"ESTIMATES OF THE ROUNDOFF ERROR IN THE SOLUTION OF A SYSTEM OF CONDITIONAL EQUATIONS"

BY V. I. GORDONOVA

TRANSLATED BY
LINDA KAUFMAN

STAN-CS-70-164
JUNE 1970

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

ESTIMATES OF THE ROUNDOFF ERROR IN THE SOLUTION OF A SYSTEM OF CONDITIONAL EQUATIONS

by

V. I. Gordonova

Translated by

Linda Kaufman

The preparation of this report was sponsored by the Office of Naval Research under grant number N0013-67-A-0112-0029, the National Science Foundation under grant number NSF GJ 408 and the Atomic Energy Commission under grant number AT (04-3) 326,PA 30.

ESTIMATES OF THE ROUNDOFF ERROR IN THE SOLUTION OF A SYSTEM OF CONDITIONAL EQUATIONS

by

V. I. Gordonova

Zh. Vychislitel. Matem. i Matem. Fiziki, Vol. 9, No. 4
July-August 1969, pp. 775-83

In the present work we will examine estimates of the equivalent perturbation of roundoff errors in the solution of a system of conditional equations by the method of least squares (Method A) and by a method which was proposed by D. K. Faddeev, V. N. Faddeeva, and V. N. Kublanovskaya in a joint report at a conference on numerical methods in Kiev in 1966 (Method B).

Let us examine the system of conditional equations:

$$Ax = f \tag{1}$$

with a rectangular matrix A having N rows and n columns, where generally $N \gg n$. Method A leads to the system of normal equations

$$A^T Ax = A^T f \tag{2}$$

with a symmetric positive &finite matrix A^TA of rank n. We will assume that the solution of (2) is found by the method of square roots, always taking advantage of the accumulation of scalar products, independently of how one computes the elements of system (2).

Method B leads to a left orthogonal transformation of (1) into

$$Px = a \tag{2'}$$

---

* The term "equivalent perturbation" seems to refer to inverse roundoff analysis.

where $P = QA$, $\ell = Qf$, matrix P has non-null elements only in the right upper triangle $\tilde{P}$ of rank n. Let $\tilde{\ell}$ be the vector whose components are the first n components of the vector Qf. The triangular system

$$\tilde{P}x = a \qquad (3)$$

is equivalent to system (2).

The total error in both methods is composed of the roundoff error in reading in the coefficients and the right-hand terms of (2) and (3) and the roundoff error during the solving of these systems. Since triangular systems may be solved very exactly ([1, Chapter 4]), we can neglect the roundoff error in the solution of (3) and in the back-solution part of the method of square roots in the solution of (2).

Because of the equivalence of (2) and (3) it does not matter whether one calculates the equivalent perturbation of roundoff errors of Methods A and B in terms of (2) or (3). We will do the calculations in terms of system (2) since this is more convenient. Everywhere below, if it is not specifically stated, we will use the symbols adopted in [1] and the Euclidean norm of the matrices and vectors.

1) Let us examine in the first place the errors of Method A. Because of the roundoff in the calculation of the scalar products, the elements of the matrix A^TA and the vector A^Tf will be obtained with a certain error; i.e., instead of (2) we obtain

$$Bx = k \qquad (4)$$

where $B = A^TA + \Delta(A^TA)$, $k = A^Tf + \Delta(A^Tf)$.

The norms of the error matrix $A(A^TA)$ and the error vector $\Delta(A^Tf)$ essentially depend on the method of calculating of scalar products in the machine.

In the carrying out of all operations in a machine with a t-digit accuracy, the elements of $\Delta(A^TA)$ and $\Delta(A^Tf)$, which we will designate respectively by Ab_{ij} and Δk_i, may be estimated on the basis of [1, Chapter 3] as

$$|\Delta b_{ij}| \le N2^{-t_1} \|a_i\| \|a_j\| , |\Delta k_j| \le N2^{-t_1} \|a_i\| \|f\| ,$$

if the calculations are executed with floating point (fl). Here and later $t_1 = t - 0.08406$, and a_i is the i-th column of the matrix A. Hence, we obtain

$$\|\Delta(A^TA)\| \le N2^{-t_1}(\sum_{i,j} \|a_i\|^2 \|a_j\|^2)^{1/2} \tag{5}$$

$$= N2^{-t_1}(\sum_i \|a_i\|^2 \sum_3 \|a_j\|^2)^{1/2} = N2^{-t_1} \|A\|^2 ,$$

$$\|\Delta(A^Tf\| \le N2^{-t_1} \|A\| \|f\| .$$

If the calculations are executed in fixed point (fi), we get correspondingly

$$\|\Delta(A^TA)\| \le Nn2^{-t-1}, \|\Delta(A^Tf)\| \le Nn^{1/2} 2^{-t-1} . \tag{6}$$

Here it is assumed $\|a_i\| \le 1-N2^{-t-1}$, $\|f\| \le 1-N2^{-t-1}$, which guarantees the possibility of calculating in fixed point.

