Journal of Computational and Applied Mathematics 117 (2000) 159–173

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Schur complement reduction in the mixed-hybrid approximation of Darcy's law: rounding error analysis ☆

J. Maryška[a,*], M. Rozložník[a], M. Tůma[b]

[a]Seminar for Applied Mathematics (SAM), Swiss Federal Institute of Technology (ETH), ETH-Zentrum, CH-8092 Zürich, Switzerland

[b]Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic

Received 3 August 1998; received in revised form 15 August 1999

Abstract

Mixed-hybrid finite element approximation of the potential fluid flow problem leads to the solution of a large symmetric indefinite system for the velocity and potential head vector components. Such discretization gives rise to a very accurate approximation of the continuity equation in every element, and for low-order discretizations, the structural properties of the discrete matrix blocks allow cheap block elimination of the positive-definite diagonal block and subsequent reduction to the Schur complement system for the pressure and Lagrangian vector components. This system is then frequently solved by the iterative conjugate gradient-type method. Whereas this approach is well known, considerably less attention has been paid to the numerical stability aspects of such transformation. It was shown in [5] that block LU factorization can be unstable even when the system matrix is symmetric positive definite. In this paper we examine this type of conditional stability for a particular application in the underground water flow modelling. We show that the actual error of the computed approximate solution depends not only on the user-defined tolerance in the conjugate gradient process but also on the spectral properties of the corresponding matrix blocks eliminated during the Schur complement reduction. It is often observed that although the backward error of the approximate solution in the iterative part is reduced to the level of machine accuracy, the total residual norm after the back-substitution process remains at certain accuracy level. We give a bound for this maximal attainable accuracy and illustrate our theoretical results on a model example.

Keywords: Potential fluid flow problem; Symmetric indefinite linear systems; Schur complement reduction; Iterative methods; Rounding error analysis

☆ This work was supported by the Grant Agency of the Czech Republic under grant No. 201/98/P108 and by the Grant Agency of the Academy of Sciences of the Czech Republic under grant No. A2030706.

* Corresponding author.

E-mail addresses: maryska@uivt.cas.cz (J. Maryška), miro@uivt.cas.cz (M. Rozložník), tuma@uivt.cas.cz (M. Tůma)

0377-0427/00/$ - see front matter
PII: S0377-0427(99)00344-1

1. Introduction

The potential fluid flow is one of the most important problems in such applications as underground water pollution modelling [20] or petroleum reservoir engineering [21]. The fluid flow in a saturated porous medium can be described by Darcy's law which relates the fluid velocity to the potential head (pressure) and by the continuity equation which represents the mass conservation law within the studied domain. A number of approximation techniques has been proposed and applied including various finite difference or finite element method schemes [20,4]. Especially successful discretizations in accurate approximation of the fluid velocity are the mixed finite element formulations resulting to large systems of linear equations with symmetric indefinite matrices [4,3,12]. To ensure the fullfilment of the continuity equation in every element of discretization (and partially also to enable the straightforward and cheap transformation of indefinite system to symmetric positive definite) hybrid version of the mixed hybrid formulation has become very popular [4]. Introducing the Lagrangian multipliers to enforce the continuity of the fluid velocity across the interior interelement boundaries and using the lowest-order Raviart–Thomas discretizations [12,13] we obtain the discrete system for the velocity components x_1 and the potential head vector components x_2 and x_3 in the form

$$\begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \tag{1.1}$$

where the discrete form of Darcy's law tensor A is element-wise block diagonal and symmetric positive definite; the outdiagonal block B^{T} that ensures the continuity equation on every element is the element-face incidence matrix (with weights equal -1); matrix C_1^{T} corresponding to the fluid velocity across the interior inter-element faces has orthogonal rows and the boundary condition-face incidence matrix C_2^{T} stands for the fullfillment of the Neumann boundary conditions (for details we refer to [13,12]). The structural pattern of the matrix obtained from a simple problem can be seen in Fig. 1. Since the matrix A is block diagonal and positive definite it can be easily inverted. We consider the following block LU factorization:

$$\begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} = \begin{pmatrix} I & & \\ L_{21}^{(1)} & -I & \\ L_{31}^{(1)} & & -I \end{pmatrix} \begin{pmatrix} U_{11}^{(1)} & U_{12}^{(1)} & U_{13}^{(1)} \\ & U_{22}^{(1)} & U_{23}^{(1)} \\ & (U_{23}^{(1)})^{\mathrm{T}} & U_{33}^{(1)} \end{pmatrix}, \tag{1.2}$$

where $L_{21}^{(1)} = (B\ C_2)^{\mathrm{T}} A^{-1}$, $L_{31}^{(1)} = C_1^{\mathrm{T}} A^{-1}$, $U_{11}^{(1)} = A$, $U_{12}^{(1)} = (B\ C_2)$, $U_{13}^{(1)} = C_1$. Then the matrix blocks $U_{22}^{(1)}$, $U_{23}^{(1)}$, $U_{32}^{(1)}$ and $U_{33}^{(1)}$ in (1.2) form the positive-definite Schur complement matrix with

$$\begin{pmatrix} U_{22}^{(1)} & U_{23}^{(1)} \\ (U_{23}^{(1)})^{\mathrm{T}} & U_{33}^{(1)} \end{pmatrix} = \begin{pmatrix} L_{21}^{(1)} \\ L_{31}^{(1)} \end{pmatrix} \begin{pmatrix} U_{11}^{(1)} & U_{12}^{(1)} \end{pmatrix}. \tag{1.3}$$

It was shown in [12,15] that the Schur complement matrix (1.3) remains sparse and due to the particular structure of system (1.1) the matrix block $U_{22}^{(1)}$ is easily invertible and the whole system can be reduced further without additional fill-in [15]. An example of the structure of nonzero elements in the Schur complement matrix (1.3) for our small example can be found in Fig. 1. Indeed, considering

