mathrm{s}_{\mathrm{k}}}, \mathrm{c}_{\mathrm{a}_{\mathrm{o}}}, \ldots, \mathrm{c}_{\mathrm{a}_{\ell}}\right) \tag{4.3.2.3}
\end{equation*}
$$

such that

$$
\begin{equation*}
z=C x, \tag{4.3.2.4}
\end{equation*}
$$

where

$$
\begin{equation*}
z=\operatorname{col}\left(h^{\prime}, p_{1}^{\prime}, s_{1}^{\prime}, \ldots, p_{k}^{\prime}, s_{k}^{\prime}, a_{0}^{\prime}, \ldots, a_{p}^{\prime}\right) \tag{4.3.2.5}
\end{equation*}
$$

Evidently, there is no need of scaling $h$ and $p_{i}$ and the respective elements of $C$ may be set to unity; $\mathbf{C}_{S_{i}}$ may be selected in such a way that all the $S_{i}^{\prime}$ be approximately equal. As to $\mathrm{c}_{\mathrm{a}_{\mathrm{j}}}$, they may be chosen in accordance with the average value of the respective degrees $q^{j}$ on the section processed. To avoid raising large numbers to high powers the polynomial origin may be suitably shifted. To prevent the accumulation of additional round-off errors all the nontrivial scaling factors should be integer degrees of the basis used (2 in our case).

Optimum scaling is a problem in itself which so far has not been successfully solved. Although we are not in a position to prove that the scaling outlined is optimal, numerical examples demonstrate its extreme effectiveness in lowering the cond $V$ down to $6-7$ decimal orders of
magnitude even in relatively simple cases (e.g. one peak with no background). The improvement of cond $V$ in more complicate spectrum sections is greater than that.

All the operators of such a scaling can easily be programmed. Since the inverse transformation is precise and error-free, it can be carried out prior to results' output. Thus the common user may be totally unaware of the whole scaling - descaling procedure, just as in the case of internal numerical constants used to compute the regularizer values $\alpha^{0}$ and $\alpha^{\mathrm{t}}$.

### 4.4 Numerical Effects of Ill-Conditioned Iteration-Step Matrices

If we were able to carry out computations with arbitrary precision (i.e. with indefinitely large number of significant digits) there wouldn't be any troubles in dealing with ill-conditioned matricies. Moreover, the very idea of ill-condition would be obsolete. Not so, however, in actual computers which usually work with $5-7$ significant digits in single precision and, in addition, have a limited range of number presentation. Under there conditions and in the presence of piling-up roud-off errors, an attempt of inverting an ill-conditioned iteration step matrix $V$ may result in:
4.4.1.Range degeneration; this means that $|\operatorname{det} V|$ falls outside the lower limit of numbers which are meaningful in the computer used. When such a situation occurs (underflow with respect to $\operatorname{det} V$ ) most systems would set $\operatorname{det} V=0$ without interupting the calculations. From the point of view of iterating according to (4.1.8), however, continuation of the process becomes meaningless. Hence, obtaining a null value of detV during the calculations is a sign of irregularities in the process used and necessitates a change of the computational regime.
4.4.2. Round-off change of the matrix type; note that by the very method of building $V$, this matrix is posi-tively-definite. Accordingly, all the diagonal elements of its inverse $\mathrm{V}^{-1}$ should be positive non-zero numbers. Hence, testing the sign of elements $\left(\mathrm{V}^{-1}\right)_{\mu \mu}$ may be an effective way of controlling the iteration process.
4.4.3. Condition divergence of the iteration process; this situation may occur when neither 4.4.1 nor 4.4.2.
are present but, due to ill-conditioned $V$, the inverted matrix is inaccurate enough to cause divergence of the iteration process, i.e. an increasing solution defect $\theta^{\mathrm{t}+1}>\theta^{\mathrm{t}}$. One should bear in mind that this may also happen by a number of other reasons (bad initial guesses $\mathrm{x}^{0}$, erroneous measurements of ${ }^{\prime} Y_{q}$, etc.). A practical rule of thumb is to ignore all these reasons and to attempt a change of the computational regime whenever the iteration process shows divergence, as if that were always due to an ill-conditioned V. A good reason to follow this rule is that an attempted regularization usually broadens the convergence domain in addition to lowering cond $\mathrm{V}^{\boldsymbol{9 /} / \text {, }}$

It is a common practice to control the iteration process by means of the solution defect only. In KATДK-F use of all the three numerical effects discussed is made to achieved this goal.

The implementation of the approach described will be the subject of another paper to appear shortly in the same JINR-series.

## REFERENCES

1. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных Задач, "Наука", М., 1974.
2. Гаджоков В. Диссертация, ъЯМЯЕ БАН, София, 1978.
3. Gadjokov V. Instr. and Exp. Technique, 1970, 5, p. 1338-1341.
4. Арльт Р. и др. ЭЧАЯ, 1974, т.5, ВЕПП. 4, с. 843-891.
5. Цупко-Ситииков В.М. ОГЯА, Д10,11-8450, Дубна, 1974, с. 285-307.
6. Alexandrov L., Gadjokov V., J. of Radioanal. Chem., 1971, 9, p.279-292.
7. Гаджоков В. Автоматиэация обработки даиньіх с помошью точечного дисплея. Ядрена Еиергия, №9, 1979.
8. Вылова Л. и др. ОКЯИ, Р10-7061, Дубна, 1973.
9. Александров Л. ОИЯИ, Б1-5-9969, Дубна, 1976.
10. Форсайт Дж., Молер К. Численное решение систем линейных алгебраических уравнений, "Мир", М., 1969.
11. Худсон Д. Статистика для ффзиков, "Мир" М., 1967.

Received by Publishing Department
on May 181979.

