объединенный
институт
ядерных
исследований
дубна

10/5-83

E11-83-71

G.A.Emelyanenko

ON METHODS OF CALCULATION WITH SPARSE MATRICES

Submitted to "Journal of Computational Physics"

1983

It is often the case in the last decade experiments on high energy physics to take out the physical information from the process of the experimental data handling via certain distributions of such physical quantities as particle energy, scattering angles, effective masses, etc. In general, these functions are not available in terms of directly measurable quantities, as, e.g., counting rates in the classical case of particle scattering. Furthermore, the efficiencies of the different parts of the data handling in general and those of the separate elements of the experimental set-up themselves are quite different. In a number of cases these problems lead to a non-negligible difference between the estimated distributions and the ideal ones. Summarizing the complexity of the above problems one can say that in general it seems to be the main task in the unfolding procedure at a real experiment to solve the inverse problem arising in most cases in the form of integral equations of the first kind:

$$f(x) = \int dx' R(x/x') E(x') \phi(x'),$$

where the ideal distribution $\phi(x')$ is connected to the observable one $f(x)$ via the efficiency function $E(x')$ and the resolution $R(x/x')$ both being characteristics of the given experimental set-up.

In solving such type of integral equations one meets characteristic troubles of the so-called ill-posed problems [2,3]. It is also a general property of these problems to have special structure of the kernel function $R(x/x') E(x')$ of the integral equation. Necessity of taking into consideration this special structure when solving such problems has led to the investigation of such linear operators (matrices) which are called in general the sparse matrices. Moreover, the results of these investigations seem to be usefull also for developing up-to-date statistical approaches of solution of inverse problems arising in the field of the quantum theory of scattering[6]. They are proved to have independent importance at solving several problems of numerical investigations of ill-conditioned boundary-value problems as well[7].

There are number of investigations devoted to the numerical and algorithmical aspects of systems with sparse matrices[8, 9, 10, 11], however a lot of earlier results can be treated as consequences of the following lemmas and theorems.

In this paper we indicate a simple technique for finding the exact numerical solution of systems of linear algebra equations. We shall describe the new methods of solving the systems $Ax = y$ and matrix inversion, where A is a matrix of quasitridiagonal and band form. Usually the method of solution is referred to as being in the class of exact methods if, in the absence of roundoffs, it yields an exact solution after a finite number of arithmetic and logical operations. The methods are based on the following.

Lemma 1. Let A is quasitridiagonal matrix of type

$$A = \begin{bmatrix} b_1 & a_2 & & & & \\ d_2 & b_2 & a_3 & & & \\ & d_3 & b_3 & a_4 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & d_{m-1} & b_{m-1} & a_m \\ & & & & d_m & b_m \end{bmatrix}, \tag{1}$$

with all principal minors being different from zero, $b_1, b_2, \dots, b_m$ are square matrices of the difference orders $[n_i, n_i]$ and $[n_{i+1}, n_{i+1}]$. Then the matrix A is represented as

$$A = D \cdot Q \cdot R = \begin{bmatrix} \omega_1 & & & & 0 \\ & \omega_2 & & & \\ & & \omega_3 & & \\ & & & \ddots & \\ 0 & & & & \omega_m \end{bmatrix} \begin{bmatrix} E_1 & & & & 0 \\ -\tilde{\beta}_2 & E_2 & & & \\ & -\tilde{\beta}_3 & E_3 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & & & -\tilde{\beta}_m & E_m \end{bmatrix} \times \tag{2}$$

$$\times \begin{bmatrix} E_1 & -C_2 & & & 0 \\ & E_2 & -C_3 & & \\ & & E_3 & -C_4 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & & & E_{m-1} & -C_m \\ & & & & E_m \end{bmatrix},$$

where

$$\begin{aligned} C_{i+1} &= \omega_i^{-1} \cdot (-a_{i+1}), \\ \tilde{\beta}_{i+1} &= \omega_{i+1}^{-1} \cdot (-d_{i+1}), \quad i = 1,2,\dots,m-1, \\ \omega_i &= b_i + d_i \cdot c_i, \quad i = 1,2,\dots,m, \end{aligned} \tag{3}$$

d_1, E_i are identity matrices and $C_1 = 0$ is matrix, or

$$A = F \cdot D \cdot R = \begin{bmatrix} E_1 & & & & 0 \\ -\beta_2 & E_2 & & & \\ & -\beta_3 & E_3 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & & & -\beta_m & E_m \end{bmatrix} \begin{bmatrix} \omega_1 & & & & 0 \\ & \omega_2 & & & \\ & & \omega_3 & & \\ & & & \ddots & \\ 0 & & & & \omega_m \end{bmatrix} \times \tag{4}$$

$$\times \begin{bmatrix} E_1 & -C_2 & & & 0 \\ & E_2 & -C_3 & & \\ & & E_3 & -C_4 & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & & & E_{m-1} & -C_m \\ & & & & E_m \end{bmatrix},$$

where

$$\begin{aligned} C_{i+1} &= \omega_i^{-1} (-a_{i+1}), \\ \beta_{i+1} &= (-d_{i+1}) \cdot \omega_i^{-1}, \quad i = 1,2,3,\dots,m-1, \\ \omega_i &= b_i + d_i \cdot c_i, \quad i = 1,2,\dots,m \end{aligned} \tag{5}$$

d_1, E_i are identity matrices and $C_1 = 0$ is matrix. Here R is right quasitriangular matrix, F and Q are left quasi-triangular matrices, D is quasidiagonal matrix.