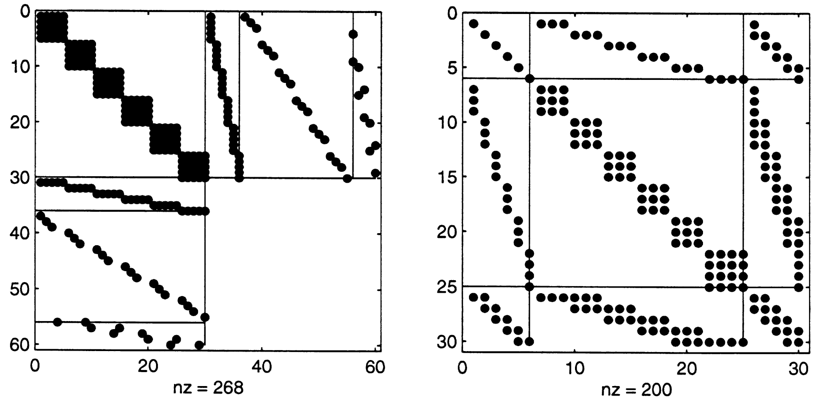

Fig. 1. Structural pattern of the indefinite matrix and its Schur complement system from a simple problem obtained in the mixed-hybrid finite element approximation of the potential fluid problem.

further the block LU factorization of matrix (1.3)

$$\begin{pmatrix} U_{22}^{(1)} & U_{23}^{(1)} \\ (U_{23}^{(1)})^{\mathrm{T}} & U_{33}^{(1)} \end{pmatrix} = \begin{pmatrix} I & \\ L_{21}^{(2)} & I \end{pmatrix} \begin{pmatrix} U_{11}^{(2)} & U_{12}^{(2)} \\ & U_{22}^{(2)} \end{pmatrix}, \tag{1.4}$$

where $L_{21}^{(2)} = (U_{23}^{(1)})^{\mathrm{T}}(U_{22}^{(1)})^{-1}$, $U_{11}^{(2)} = U_{22}^{(1)}$, $U_{12}^{(2)} = U_{23}^{(1)}$, the matrix block $U_{22}^{(2)}$ forms the second Schur complement matrix and it is equal to $U_{22}^{(2)} = U_{33}^{(1)} - L_{21}^{(2)}U_{12}^{(2)}$.

Given the block LU factorizations (1.2) and (1.4) the solution of the whole indefinite system (1.1) can be obtained using the following steps. First, we transform the right-hand vector using the forward substitution solving the triangular systems

$$\begin{pmatrix} I & & \\ L_{21}^{(1)} & -I & \\ L_{31}^{(1)} & & -I \end{pmatrix} \begin{pmatrix} z_1^{(1)} \\ z_2^{(1)} \\ z_3^{(1)} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \qquad \begin{pmatrix} I & \\ L_{21}^{(2)} & I \end{pmatrix} \begin{pmatrix} z_1^{(2)} \\ z_2^{(2)} \end{pmatrix} = \begin{pmatrix} z_2^{(1)} \\ z_3^{(1)} \end{pmatrix}. \tag{1.5}$$

The vector x_3 is then computed, using some iterative technique, from the Schur complement system

$$U_{22}^{(2)} x_3 = z_2^{(2)}. \tag{1.6}$$

Finally, we get the vectors x_1 and x_2 via the block back-substitution process in the form

$$\begin{pmatrix} U_{11}^{(1)} & U_{12}^{(1)} \\ & U_{11}^{(2)} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} z_1^{(1)} \\ z_1^{(2)} \end{pmatrix} - \begin{pmatrix} U_{13}^{(1)} \\ U_{12}^{(2)} \end{pmatrix} x_3. \tag{1.7}$$

The outline of the paper is as follows. In Section 2, we analyse the roundoff error propagation in the Schur complement reduction, and in Section 3, we give a bound for the ultimate accuracy level of the approximate solutions actually computed in the finite precision run of some conjugate gradient-type iterative method applied to system (1.6) and in the subsequent back-substitution process (1.7). In Section 4, we present the results from our numerical example that illustrate our theoretical analysis. Finally, in Section 5 we give some concluding remarks and mention some possible directions for future work.

2. Rounding error analysis of the reduction to the Schur complement systems

While the concept of reduction to the Schur complement systems is standard and widely used in many applications considerably less attention has been devoted to the numerical stability analysis of such an approach. Thorough rounding error analysis of the block LU factorization has been given by Demmel et al. [5]. They showed that such a block method is stable for symmetric positive systems only if the system matrix is well conditioned. In this section we follow their approach for our particular indefinite system (1.1). We estimate the backward error associated with the approximate solution in terms of the user tolerance prescribed for the iterative part and in terms of the spectral norms of certain matrix blocks that appear during the block elimination process.

In the following, we assume the standard IEEE model of floating point arithmetic [6]. The quantities computed in the finite precision arithmetic are denoted by a bar; ε denotes the computer arithmetic machine precision. By $\mathrm{O}(m)$ we denote the quantities proportional to the argument m and by N we denote the dimension of the system matrix (1.1). Since the dimensions of the individual matrix blocks in the system are all proportional to N, in our formulae we will not distinguish between their actual dimensions and the dimension of the whole system. The corresponding proportionality coefficients will be considered as a part of the $\mathrm{O}(N)$ notation. Throughout the paper by $\|\cdot\|$ we will denote the spectral (euclidean) matrix or vector norm.

The vector $\boldsymbol{x}_3$ is a computed result of a finite precision run of some iterative conjugate gradient-type method applied to the Schur complement system with the matrix $\boldsymbol{U}_{22}^{(2)}$ and the right-hand-side vector $\boldsymbol{z}_2^{(2)}$. For the termination of this iterative process we shall use the stopping criterion based on the backward error with the prescribed tolerance tol

$$\frac{\|\boldsymbol{z}_2^{(2)} - \boldsymbol{U}_{22}^{(2)}\boldsymbol{x}_3\|}{\|\boldsymbol{U}_{22}^{(2)}\|\|\boldsymbol{x}_3\|} \leqslant \mathrm{tol}. \tag{2.1}$$

