ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ"

ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Книга 1 — Математика
Том 83, 1989

ANNUAIRE DE L'UNIVERSITE DE SOFIA „ST. KLIMENT OHRIDSKI"

FACULTE DE MATHEMATIQUES ET INFORMATIQUE
Livre 1 — Mathématiques
Tome 83, 1989

ROUNDING ANALYSIS OF PARALLEL ALGORITHM FOR THE SOLUTION OF A TRIDIAGONAL LINEAR SYSTEM OF EQUATIONS

Plamen Piskulijski

Пламен Пискулийски. АНАЛИЗ ОШИБКИ ОКРУГЛЕНИЯ В ПАРАЛЛЕЛЬНОМ АЛГОРИТМЕ РЕШЕНИЯ ТРЕХДИАГОНАЛЬНЫХ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

В этой статье мы представляем параллельный алгоритм решения систем линейных алгебраических уравнений с трехдиагональными матрицами. Тоже проводим детальный анализ ошибки округления. Наконец, приведены результаты некоторых численных экспериментов и сделано сравнение между ошибками параллельных и последовательных алгоритмов.

Plamen Piskulijski. ROUNDING ANALYSIS OF PARALLEL ALGORITHM FOR THE SOLUTION OF A TRIDIAGONAL LINEAR SYSTEM OF EQUATIONS

In this paper we present a parallel algorithm for the solution of a tridiagonal linear system of equations.We also give a detailed rounding error analysis. Finaly we present some numerical experiments and give a comparison between the errors of parallel and consecutive algorithms.

1. INTRODUCTION

In this paper we examin the solution of tridiagonal systems of linear equations and estimate the rounding error. It is well known that such systems can be solved using a conventional serial computer in a time proportional to N where N is the number of equations.

Stone [13] first discussed the solution of a tridiagonal system on a parallel computer, relating the LDU decomposition using the first and second order linear recurrences. He developed recursive doubling algorithms to compute the necessary terms in $O(\log N)$ steps with $O(N)$ processors. The methods of odd-even elimination [14], [9] are another class of parallel algorithms with some quite different characteristics. There are many other algorithms for solving tridiagonal systems [7], [10], [5], [6], [3], which solve them in $O(\log N)$ steps with $O(N)$ processors.

Several authors have noted that cyclic reduction is just Gaussian elimination applied to PAP^T for a particular permutation matrix P (see for example [9]). Thus the algorithm is numerically stable for matrices for which Gaussian elimination is stable without pivoting, for example, symmetric positive definite or diagonally dominant matrices. The situation is not as attractive for Stone's algorithm. Using a stability analysis technique for recurrence relations introduced in [8], [11] the authors have shown that the algorithm is in general unstable, suffering from exponential error growth.

Sameh and Kuck [12] present two algorithms for tridiagonal systems using $O(N)$ processors. One of the algorithms requires $\log N$ steps but can suffer from exponential growth of errors; the more stable version requires $O[(\log\log N)(\log N)]$ steps.

There are consistent algorithms different from cyclic reduction. Swarztrauber [15], [16] introduced an algorithm for tridiagonal systems based on an efficient implementation of Cramer's rule. The algorithm requires $O(\log N)$ steps on $O(N)$ processors but only $O(N)$ total operations are performed. Unlike cyclic reduction the algorithm is well defined for general nonsingular systems.

Our goal in this paper is an error analysis of the parallel algorithm given in [3] using the approach from [2]. In section 2 we present the algorithm for solving tridiagonal systems of linear equations, which solve the system of order N in $O(n \log_n N)$ steps, where $N = n^s$. In section 3 we analize the rounding error in numerical solution of the system, using forward analysis. In section 4 we have performed some numerical experiments on an IBM–PC in machine precision $\approx 10^{-7}$ and we have given a comparison between the errors of parallel and consecutive algorithms.

2. PARALLEL ALGORITHM

The parallel algorithm from [3] will be applied for solving the tridiagonal system of equations

(2.1) $\quad A_i u_{i-1} + B_i u_i + C_i u_{i+1} = f_i, \quad u_0 = \alpha, \quad u_N = \beta, \quad i = 1, 2, \ldots, N-1$

under the assumption that the following conditions hold

(2.2) $$|A_i| + |C_i| \leqq |B_i|, \quad i = 1, 2, \ldots, N-1.$$

Conditions (2.2) ensure the existence of the algorithm of the system (2.1) and the stability of the problem [4]. For simplicity the problem (2.1), (2.2) can be written in the form:

(2.3) $$\Lambda u_i = f_i, \quad i = 1, 2, \ldots, N-1, \quad u_0 = \alpha, \quad u_N = \beta.$$

We represent the solution of the system (2.3) in the form

(2.4) $$u = \alpha u^{1,0} + \beta u^{0,1} + u^{0,0}$$

where the vectors $u^{1,0}$, $u^{0,1}$, $u^{0,0}$ are solutions of the systems :

$$\begin{aligned} &\Lambda u_i^{1,0} = 0, \quad \Lambda u_i^{0,1} = 0, \quad \Lambda u_i^{0,0} = f_i, \qquad i = 1, 2, \ldots, N-1, \\ &u_0^{1,0} = 1, \quad u_0^{0,1} = 0, \quad u_0^{0,0} = 0, \\ &u_N^{1,0} = 0, \quad u_N^{0,1} = 1, \quad u_N^{0,0} = 0 \end{aligned}$$

respectively.

