Projected Krylov Methods for Solving Non-Symmetric Two-by-Two Block Linear Systems Arising from Fictitious Domain Formulations

Radek KUCERA, Tomas KOZUBEK, Alexandros MARKOPOULOS, Jaroslav HASLINGER, Lukas MOCEK

Centre of Excellence IT4Innovations, National Supercomputing Center, VSB–Technical University of Ostrava, 17. listopadu 15, 708 33 Ostrava–Poruba, Czech Republic

radek.kucera@vsb.cz, tomas.kozubek@vsb.cz, alexandros.markopoulos@vsb.cz, jaroslav.haslinger@mff.cuni.cz, minkos@seznam.cz

Abstract. *The paper deals with the solution of large non-symmetric two-by-two block linear systems with a singular leading submatrix. Our algorithm consists of two levels. The outer level combines the Schur complement reduction with the orthogonal projectors that leads to the linear equation on subspaces. To solve this equation, we use a Krylov-type method representing the inner level of the algorithm. We propose a general technique how to get from the standard Krylov methods their projected variants generating iterations on subspaces. Then we derive the projected GMRES. The efficiency of our approach is illustrated by examples arising from the combination of the fictitious domain and FETI method.*

Keywords

Domain decomposition, fictitious domain, GMRES, Krylov method, null-space method, orthogonal projector, Schur complement.

1. Introduction

We consider two-by-two block linear systems

$$\mathcal{A}\begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1}$$

where

$$\mathcal{A} = \begin{pmatrix} A & B_1^\top \\ B_2 & -C \end{pmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

with $A \in \mathbb{R}^{n\times n}$ singular, $C \in \mathbb{R}^{m\times m}$, $B_1, B_2 \in \mathbb{R}^{m\times n}$, $f, u \in \mathbb{R}^n$, $g, \lambda \in \mathbb{R}^m$, and $m \ll n$. Systems of this type arise in a variety of scientific and engineering applications [2]. For instance, when the FETI (Finite Element Tearing and Interconnecting) domain decomposition method [8], [6], [20] is used for the numerical solution of elliptic PDEs, we get a saddle-point linear system, i.e., Eq. (1) with A being symmetric, positive semidefinite, $B_1 = B_2$, and $C = 0$. The FETI algorithm consists of two levels. The *outer* level combines the Schur complement reduction requiring a generalized inverse to A with the null-space method performed by orthogonal projectors. It results in the linear equation with the singular matrix that is, fortunately, the symmetric and positive definite operator in a subspace $\mathbb{V} \subset \mathbb{R}^m$. Therefore, this equation can be solved in $\mathbb{V}$ by the projected CGM representing the *inner* level of the FETI algorithm. Note that the projected CGM was developed also in context of optimization problems [12].

The extension of the FETI algorithm for solving Eq. (1) called the PSCM (Projected Schur Complement Method) was proposed in [16] (for $C = 0$). Although the outer level is based on the same ideas as in the FETI algorithm, we arrive at the linear equation whose matrix is the invertible operator between two different subspaces $\mathbb{V}_1$ and $\mathbb{V}_2$ in $\mathbb{R}^m$. Therefore, the projected iterative method for non-symmetric, indefinite operators is needed in the inner level of the PSCM. In this paper, we develop a general technique enabling us to derive from the standard Krylov methods their projected variants generating iterations on subspaces. Then we derive the projected GMRES (ProjGMRES). The projected Krylov methods for solving Eq. (1) with A being non-symmetric, $B_1 = B_2$, and $C = 0$ have been developed also in [23] (without the Schur complement reduction).

Our research is motivated by the development of the fictitious domain (FD) method for solving elliptic PDEs. The main idea consists in an extension of the original PDE problem defined in a domain ω to a larger

domain $\Omega \supset \omega$ with a simple shape, e.g. a box. The new problem is chosen in such a way that its solution restricted to ω coincides with the solution of the original problem. Since Ω has a simple shape, one can use specific partitions based on non-fitted uniform meshes that does not respect the geometry of ω. Therefore, the resulting stiffness matrix, represented in Eq. (1) by A, does not depend on ω. Moreover, one can combine the FD method with other techniques, e.g. with the FETI domain decomposition method that enables us to perform an efficient parallelization of computations. There are several ways how to define the new problem in Ω. One of them uses boundary Lagrange multipliers introduced on the boundary of $\partial\omega$ [9], [10], [13], [14]. The linear system Eq. (1) arising from the finite element approximation is typically symmetric. Unfortunately, this approach suffers from a serious drawback: the computed solution has a generally non-zero jump of its normal derivative across γ. If non-fitted meshes are used, the singularity appears inside of some elements of the partition, namely those ones the interior of which is cut by γ. Consequently, the theoretical rate of convergence of approximate solutions is slow, at most $1/2$ [16], [15]. To improve the accuracy in ω, the authors of [16] proposed the *smooth* FD method. Instead of Lagrange multipliers on γ, control variables defined on a closed curve Γ in Ω having a positive distance from ω are used. The solution is still singular in Ω but the singularity is shifted away from ω to Γ and as a result, convergence in ω becomes faster. The algebraic formulation leads to the non-symmetric linear system Eq. (1), in which B_1 and B_2 play the role of the "trace matrices" (i.e., approximations of boundary operators) on Γ and γ, respectively, and $C = 0$.