Considering the result shown by Rigal and Gaches in [17]

$$\frac{\|\boldsymbol{z}_2^{(2)} - \boldsymbol{U}_{22}^{(2)}\boldsymbol{x}_3\|}{\|\boldsymbol{U}_{22}^{(2)}\|\|\boldsymbol{x}_3\|} = \min\{v|(\boldsymbol{U}_{22}^{(2)} + \delta U_{22}^{(2)})\boldsymbol{x}_3 = \boldsymbol{z}_2^{(2)},\ \|\delta U_{22}^{(2)}\|/\|\boldsymbol{U}_{22}^{(2)}\| \leqslant v\}, \tag{2.2}$$

it follows that after successful termination of the iterative method there exists a perturbation matrix $\delta U_{22}^{(2)}$ such that

$$(\boldsymbol{U}_{22}^{(2)} + \delta U_{22}^{(2)})\boldsymbol{x}_3 = \boldsymbol{z}_2^{(2)}, \quad \|\delta U_{22}^{(2)}\|/\|\boldsymbol{U}_{22}^{(2)}\| \leqslant \mathrm{tol}. \tag{2.3}$$

The best one can hope for is that the backward error of the actually computed approximate solution gains an order equal to machine precision multiplied by a low-degree polynomial in the system dimension N. We leave now the discussion on the maximal attainable accuracy of the particular iterative solvers for the next section. The vector $\boldsymbol{x}_2$ is a computed result of a block back-substitution process (1.7). From the stability analysis of this block back-substitution as well as from the analysis of the Cholesky factorization of the positive-definite block $\boldsymbol{U}_{11}^{(2)}$ (see [22, pp. 9–27]) it follows that

$$\left\{\begin{pmatrix} \boldsymbol{U}_{11}^{(2)} & \boldsymbol{U}_{12}^{(2)} \\ & \boldsymbol{U}_{22}^{(2)} \end{pmatrix} + \begin{pmatrix} \delta U_{11}^{(2)} & \delta U_{12}^{(2)} \\ & \delta U_{22}^{(2)} \end{pmatrix}\right\} \begin{pmatrix} \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{z}_1^{(2)} \\ \boldsymbol{z}_2^{(2)} \end{pmatrix}, \tag{2.4}$$

where the perturbation matrix satisfies

$$\left\| \begin{pmatrix} \delta U_{11}^{(2)} & \delta U_{12}^{(2)} \\ & \delta U_{22}^{(2)} \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \left\| \begin{pmatrix} \boldsymbol{U}_{11}^{(2)} & \boldsymbol{U}_{12}^{(2)} \\ & \boldsymbol{U}_{22}^{(2)} \end{pmatrix} \right\|. \tag{2.5}$$

We note here that it is essential to keep the matrix $\boldsymbol{U}_{11}^{(2)}$ factorized and to get its inverse via triangular solves in the last stage of block back substitution. Computing the inverse of the matrix $\boldsymbol{U}_{11}^{(2)}$ explicitly would lead to the extra factor $\kappa(\boldsymbol{U}_{11}^{(2)})$ in bound (2.5). Using the backward error analysis of the forward substitution (see, e.g., [7, pp. 87–89]) we can write

$$\left\{ \begin{pmatrix} I & \\ \boldsymbol{L}_{21}^{(2)} & I \end{pmatrix} + \begin{pmatrix} \delta L_{11}^{(2)} & \\ \delta L_{21}^{(2)} & \delta L_{22}^{(2)} \end{pmatrix} \right\} \begin{pmatrix} \boldsymbol{z}_1^{(2)} \\ \boldsymbol{z}_2^{(2)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{z}_2^{(1)} \\ \boldsymbol{z}_3^{(1)} \end{pmatrix}, \tag{2.6}$$

where

$$\left\| \begin{pmatrix} \delta L_{11}^{(2)} & \\ \delta L_{21}^{(2)} & \delta L_{22}^{(2)} \end{pmatrix} \right\| \leqslant \mathrm{O}(N\varepsilon) \left\| \begin{pmatrix} I & \\ \boldsymbol{L}_{21}^{(2)} & I \end{pmatrix} \right\|. \tag{2.7}$$

Assuming again the matrix $\boldsymbol{U}_{22}^{(1)}$ having been factorized by the Cholesky decomposition, for the computed matrix block $\boldsymbol{L}_{21}^{(2)}$ it follows (see also [22]) that

$$\boldsymbol{L}_{21}^{(2)}\boldsymbol{U}_{22}^{(1)} = (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} + \delta E_{21}^{(2)}, \quad \|\delta E_{21}^{(2)}\| \leqslant \mathrm{O}(N^{3/2}\varepsilon)\|\boldsymbol{L}_{21}^{(2)}\|\|\boldsymbol{U}_{22}^{(1)}\|. \tag{2.8}$$

The standard rounding error formulae for the matrix–matrix multiplication [7, p. 66] give the bound for the error in the computation of the Schur complement matrix $\boldsymbol{U}_{22}^{(2)}$ in the form

$$\boldsymbol{U}_{22}^{(2)} = \boldsymbol{U}_{33}^{(1)} - \boldsymbol{L}_{21}^{(2)}\boldsymbol{U}_{23}^{(1)} + \delta E_{22}^{(2)}, \quad \|\delta E_{22}^{(2)}\| \leqslant \mathrm{O}(N\varepsilon)\{\|\boldsymbol{U}_{33}^{(1)}\| + \|\boldsymbol{L}_{21}^{(2)}\|\|\boldsymbol{U}_{23}^{(1)}\|\}. \tag{2.9}$$

Substituting (2.4) into (2.6), considering the equalities $\boldsymbol{U}_{11}^{(2)} = \boldsymbol{U}_{22}^{(1)}$, $\boldsymbol{U}_{12}^{(2)} = \boldsymbol{U}_{23}^{(1)}$, and using bounds (2.8), (2.9) and the same approach as in [5] we can show that the computed approximate solutions $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ satisfy the perturbed Schur complement system

$$\left\{ \begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} + \begin{pmatrix} \delta U_{22}^{(1)} & \delta U_{23}^{(1)} \\ \delta U_{32}^{(1)} & \delta U_{33}^{(1)} \end{pmatrix} \right\} \begin{pmatrix} \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{z}_2^{(1)} \\ \boldsymbol{z}_3^{(1)} \end{pmatrix}, \tag{2.10}$$

where the perturbation matrix can be bounded as follows:

$$\begin{aligned} &\left\| \begin{pmatrix} \delta U_{22}^{(1)} & \delta U_{23}^{(1)} \\ \delta U_{32}^{(1)} & \delta U_{33}^{(1)} \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \\ &\left\{ \left\| \begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} \right\| + \left\| \begin{pmatrix} I & \\ \boldsymbol{L}_{21}^{(2)} & I \end{pmatrix} \right\| \left\| \begin{pmatrix} \boldsymbol{U}_{11}^{(2)} & \boldsymbol{U}_{12}^{(2)} \\ & \boldsymbol{U}_{22}^{(2)} \end{pmatrix} \right\| \right\}. \end{aligned} \tag{2.11}$$