Choosing the following knots of parallelism of the problem (2.3) $0 = i_0 < i_1 < \ldots < i_k = N$, we set

$$\alpha_m = u_{i_m}, \ m = 0, 1, \ldots, k; \tag{2.5}$$
$$u^m = \alpha_{m-1} u^{m,1,0} + \alpha_m u^{m,0,1} + u^{m,0,0}, \ m = 1, \ldots, k$$

where u^m, $u^{m,1,0}$, $u^{m,0,1}$, $u^{m,0,0}$ are solutions of linear systems

$$\Lambda u_i^m = f_i, \ i = i_{m-1} + 1, \ldots, i_m - 1, \quad u_{i_{m-1}}^m = \alpha_{m-1}, \quad u_{i_m}^m = \alpha_m, \tag{2.6}$$

$$\Lambda u_i^{m,1,0} = 0, \ i = i_{m-1} + 1, \ldots, i_m - 1, \quad u_{i_{m-1}}^{m,1,0} = 1, \quad u_{i_m}^{m,1,0} = 0, \tag{2.7}$$

$$\Lambda u_i^{m,0,1} = 0, \ i = i_{m-1} + 1, \ldots, i_m - 1, \quad u_{i_{m-1}}^{m,0,1} = 0, \quad u_{i_m}^{m,0,1} = 1, \tag{2.8}$$

$$\Lambda u_i^{m,0,0} = f_i, \ i = i_{m-1} + 1, \ldots, i_m - 1, \quad u_{i_{m-1}}^{m,0,0} = 0, \quad u_{i_m}^{m,0,0} = 0. \tag{2.9}$$

It follows from (2.5)–(2.9) that the solution u of (2.3) satisfies $u_i = u_i^m$ for $i = i_{m-1}, i_m$, i.e. the solution of the system (2.3) consists of the steps:

a) to solve the systems (2.7), (2.8), (2.9) in each interval (i_{m-1}, i_m);

b) to obtain α_i, $i = 0, 1, \ldots, k$;

c) to derive the solution by formulas (2.5).

The unknown quantities α_i are defined in the following way: from (2.1) and (2.5) we obtain

$$u^m = \alpha_{m-1} u^{m,1,0} + \alpha_m u^{m,0,1} + u^{m,0,0}, \tag{2.10}$$
$$u^{m+1} = \alpha_m u^{m+1,1,0} + \alpha_{m+1} u^{m+1,0,1} + u^{m+1,0,0},$$

$$A_{i_m} u_{i_m - 1} + B_{i_m} u_{i_m} + C_{i_m} u_{i_m+1} = A_{i_m} u_{i_m-1}^m + B_{i_m} \alpha_m + C_{i_m} u_{i_m+1}^{m+1} = f_{i_m} \tag{2.11}$$

and substituting (2.10) in (2.11) we get

$$a_m \alpha_{m-1} + b_m \alpha_m + c_m \alpha_{m+1} = \varphi_m, \ m = 1, \ldots, k-1, \quad \alpha_0 = \alpha, \quad \alpha_k = \beta \tag{2.12}$$

where

$$a_m = A_{i_m} u_{i_m-1}^{m,1,0}, \quad b_m = B_{i_m} + A_{i_m} u_{i_m-1}^{m,0,1} + C_{i_m} u_{i_m+1}^{m+1,1,0}, \tag{2.13}$$
$$c_m = C_{i_m} u_{i_m+1}^{m+1,0,1}, \quad \varphi_m = f_{i_m} - A_{i_m} u_{i_m-1}^{m,0,0} - C_{i_m} u_{i_m+1}^{m+1,0,0}.$$

So the obtained parallel algorithm for solving the system (2.3) is:

a) to solve in parallel the systems (2.7) – (2.9) in each interval (i_m, i_{m+1});

b) to obtain a_m, b_m, c_m, φ_m for $m = 1, \ldots, k-1$ in parallel in each interval by formulas (2.13);

c) to solved the system (2.12) by some method;

d) to derive in parallel the solution of the problem by formulas (2.5).

Let us suppose for convenience, that $N = n^s$, where n and s are integers, such that $n, s \geq 2$. Choosing $k = N/n = n^{s-1}$ equidistant knots of parallelism i_m, i.e. $i_m = mn$, $m = 0, 1, \ldots, k$ and solving (9) by the described method, we come to the problem (2.12) for the determination of α_m, $m = 0, 1, \ldots, k$. The problem (2.12) is in the same form as (2.3) but it is of size n-times smaller than (2.3). We again apply the same parallel scheme for the system (2.12), but this time with

$k/n = n^{s-2}$ equidistant points j_l, $l = 0, \ldots, n^{s-2}$ and get the tridiagonal system of size n^2 times smaller than (2.3). Finaly we obtain the system of the kind (2.12) which is of size n and we solve it by the same method. After that we deduce the solution of the system (2.3) from formulas (2.5).

3. ERROR ANALYSIS

Let us have the system

$$Au = f \tag{3.1}$$

and instead of this system let us solve the system

$$\tilde{A}\tilde{u} = \tilde{f} \tag{3.2}$$

where $\tilde{A} = A + \delta A$, $\tilde{f} = f + \delta f$, $\overline{u} = u + \delta u$. Then the relative data error satisfies

$$\frac{\|\delta u\|}{\|u\|} \leq \nu(\tilde{A})\frac{\|\delta A\|}{\|\tilde{A}\|} + \nu(\tilde{A})\frac{\|\delta A\|\,\|\delta f\|}{\|\tilde{A}\|\,\|f\|} + \nu(\tilde{A})\frac{\|\delta f\|}{\|f\|} \tag{3.3}$$

where $\nu(\tilde{A}) = \|\tilde{A}\| \cdot \|\tilde{A}^{-1}\|$ is the condition number of $\tilde{A}$. Consequently, the full error Δu satisfies $\|\Delta u\| \leq \|u - \tilde{u}\| + \|\tilde{u} - \overline{u}\|$ where the first member is the error from inexactness of coeficients (data error) and the second member is rounding error.