Let us introduce some conventions that we use throughout the whole paper. If $M \in \mathbb{R}^{k\times l}$ is a matrix, then $\mathcal{N}(M)$ and $\mathcal{R}(M)$ denote its null-space and range-space defined by $\mathcal{N}(M) = \{v \in \mathbb{R}^l : Mv = 0\}$ and $\mathcal{R}(M) = \{v \in \mathbb{R}^k : v = Mw,\ w \in \mathbb{R}^l\}$, respectively. $M^\top$ is the transpose of M. The Moore-Penrose inverse of M will be denoted by $M^\dagger$. The symbol $\|\cdot\|$ stands for the Euclidean norm of vectors, i.e., $\|v\| = (v^\top v)^{1/2}$ for $v \in \mathbb{R}^k$. The spectral condition number of the symmetric, positive definite matrix $N \in \mathbb{R}^{k\times k}$ is given by $\kappa(N) = \sigma_{\max}(N)/\sigma_{\min}(N)$, where $0 < \sigma_{\min}(N) \le \sigma_{\max}(N)$ are the smallest and largest eigenvalues of N, respectively. Finally, I and 0 denote the identity and zero matrices of an appropriate order, respectively.

2. Preliminaries

In this section, we summarize results of [16], [19]. Our aim is to find $(u, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying Eq. (1). We will assume:

(A1) the matrix $\mathcal{A}$ in Eq. (1) is nonsingular;
(A2) the block A of $\mathcal{A}$ is singular.

It is easy to show that (A1) implies

$$\mathcal{N}(A^\top) \cap \mathcal{N}(B_1) = \{0\}, \quad \mathcal{N}(A) \cap \mathcal{N}(B_2) = \{0\}. \quad (2)$$

Further, (A2) yields the existence of $R_A, R_{A^\top} \in \mathbb{R}^{n\times l}$ whose columns span $\mathcal{N}(A)$, and $\mathcal{N}(A^\top)$, respectively, where $l := l(A)$ is the nullity of A. By $X \in \mathbb{R}^{n\times n}$ we denote an arbitrary generalized inverse of A satisfying $A = AXA$. Our algorithm requires that X, R_A, and $R_{A^\top}$ are to our disposal so that one can efficiently compute actions of these matrices on vectors; see [19] for computational details.

To describe the outer level of the PSCM, we start with the Schur complement reduction. The first block equation in Eq. (1) is satisfied, if $f - B_1^\top\lambda \in \mathcal{R}(A)$ or equivalently

$$R_{A^\top}^\top(f - B_1^\top\lambda) = 0. \quad (3)$$

The first component of the solution is given by

$$u = X(f - B_1^\top\lambda) + R_A\alpha, \quad (4)$$

where $\alpha \in \mathbb{R}^l$ is an appropriate vector. Substituting Eq. (4) into the second block equation in Eq. (1), we arrive at

$$(B_2XB_1^\top + C)\lambda - B_2R_A\alpha = B_2Xf - g. \quad (5)$$

Let us denote $F = B_2XB_1^\top + C$, $G_1 = -R_A^\top B_2^\top$, $G_2 = -R_{A^\top}^\top B_1^\top$, $d = B_2Xf - g$, $e = -R_{A^\top}^\top f$, and

$$\mathcal{S} = \begin{pmatrix} F & G_1^\top \\ G_2 & 0 \end{pmatrix} \in \mathbb{R}^{(m+l)\times(m+l)}. \quad (6)$$

The matrix $\mathcal{S}$ is the (negative) Schur complement of the block A in $\mathcal{A}$. Obviously, $\mathcal{S}$ is not unique, since it depends on the choice of X, R_A, and $R_{A^\top}$. Nevertheless, one can prove that each $\mathcal{S}$ is non-singular due to (A1). Summarizing Eq. (3) and Eq. (5), we see that the pair $(\lambda, \alpha) \in \mathbb{R}^m \times \mathbb{R}^l$ satisfies

$$\mathcal{S}\begin{pmatrix} \lambda \\ \alpha \end{pmatrix} = \begin{pmatrix} d \\ e \end{pmatrix}. \quad (7)$$

To solve Eq. (1), one computes (λ, α) from Eq. (7) and then u by Eq. (4).

Assuming λ to be known, α is given by

$$\alpha = (G_1G_1^\top)^{-1}G_1(d - F\lambda), \quad (8)$$

as follows from the first block equation in Eq. (7). The inverse $(G_1G_1^\top)^{-1}$ in Eq. (8) is well-defined, since the second relation in Eq. (2) guarantees the full row-rank of G_1 (similarly for G_2 due to the first relation in Eq. (2)). It remains to show how to compute λ.