Moreover, due to the fact that the Schur complement system matrix (1.3) is symmetric positive definite, the computed lower triangular factor in the decomposition (1.4) can be bounded by a small multiple of the square root of the condition number of matrix (1.3) (for details we refer to [5]). The computed upper triangular factor in (2.11) can be bounded, up to the quantities proportional to the machine precision, by the norm of matrix (1.3). Detailed analysis can be found in [5] or [11]. Consequently, bound (2.11) can be written in the form

$$\left\| \begin{pmatrix} \delta U_{22}^{(1)} & \delta U_{23}^{(1)} \\ \delta U_{32}^{(1)} & \delta U_{33}^{(1)} \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \left\| \begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} \right\| \sqrt{\kappa \begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix}}.$$

The computed vector $\boldsymbol{x}_1$ is obtained from the first equation in the block back-substitution process (1.7). Considering the previous bound and using again the fact that the matrix block $\boldsymbol{U}_{11}^{(1)} = A$ is during the backsubstitution factorized in triangular factors it can be shown that the whole computed approximate solution satisfies the block triangular system

$$\left\{ \begin{pmatrix} \boldsymbol{U}_{11}^{(1)} & \boldsymbol{U}_{12}^{(1)} & \boldsymbol{U}_{13}^{(1)} \\ & \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ & (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} + \begin{pmatrix} \delta U_{11}^{(1)} & \delta U_{12}^{(1)} & \delta U_{13}^{(1)} \\ & \delta U_{22}^{(1)} & \delta U_{23}^{(1)} \\ & \delta U_{32}^{(1)} & \delta U_{33}^{(1)} \end{pmatrix} \right\} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{z}_1^{(1)} \\ \boldsymbol{z}_2^{(1)} \\ \boldsymbol{z}_3^{(1)} \end{pmatrix},$$

$$\frac{\left\| \begin{pmatrix} \delta U_{11}^{(1)} & \delta U_{12}^{(1)} & \delta U_{13}^{(1)} \\ & \delta U_{22}^{(1)} & \delta U_{23}^{(1)} \\ & \delta U_{32}^{(1)} & \delta U_{33}^{(1)} \end{pmatrix} \right\|}{\left\| \begin{pmatrix} \boldsymbol{U}_{11}^{(1)} & \boldsymbol{U}_{12}^{(1)} & \boldsymbol{U}_{13}^{(1)} \\ & \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ & \boldsymbol{U}_{32}^{(1)} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} \right\|} \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \sqrt{\kappa \begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix}}. \tag{2.12}$$

From the standard analysis the computed right-hand-side vector in the forward substitution (1.5) satisfies (see [7, p. 87])

$$\left\{ \begin{pmatrix} I & & \\ \boldsymbol{L}_{21}^{(1)} & -I & \\ \boldsymbol{L}_{31}^{(1)} & & -I \end{pmatrix} + \begin{pmatrix} \delta L_{11}^{(1)} & & \\ \delta L_{21}^{(1)} & \delta L_{22}^{(1)} & \\ \delta L_{31}^{(1)} & \delta L_{32}^{(1)} & \delta L_{33}^{(1)} \end{pmatrix} \right\} \begin{pmatrix} \boldsymbol{z}_1^{(1)} \\ \boldsymbol{z}_2^{(1)} \\ \boldsymbol{z}_3^{(1)} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

$$\left\| \begin{pmatrix} \delta L_{11}^{(1)} & & \\ \delta L_{21}^{(1)} & \delta L_{22}^{(1)} & \\ \delta L_{31}^{(1)} & \delta L_{32}^{(1)} & \delta L_{33}^{(1)} \end{pmatrix} \right\| \leqslant \mathrm{O}(N\varepsilon) \left\| \begin{pmatrix} I & & \\ \boldsymbol{L}_{21}^{(1)} & -I & \\ \boldsymbol{L}_{31}^{(1)} & & -I \end{pmatrix} \right\|. \tag{2.13}$$

Similar to (2.6), the matrix blocks $\boldsymbol{L}_{21}^{(1)}$ and $\boldsymbol{L}_{31}^{(1)}$ are computed results of the Cholesky process applied to the multiple right-hand-side system with the positive-definite matrix A. From the rounding error analysis of the Cholesky factorization [22] we have

$$\begin{pmatrix} \boldsymbol{L}_{21}^{(1)} \\ \boldsymbol{L}_{31}^{(1)} \end{pmatrix} (A + \delta A) = \begin{pmatrix} (B\ C_2)^{\mathrm{T}} \\ C_1^{\mathrm{T}} \end{pmatrix}, \quad \|\delta A\| \leqslant \mathrm{O}(N^{3/2}\varepsilon)\|A\|, \tag{2.14}$$

or equivalently,

$$\begin{pmatrix} \boldsymbol{L}_{21}^{(1)} \\ \boldsymbol{L}_{31}^{(1)} \end{pmatrix} A = \begin{pmatrix} (B\ C_2)^{\mathrm{T}} \\ C_1^{\mathrm{T}} \end{pmatrix} + \begin{pmatrix} \delta E_{21}^{(1)} \\ \delta E_{31}^{(1)} \end{pmatrix},$$
$$\|\delta E_{21}^{(1)}\| \leqslant \mathrm{O}(N^{3/2}\varepsilon)\|A\|\|\boldsymbol{L}_{21}^{(1)}\|, \quad \|\delta E_{31}^{(1)}\| \leqslant \mathrm{O}(N^{3/2}\varepsilon)\|A\|\|\boldsymbol{L}_{31}^{(1)}\|. \tag{2.15}$$