If the scalar products are calculated with double precision, then the estimate under consideration is practically independent of N. In particular, in the case of floating point $(f\ell_2)$, according to [1, Chapter 3],

$$|\Delta b_{ij}| \leq 2^{-t} (a_i^T a_j) + \frac{3}{2} N 2^{-2t+0.08406} \|a_i\| \|a_j\| .$$

Assuming $\frac{3}{2} N2^{-t} < 0.1$, we obtain

$$|\Delta b_{ij}| \leq 2^{-t} (a_i^T a_j) + 0.11 \cdot 2^{-t} \|a_i\| \|a_j\| .$$

Using the relation $|(a_i^T a_j)| \leq \|a_i\| \|a_j\|$, we find

$$\|\Delta(A^T A) < 1.11 \cdot 2^{-t} \|A\|^2 . \tag{7}$$

In the same way,

$$\|\Delta(A^T f)\| \leq 1.11' 2^{-t} \|A\| \|f\| . \tag{8}$$

In the case of fixed point (fi_2), we have

$$\|\Delta(A^T A)\| \leq n2^{-t-1}, \ \|\Delta(A^T f)\| \leq n^{1/2} 2^{-t-1} \tag{9}$$

with the assumption that $\|a_i\| \leq 1-2^{-t-1}$, $\|f\| \leq 1-2^{-t-1}$.

Let us now estimate the equivalent perturbation due to the roundoff error in the application of the forward step in the method of square roots, i.e., in the decomposition of the matrix of system (4) into the product of two triangular matrices. It is known that the triangular factors S and S^T of the matrix B that are really obtained in the machine are the exact factors of a certain matrix B+E, that is

$$B + E = SS^T. \tag{10}$$

The following estimates are verifiable for the elements e_{ij} of matrix E:

$$|e_{ji}| \leq \begin{cases} |s_{ij}s_{jj}|2^{-t}, & i > j \\ |s_{ji}s_{ii}|2^{-t}, & i < j \\ s_{ii}^2\ 2^{-t}, & i = j \end{cases} \tag{11}$$

with an accuracy up to terms of $O(2^{-2t})$ in calculations with floating point and

$$|e_{ij}| \leq \begin{cases} 0.5s_{ii}\ 2^{-t}, & i > j \\ 0.5s_{jj}2.^{-t}, & i < j \\ 1.00001s_{ii}\ 2^{-t}, & i = j \end{cases} \tag{12}$$

in calculations with fixed point. In the latter case, if $|b_{ij}| \leq 1-1.00001\cdot 2^{-t}$ for all i,j and if matrix B is not very badly conditioned, then $|s_{ij}| < 1$ for all i,j.

Considering (4) and (10), we get that the numerical decomposition is exactly the decomposition of a perturbed matrix, i.e., $A^TA + C = SS^T$, where

$$C = \Delta(A^TA) + E. \tag{13}$$

The norm of C is indeed of interest to us as the norm of the total error in the coefficients of system (2), while the norm of the vector $\Delta(A^Tf)$ is the norm of the error in the right-hand side of the system.

From (11) in calculations with floating point, neglecting terms of order 2^{-2t}, we have

$$\sum_{i,j=1}^{n} e_{ij}^2 \leq 2^{-2t} \sum_{i=1}^{n} \left(\sum_{j=1}^{i-1} s_{ij}^2 s_{jj}^2 + 4s_{ii}^4 + \sum_{j=i+1}^{n} s_{ji}^2 s_{ii}^2 \right)$$

$$\leq 2 \cdot 2^{-2t} \left(\sum_{i,j=1}^{n} s_{ij}^2 s_{jj}^2 + \sum_{i,j=1}^{n} s_{ji}^2 s_{ii}^2 \right)$$

$$= 4 \cdot 2^{-2t} \sum_{i,j=1}^{n} s_{ij}^2 s_{jj}^2$$

$$\leq 4 \cdot 2^{-2t} \max_j s_{jj}^2 \sum_{i,j=1}^{n} s_{ij}^2 = 4 \cdot 2^{-2t} \|S\|^4 .$$