First we will examine the tridiagonal systems (2.7) and (2.8). The system (2.7) has the form

$$A_j u_{j-1} + B_j u_j + C_j u_{j+1} = 0,\ j = 2, \ldots, n-1, \quad u_1 = 1, \quad u_n = 0. \tag{3.4}$$

We will apply for it the following formulas:

$$\begin{aligned} &\alpha_j = -\frac{A_j}{B_j - C_j\alpha_{j+1}}, \quad \beta_j = \frac{f_j - C_j\beta_{j+1}}{B_j - C_j\alpha_{j+1}}, \quad j = n, n-1, \ldots, 1, \\ &u_j = \alpha_j u_{j-1} + \beta_j, \quad j = 1, \ldots, n, \end{aligned} \tag{3.5}$$

in order to evaluate a solution of the system. Since $|\alpha_n| = 0$ and $|\alpha_{n-1}| = |C_{n-1}/B_{n-1}| \leq 1$, we obtain by induction that

$$|\alpha_j| \leq |C_j|/(|B_j| - |A_j|) \leq 1, \quad j = n-2, \ldots, 1. \tag{3.6}$$

Since $\beta_n = \beta_{n-1} = \cdots = \beta_2 = 0$, $\beta_1 = 1$ and $u_1 = \beta_1 = 1$, it is easily derived that

$$|u_j| \leq 1,\ j = 1, \ldots, n. \tag{3.7}$$

For the system (2.8), which is in the form

$$A_j u_{j-1} + B_j u_j + C_j u_{j+1} = 0,\ j = 2, \ldots, n-1, \quad u_1 = 0, \quad u_n = 1, \tag{3.8}$$

we apply the following formulas

$$\begin{aligned} &\alpha_j = -\frac{C_j}{B_j + A_j\alpha_{j-1}}, \quad \beta_j = \frac{f_j - A_j\beta_{j-1}}{B_j + A_j\alpha_{j-1}}, \qquad j = 1, \ldots, n, \\ &u_j = \alpha_j u_{j+1} + \beta_j, \quad j = n, \ldots, 1, \end{aligned} \tag{3.9}$$

in order to evaluate a solution of the system. By analogy with (3.6) we deduce that $|\alpha_j| \leq 1$, $j = 1, \ldots, n$, $\beta_1 = \cdots = \beta_{n-1} = 0$, $\beta_n = 1$, and since $|u_n| = 1$ it is easily seen that

$$|u_j| \leq 1, \quad j = n, \ldots, 1, \tag{3.10}$$

We obtained that the solutions of the systems (2.7) and (2.8) do not exceed unity in absolute value.

Using the results of [2] we get that the rounding error of the system (2.7) satisfies

$$\begin{aligned}\left|\delta u_{n-i}^{m,1,0}\right| &\leqq \varepsilon\{i + C\kappa[n(n+1)-(n-i)(n-i+1)]/2 \\ (3.11)\quad &+ C_1\kappa[(n+1)(n+2)(2n+3)-(n-i+1)(n-i+2)(2n-2i+3)]/12\}, \\ \left|\delta u_{n}^{m,1,0}\right| &\leqq C_1\kappa\frac{n(n+1)}{2}\varepsilon, \quad i=1,\dots,n-1\end{aligned}$$

where ε is an unit roundoff, C does not depend on ε and n, and

(3.12) $$\varepsilon\kappa n(n-1)/2 \leqq 1-1/C$$

where

$$\kappa = \max_{1\leqq j\leqq s}\{\kappa_j\}, \quad \kappa_j = \max_{2\leqq k\leqq n-1}\left\{\prod_{i=2}^{k}|\gamma_i|\right\}, \quad \gamma_j = \alpha_j A_j/C_j,$$

and

$$C_1 = 1 + \max_{2\leqq j\leqq n-1}\{|A_j\overline{\beta}_{j-1}/C_j|\}.$$

It turns out that $|\delta u_s^{m,1,0}| = O(\varepsilon n^3)$, if the conditions (2.2) are valid.

Now we examine the system (2.8) which is of order n. It is of the same type (3.8) and condition (2.2) holds for it. Using [2] and since $\beta_1 = \beta_2 = \cdots = \beta_{n-1} = 0$, $\beta_n = 1$ for the rounding error of the system (2.8) it follows

$$\begin{aligned}(3.13)\quad \left|\delta u_{n-i}^{m,0,1}\right| &\leqq \varepsilon\left\{i + C\frac{\kappa}{2}[n(n+1)-(n-i)(n-i+1)]\right\}, \quad i=1,\dots,n-1, \\ \left|\delta u_{n}^{m,0,1}\right| &\leqq \varepsilon,\end{aligned}$$

i.e. $|\delta u_s^{m,0,1}| = O(\varepsilon n^2)$.

Now it remains to evaluate the error for the system (2.9)

(3.14) $$A_j u_{j-1} + B_j u_j + C_j u_{j+1} = f_j, \quad j=2,\dots,n-1, \quad u_1 = 0, \quad u_n = 0.$$

Conditions (2.2) are valid for it, too. Similarly, for this system (3.14) we find the following estimations

$$\begin{aligned}\left|\delta u_{n-i}^{m,0,0}\right| &\leqq \varepsilon\{i + C\kappa[n(n+1)-(n-i)(n-i+1)]\overline{u}/2 \\ (3.15)\quad &+ C_1\kappa[(n+1)(n+2)(2n+3)-(n-i+1)(n-i+2)(2n-2i+3)]/12\}. \\ \left|\delta u_{n}^{m,0,0}\right| &\leqq C_1\kappa\frac{n(n+1)}{2}\varepsilon, \quad i=1,\dots,n-1,\end{aligned}$$

under the assumption that $\overline{u}_j^{m,0,0}$ are bounded, i.e. $\overline{u} = \max_j\left\{|\overline{u}_j^{m,0,0}|\right\}$. We obtained that $|\delta u_j^{m,0,0}| = O(\varepsilon n^3)$.

On the next stage of the algorithm the system (2.12) must be solved with coeficients (2.13). This system is tridiagonal and is of size n times smaller than (2.3). Conditions (2.2) hold for it.