Proof. As a starting point for our proof we use the idea of the Gauss elimination method. It is like that of the theoremes[12] for a non-singular symmetric matrices.

From the assumption that all principle minors of A are different from zero we have the non-singular matrices $\{\omega_i\}_{i=1}^{m}$.

Lemma 2. If A is non-singular quasitridiagonal matrix (Lemms 1) then the elements-blocks B_{ij} of inverse matrix A^{-1} are represented as

$$B_{ij} = \sum_{k=\max(i,j)}^{m} \prod_{\mu=i+1}^{k} C_{\mu}\cdot \omega_k^{-1} \cdot \prod_{\eta=j+1}^{k} \beta_{\eta}, \tag{6}$$

or

$$B_{ij} = \sum_{k=\max(i,j)}^{m} \prod_{\mu=i+1}^{k} C_{\mu}\cdot \prod_{\eta=j+1}^{k} \tilde{\beta}_{\eta}\omega_j^{-1}, \tag{7}$$

where

$$\prod_{\mu=q+1}^{p} C_{\mu} = \begin{cases} C_{q+1}\cdot C_{q+2} \cdots C_p, & \text{if } p > q \\ E_p, & \text{if } p \le q, \end{cases}$$

$$\prod_{\mu=q+1}^{p} \beta_{\mu} = \begin{cases} \beta_p \cdot \beta_{p+1} \cdots \beta_{q+1}, & \text{if } p > q \\ E_p, & \text{if } p \le q. \end{cases} \tag{8}$$

Proof. It is easy to see that the formulas (6) and (7) are equivalent, since

$$\omega_k^{-1} \cdot \prod_{\eta=j+1}^{k} \beta_{\eta} = \prod_{\eta=j+1}^{k} \tilde{\beta}_{\eta} \cdot \omega_j^{-1}. \tag{9}$$

From (4), (5), and (8) we obtain

$$\omega_k^{-1} \prod_{\eta=j+1}^{k} \beta_{\eta} = \omega_k^{-1}[\beta_k \cdot \beta_{k-1} \cdots \beta_{j+1}] =$$

$$= \omega_k^{-1}\{[(-d_k)\cdot\omega_{k-1}^{-1}] \;\ldots\; [(-d_{j+1})\cdot\omega_j^{-1}]\} =$$

$$= \{[\omega_k^{-1}(-d_k)]\cdot[\omega_k^{-1}\cdot(-d_{k-1})] \ldots [\omega_{j+1}^{-1}(-d_{j+1})]\}\cdot\omega_j^{-1} =$$

$$= \prod_{\eta=j+1}^{k} \tilde{\beta}_{\eta} \cdot \omega_j^{-1}.$$

Similarly from the right part of formula (9) we obtain the left one. From the formula (4), (5) for elements-blocks of inverse matrices R^{-1}, D^{-1}, F^{-1} we have

$$(R^{-1})_{ij} = \begin{cases} 0, & i > j,\; j = 1,2,\ldots,m \\ \prod_{k=i+1}^{j} C_k, & i \le j \le m,\; i = 1,2,\ldots,m \end{cases}, \quad (D^{-1})_{ij} = \begin{cases} 0, & i \ne j \\ \omega_i^{-1}, & i = j, \end{cases}$$

$$(F^{-1})_{ij} = \begin{cases} 0, & j > i,\; i = 1,2,\ldots,m \\ \prod_{k=j+1}^{i} \beta_k, & j \le i \le m,\; j = 1,2,\ldots,m. \end{cases} \tag{10}$$

Thus the equation (6) of the Lemma 2 can be obtained by premultiplying of matrices R^{-1}, D^{-1}, and F^{-1} from (10).

Theorem 1

Let A is quasitridiagonal matrix of the same type as in Lemma 1. Then the elements-blocks B_{ij} of inverse matrix A^{-1} are represented as

$$B_{ij} = \begin{cases} B_{ii} \cdot \prod_{k=j+1}^{i} \beta_k, & i \ge j \\ \prod_{k=i+1}^{j} C_k \cdot B_{jj}, & j \ge i. \end{cases} \tag{11}$$

Proof of theorem 1 may be obtained by using results of Lemma 2 and relations (8). Then we have for elements-blocks of left triangular matrix B:

$$B_{ij} \overset{i \ge j}{=} \sum_{k=i}^{m} \prod_{\mu=i+1}^{k} C_{\mu} \cdot \omega_k^{-1} \cdot \prod_{\eta=j+1}^{k} \beta_{\eta} =$$

$$= \sum_{k=i}^{m} \left[\prod_{\mu=i+1}^{k} C_{\mu} \cdot \omega_k^{-1} \cdot \prod_{\eta=i+1}^{k} \beta_{\eta} \right] \cdot \prod_{\eta=j+1}^{i} \beta_{\eta}.$$