Considering that

$$\sum_{i=1}^{n} s_{ij}^2 = \sum_{j=1}^{i} s_{ij}^2 = b_{ii} + e_{ii} = \sum_{j=1}^{N} a_{ji}^2 + \Delta a_{ii} + e_{ii}$$

$$= \sum_{j=1}^{N} a_{ji}^2 [1 + O(N2^{-t})] = \|a_i\|^2 [1 + O(N2^{-t})],$$

we obtain $\|S\| = \|A\|[1 + O(N2^{-t})]$, where

$$\|E\| \leq 2 \cdot 2^{-t} \|A\|^2 (1 + O(N2^{-t})). \tag{14}$$

As V. V. Voyevodin observed, these considerations permit us to obtain an estimate of the equivalent perturbation for the method of square roots which is n times better than that suggested in [2], without the assumption of accumulation.

Actually from the above explanation. it follows that with an accuracy up to quantities of order $O(2^{-2t})$

$$\|E\| < 2\cdot 2^{-t} [\sum_j (\sum_i s_{ij}^2)^2]^{1/2} \leq 2\cdot 2^{-t} \|S^T S\|$$

$$= 2\cdot 2^{-t} \|SS^T\| = 2\cdot 2^{-t} \|B\|.$$

Passing from the Euclidean norm to the spectral norm, we obtain

$$\|E\| \leq 2\cdot 2^{-t}(\mathrm{Sp}\ B)^{1/2} \max_i s_{ii} \leq 2\cdot 2^{-t}(n \max_i \lambda_i^2)^{1/2}$$

$$= 2\cdot 2^{-t}\, n^{1/2}\, \|B\|_2 .$$

This estimate is n times better than the one obtained in [2], for example. For fixed point, an estimate analogous to (14), derived from (12) with the assumption that $|s_{ij}| < 1$, has the form

$$\|E\| \leq n2^{-t-1}(1 + O(\tfrac{1}{n})). \tag{15}$$

Using the relations (5)-(15), we obtain finally

$$\|C\| \leq N2^{-t_1} \|A\|^2(1 + O(\tfrac{1}{N})), \|\Delta(A^T f)\| \leq N2^{-t_1} \|A\| \|f\| ; \qquad (f\ell)$$

$$\|C\| \leq 2.71\cdot 2^{-t} \|A\|^2, \|\Delta(A^T f)\| \leq 1.11\cdot 2^{-t} \|A\| \|f\| ; \qquad (f\ell_2)$$

$$\|C\| \leq Nn2^{-t-1}(1 + O(\tfrac{1}{N})), \|\Delta(A^T f)\| \leq Nn^{1/2}\, 2^{-t-1} ; \qquad (fi)$$

$$\|C\| \leq n2^{-t}(1 + O(\tfrac{1}{n})), \|\Delta(A^T f)\| < n^{1/2}\, 2^{-t-1} , \qquad (fi_2)$$

respectively, for the calculation of the elements of $A^T A$ and $A^T f$ in the cases of $f\ell$, $f\ell_2$, fi, fi_2.

2) We will now estimate the equivalent perturbation for the errors in Method B, which is equivalently an estimate of the errors in the elements of the system

$$P^T P x = P^T \ell \tag{16}$$

which were obtained because of the inaccurate calculation of P and ℓ. Let us denote by AP and $\Delta\ell$, respectively, the matrix and the vector error. Because of these errors, instead of (16) we obtain the perturbed system $(P + \Delta P)^T(P + \Delta P)x = (P + \Delta P)^T(\ell + \Delta\ell)$. Neglecting the products $\Delta P^T \Delta P$ and $(\Delta P)^T \Delta\ell$, we obtain for the perturbations the approximate equalities

$$\Delta(P^T P) = P^T \Delta P + (\Delta P)^T P, \quad \Delta(P^T \ell) = P^T \Delta\ell + (\Delta P)^T \ell,$$

from which

$$\|\Delta(P^T P)\| \leq 2\|P\| \, \|\Delta P\| \, , \|\Delta(P^T \ell)\| \leq \|P\| \, \|\Delta\ell\| + \|\Delta P\| \, \|\ell\| \, .$$