Now we will derive rounding error of the coeficients (2.13). Instead of the exact value $a_m^{(2)}$, where the superscripts denote the stage of the algorithm, let some approximation value $\overline{a}_m^{(2)}$ be taken, such that $\overline{a}_m^{(2)} = A_{i_m}\overline{u}_{i_m-1}^{m,1,0} + \varepsilon$. Then

$$\left|\delta a_m^{(2)}\right| = \left|\overline{a}_m^{(2)} - a_m^{(2)}\right| \leqq |A_{i_m}| \left|\delta u_{i_m-1}^{m,1,0}\right| + \varepsilon,$$

and from (3.11) it follows

$$\left|\delta a_m^{(2)}\right| \leqq \varepsilon(1+|A_{i_m}|) + |A_{i_m}|\varepsilon C\kappa n + |A_{i_m}|\varepsilon C_1\kappa(n^2+2n+1)/2. \tag{3.16}$$

By analogy with (3.16), we derive the following estimations for $|\delta c_m^{(2)}|$, $|\delta b_m^{(2)}|$, $|\delta\varphi_m^{(2)}|$:

$$\left|\delta c_m^{(2)}\right| \leqq \varepsilon(n-2+|C_{i_m}|) + |C_{i_m}|\varepsilon C\frac{\kappa}{2}(n^2+n-6) \tag{3.17}$$

$$\left|\delta b_m^{(2)}\right| \leqq \varepsilon(1+|A_{i_m}|+(n-2)|C_{i_m}|) + |A_{i_m}|(n^2+3n-6)\varepsilon C\frac{\kappa}{2} + |C_{i_m}|\varepsilon C_1\kappa(2n^3+9n^2+13n-78)/12 \tag{3.18}$$

$$\left|\delta\varphi_m^{(2)}\right| \leqq \varepsilon(1+|A_{i_m}|+(n-2)|C_{i_m}|) + |A_{i_m}|(n^2+3n-6)\varepsilon C\frac{\kappa}{2}\overline{u} + \left(|A_{i_m}|6(n^2+2n+1) + |C_{i_m}|(2n^3+9n^2+13n-78)\right)\varepsilon C_1\kappa/12. \tag{3.19}$$

We again apply the same parallel scheme for the system (2.12), but this time with $k/n = n^{s-2}$ equidistant points j_l, $l = 0, \ldots, n^{s-1}$. Thus, the obtained tridiagonal systems is of the same kind as (2.7) – (2.9), but the coeficients of these systems of order n are the coeficients of the system (2.12). And they already have some error and an error in the new systems appears from the inexactness of coeficients (data error). This error must be estimated.

The full error Δu satisfies $\|\Delta u\| \leqq \|u-\tilde{u}\|+\|\tilde{u}-\overline{u}\|$ and now we have to estimate the full error Δu. The rounding error is evaluated similarly to the rounding error of the systems (2.7) – (2.9). So we evaluate the data error by (3.3) using that the condition number for tridiagonal matrices is $O(n^2)$ [2].

Throughout this section the superscripts in brackets denote the stage of the algorithm and $\|\varphi\|$, $\|\tilde{A}\|$ are a maximum of the norms of φ and the matrix of the system (2.9) at different steps respectively. First, let us examine the system of the kind (2.7). For it $\|\delta\varphi\| = 0$ and then

$$\frac{\|\overline{\delta}\alpha^{m,1,0}\|^{(2)}}{\|\alpha^{m,1,0}\|^{(2)}} \leqq \nu(\tilde{A})\frac{\|\delta A\|^{(2)}}{\|\tilde{A}\|} \leqq \frac{\nu(\tilde{A})}{\|\tilde{A}\|}A_n,$$

where

$$A_n = |C_{i_m}|\varepsilon\frac{\kappa}{2}\left[C_1(2n^3+9n^2+13n-78)/6 + 2C(n^2+n-6)\right] + |C_{i_m}|\varepsilon(n-2) + |A_{i_m}|\varepsilon\frac{\kappa}{2}\left[C_1(n^2+2n+1)+4Cn\right] + \varepsilon\left[2+2|A_{i_m}|+|C_{i_m}|\right],$$

and A is the matrix of corresponding system. Since the rounding error of this system is evaluated by (3.11), then

$$\|\Delta\alpha^{m,1,0}\|^{(2)} \leqq \|\overline{\delta}\alpha^{m,1,0}\|^{(2)} + \|\delta\alpha^{m,1,0}\|^{(2)} \leqq \nu(\tilde{A})A_n/\|\tilde{A}\| + \overline{A}_n, \tag{3.20}$$

$$\overline{A}_n = \varepsilon[n - 1 + C\frac{\kappa}{2}(n^2 + n - 2) + C_1\kappa(2n^3 + 9n^2 + 13n - 24)/12].$$

For the system (2.8) also $\|\delta\varphi\| = 0$ and the corresponding error is

$$(3.21) \qquad \|\Delta\alpha^{m,0,1}\|^{(2)} \leq \|\overline{\delta}\alpha^{m,0,1}\|^{(2)} + \|\delta\alpha^{m,0,1}\|^{(2)} \leq \nu(\tilde{A})A_n/\|\tilde{A}\| + \tilde{A}_n,$$

$$\tilde{A}_n = \varepsilon[n - 1 + C\frac{\kappa}{2}(n^2 + n - 2)].$$

It remains to examine the system of kind (2.9). For it $\|\delta\varphi\| \neq 0$ and then from (3.3) it is derived

$$\frac{\|\overline{\delta}\alpha^{m,0,0}\|^{(2)}}{\|\alpha^{m,0,0}\|^{(2)}} \leq \nu(\tilde{A})\frac{\|\delta A\|^{(2)}}{\|\tilde{A}\|} + \nu(\tilde{A})\frac{\|\delta A\|^{(2)}\|\delta\varphi\|^{(2)}}{\|\tilde{A}\|\,\|\varphi\|} + \nu(\tilde{A})\frac{\|\delta\varphi\|^{(2)}}{\|\varphi\|}$$

$$\leq \nu(\tilde{A})\frac{A_n}{\|\tilde{A}\|} + \nu(\tilde{A})\frac{A_n B_n}{\|\tilde{A}\|\,\|\varphi\|} + \nu(\tilde{A})\frac{B_n}{\|\varphi\|},$$

$$B_n = \varepsilon\,(1 + |A_{i_m}| + (n-2)\,|C_{i_m}|) + \left(|A_{i_m}|\,2n + |C_{i_m}|\,(n^2 + n - 6)\right)\varepsilon C\,\overline{u}\kappa/2$$
$$+ \left(|A_{i_m}|\,6(n^2 + 2n + 1) + |C_{i_m}|\,(2n^3 + 9n^2 + 13n - 78)\right)\varepsilon C_1\kappa/12.$$