Using the identities $\boldsymbol{U}_{12}^{(1)} = (B\ C_2)$, $\boldsymbol{U}_{13}^{(1)} = C_1$, having the computed Schur complement matrix (1.3) expressed in the form

$$\begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{L}_{21}^{(1)} \\ \boldsymbol{L}_{31}^{(1)} \end{pmatrix} (\ \boldsymbol{U}_{12}^{(1)} \quad \boldsymbol{U}_{13}^{(1)}\ ) + \begin{pmatrix} \delta E_{22}^{(1)} & \delta E_{23}^{(1)} \\ \delta E_{32}^{(1)} & \delta E_{33}^{(1)} \end{pmatrix},$$
$$\left\| \begin{pmatrix} \delta E_{22}^{(1)} & \delta E_{23}^{(1)} \\ \delta E_{32}^{(1)} & \delta E_{33}^{(1)} \end{pmatrix} \right\| \leqslant \mathrm{O}(N\varepsilon) \left\| \begin{pmatrix} \boldsymbol{L}_{21}^{(1)} \\ \boldsymbol{L}_{31}^{(1)} \end{pmatrix} \right\| \|(\ \boldsymbol{U}_{12}^{(1)} \quad \boldsymbol{U}_{13}^{(1)}\ )\| \tag{2.16}$$

and substituting equality (2.12) into (2.13) it can be shown that the computed approximate solution is an exact solution of the perturbed problem

$$\left\{ \begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} + \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\} \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \boldsymbol{x}_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}. \tag{2.17}$$

The norm of the matrix perturbation in (2.17) can be bounded as follows:

$$\left\| \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \left[\left\| \begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} \right\| \right.$$
$$+ \left\| \begin{pmatrix} I & & \\ \boldsymbol{L}_{21}^{(1)} & -I & \\ \boldsymbol{L}_{31}^{(1)} & & -I \end{pmatrix} \right\| \left\| \begin{pmatrix} U_{11}^{(1)} & \boldsymbol{U}_{12}^{(1)} & \boldsymbol{U}_{13}^{(1)} \\ & \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ & (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix} \right\|$$
$$\left. \times \sqrt{\kappa \begin{pmatrix} \boldsymbol{U}_{22}^{(1)} & \boldsymbol{U}_{23}^{(1)} \\ (\boldsymbol{U}_{23}^{(1)})^{\mathrm{T}} & \boldsymbol{U}_{33}^{(1)} \end{pmatrix}} \right]. \tag{2.18}$$

We have shown that the approximate solution computed in the finite precision arithmetic is an exact solution of a nearby problem. The norm of the matrix perturbation can be bounded in terms of the user tolerance prescribed for the iterative solver and in terms of the norms of certain matrices which appearing during the reduction to Schur complement systems. The bounds for the norm of the corresponding residual and forward error vector can be obtained from (2.17) considering the equalities

$$\begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \begin{pmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{pmatrix},$$

$$\begin{pmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{pmatrix} = \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix}, \tag{2.19}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} = \begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix}^{-1} \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix}.$$

3. Maximal attainable accuracy of the computed approximate solution

The bound for the backward error associated with the computed approximate solution is dominated by the prescribed user tolerance in the iterative part of the computation. Roughly, the higher the stopping criterion, the higher is the backward error associated with the computed solution in the finite precision arithmetic. This prescribed level is magnified by the quantities that play a similar role as the growth factor in Gaussian elimination with partial pivoting (see, e.g., [7]). In the following we examine the spectral properties (the norms and the condition numbers) of these matrices and give a bound for the backward error of the computed solution in terms of extremal eigenvalues or the extremal singular values of corresponding matrix blocks in system (1.1).

The bound for the norm of lower triangular factor in (2.18) can be obtained from (2.14). Substituting (2.15) into (2.16) for the computed Schur complement matrix (1.3) we have

$$\begin{pmatrix} \bar{U}_{22}^{(1)} & \bar{U}_{23}^{(1)} \\ (\bar{U}_{23}^{(1)})^{\mathrm{T}} & \bar{U}_{33}^{(1)} \end{pmatrix} = \begin{pmatrix} (B\ C_2)^{\mathrm{T}} \\ C_1^{\mathrm{T}} \end{pmatrix} A^{-1}((B\ C_2)\ C_1) + \begin{pmatrix} \delta E_{21}^{(1)} \\ \delta E_{31}^{(1)} \end{pmatrix} A^{-1}((B\ C_2)\ C_1) + \begin{pmatrix} \delta E_{22}^{(1)} & \delta E_{23}^{(1)} \\ \delta E_{32}^{(1)} & \delta E_{33}^{(1)} \end{pmatrix}.$$

Using these results in (2.18) and including the terms with the second order in ε and tol into the $\mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\})$ notation we can write for the norm of the matrix perturbation

$$\left\| \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\|$$

$$\leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \left[\left\| \begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} \right\| \right.$$

$$+ \left\| \begin{pmatrix} I & & \\ (B\ C_2)^{\mathrm{T}} A^{-1} & -I & \\ C_1^{\mathrm{T}} A^{-1} & & -I \end{pmatrix} \right\| \left\| \begin{pmatrix} A & (B\ C_2) & C_1 \\ & (B\ C_2)^{\mathrm{T}} A^{-1} (B\ C_2) & (B\ C_2)^{\mathrm{T}} A^{-1} C_1 \\ & C_1^{\mathrm{T}} A^{-1} (B\ C_2) & C_1^{\mathrm{T}} A^{-1} C_1 \end{pmatrix} \right\|$$

$$\left. \times \sqrt{\kappa\left(\begin{pmatrix} (B\ C_2)^{\mathrm{T}} A^{-1} (B\ C_2) & (B\ C_2)^{\mathrm{T}} A^{-1} C_1 \\ C_1^{\mathrm{T}} A^{-1} (B\ C_2) & C_1^{\mathrm{T}} A^{-1} C_1 \end{pmatrix} \right)} \right]. \tag{3.20}$$

Considering the estimate for the condition number of the Schur complement matrix (1.3)

$$\kappa\left(\begin{pmatrix} (B\ C_2)^{\mathrm{T}} A^{-1} (B\ C_2) & (B\ C_2)^{\mathrm{T}} A^{-1} C_1 \\ C_1^{\mathrm{T}} A^{-1} (B\ C_2) & C_1^{\mathrm{T}} A^{-1} C_1 \end{pmatrix} \right) \leqslant \kappa^2((B\ C_2\ C_1))\kappa(A) \tag{3.21}$$

and applying the standard techniques for bounding the norm of a block matrix one can get the bound for the backward error in terms of the user tolerance tol and of the norms or the condition numbers of the matrix blocks A and $(B\ C_2\ C_1)$.