We use the null-space method performed by the orthogonal projectors P_i onto $\mathcal{N}(G_i)$:

$$P_i = I - G_i^\top (G_i G_i^\top)^{-1} G_i, \quad i = 1, 2. \tag{9}$$

Applying P_1 to the first block equation in Eq. (7) and using $P_1 G_1^\top = 0$, we arrive at the problem in terms of λ:

$$P_1 F \lambda = P_1 d, \quad G_2 \lambda = e. \tag{10}$$

To get a linear equation in $\mathcal{N}(G_2)$, we decompose λ on two orthogonal components $\lambda_\mathcal{N} \in \mathcal{N}(G_2)$ and $\lambda_\mathcal{R} \in \mathcal{R}(G_2^\top)$ so that $\lambda = \lambda_\mathcal{N} + \lambda_\mathcal{R}$. As $I - P_2$ is the orthogonal projector onto $\mathcal{R}(G_2^\top)$, we get

$$\lambda_\mathcal{R} = (I - P_2)\lambda = G_2^\top (G_2 G_2^\top)^{-1} e.$$

Hence, $\lambda_\mathcal{R}$ is easily computable. Assuming that $\lambda_\mathcal{R}$ is known, we find that $\lambda_\mathcal{N}$ satisfies the linear equation

$$P_1 F \lambda_\mathcal{N} = q, \tag{11}$$

where $q = P_1(d - F\lambda_\mathcal{R})$. Although the matrix $P_1 F$ is singular, the following theorem guarantees the solvability of Eq. (11).

Theorem 1 [19] *The linear operator $P_1 F$ is invertible between $\mathcal{N}(G_2)$ and $\mathcal{N}(G_1)$.*

ALGORITHM PSCM

Input: X, R_A, $R_{A^\top}$, B_1, B_2, C, f, g.
Step 1: Assemble G_i, $H_i = (G_i G_i^\top)^{-1}$, $i = 1, 2$, d, e.
Step 2: Assemble $\lambda_\mathcal{R} = G_2^\top H_2 e$.
Step 3: Assemble $q = P_1(d - F\lambda_\mathcal{R})$.
Step 4: Compute $\lambda_\mathcal{N}$ by solving Eq. (11).
Step 5: Assemble $\lambda = \lambda_\mathcal{R} + \lambda_\mathcal{N}$.
Step 6: Assemble $\alpha = H_1 G_1 (d - F\lambda)$.
Step 7: Assemble $u = X(f - B_1^\top \lambda) + R_A \alpha$.

The matrices F, P_1, and P_2 need not be assembled explicitly, if an approprite iterative method in Step 4 is used. We need only actions on $\mu \in \mathbb{R}^m$ evaluated successively as indicated by the parentheses: $F\mu = B_2(X(B_1^\top \mu)) + C\mu$, $P_i\mu = \mu - G_i(H_i(G_i^\top \mu))$, $i = 1, 2$. The following theorem shows that $P_1 F$ is invariant with respect to the generalized inverse X.

Theorem 2 [19] *Let X be an arbitrary generalized inverse of A and let $A^\dagger$ be the Moore-Penrose one. Let us denote $\mu_X = P_1 B_2 X B_1^\top \mu$ and $\mu_{A^\dagger} = P_1 B_2 A^\dagger B_1^\top \mu$ for $\mu \in \mathcal{N}(G_2)$. It holds that $\mu_X = \mu_{A^\dagger}$.*

3. Projected Krylov Methods

In this section, we present a general technique enabling us to derive projected Krylov methods for solving Eq. (11). Since our approach is general, we replace Eq. (11) by the following abstract problem: find $x \in \mathbb{V}_2$ such that

$$Mx = q, \tag{12}$$

where $M \in \mathbb{R}^{m \times m}$ represents the invertible operator $M : \mathbb{V}_2 \mapsto \mathbb{V}_1$ between subspaces $\mathbb{V}_1, \mathbb{V}_2 \subset \mathbb{R}^m$ with $\dim \mathbb{V}_1 = \dim \mathbb{V}_2 = m - l$, $1 \le l < m$, and $q \in \mathbb{V}_1$. These assumptions guarantee that there is the unique solution to Eq. (12). Note that the matrix M may be singular (on $\mathbb{R}^m$).

Let $Z_1, Z_2 \in \mathbb{R}^{m \times (m-l)}$ be matrices whose columns span $\mathbb{V}_1$, $\mathbb{V}_2$, respectively. Let $\bar{x}, \bar{q} \in \mathbb{R}^{m-l}$ be such that $q = Z_1(Z_1^\top Z_1)^{-1}\bar{q}$, $x = Z_2\bar{x}$. Substituting these vectors into Eq. (12) and multiplying by $Z_1^\top$, we find that Eq. (12) reduces to the system of linear equations:

$$N\bar{x} = \bar{q}, \tag{13}$$

where $N = Z_1^\top M Z_2 \in \mathbb{R}^{(m-l)\times(m-l)}$.