The last element of this product is independent of k. The first element of this product defines, according to (6), the element-block B_{ii}. And so we obtain that

$$B_{ij} \overset{i\geq j}{=} B_{ii} \cdot \prod_{k=j+1}^{i} \beta_k \ .$$

If we use the procedure just described for the right triangular of the matrix B, then we have

$$B_{ij} \overset{i\leq j}{=} \sum_{k=j}^{m} \prod_{\mu=i+1}^{j} C_\mu \left[\prod_{\mu=j+1}^{k} C_\mu \cdot \omega_k^{-1} \cdot \prod_{\eta=j+1}^{k} \beta_\eta \right] =$$

$$= \prod_{\mu=i+1}^{j} C_\mu \cdot B_{jj} \ .$$

From the resulting relations follows eq. (11), which proves the theorem 1.

Theorem 2

Let A is quasitridiagonal matrix of the same type as in theorem 1 and let $Ax=y$ is the system of linear equations, then its exact solution is found as:

$$\begin{aligned}
x_i &= B_{ii} \cdot \gamma_i + \delta_i \ , \quad i = 1,2,\dots,m \ , \\
\gamma_i &= \beta_i \cdot \gamma_{i-1} + y_i \ , \quad i = 2,3,\dots,m \ , \\
\gamma_1 &= y_1 \ , \\
\delta_{i-1} &= C_i \cdot (\delta_i + B_{ii} \cdot y_i) \ , \quad i = m, m-1,\dots,2 \ , \\
\delta_m &= 0 \ .
\end{aligned} \tag{12}$$

Proof of theorem 2 may be obtained from the resulting relations for the elements-blocks B_{ij} (11) of inverse matrix A^{-1}. For the performance of these computations one may get a system of recurrence relations for determining the elements B_{ij}:

$$\begin{aligned}
B_{mm} &= \omega_m^{-1} \ , \\
B_{i-1\, i-1} &= \omega_{i-1}^{-1} + C_i \cdot B_{ii} \cdot \beta_i \ , \quad i = m, m-2,\dots,2, \\
B_{i\, j-1} &= B_{ij} \cdot \beta_j \ , \quad 1 \leq j \leq i \ , \\
B_{i-1\, j} &= C_i \cdot B_{ij} \ , \quad 1 \leq i \leq j \ .
\end{aligned} \tag{13}$$

where

$\{\omega_i, C_i, \beta_i\}_{i=1}^{m}$ are (5).

The method (12), (13) is convenient because of its uniform computation scheme and stability to computational error.

ACKNOWLEDGEMENT

The author would like to thank Dr. A.Kondor (HPR) for helpfil discussions.

REFERENCES

1. Zhigunov V.P. IHEP 81-5, Serpukhov, 1981.
2. Tikhonov A.N., Arsenin V.Ya. Metody resheniya nekorrektnych zadach (Methods for Solution of Ill-Posed Problems). "Nauka", M., 1979.
3. Emelyanenko G.A. JINR, D10-6142, Dubna, 1972.
4. Zhigunov V.P., Sokolov S.N. Sov.J.Part.Nucl., 1982, v.13, No. 5. p. 1024-1069.
5. Emelyanenko G.A. JINR, 11-6076, Dubna, 1971.
6. Bakhalov N.S. Numerical Methods, MIR Publ., Moscow, 1977.
7. Tewarson G. SIAM Review, 1970, v.12, No.5.
8. Allgower E.L. Numer.Math., 1974, 21, p. 279-284.
9. Wilkinson J.H., Martin R.S. Numer.Math., 1963, 7, pp. 355-361.
10. Baranger J., Duc-Jacquet M., R.I.R.O. NR-3, 1971.
11. Emelyanenko G.A. JINR, P11-6933, Dubna, 1973.

Received by Publishing Department
on February 7,1983.

WILL YOU FILL BLANK SPACES IN YOUR LIBRARY?

You can receive by post the books listed below. Prices - in US $, including the packing and registered postage

Емельяненко Г.А. E11-83-71
Методы вычислений с разреженными матрицами

Известно большое количество исследований, посвященных численным и алгебраическим аспектам систем с разреженными матрицами /8, 9, 10, 11/, тем не менее многие из ранее полученных результатов могут быть выведены как следствие лемм и теорем этой статьи.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна 1983

Emelyanenko G.A. E11-83-71
On Methods of Calculation with Sparse Matrices

There are number of investigations devoted to the numerical and algorithmical aspects of systems with sparse matrices /8, 9, 10, 11/, however a lot of earlier results can be treated as consequences of the lemmas and theorems of this paper.

The investigation has been performed at the Laboratory of Computing Technique and Automations, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna 1983