$$\left\| \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\|$$

$$\leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \Big[\|A\| + \|(B\ C_2\ C_1)\| + (1 + \|(B\ C_2\ C_1)\| \|A^{-1}\|)$$

$$\times \max\{\|A\|, \|(B\ C_2\ C_1)\| + \|(B\ C_2\ C_1)\|^2 \|A^{-1}\|\} \sqrt{\kappa(A)}\, \kappa((B\ C_2\ C_1)) \Big]. \tag{3.22}$$

It was shown in [13] that for our particular potential fluid flow application (1.1) the spectrum of the symmetric positive-definite block A satisfies

$$\sigma(A) \subset [c_1 N^{1/3}, c_2 N^{1/3}], \tag{3.23}$$

where c_1 and c_2 are positive constants independent of the system parameters and are dependent on the properties of Darcy's tensor and on the geometry of considered domain. Assuming at least one Dirichlet boundary condition it was also shown that the singular values of the off-diagonal block $(B\ C_2\ C_1)$ are included in the interval

$$\mathrm{sv}((B\ C_2\ C_1)) \subset [c_3 N^{-1/3}, c_4] \tag{3.24}$$

for some positive constants c_3 and c_4, again independent of the parameters of the linear system (1.1). The constants c_1, c_2, c_3 and c_4 describe the underlying potential fluid flow problem and as we will show later they play a substantial role in the estimates for the final accuracy of the actual computation. For details we refer to [14], see also [13]. Considering the inclusion sets (3.23) and (3.24) in (3.22) we obtain

$$\left\| \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\}) \Big[c_2 N^{1/3} + c_4 + \sqrt{c_2/c_1}\,(c_4/c_3) N^{1/3} \times (1 + (c_4/c_1)N^{-1/3}) \max\{c_2 N^{1/3}, c_4 + (c_4^2/c_1)N^{-1/3}\} \Big]. \tag{3.25}$$

Bound (3.25) can be slightly simplified. For realistic problems in our application one usually has $c_1 N^{1/3} < c_2 N^{1/3} \ll 1 \leqslant c_4 \leqslant \sqrt{5} + \sqrt{2}$ and $c_3 N^{-1/3} \ll c_2 N^{1/3}$. We note here that these inequalities are true for all practically available problem sizes N. In this case for the eigenvalues of the symmetric indefinite system matrix (1.1) we have (see [15,14])

$$\sigma\left(\begin{pmatrix} A & (B\ C_2) & C_1 \\ (B\ C_2)^{\mathrm{T}} & & \\ C_1^{\mathrm{T}} & & \end{pmatrix} \right) \subset [-c_4, -(c_3^2/c_2)N^{-1}] \cup [c_1 N^{1/3}, c_4]. \tag{3.26}$$

The bound for the matrix perturbation (3.25) then becomes of the order

$$\left\| \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\} N^{-1/3}) \sqrt{\frac{c_2}{c_1^5}\frac{c_4^4}{c_3}} \tag{3.27}$$

and, consequently, the residual norm and the error norm of the computed approximate solution can be estimated in the form

$$\left\| \begin{pmatrix} \delta b_1 \\ \delta b_2 \\ \delta b_3 \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\} N^{-1/3}) \sqrt{\frac{c_2}{c_1^5}\frac{c_4^4}{c_3}} \left\| \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} \right\|, \tag{3.28}$$

$$\left\| \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \right\| \leqslant \mathrm{O}(\max\{\mathrm{tol}, N^{3/2}\varepsilon\} N^{2/3}) \sqrt{\frac{c_2^3}{c_1^5}\frac{c_4^4}{c_3^3}} \left\| \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix} \right\|. \tag{3.29}$$

The Schur complement system (1.6) is symmetric positive definite. Several possible conjugate gradient-type methods for solving this system are available. From a number of methods we can use some implementation of the conjugate gradient method [10] characterized by minimization of the approximate solution energy norm or we can use the residual norm minimizing conjugate residual [19] (minimal residual [16]) method existing also in several variants. It is a well-known fact that there is a limitation on the accuracy of the approximate solutions generated by conjugate gradient-type methods. It is also often observed that while the recursively updated residual vector converges to zero, the actual residual norm of the approximate solution remains at a certain accuracy. The best one can then hope for is that the backward error of the computed approximate solution becomes of the order of machine precision, i.e. the finite precision run of the corresponding iterative method reaches the user tolerance given as the small multiple of the machine precision. More precisely, we want the backward error of the computed approximate solution to be bounded by the low degree polynomial in the matrix dimension N times machine precision ε. The limiting accuracy of the conjugate gradient method has been studied thoroughly in several papers [8,9,18]. It was shown in [9] that the ultimate level of accuracy depends on the chosen implementation even in the symmetric positive-definite case and that the conjugate gradient implementation based on three-term recurrences is potentially less accurate than the classical implementation based on coupled two-term recurrences. Analysis of Greenbaum [8] (see also [18]) shows that for the two-term recurrence conjugate gradient implementation the maximum expectable accuracy (measured by relative residual norm of the computed approximate solution and assuming the termination within N steps) can be bounded by the term $O(N^{5/2}\varepsilon)\kappa(A)$. Consequently, the backward error associated with the computed solution can be bounded by $O(N^{5/2}\varepsilon)$ and one cannot hope for anything better when using the tolerance below this level. A similar type of result was given for the conjugate residual method (implementation with coupled two-term recurrences) [18]. Moreover, it was pointed out in [18] that the minimal residual (MINRES) implementation of the same method may be unstable for ill-conditioned problems due to the computation of certain nonorthogonal basis of Krylov subspace. This result also resembles the behaviour described in [5] and called as the conditional stability behaviour of the algorithm.