Lemma 1 *The matrix N in Eq. (13) is non-singular.*

Proof. Let $Z_1^\top M Z_2 y = 0$ be the homogeneous system. Denoting $y_1 = MZ_2 y$, we obtain $Z_1^\top y_1 = 0$. As $y_1 \in \mathbb{V}_1$ is orthogonal to all basis vectors of $\mathbb{V}_1$, we get $y_1 = 0$. In $MZ_2 y = 0$, we set $y_2 = Z_2 y$. Then, $My_2 = 0$ implies $y_2 = 0$ due to the invertibility of M. Finally, $Z_2 y = 0$ yields $y = 0$, as Z_2 has full column-rank. Hence, the solution to the homogeneous system is trivial. □

To propose a projected method for solving Eq. (12), we start from a (standard) Krylov method applied to Eq. (13) that generates approximations to the solution of Eq. (13) in $\mathbb{R}^{m-l}$. Our aim is to transform these approximations to get directly approximations of the solution to Eq. (12). We will also replace Z_1 and Z_2 by the respective orthogonal projectors so that Z_1 and Z_2 are not explicitly needed in computations.

First of all we show how to transform three vector operations performed by the Krylov methods from $\mathbb{R}^{m-l}$ to $\mathbb{V}_2$, namely, the linear combination and the scalar product of vectors and the matrix-vector multiplication. We will use notation introduced already in Eq. (12) and Eq. (13): if $\bar{x} \in \mathbb{R}^{m-l}$ is a "short" vector, then $x \in \mathbb{V}_2$ denotes the "extended" vector given by $x = Z_2\bar{x}$. As $\bar{x}$ is the vector of coordinates of x with respect to the basis of Z_2, there is the one-to-one correspondence between $\bar{x}$ and x (similarly for $\bar{y}, \bar{z} \in \mathbb{R}^{m-l}$ and $y, z \in \mathbb{V}_2$, respectively).

Let $a, b \in \mathbb{R}$. The linear combination is easy to transform, since

$$\bar{z} = a\bar{x} + b\bar{y} \quad \Longleftrightarrow \quad z = ax + by. \tag{14}$$

Without lost of generality, we may assume that Z_2 is orthogonal, i.e.,

$$Z_2^\top Z_2 = I. \tag{15}$$

Using Eq. (15) in the scalar product of $\bar{x}$ and $\bar{y}$, we get

$$\bar{x}^\top \bar{y} = \bar{x}^\top Z_2^\top Z_2 \bar{y} = x^\top y. \tag{16}$$

From Eq. (14) and Eq. (16) we see that linear combinations and scalar products are identical for "short" and "extended" vectors, i.e., the respective formulas will be the same in the Krylov method as well as in its projected variant.

Let us discuss the matrix-vector multiplication by N from Eq. (13). We get:

$$\begin{aligned} \bar{y} = Z_1^\top M Z_2 \bar{x} \quad &\Longleftrightarrow \quad Z_2 \bar{y} = Z_2 Z_1^\top M Z_2 \bar{x} \\ &\Longleftrightarrow \quad y = Z_2 Z_1^\top M x. \end{aligned} \tag{17}$$

To replace $Z_2 Z_1^\top$ by an orthogonal projector, we consider $M_P \in \mathbb{R}^{m\times m}$ representing another invertible operator $M_P : \mathbb{V}_2 \mapsto \mathbb{V}_1$. Then we introduce Z_1 by

$$Z_1 = M_P Z_2. \tag{18}$$

It follows immediately from the invertibility of M_P that the columns of Z_1 span $\mathbb{V}_1$. Therefore, Z_1 is well-defined by Eq. (18). Inserting Eq. (18) into the last equality in Eq. (17), we arrive at

$$y = P_2 M_P^\top M x, \tag{19}$$

where

$$P_2 = Z_2 Z_2^\top, \tag{20}$$

is the orthogonal projector from $\mathbb{R}^m$ onto $\mathbb{V}_2$ due to Eq. (15). When solving Eq. (11), one can replace Eq. (20) by Eq. (9). This avoids the explicite knowledge of Z_2.

It remains to show how to choose M_P in practice. We propose two variants.

Variant 1: $M_P = M$.

Variant 2: $M_P = P_1$,

where P_1 is the orthogonal projector from $\mathbb{R}^m$ onto $\mathbb{V}_1$. One can compute P_1 again by Eq. (9) .

Below we will analyze the matrix N in Eq. (13) for both variants using the smallest, largest singular values of M on $\mathbb{V}_2$ defined by:

$$\begin{aligned} \sigma_{\min}(M|\mathbb{V}_2) &= \min_{x\in\mathbb{V}_2,\, x\neq 0} \frac{\|Mx\|}{\|x\|}, \\ \sigma_{\max}(M|\mathbb{V}_2) &= \max_{x\in\mathbb{V}_2,\, x\neq 0} \frac{\|Mx\|}{\|x\|}, \end{aligned}$$

respectively. As $0 < \sigma_{\min}(M|\mathbb{V}_2) \le \sigma_{\max}(M|\mathbb{V}_2)$, the condition number of M on $\mathbb{V}_2$ is defined by

$$\kappa(M|\mathbb{V}_2) = \frac{\sigma_{\max}(M|\mathbb{V}_2)}{\sigma_{\min}(M|\mathbb{V}_2)}.$$

3.1. Variant 1

In this case, $N = Z_2^\top M^\top M Z_2$ is symmetric, positive definite so that the CGM can be applied to Eq. (13). It is well-known that its convergence rate is determined by the spectral condition number $\kappa(N)$ of N [11].