Because of the orthogonality of the matrix of transformation Q, we have

$$\|P\| = \|QA\| = \|A\| \text{ and } \|\ell\| = \|Qf\| = \|f\| \, ,$$

whence

$$\|\Delta(P^T P)\| \leq 2\|A\| \, \|\Delta P\| \, , \|\Delta(P^T \ell)\| \leq \|A\| \, \|\Delta\ell\| + \|\ell\| \, \|\Delta P\| \, . \tag{17}$$

In order to obtain final results it is necessary to estimate the norms of ΔP and $\Delta\ell$. These estimates essentially depend on the actual method of obtaining P, i.e., the method of transforming the system of simultaneous equations into system (2'). To obtain the matrix 'P we will eliminate the elements a_{ij} of matrix A for which $i > j$. We will perform the elimination with the help of a matrix of rotation or reflection [3]. Moreover, we will designate by $\alpha_1, \alpha_2, \ldots$ constants, which depend on the actual method of rounding in the machine. According to the assumptions of [1], these constants are not more than a few units or 1–2 tens.

(1) The transformation of matrix A is accomplished with the help of a succession of elementary rotation matrices T_{ij} in a cyclic order (Method B_1). Each of these rotations eliminates the element standing in the (i,j)-th position.

The roundoff error during the corresponding process of eliminating the subdiagonal elements of the square matrix was investigated in [1, Chapter 3], where elimination by columns was examined. In our case it is more convenient and necessary to eliminate elements by rows, i.e., in the order (2,1), (3,1), (3,2), (4,1),...,(n,n-1), (n+1,1),...,(n+1,n),..., (N,n). It can be shown that the roundoff error in the elimination of elements by rows and columns is the same.

Without stating the calculations, which are like those examined in [1, Chapter 3], but which are even more cumbersome, let us write the final result for the i-th column Δ_i of the error matrix ΔP:

$$\|\Delta_i\| \leq \alpha_i 2^{-t}\left[n(N-n) + \frac{n(n-1)}{2}\right]^{1/2}(N+n-2)^{1/2}(1+6\cdot 2^{-t})^{(N+n-3)}\|a_i\| \tag{18}$$

in the case of computing with floating point. In the same way an estimate, with the substitution of $\|f\|$ for $\|a_i\|$, is verifiable for, the error of transforming the column of the right-hand side. Here the calculation of scalar products with double precision has not been assumed. This cannot essentially change the estimate since, in the process under consideration, we do not encounter the calculation of scalar products of a vector of more than the second order.

In computing with fixed point.

$$\|\Delta_i\| \le \alpha_2 \, 2^{-t}[n(N-n) + \frac{n(n-1)}{2}] \; ; \qquad (19)$$

moreover, for it to be possible to compute with fixed point it is sufficient that

$$\|a_{i\cdot}\| \le 1 - \alpha_2 \, 2^{-t}[n(N-n) + \frac{n(n+1)}{2}] \; .$$

The same estimate is correct for the error of rotating the right-hand side.

The estimate obtained is exactly like that given in [1, Chapter 3], where actually the fact that the transformed matrix is square is not used.

Considering that $\|\Delta P\| = (\sum_{i=1}^{n} \|\Delta_i\|^2)^{1/2}$, we obtain from (18)

$$\|\Delta P\| \le \alpha_1 \, Nn^{1/2} \, 2^{-t} \, \|A\| \; , \; \|\Delta \ell\| \le \alpha_1 \; Nn^{1/2} \; 2^{-t} \; \|f\|$$

for floating point. In the same way from (19) we obtain

$$\|\Delta P\| \le \alpha_2 \; Nn^{1/2} \; 2^{-t}\|A\|, \; \|\Delta \ell\| \; \le \; \alpha_2 \; Nn2^{-t}$$

for fixed point.

(2) Errors can be reduced essentially if one uses rotation matrices with the order of elimination of the unknowns that is suggested in [4] (Method B_2).