Then for the system (2.9) we find

$$(3.22) \quad \|\Delta\alpha^{m,0,0}\|^{(2)} \leq \left(\nu(\tilde{A})\frac{A_n}{\|\tilde{A}\|} + \nu(\tilde{A})\frac{A_n B_n}{\|\tilde{A}\|\,\|\varphi\|} + \nu(\tilde{A})\frac{B_n}{\|\varphi\|}\right)\|\alpha^{m,0,0}\|^{(2)} + \overline{B}_n,$$

$$\overline{B}_n = \varepsilon[n - 1 + C\frac{\kappa}{2}(n^2 + n - 2)\overline{u} + C_1\kappa(2n^3 + 9n^2 + 13n - 24)/12].$$

Now we derive the rounding error of coeficients (2.13) at the third stage. Let instead of $a_m^{(3)}$ an approximation value $\overline{a}_m^{(3)}$ be taken, such that

$$\overline{a}_m^{(3)} = \overline{a}_{i_m}^{(2)}\left(\overline{\alpha}_{i_m-1}^{m,1,0}\right)^{(2)} + \varepsilon.$$

Then

$$\left|\delta a_m^{(3)}\right| = \left|\overline{a}_m^{(3)} - a_m^{(3)}\right| \leq \left|a_m^{(2)}\right|\,\|\delta\alpha^{m,1,0}\|^{(2)} + \left|\delta a_m^{(2)}\right|\,\left|\overline{\alpha}_{i_m-1}^{m,1,0}\right|^{(2)} + \varepsilon,$$

and since (3.7) holds, then

$$\left|\delta a_m^{(3)}\right| \leq \left|a_m^{(2)}\right|\,\|\Delta\alpha^{m,1,0}\|^{(2)} + \left|\delta a_m^{(2)}\right| + \varepsilon.$$

Then from (3.20) it follows that

$$(3.23) \qquad \left|\delta a_m^{(3)}\right| \leq \left|a_m^{(2)}\right|\left(\nu(\tilde{A})A_n/\|\tilde{A}\| + \overline{A}_n\right) + \left|\delta a_m^{(2)}\right| + \varepsilon.$$

By analogy with (3.23), we derive the following estimations for $\left|\delta c_m^{(3)}\right|$, $\left|\delta b_m^{(3)}\right|$, $\left|\delta\varphi_m^{(3)}\right|$

$$(3.24) \qquad \left|\delta c_m^{(3)}\right| \leq \left|c_m^{(2)}\right|\left(\nu(\tilde{A})A_n/\|\tilde{A}\| + \tilde{A}_n\right) + \left|\delta c_m^{(2)}\right| + \varepsilon,$$

$$(3.25)\quad \left|\delta b_m^{(3)}\right| \le \left(\left|c_m^{(2)}\right| + \left|a_m^{(2)}\right|\right)\nu(\tilde{A})\frac{A_n}{\|\tilde{A}\|} + \left|a_m^{(2)}\right|\tilde{A}_n + \left|c_m^{(2)}\right|\overline{A}_n + \left|\delta a_m^{(2)}\right| + \left|\delta c_m^{(2)}\right| + \left|\delta b_m^{(2)}\right| + \varepsilon,$$

$$(3.26)\quad \left|\delta\varphi_m^{(3)}\right| \le \left(\left|c_m^{(2)}\right| + \left|a_m^{(2)}\right|\right)\nu(\tilde{A})M_1 + \left(\left|\delta a_m^{(2)}\right| + \left|\delta c_m^{(2)}\right|\right)\overline{\alpha} + \left|\delta\varphi_m^{(2)}\right| + \varepsilon,$$

$$(3.27)\quad M_1 = A_n/\|\tilde{A}\| + A_n B_n/\|\tilde{A}\|\,\|\varphi\| + B_n/\|\varphi\| + \overline{B}_n.$$

On the next stage from the formulas (3.20) – (3.22) it follows

$$\|\Delta\alpha^{m,1,0}\|^{(3)} \le \frac{\nu(\tilde{A})}{\|\tilde{A}\|}\left[\frac{\nu(\tilde{A})}{\|\tilde{A}\|}A_n\left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right) + \left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right)(\overline{A}_n + \tilde{A}_n) + 2\left|\delta a_m^{(2)}\right| + 2\left|\delta c_m^{(2)}\right| + \left|\delta b_m^{(2)}\right| + 3\varepsilon\right] + \overline{A}_n,$$

$$\|\Delta\alpha^{m,0,1}\|^{(3)} \le \frac{\nu(\tilde{A})}{\|\tilde{A}\|}\left[\frac{\nu(\tilde{A})}{\|\tilde{A}\|}A_n\left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right) + \left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right)(\overline{A}_n + \tilde{A}_n) + 2\left|\delta a_m^{(2)}\right| + 2\left|\delta c_m^{(2)}\right| + \left|\delta b_m^{(2)}\right| + 3\varepsilon\right] + \tilde{A}_n,$$

$$\|\Delta\alpha^{m,0,0}\|^{(3)} \le \nu(\tilde{A})^2\left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right)M_2 + 2\frac{\nu(\tilde{A})^3}{\|\tilde{A}\|^2\|\varphi\|}\left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right)A_n M_1 + \nu(\tilde{A})\left\{\frac{1}{\|\varphi\|}\left[\left(\left|\delta a_m^{(2)}\right| + \left|\delta c_m^{(2)}\right|\right)\overline{\alpha} + \left|\delta\varphi_m^{(2)}\right| + \varepsilon\right]\right.$$
$$\left. + \frac{1}{\|A\|}\left[\left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right)\left(\overline{A}_n + \tilde{A}_n\right) + 2\left|\delta a_m^{(2)}\right| + 2\left|\delta c_m^{(2)}\right| + \left|\delta b_m^{(2)}\right| + 3\varepsilon\right]\right\} + O(\varepsilon n^3),$$

where $M_2 = M_1/\|\varphi\| + 2A_n/\|\tilde{A}\|^2$.