Theorem 3 *Let Z_2 be orthogonal. It holds:*

$$\kappa(N) = \kappa(M|\mathbb{V}_2)^2.$$

Proof. Recall that $\kappa(N) = \sigma_{\max}(N)/\sigma_{\min}(N)$, where $\sigma_{\min}(N)$, $\sigma_{\max}(N)$ are the smallest, largest eigenvalues of N, respectively. We get:

$$\begin{aligned} \sigma_{\min}(N) &= \min_{\bar{x}\in\mathbb{R}^{m-l},\, \bar{x}\neq 0} \frac{\bar{x}^\top N \bar{x}}{\bar{x}^\top \bar{x}} \\ &= \min_{\bar{x}\in\mathbb{R}^{m-l},\, \bar{x}\neq 0} \frac{\bar{x}^\top Z_2^\top M^\top M Z_2 \bar{x}}{\bar{x}^\top Z_2^\top Z_2 \bar{x}} \\ &= \min_{x\in\mathbb{V}_2,\, x\neq 0} \frac{x^\top M^\top M x}{x^\top x} \\ &= \sigma_{\min}(M|\mathbb{V}_2)^2. \end{aligned}$$

Analogously, $\sigma_{\max}(N) = \sigma_{\max}(M|\mathbb{V}_2)^2$. □

As $M_P = M$ is invertible, N is non-singular for any choice of the input data. On the other hand, two expensive matrix-vector multiplications by M and its transpose are needed in Eq. (19). Moreover, as $\kappa(N)$ is the square of $\kappa(M|\mathbb{V}_2)$, it is usually too high so that the convergence rate of the CGM (and other Krylov methods) may be slow.

3.2. Variant 2

Now the invertibility of $M_P = P_1$ is not guaranteed. To get this property we need the following theorem.

Theorem 4 *Let P_1 be the orthogonal projector onto $\mathbb{V}_1$. The restriction $P_1 : \mathbb{V}_2 \mapsto \mathbb{V}_1$ is invertible iff*

$$\mathbb{V}_2 \cap \mathbb{V}_1^\perp = \{0\}, \tag{21}$$

where $\mathbb{V}_1^\perp$ is the orthogonal complement to $\mathbb{V}_1$ in $\mathbb{R}^m$.

Proof. First we prove the implication "⇐". Any $x \in \mathbb{V}_2$ can be split into two orthogonal components: $x = x_{\mathbb{V}_1^\perp} + x_{\mathbb{V}_1}$, where $x_{\mathbb{V}_1^\perp} \in \mathbb{V}_1^\perp$ and $x_{\mathbb{V}_1} \in \mathbb{V}_1$. If $x \neq 0$, then Eq. (21) yields $x_{\mathbb{V}_1} \neq 0$ and $P_1 x = x_{\mathbb{V}_1}$. Therefore, the only solution of the homogeneous equation $P_1 x = 0$ on $\mathbb{V}_2$ is trivial so that the invertibility of P_1 on $\mathbb{V}_2$ is proved. To prove the opposite implication "⇒", we assume that there is $x \in \mathbb{V}_2 \cap \mathbb{V}_1^\perp$, $x \neq 0$. Then, x is the non-zero solution of the homogeneous

equation $P_1 x = 0$ on $\mathbb{V}_2$. This contradicts to the invertibility of P_1 on $\mathbb{V}_2$. □

Condition Eq. (21) is equivalent to the fact that the angle θ between the subspaces $\mathbb{V}_2$ and $\mathbb{V}_1^\perp$ is non-zero. Recall that $\theta = \arccos\gamma$, where

$$\gamma = \max_{\substack{x\in\mathbb{V}_2,\, y\in\mathbb{V}_1^\perp \\ x\neq 0,\, y\neq 0}} \frac{x^\top y}{\|x\|\|y\|}.$$

Since Eq. (21) yields $0 \le \gamma < 1$, the *strengthened Cauchy-Schwarz inequality* holds:

$$x^\top y \le \gamma\|x\|\|y\| \qquad \forall x \in \mathbb{V}_2 \;\; \forall y \in \mathbb{V}_1^\perp.$$

The angle between the subspaces $\mathbb{V}_2$, $\mathbb{V}_1^\perp$ and between their orthogonal complements $\mathbb{V}_2^\perp$, $\mathbb{V}_1$ is the same [17]. Therefore,

$$x^\top y \le \gamma\|x\|\|y\| \qquad \forall x \in \mathbb{V}_2^\perp \;\; \forall y \in \mathbb{V}_1, \tag{22}$$

with γ as above. Moreover, the equality in Eq. (22) is achieved for nonzero $x^* \in \mathbb{V}_2^\perp$, $y^* \in \mathbb{V}_1$ such that

$$x^* = Q_2 y^*, \tag{23}$$

where $Q_2 = I - P_2$ is the orthogonal projector from $\mathbb{R}^m$ onto $\mathbb{V}_2^\perp$.