Therefore, the lowest realistic user tolerance (when using the most appropriate implementation of the iterative method) is of the order $\mathrm{tol} = O(N^{5/2}\varepsilon)$. Consequently, for our particular implementation from (3.27) we can write

$$\left\| \begin{pmatrix} \delta E_{11} & \delta E_{12} & \delta E_{13} \\ \delta E_{21} & \delta E_{22} & \delta E_{23} \\ \delta E_{31} & \delta E_{32} & \delta E_{33} \end{pmatrix} \right\| \leqslant O(N^{5/2-1/3}\varepsilon)\sqrt{\frac{c_2}{c_1^5}\frac{c_4^4}{c_3}}. \tag{3.30}$$

4. Numerical experiments

In the following we present experimental results in a numerical example that illustrates our theoretical analysis.

We considered a model potential flow problem in a rectangular domain with homogeneous Neumann conditions and with Dirichlet condition imposed only on one part of the domain boundary. The lowest-order prismatic discretization of the domain with several values of the mesh size h was used. In our experiments we generated linear systems for several values of the discretization parameter h. The dimensions of corresponding system matrices for these values are approximately $N \sim 22/h^3$. The inclusion sets for the spectrum of the matrix block A and for the singular value set of the matrix block $(B\ C_2\ C_1)$ were computed in the following way. The extremal eigenvalues of the block diagonal matrix A were computed directly by the LAPACK eigenvalue solver element by element. The extremal singular values of the block $(B\ C_2\ C_1)$ were as the squared roots of the extremal eigenvalues of the matrix $(B\ C_2\ C_1)^{\mathrm{T}}(B\ C_2\ C_1)$ approximated by the reduction of this matrix to the symmetric tridiagonal form using 1000 steps of the symmetric Lanczos algorithm [7] and by the subsequent eigenvalue computation of the resulting tridiagonal matrix using the

Table 1
Spectral properties of matrix blocks and the residual norm

h	Spectral properties of matrix blocks		The norm of computed solution	True residual norm, tol $= 10^{-14}$
	Spectrum of A	Sing. values of $(B\,C)$		
1/5	[0.16e−2, 0.1e−1]	[0.179e−1, 2.63]	12602.08	0.9455e−8
1/10	[0.33e−2, 0.2e−1]	[0.997e−2, 2.64]	37302.81	0.1051e−7
1/15	[0.50e−2, 0.3e−1]	[0.691e−2, 2.64]	69515.73	0.1481e−7
1/20	[0.66e−2, 0.4e−1]	[0.528e−2, 2.64]	107777.54	0.2109e−7
1/30	[0.10e−1, 0.6e−1]	[0.358e−2, 2.65]	199370.25	0.3202e−7
1/40	[0.13e−1, 0.8e−1]	[0.271e−2, 2.65]	307999.99	0.4291e−7

LAPACK double-precision subroutine DSYEV [1]. The corresponding numerical values are included in Table 1.

In the iterative part of the solution, the coupled two-term conjugate residual implementation was used [19]. As the stopping criterion we took the backward error associated with the computed approximate solution. The spectral norm of the Schur complement matrix $\bar{U}_{22}^{(2)}$ in (2.1) was replaced by the Frobenius norm and the corresponding residual norm was in the actual computation was approximated by the norm of the updated residual which is usually converging to zero, or at least, reaches the level much less than the true residual of the actually computed approximate solution. The user tolerance 10^{-14} chosen for this experiment was very optimistic and actually it was below the maximally attainable accuracy level of the conjugate residual method. In Table 1 we consider also the true residual norms of the whole computed approximate solution. We can observe that ultimate residual accuracy is almost on the same level for all values of the discretization parameter h. This level is given essentially by the values of the positive constants c_1, c_2, c_3 and c_4 (see the discussion in the previous section). The slight increase in the residual norm can be explained by the increase of the system dimension N which is included in the term $\mathrm{O}(N^{5/2-1/3}\varepsilon)$ and which also contributes to the increase of the norm of the approximate solution (see also Table 1 and Eq. (2.19)).

In the second experiment, we have examined the dependence of the actual residual norm of the computed approximates on the user tolerance tol chosen in the iterative part of the computation. We have also investigated the norms of three residual vector blocks that correspond to three blocks of right-hand-side vector in the process of reduction to the Schur complement systems. From Table 2 it is clear that the accuracy of the block related to the Schur complement system (1.6) is given roughly by the residual norm corresponding to the user tolerance value tol. The accuracy levels in the other two blocks remain the same for the different values of the tolerance tol and their absolute values are roughly given by the multiplication factors that appear in the analysis of a corresponding step in the block elimination process. In Fig. 2 we plotted the true and recursive residual norm history for the example with the discretization parameter $h = 1/15$ together with the levels of accuracies obtained for different values of the prescribed tolerance level tol. Here we can also observe that the user tolerance tol $= 10^{-14}$ is actually below the attainable accuracy of the conjugate residual method applied to the Schur complement system (1.6).

Table 2
Residual of the computed solution versus user tolerance

tol	True residual whole system	True residuals of system blocks, $h = 1/15$		
		$\|\delta b_1\|_\infty$	$\|\delta b_2\|_\infty$	$\|\delta b_3\|_\infty$
10^{-6}	0.5879e0	0.5684e−13	0.2392e−9	0.2138e−1
10^{-8}	0.5865e−2	0.5684e−13	0.2375e−9	0.2694e−3
10^{-10}	0.6112e−4	0.5683e−13	0.2106e−9	0.3124e−5
10^{-12}	0.5971e−6	0.4916e−13	0.2297e−9	0.2241e−7
10^{-14}	0.1566e−7	0.3798e−13	0.2046e−9	0.6595e−9