Finally, at the $(s-1)$-th stage the system of order n and of type (2.12) is obtained and it can be written in the form

$$(3.28)\quad \begin{aligned} &a_m^{(s)}\alpha_{m-1} + b_m^{(s)}\alpha_m + c_m^{(s)}\alpha_{m+1} = \varphi_m^{(s)}, \quad m = 1,\dots,n-1,\\ &\alpha_0 = \alpha, \quad \alpha_n = \beta, \end{aligned}$$

where the error for the coeficients is:

$$\begin{aligned}
\left|\delta a_m^{(s)}\right| &\leqq \left|a_m^{(s-1)}\right| \|\Delta\alpha^{m,1,0}\|^{(s-1)} + \left|\delta a_m^{(s-1)}\right| + \varepsilon,\\
\left|\delta c_m^{(s)}\right| &\leqq \left|c_m^{(s-1)}\right| \|\Delta\alpha^{m,0,1}\|^{(s-1)} + \left|\delta c_m^{(s-1)}\right| + \varepsilon,\\
\left|\delta b_m^{(s)}\right| &\leqq \left|c_m^{(s-1)}\right| \|\Delta\alpha^{m,1,0}\|^{(s-1)} + \left|a_m^{(s-1)}\right| \|\Delta\alpha^{m,0,1}\|^{(s-1)}\\
&\quad + \left|\delta a_m^{(s-1)}\right| + \left|\delta c_m^{(s-1)}\right| + \left|\delta b_m^{(s-1)}\right| + \varepsilon,\\
\left|\delta \varphi_m^{(s)}\right| &\leqq \left(\left|c_m^{(s-1)}\right| + \left|a_m^{(s-1)}\right|\right) \|\Delta\alpha^{m,0,0}\|^{(s-1)}\\
&\quad + \left(\left|\delta a_m^{(s-1)}\right| + \left|\delta c_m^{(s-1)}\right|\right)\overline{\alpha} + \left|\delta \varphi_m^{(s-1)}\right| + \varepsilon,
\end{aligned}$$

$$\begin{aligned}
\|\Delta\alpha^{m,1,0}\|^{(s-1)} \leqq \frac{\nu(\tilde{A})}{\|\tilde{A}\|}\Bigg(& 2^{s-3}\frac{\nu(\tilde{A})^{s-3}}{\|\tilde{A}\|^{s-3}}A_n C_{s-2}\\
&+ 2^{s-5}3\,\frac{\nu(\tilde{A})^{s-4}}{\|\tilde{A}\|^{s-4}}A_n C_{s-3} + O(\varepsilon n^{2s-7})\Bigg) + \overline{A}_n,
\end{aligned}$$

$$\begin{aligned}
\|\Delta\alpha^{m,0,1}\|^{(s-1)} \leqq \frac{\nu(\tilde{A})}{\|\tilde{A}\|}\Bigg(& 2^{s-3}\frac{\nu(\tilde{A})^{s-3}}{\|\tilde{A}\|^{s-3}}A_n C_{s-2}\\
&+ 2^{s-5}3\,\frac{\nu(\tilde{A})^{s-4}}{\|\tilde{A}\|^{s-4}}A_n C_{s-3} + O(\varepsilon n^{2s-7})\Bigg) + \tilde{A}_n,
\end{aligned}$$

$$\|\Delta\alpha^{m,0,0}\|^{(s-1)} \leqq \nu(\tilde{A})^{s-2}M_{s-2}C_{s-2} + 2^{s-3}\frac{\nu(\tilde{A})^{2s-5}}{\|\tilde{A}\|^{s-2}\|\varphi\|}A_n M_{s-3}C_{s-2}^2 + O(\varepsilon n^{2s-3}),$$

$$C_{s-i} = \left(\left|a_m^{(s-i)}\right| + \left|c_m^{(s-i)}\right|\right)\dots\left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right), \quad i = 1,\dots,s-2,$$

$$M_{s-i} = M_{s-i-1}/\|\varphi\| + A_n 2^{s-i+1}/\|\tilde{A}\|^{s-i}, \quad i = 1,\dots,s-2,$$

and M_1 is defined by (3.28).

Now we have to evaluate the solution of the system (3.28) in order to find the rounding error of the system (2.3). The total error can be estimated by $\|\Delta\alpha\|^{(s)} \leqq \|\alpha - \tilde{\alpha}\| + \|\tilde{\alpha} - \overline{\alpha}\|$ and consequently

$$\|\Delta\alpha\|^{(s)} \leqq \nu(\tilde{A})^{s-1}M_{s-1}C_{s-1} + 2^{s-2}\frac{\nu(\tilde{A})^{2s-3}}{\|\tilde{A}\|^{s-1}\|\varphi\|}A_n M_{s-2}C_{s-1}^2 + O(\varepsilon n^{2s-1}).$$

To estimate the product C_{s-i} let us set

(3.29) $$M = \max_{1 \leqq k \leqq N}(|A_k|, |C_k|).$$

Then from (3.7), (3.10) and (3.29) we get

$$\begin{aligned} C_{s-i} &= \left(\left|a_m^{(s-i)}\right| + \left|c_m^{(s-i)}\right|\right) \dots \left(\left|a_m^{(2)}\right| + \left|c_m^{(2)}\right|\right) \\ &\leq \left(M \left|u_{i_m-1}^{m,1,0}\right|^{(1)} \dots \left|\alpha_{i_m-1}^{m,1,0}\right|^{(s-i)} + M \left|u_{i_m+1}^{m+1,0,1}\right|^{(1)}\right. \\ &\quad \left. \dots \left|\alpha_{i_m+1}^{m+1,0,1}\right|^{(s-i)}\right) \dots \left(M \left|u_{i_m-1}^{m,1,0}\right|^{(1)} + M \left|u_{i_m+1}^{m+1,0,1}\right|^{(1)}\right) \\ &\leq M^{s-2} 2^{s-2}, \qquad i = 1, \dots, s-2. \end{aligned}$$