Note that $N = Z_2^\top M Z_2$ in Eq. (13). The following theorem proves, among others, that this N is close to a singular matrix when θ is small.

Theorem 5 *Let Z_2 be orthogonal. Let θ be the angle between the subspaces $\mathbb{V}_2$ and $\mathbb{V}_1^\perp$. The smallest, largest singular values $\sigma_{\min}(N)$, $\sigma_{\max}(N)$ of N on $\mathbb{R}^{m-l}$, respectively, satisfy:*

$$\sin\theta\cdot\sigma_{\min}(M|\mathbb{V}_2) \le \sigma_{\min}(N) \le \sin\theta\cdot\sigma_{\max}(M|\mathbb{V}_2), \tag{24}$$

$$\sin\theta\cdot\sigma_{\min}(M|\mathbb{V}_2) \le \sigma_{\max}(N) \le \sigma_{\max}(M|\mathbb{V}_2). \tag{25}$$

Proof. As P_2, Q_2 are the orthogonal projectors from $\mathbb{R}^m$ onto $\mathbb{V}_2$, $\mathbb{V}_2^\perp$, respectively, it holds:

$$\|y\|^2 - \|P_2 y\|^2 = \|Q_2 y\|^2. \tag{26}$$

Assuming $y \in \mathbb{V}_1$ and using Eq. (22), we have

$$\|Q_2 y\|^2 = y^\top Q_2 y \le \gamma\|Q_2 y\|\|y\|$$

implying $\|Q_2 y\| \le \gamma\|y\|$. Together with Eq. (26), we obtain

$$\|y\|^2 - \|P_2 y\|^2 \le \gamma^2\|y\|^2$$

and, consequently,

$$\|P_2 y\| \ge \sqrt{1-\gamma^2}\cdot\|y\| = \sin\theta\cdot\|y\|.$$

Since the bound is achieved for y^* from Eq. (23), we get

$$\min_{y\in\mathbb{V}_1,\, y\neq 0} \frac{\|P_2 y\|}{\|y\|} = \frac{\|P_2 y^*\|}{\|y^*\|} = \sin\theta. \tag{27}$$

Now

$$\begin{aligned}
\sigma_{\min}(N) &= \min_{\bar{x}\in\mathbb{R}^{m-l},\, \bar{x}\neq 0} \frac{\|N\bar{x}\|}{\|\bar{x}\|} \\
&= \min_{\bar{x}\in\mathbb{R}^{m-l},\, \bar{x}\neq 0} \frac{(\bar{x}^\top Z_2^\top M^\top Z_2 Z_2^\top M Z_2 \bar{x})^{1/2}}{(\bar{x}^\top Z_2^\top Z_2 \bar{x})^{1/2}} \\
&= \min_{x\in\mathbb{V}_2,\, x\neq 0} \frac{(x^\top M^\top P_2 M x)^{1/2}}{(x^\top x)^{1/2}} \\
&= \min_{x\in\mathbb{V}_2,\, x\neq 0} \frac{\|P_2 M x\|}{\|x\|} \\
&= \min_{x\in\mathbb{V}_2,\, x\neq 0 y = Mx} \frac{\|P_2 y\|}{\|y\|}\cdot\frac{\|Mx\|}{\|x\|}. \qquad (28)
\end{aligned}$$

Let x in the last expression of Eq. (28) be chosen as $\hat{x} \in \mathbb{V}_2$ satisfying $y^* = M\hat{x}$, where y^* is from Eq. (27). Then

$$\begin{aligned}
\sigma_{\min}(N) &\le \frac{\|P_2 y^*\|}{\|y^*\|}\cdot\frac{\|M\hat{x}\|}{\|\hat{x}\|} \\
&\le \sin\theta\cdot\sigma_{\max}(M|\mathbb{V}_2),
\end{aligned}$$

proves the upper bound in Eq. (24). Further Eq. (28) yields

$$\begin{aligned}
\sigma_{\min}(N) &\ge \min_{y\in\mathbb{V}_1,\, y\neq 0} \frac{\|P_2 y\|}{\|y\|}\cdot\min_{x\in\mathbb{V}_2,\, x\neq 0} \frac{\|Mx\|}{\|x\|} \\
&= \sin\theta\cdot\sigma_{\min}(M|\mathbb{V}_2),
\end{aligned}$$

the lower bound in Eq. (24). To prove Eq. (25), we start from

$$\sigma_{\max}(N) = \max_{\substack{x\in\mathbb{V}_2,\, x\neq 0 \\ y = Mx}} \frac{\|P_2 y\|}{\|y\|}\cdot\frac{\|Mx\|}{\|x\|}.$$

The upper bound follows immediately from the fact that $\|P_2 y\| \le \|y\|$. Using the same $\hat{x}$ and y^* as above, we obtain the lower bound:

$$\begin{aligned}
\sigma_{\max}(N) &\ge \frac{\|P_2 y^*\|}{\|y^*\|}\cdot\frac{\|M\hat{x}\|}{\|\hat{x}\|} \\
&\ge \sin\theta\cdot\sigma_{\min}(M|\mathbb{V}_2).
\end{aligned}$$

□

4. Projected GMRES

In this section, we derive the projected GMRES method (ProjGMRES) for solving Eq. (11) (and Eq. (12)). We use the same notation for "short"/"extended" vectors, respectively, as in Section 3. We start from the GMRES method [24], [11] introduced below for solving Eq. (13).