Let us denote by M the number of cycles required for the transformation of A into triangular form. The estimate computed in [4] for our case takes on the form

$$\|\Delta P\| \leq \alpha_3 2^{-t} M(1 + 6 \cdot 2^{-t})^{M-1} \|A\| \; , \; \|\Delta \ell\| \leq \alpha_3 2^{-t}_M (1 + 6 \cdot 2^{-t})^{M-1} \|f\| \tag{20}$$

for floating point, and

$$\|\Delta P\| \leq \alpha_4 2^{-t} M^{1/2} \; n^{1/2} [n(N-n) + \frac{n(n-1)}{2}]^{1/2} \; ,$$

$$\|\Delta \ell\| < \alpha_4 2^{-t} M^{1/2} [n(N-n) + \frac{n(n-1)}{2}]^{1/2} \tag{21}$$

for fixed point.

For an estimate of the value of M let us note that the number of cycles is independent of the actual realization of the process suggested in [4] if one does not consider zero elements of the initial matrix or any elements accidentally zeroed in one elementary transformation. For the elimination of the m-1 elements of the matrix consisting of m rows and one column, $[\log_2(m-1)] + 1$ cycles are required. Here the square brackets denote the integer part.

Let the matrix have N rows and n columns. For the elimination of all the elements of the first column except the first element, one requires $[\log_2(N-1)] + 1$ cycles. With these it could happen that some of the elements of other columns are eliminated. However, even if one disregards the last situation for the elimination of elements of the second column, $[\log_2(N-2)] + 1$ cycles are required, etc.

Finally, we obtain

$$M \leq \sum_{k=1}^{n} [\log_2(N-k) + n \leq n\{[\log_2(N-1)] + 1\} \; .$$

This estimate is a little excessive, but not by more than 4-5 times for $N \leq 100000$.

Using this estimate for M, we find from (20) and (21)

$$\|\Delta P\| \leq \alpha_3 \, n \log_2 N \cdot 2^{-t}\|A\| \ , \ \|\Delta \ell\| \leq \alpha_3 \, n \log_2 N \cdot 2^{-t}\|f\|$$

for floating point and

$$\|\Delta P\| \leq \alpha_4 \, n^{3/2}(N \log_2 N)^{1/2}, \quad \|\Delta \ell\| \leq \alpha_4 \, n(N \log_2 N)^{1/2}$$

for fixed point.

(3) Using a matrix of rotation (Method B_3) for the elimination of the elements of A appears most expedient in that case where the scalar products are calculated with double precision. Moreover, the estimates for ΔP and $\Delta \ell$ are practically independent of N. Let us assume here that the calculation is. carried out in floating point. The results obtained in [1, Chapter 3] go for rectangular matrices A and give

$$\|\Delta P\| \leq \alpha_5(n-1)2^{-t}\|A\|, \ \|\Delta \ell\| \leq \alpha_5(n-1)2^{-t}\|f\|.$$

Having substituted the estimate received for AP and $\Delta \ell$ into (17), we obtain a final estimate of the norm of the error matrix $A(P^T P)$ and the-error vector $\Delta(P^T \ell)$; namely,

for method B_1:

$$\|\Delta(P^T P)\| \leq \alpha_1 N n^{1/2} 2^{-t}\|A\|^2, \ \|\Delta(P^T \ell)\| \leq \alpha_1 N n^{1/2} 2^{-t}\|A\| \ \|f\| \qquad (f\ell)$$

$$\|\Delta(P^T P)\| \leq \alpha_2 N n^2 2^{-t}, \ \|\Delta(P^T \ell\| \leq \alpha_2 N n^{3/2} 2^{-t} \ ; \qquad (fi)$$

for method B_2:

$$\|\Delta(P^TP)\| \le \alpha_3\, n \log_2 N \cdot 2^{-t}\|A\|^2, \quad \|\Delta(P^T\ell)\| \le \alpha_3 n \log_2 N\cdot 2^{-t}\|A\|\,\|f\| \qquad (f\ell)$$

$$\|\Delta(P^TP)\| \le \alpha_4 n^2 (N \log_2 N)^{1/2}, \quad \|\Delta(P^T\ell)\| \le \alpha_4 n^{3/2}(N \log_2 N)^{1/2}; \qquad (fi)$$

for method B_3:

$$\|\Delta(P^TP)\| \le \alpha_5(n-1)2^{-t}\|A\|^2, \quad \|\Delta(P^T\ell)\| \le \alpha_5(n-1)2^{-t}\|A\|\,\|f\|. \qquad (f\ell_2).$$

Comparing the obtained results, we see that the estimates of the equivalent perturbations for the matrix of system 2 have the form $2^{-t}\varphi(N,n)\|A\|^2$ and $2^{-t}\psi(N,n)$, respectively, for the different methods of calculation. In the table-the order of magnitude of the functions $\varphi(N,n)$ and $\psi(N,n)$ are set forth $(N \gg n)$.