Developing M_{s-2} and M_{s-1} we obtain that the error of the system (2.3) can be estimated by

$$\begin{aligned} \|\Delta\alpha\|^{(s)} &\leq \nu(\tilde{A})^{s-1} M^{s-2} C_1 n^3 2^{s-2} K_1 \frac{\kappa}{6} \\ &\quad + \nu(\tilde{A})^{s-1} M^{s-1} 2^{s-2} \varepsilon n^2 \frac{\kappa}{12} (6CK_2 + 15 C_1 K_3) \\ &\quad + 2^s \nu(\tilde{A})^{2s-3} \left(M^{s-1} 2^{s-2} \varepsilon C_1 n^3 \frac{\kappa}{12}\right)^2 K_4 / \|\tilde{A}\|^{s-1} \|\varphi\|^2 \\ &\quad + O(\varepsilon n^{2s-1}), \end{aligned}$$

$$\begin{aligned} K_1 &= (\|\tilde{A}\| M + \|\varphi\| M + \|\varphi\| \, \|\tilde{A}\|) / \|\varphi\|^{s-1} \|\tilde{A}\| + P_{s-1} M / 2 \|\varphi\|, \\ K_2 &= (2\|\varphi\| + \|\varphi\| \, \|\tilde{A}\| \overline{u} / M) / \|\varphi\|^{s-1} \|\tilde{A}\| + P_{s-1} / \|\varphi\|, \\ K_3 &= (\|\varphi\| + \|\tilde{A}\|^2 12 \overline{u} / C_1 \kappa) / \|\varphi\|^{s-1} \|\tilde{A}\| + P_{s-1} / 2 \|\varphi\|, \\ K_4 &= (\|\tilde{A}\| + \|\varphi\| + \|\varphi\| \, \|\tilde{A}\| / M) / \|\varphi\|^{s-3} \|\tilde{A}\| + P_{s-2}/2, \\ P_{s-i} &= 4(\|\tilde{A}\|^{s-i-1} - 2^{s-i-1}) / \|\tilde{A}\|^{s-i} (\|\tilde{A}\|^2 - 2), \quad i = 1, 2. \end{aligned}$$

And finally we have to estimate the rounding error from (2.5) at each stage. By analogy with (3.23)–(3.26) it follows for the rounding error of the solution of the system (2.3):

$$\begin{aligned} \text{(3.30)} \quad \|\delta u\| &\leq 2^{s-1} \|\delta\alpha\|^{(s)} + \|\overline{\alpha}\| \sum_{i=1}^{s-2} 2^{s-i-1} \left(\|\delta\alpha^{m,1,0}\|^{(s-i)} + \|\delta\alpha^{m,0,1}\|^{(s-i)}\right) \\ &\quad + \sum_{i=1}^{s-2} 2^{s-i-1} \|\delta\alpha^{m,0,0}\|^{(s-i)}, \end{aligned}$$

where $\|\overline{\alpha}\|$ is a maximum of the norms of solutions α at the systems of kind (2.12) at the different steps. And consequently

$$\begin{aligned} \|\Delta u\| &\leq \nu(\tilde{A})^{s-1} M^{s-2} C_1 n^3 2^{2s-3} K_1 \frac{\kappa}{6} \\ &\quad + \nu(\tilde{A})^{s-1} M^{s-1} 2^{2s-3} \varepsilon n^2 \frac{\kappa}{12} (6CK_2 + 15 C_1 K_3) \\ &\quad + 2^{2s-1} \nu(\tilde{A})^{2s-3} \left(M^{s-1} 2^{s-2} \varepsilon C_1 n^3 \frac{\kappa}{12}\right)^2 K_4 + O(\varepsilon n^{2s-1}) / \|\tilde{A}\|^{s-1} \|\varphi\|^2. \end{aligned}$$

N	$\varepsilon \approx 10^{-7}$, $\max \lvert x_i - \overline{x}_i \rvert$	
	parallel alg.	conseq. alg.
$3^6 + 1 = 730$	79,64	
$9^3 + 1 = 730$	12,77344	1,539
$27^2 + 1 = 730$	5,09375	
$6^4 + 1 = 1297$	188,6	7,09375
$36^2 + 1 = 1297$	51,5	
$2^{12} + 1 = 4097$	0	
$4^6 + 1 = 4097$	0	
$8^4 + 1 = 4097$	0	5226,626
$16^3 + 1 = 4097$	2506,875	
$64^2 + 1 = 4097$	1712	
$3^8 + 1 = 6562$	574586	
$9^4 + 1 = 6562$	84836,5	45210,5
$81^2 + 1 = 6562$	34623,5	
$10^4 + 1 = 10001$	309374	1365653
$100^2 + 1 = 10001$	167126	
$5^6 + 1 = 15626$	9342526	
$25^3 + 1 = 15626$	117528	$1,134279 . 10^7$
$125^2 + 1 = 15626$	926624	
$12^4 + 1 = 20737$	3919304	$2,955723 . 10^7$
$144^2 + 1 = 20737$	463156	

Table 1

We have from [2] that

(3.31)
$$|\delta u_1| \leq \varepsilon\{N - 1 + C\kappa(N^2 + N - 2)\overline{u}/2 + C_1\kappa(2N^3 + 9N^2 + 13N - 24)/12\}.$$

Since $\nu(A) = O(n^2)$ for tridiagonal matrices [2], it is easily seen that if

$$2^{2s-3}M^{s-2}K_1 \approx n^{s-1} \quad \text{and} \quad 2^{2s-3}M^{s-1}(K_2 + K_4/18 + 15K_3/9) \approx n^s,$$

then the errors of parallel and consecutive algorithms will be approximately equal.

Note that if instead of (2.2) we have

(3.32)
$$|A_i| + |C_i| < |B_i|, \quad i = 1, \ldots, N - 1,$$

then the rounding error for solving the tridiagonal systems is $|\delta u_i| \leq O(\varepsilon N)$, $i = 0, \ldots, N$. Since the condition number for such tridiagonal matrices is constant [1], the rounding error of the parallel algorithm given in [3] satisfies $\|\delta u\| \leq O(\varepsilon n)$.