The GMRES method is initialized by $\bar{x}^0 \in \mathbb{R}^{m-l}$ and stopped, if the terminating tolerance $\varepsilon > 0$ or the maximal number of iterations $k_{\max} \geq 1$ is achieved. The obtained result $\bar{x}^{\hat{k}} \in \mathbb{R}^{m-l}$ approximates the solution of Eq. (13). Here and in what it follows, $\hat{k}$ stands for the final number of iterations. In step 16° of the scheme below, $H = (h_{i,j}) \in \mathbb{R}^{(k+1)\times k}$ denotes the upper Hessenberg matrix and $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{k+1}$ for $k = \hat{k}$. The minimizer $v \in \mathbb{R}^{\hat{k}}$ in 16° can be efficiently computed using the Givens rotations and, then, solving the resulting triangular linear system [24, 11]. The columns of $\bar{Q}$ are the Arnoldi vectors $\bar{q}^k \in \mathbb{R}^{m-l}$ so that $\bar{Q} = (\bar{q}^1, \ldots, \bar{q}^{\hat{k}}) \in \mathbb{R}^{(m-l)\times\hat{k}}$ represents the orthogonal basis of the respective Krylov space. Note that the norm of the residua *err* in step 13° may be obtained without any explicit knowledge of $\bar{x}^k$ [21]. The computation of *err* is based on applying the Givens rotations to the extended matrix $\tilde{H} = (H|\beta e_1) \in \mathbb{R}^{(k+1)\times(k+1)}$. Eliminating the subdiagonal we $\tilde{H}$, we get *err* as the last diagonal entry in a triangular matrix:

$$\tilde{H} = \begin{pmatrix} h_{11} & h_{12} & \ldots & h_{1,k} & \beta \\ h_{21} & h_{22} & \ldots & h_{2,k} & 0 \\ & h_{32} & \ldots & h_{3,k} & 0 \\ & & \ddots & \vdots & \vdots \\ & & & h_{k+1,k} & 0 \end{pmatrix}$$

$$\overset{\text{Givens rotations}}{\longrightarrow} \begin{pmatrix} \times & \times & \ldots & \times & \times \\ 0 & \times & \ldots & \times & \times \\ & 0 & \ldots & \times & \times \\ & & \ddots & \vdots & \times \\ & & & 0 & err \end{pmatrix}. \quad (29)$$

The GMRES method for solving Eq. (13) is introduced below. Our implementation of the ProjGMRES for solving Eq. (12) is derived from the GMRES using the following results of Section 3: *(i)* the linear combination of vectors in the corresponding steps of the GMRES and the ProjGMRES are given by the same formulas; *(ii)* the scalar products, including the Euclidean norms of vectors, in GMRES and ProjGMRES give the same values; *(iii)* the matrix-vector multiplication by N in the GMRES is replaced by $P_2 M_P^\top M$ in ProjGMRES. Acording to *(i)-(iii)*, one can insert the most of the GMRES formulas into the ProjGMRES removing only the bars over vectors. The remaining formulas are discussed below. The computation of the residua in step 1° of the ProjGMRES is based on the fact that $r^0 = Z_2\bar{r}^0$, $\bar{q} = Z_1^\top q$, and Eq. (18):

$$r^0 = Z_2\bar{r}^0 = Z_2 Z_1^\top q - P_2 M_P^\top M x^0 = P_2 M_P^\top (q - M x^0).$$

The analogous formula is formally used in step 13°, however, the efficient computation of *err* in the ProjGMRES is given again by the principle represented by Eq. (29). Similarly, the minimization problem in step 16° can be realized by the same way as in the GMRES. The matrix $Q = (q^1, \ldots, q^{\hat{k}}) \in \mathbb{R}^{m\times\hat{k}}$ consists of the extended Arnoldi vectors $q^k \in \mathbb{V}_2$. The inputs M, q, P_2, and M_P are introduced in Section 3 and $x^0 \in \mathbb{V}_2$. The meaning of ε and k_{max} is the same as in the GMRES. Obviously, if $x^0 = Z_2\bar{x}^0$, then the GMRES and the ProjGMRES converge for the same number of iterations and $x^{\hat{k}} = Z_2\bar{x}^{\hat{k}}$ (in the exact arithmetic).