	Type of Computation			
Method	$f\ell$	$f\ell_2$	fi	fi_2
A	N	const	nN	n
B_1	$n^{1/2}N$	$n^{1/2}N$	n^2N	n^2N
B_2	$n \log_2 N$	$n \log_2 N$	$n^2(N \log_2 N)^{1/2}$	$n^2(N \log_2 N)^{1/2}$
B_3	nN	n	n^-N^2	$n^2N^{1/2}$

In this table it is seen that a comparison of Methods A and B, in the sense of majorizing the estimate, gqes as a rule in favor of Method A. Method B_2 is the elimination method.

Let us go now from the equivalent perturbations to the error in the solution of the system. It is not difficult to construct an example in which with Method B_1 one obtains an order of the norm of the error in the solution which is equal to the largest estimate of Method A without accumulation. Let us examine, for example, the system with a matrix of coefficients and a right-hand vector, respectively, of the form

$$a_{ij} = \begin{cases} 0.5 & i = j, \\ 0 & i \neq j,\ i \leq n, \\ \epsilon_1 & i > n, \end{cases} \qquad f_i = \begin{cases} 1/n & i \leq n; \\ 0 & i > n; \end{cases}$$

where $\epsilon \ll 1$, so that $n(N-n)\,\epsilon < 1$.

Let us consider that computations are carried out with fixed point, and that the elementary matrix rotations are computed exactly. Assume that multiplication by these matrices is equally exact. After each multiplication by an elementary matrix of rotation, one rounds off the elements obtained up to a t digit number with fixed point, which gives an error of 2^{-t-1}. It is possible to assume that in this situation the elements of AP, which stand on the main diagonal and above, have the form $(N-n)2^{-t-1} + O(n(N-n)\,\epsilon\, 2^{-t})$. Also, the components of the vector $\Delta\ell$ have this form with numbers which are not larger than n.

Let us designate by Ax the vector of the error of the solution. When $(\tilde{P} + \Delta\tilde{P})(x + Ax) = \tilde{\ell} + \Delta\tilde{\ell}$, then, neglecting the product $\Delta\tilde{P}\Delta x$, we obtain $Ax = \tilde{P}^{-1}(\Delta\tilde{\ell} - \Delta\tilde{P}x)$. Having computed $\tilde{P}^{-1}$ and x, we obtain $\|\Delta x\| = O((N-n)^{1/2}2^{-t})$. The same order for $\|\Delta x\|$ is obtained in Method A if one uses the identity $Ax = (A^TA)^{-1}(\Delta(A^Tf) - Cx)$ and the maximum estimates for $\Delta(A^T\ell)$ and C.

In conclusion, let us take note of a fact which is connected to the practical application of the methods under consideration. The application of Methods B_2 and B_3 requires storage in memory of all the elements of the matrix A, while the application of Methods A and B_1 permits a row-by-row introduction of the information. The latter allows one a practically limitless way to increase the values of N. In the row-by-row introduction of information in Method A with accumulation of scalar products, one demands in addition $n^2 + n$ work cells for the storage of intermediate values during the calculation of the elements of $A^T A$ and $A^T f$. Actually, in this case the coefficients (and the right-hand side) of system (2) can be considered in a parallel fashion and each of these intermediate values, written down in 2t digits, can be stored in 2 cells of memory.

The author wishes to thank V. V. Voyevodin for posing the problem and for guidance.

REFERENCES

[1] J. H. Wilkinson. The Algebraic Eigenvalue Problem. Oxford, Clarendon Press, 1965.

[2] J. H. Wilkinson. Apriori error analysis of algebraic processes. Reports of the Proceedings of the International Congress of Mathematics, Moscow, 1966.

[3] D. K. Faddeev, V. N. Faddeeva. Computational Methods of Linear Algebra. Translated by R. C. Williams. San Francisco: W. H. Freeman, 1963.

[4] V. V. Voevodin (Voyevodin). On the order of elimination of unknowns. USSR Computational Mathematics and Mathematical Physics, 1966, vol. 6, no. 4, pp. 203-06. R. C. Glass, translation editor. Oxford, Pergamon Press, Ltd.