4. NUMERICAL EXPERIMENTS

In this section a comparison of error bounds is given between the parallel and consequtive algorithms. For illustration of the above estimations with respect to the rounding error we have performed some numerical experiments for solving the

N	$\varepsilon \approx 10^{-7}$, $\max \lvert x_i - \overline{x}_i \rvert$	
	parallel alg.	conseq. alg.
$3^6+1=730$ $9^3+1=730$ $27^2+1=730$	$1{,}43.10^{-3}$ $2{,}30.10^{-4}$ $9{,}19.10^{-5}$	$2{,}60.10^{-5}$
$6^4+1=1297$ $36^2+1=1297$	$1{,}08.10^{-3}$ $2{,}94.10^{-4}$	$4{,}82.10^{-5}$
$2^{12}+1=4097$ $4^6+1=4097$ $8^4+1=4097$ $16^3+1=4097$ $64^2+1=4097$	0 0 0 $1{,}43.10^{-3}$ $9{,}79.10^{-4}$	$3{,}96.10^{-3}$
$3^8+1=6562$ $9^4+1=6562$ $81^2+1=6562$	$1{,}28.10^{-1}$ $1{,}89.10^{-2}$ $7{,}72.10^{-3}$	$1{,}08.10^{-2}$
$10^4+1=10001$ $100^2+1=10001$	$2{,}95.10^{-2}$ $1{,}58.10^{-2}$	$1{,}45.10^{-1}$
$5^6+1=15626$ $25^3+1=15626$ $125^2+1=15626$	$3{,}69.10^{-1}$ $4{,}62.10^{-3}$ $3{,}63.10^{-2}$	$4{,}24.10^{-1}$
$12^4+1=20737$ $144^2+1=20737$	$8{,}73.10^{-2}$ $1{,}03.10^{-2}$	$6{,}19.10^{-1}$

Table 2

following tridiagonal systems of equations

$$(4.1)\qquad \begin{aligned} &u_{i-1}-2u_i+u_{i+1}=-1, \quad i=2,\ldots,N-1,\\ &u_1=0, \quad u_N=N-1,\end{aligned}$$

$$(4.2)\qquad \begin{aligned} &-u_{i-1}+2u_i-u_{i+1}=0, \quad i=2,\ldots,N-1,\\ &2u_1-u_2=1, \quad -u_{N-1}+2u_N=1\end{aligned}$$

for different choices of N. Let $\overline{u}_i$ denote the computed solution of the systems. The exact solutions of the systems (4.1) and (4.2) are $u_{N+1-k}=k(n-k)/2$, $k=1,\ldots,N$ and $u_i=1$, $i=1,\ldots,N$, respectively.

In most of the tested cases, the absolute error of the parallel algorithm is better than this of the consequtive one. It turns out to be advisable to use parallel algorithm when N is large. Also if n_1 and n_2 are such that $N=n_1^a=n_2^b$ and $n_1<n_2$, then the error of parallel algorithm is better for n_2.

Finally, in view of the above, it seems to be advisable to use a double lenght accumulator (see [17]). This will have a favorable practical effect on stabilizing the algorithm and reducing the error bounds.

Here are given two tables illustrating the results of this paper. The table 1 and table 2 concern the systems (4.1) and (4.2), respectively.

REFERENCES

1. Завьялов, Ю. С., Б. И. Квасов, В. Л. Мирошниченко. Методы сплайн-функции. М., 1980.
2. Ильин, В. П., Ю. И. Кузнецов. Трехдиагональные матрицы и их приложения. М., Наука, 1985, 52–59.
3. Коновалов, А. Н., Н. Н. Яненко. Некоторые вопросы теории модульного анализа параллельного программирования для задач математической физики и механики сплошной среды. Современные проблемы мат. физики и вычислительной математики, Наука, 1982.
4. Самарский, А. А. Теория разностных схем. М., Наука, 1977.
5. Buneman, O. A compact non-iterative Poisson solver. Inst. for Plasma Res., Stanford Univ., Rep. SU-IPR-294, Stanford, 1969.
6. Buzbee, B. L., G. H. Golub, C. W. Nielson. On direct methods for solving Poisson's equations. *SIAM J. Numer. Anal.*, vol. 7, № 4, 627–656.
7. Csanky, L. On the parallel complexity of some computational problems. Diss. Comput. Sci. Div., Univ. of California, Berkeley, 1974.
8. Dubois, P., G. Rodrigue. An analysis of the recursive doubling algorithm. *In*: Kuck et al., 1977, 299–305.
9. Lambiotte, J., R. Voigt. The solution of tridiagonal systems on the CDS STAR-100 computer. *ACM Trans. Math. Software*, 1, 1975, 308–329.
10. Orcutt, S. E. Computer organisation and algorithms for very high-speed computations. Diss., Stanford Univ., 1974.
11. Sameh, A., D. Kuck. A parallel QR algorithm for symmetric tridiagonal matrices. *IEEE Tras. Comput.*, 1977, C-26, 147–153.
12. Sameh, A., D. Kuck. On stable parallel linear systems solvers. *J. ACM*, 25, 1978, 81–91.
13. Stone, H. S. An efficient parallel algorithm for the solution of a tridiagonal system of equations. *J. ACM*, 20, 1973, 27–38.
14. Stone, H. S. Parallel tridiagonal solvers. *ACM Tras. Math. Software*, vol. 1, № 4, 1975, 289–307.
15. Swarztrauber, P. A parallel algorithm for solving general tridiagonal equations. *Math. Comp.*, 33, 1979, 189–199.
16. Swarztrauber, P. The solution of tridiagonal systems on the CRAY-1. *In*: Jesshope and Hockney, vol. 2, 1979, 343–358.
17. Wilkinson, J. H. Rounding errors in algebraic processes. Prentice-Hall, 1963.

Received 23.05.1990