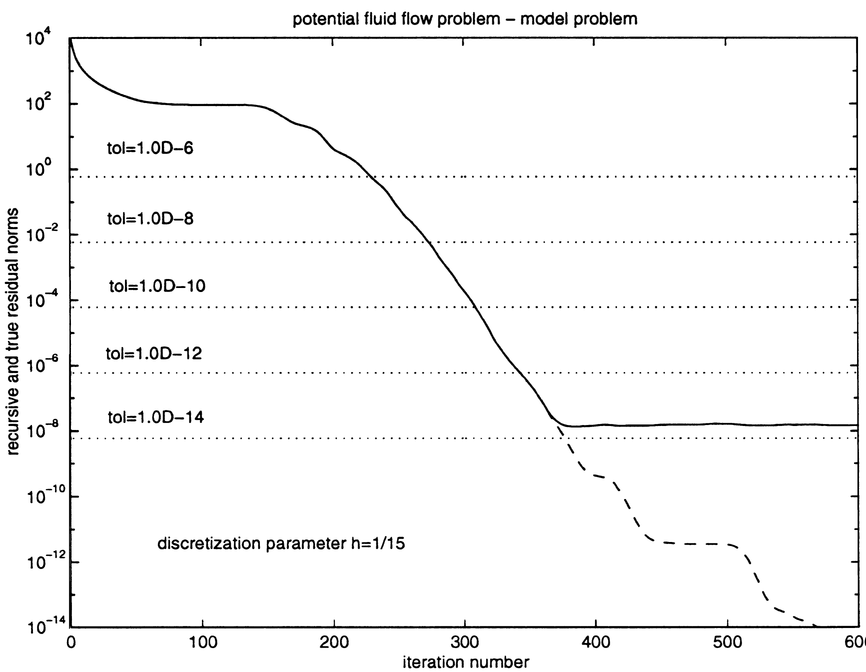

Fig. 2. Different values of accuracy level with respect to the user tolerance level.

5. Conclusions

In this paper we examined the rounding error propagation in the process of reduction of the augmented symmetric indefinite system to the system with the positive-definite Schur complement matrix. Using a similar approach as in [5] we have shown that such a transformation is conditionally stable, i.e., the stability is guaranteed only if certain matrices in this block elimination process are well conditioned. We have analysed the condition number of corresponding matrix blocks for our particular application in the underground water flow modelling. It was shown that the level of accuracy in solution of linear system can be bounded in terms of parameters which are dependent on the underlying problem characteristics such as the properties of Darcy's tensor or as the geometry of the domain. Using these quantities we gave a bound for the maximally attainable accuracy of the approximate solution computed in the finite precision arithmetic.

Similar rounding error analysis was also considered recently by Arioli in [2] for the solution of augmented systems arising in sparse quadratic programming. His approach is based on the null space method combining a direct QR factorization of the out-diagonal block in the system matrix with an iterative solver on the corresponding null space. Since the Householder QR decomposition is backward stable [7] it is shown in [2] that in the case of the termination of the iterative solver on the level of its limiting accuracy, the backward error of the computed approximate solution maybe bounded by the term proportional to the machine precision ε. The computational comparison of these two approaches as well as of the other variants of the null space approach remains to be done in the future and will be published elsewhere.

References

[1] E. Anderson et al., LAPACK User's Guide, SIAM, Philadelphia, 1992.
[2] M. Arioli, The use of QR factorization in sparse quadratic programming, IAN Report No., IAN C.N.R., Pavia, 1998.
[3] L. Bergamaschi, S. Mantica, F. Saleri, Mixed finite element approximation of Darcy's law in porous media, Technical Report CRS4, Cagliari, 1994.
[4] F. Brezzi, M. Fortin, Mixed and Hybrid Finite Element Methods, Springer, Berlin, 1991.
[5] J.W. Demmel, N.J. Higham, R.S. Schreiber, Block LU factorization, Numer. Linear Algebra Appl. 2 (2) (1995) 173–190.
[6] D. Goldberg, What every computer scientist should know about floating-point arithmetic, ACM Comput. Surveys 23 (1) (1991) 5–47.
[7] G.H. Golub, C.F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, MD, 1989.
[8] A. Greenbaum, Estimating the attainable accuracy of recursively computed residual methods, SIAM J. Matrix Anal. Appl. 18 (1997) 535–551.
[9] M.H. Gutknecht, Z. Strakoš, Accuracy of three-term and two-term recurrences for Krylov space solvers, Technical Report TR-97-21, CSCS/SCSC, ETH, Zürich, 1997, SIAM J. Sci. Comput., submitted for publication.
[10] M.R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bureau Standards 49 (1952) 409–435.
[11] N.J. Higham, Accuracy and Stability of Numerical Algorithms, SIAM, Philadelphia, 1996.
[12] E.F. Kaasschieter, A.J.M. Huijben, Mixed-hybrid finite elements and streamline computation for the potential flow problem, Numer. Methods Partial Differential Equations 8 (1992) 221–266.
[13] J. Maryška, M. Rozložník, M. Tůma, Mixed-hybrid finite element approximation of the potential fluid flow problem, J. Comput. Appl. Math. 63 (1995) 383–392.
[14] J. Maryška, M. Rozložník, M. Tůma, The potential fluid flow problem and the convergence rate of the minimal residual method, Numer. Linear Algebra Appl. 3 (6) (1996) 525–542.
[15] J. Maryška, M. Rozložník, M. Tůma, Schur complement systems in the mixed-hybrid finite element approximation of the potential fluid flow problem, Technical Report No. 730, ICS AS CR, Prague, pp. 18, 1997, SIAM J. Sci. Comput., submitted for publication.
[16] C.C. Paige, M.A. Saunders, Solution of sparse indefinite systems of linear equations, SIAM J. Numer. Anal. 12 (1975) 617–629.
[17] J. Rigal, J. Gaches, On the compatibility of a given solution with the data of a linear system, J. ACM 14 (1967) 543–546.
[18] G.L.G. Sleijpen, H.A. van der Vorst, J. Modersitzki, The main effects of rounding errors in Krylov solvers for symmetric linear systems, Preprint nr. 1006, Utrecht University, 1997, SIAM J. Matrix Anal. Appl., submitted for publication.

[19] E. Stiefel, Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme, Comm. Math. Helv. 29 (1955) 157–179.

[20] H.F. Wang, M.P. Anderson, Introduction to Groundwater Modelling, Finite Difference and Finite Element Methods, W.H. Freeman and Company, San Francisco, 1982.

[21] M.F. Wheeler, R. Gonzales, Mixed finite element methods for petroleum reservoir engineering problems, Proceedings of Sixth International Conference on Computing Methods in Applied Sciences and Engineering, INRIA, Versailles, France, 1983, Computing Methods in Engineering and Applied Sciences VI, North-Holland, Amsterdam, 1984, pp. 639–658.

[22] J.H. Wilkinson, C. Reinsch (Eds.), Handbook for Automatic Computation, Vol. 2, Linear Algebra, Springer, New York, 1971.