GMRES(N,$\bar{q}$,$\bar{x}^0$,ε,k_{max})$\mapsto \bar{x}^{\hat{k}}$

1° $\bar{r}^0 := \bar{q} - N\bar{x}^0$
2° $\beta := \|\bar{r}^0\|$, $err := \beta$
3° $k := 0$, $\varepsilon := \varepsilon \times \beta$
4° *while* $err > \varepsilon$ & $k \leq k_{max}$
5° $\quad \bar{q}^{k+1} := \bar{r}^k / \|\bar{r}^k\|$
6° $\quad k := k + 1$
7° $\quad \bar{r}^k := N\bar{q}^k$
8° $\quad$ *for* $i := 1, \ldots, k$
9° $\quad\quad h_{ik} := (\bar{q}^i)^\top \bar{r}^k$
10° $\quad\quad \bar{r}^k := \bar{r}^k - h_{ik}\bar{q}^i$
11° $\quad$ *end*
12° $\quad h_{k+1,k} := \|\bar{r}^k\|$
13° $\quad err := \|\bar{q} - N\bar{x}^k\|$
14° *end*
15° $\hat{k} := k$
16° $v := \arg\min_{w\in\mathbb{R}^{\hat{k}}} \|Hw - \beta e_1\|$
17° $\bar{x}^{\hat{k}} := \bar{x}^0 + \bar{Q}v$

ProjGMRES(M,q,P_2,M_P,x^0,ε,k_{max})$\mapsto x^{\hat{k}}$

1° $r^0 := P_2 M_P^\top (q - M x^0)$
2° $\beta := \|r^0\|$, $err := \beta$
3° $k := 0$, $\varepsilon := \varepsilon \times \beta$
4° *while* $err > \varepsilon$ & $k \leq k_{max}$
5° $\quad q^{k+1} := r^k / \|r^k\|$
6° $\quad k := k + 1$
7° $\quad r^k := P_2 M_P^\top M q^k$
8° $\quad$ *for* $i := 1, \ldots, k$
9° $\quad\quad h_{ik} := (q^i)^\top r^k$
10° $\quad\quad r^k := r^k - h_{ik} q^i$
11° $\quad$ *end*
12° $\quad h_{k+1,k} := \|r^k\|$
13° $\quad err := \|P_2 M_P^\top (q - M x^k)\|$
14° *end*
15° $\hat{k} := k$
16° $v := \arg\min_{w\in\mathbb{R}^{\hat{k}}} \|Hw - \beta e_1\|$
17° $x^{\hat{k}} := x^0 + Qv$

The Arnoldi vectors q^k generated in the ProjGMRES belong to $\mathbb{V}_2$. This property may be violated due to round-off errors so that q^k may be deflected from $\mathbb{V}_2$. In order to remove this kind of instability, we propose to project each r^k into $\mathbb{V}_2$, i.e., we add $r^k := P_2 r^k$ in step 12°.

Let us turn our attention to problem Eq. (11), i.e., $M = P_1F$, $x = \lambda_{\mathcal{N}}$, $\mathbb{V}_1 = \mathcal{N}(G_1)$, and $\mathbb{V}_2 = \mathcal{N}(G_2)$. The formulas in steps 1°, 7° of the ProjGMRES take the form

$$r^0 := P_2F^\top(q - P_1Fx^0), \quad r^k := P_2F^\top P_1Fq^k \tag{30}$$

for the *Variant 1*, or

$$r^0 := P_2(q - P_1Fx^0), \quad r^k := P_2P_1Fq^k \tag{31}$$

for the *Variant 2*, respectively. The orthogonal projectors P_1, P_2 are given by Eq. (9).

5. Numerical Experiments

To test our algorithms, we shall solve linear systems Eq. (1) arising from the combination of the FD and FETI method applied to finite element approximations of linear elasticity problems.

5.1. Formulation

Let us consider an elastic body represented by a bounded domain $\omega \subset \mathbb{R}^2$ with the sufficiently smooth boundary $\partial\omega$ consisting of two disjoint parts γ_u and γ_p, $\partial\omega = \overline{\gamma}_u \cup \overline{\gamma}_p$ (see Fig. 1). Displacements $\boldsymbol{g} \in (L^2(\gamma_u))^2$ are prescribed on γ_u, while surface tractions of density $\boldsymbol{p} \in (L^2(\gamma_p))^2$ act on γ_p. Finally we suppose that the body ω is subject to volume forces of density $\boldsymbol{f}_{|\omega}$, where $\boldsymbol{f} \in (L^2_{loc}(\mathbb{R}^2))^2$. We seek a displacement field $\boldsymbol{u}$ in ω satisfying the equilibrium equation in ω and the Dirichlet and Neumann conditions on the boundary:

$$\left.\begin{aligned} -\operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}) &= \boldsymbol{f} \quad \text{in} \quad \omega, \\ \boldsymbol{u} &= \boldsymbol{g} \quad \text{on} \quad \gamma_u, \\ \boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{\nu} &= \boldsymbol{p} \quad \text{on} \quad \gamma_p, \end{aligned}\right\} \tag{32}$$

where $\boldsymbol{\sigma}(\boldsymbol{u})$ is the stress tensor in ω corresponding to $\boldsymbol{u}$ and $\boldsymbol{\nu}$ stands for the unit outward normal vector to γ. The stress tensor $\boldsymbol{\sigma}(\boldsymbol{u})$ is related to the linearized strain tensor $\boldsymbol{\epsilon}(\boldsymbol{u}) = 1/2(\nabla\boldsymbol{u} + \nabla^\top\boldsymbol{u})$ by Hooke's law for elastic, homogeneous, and isotropic materials:

$$\boldsymbol{\sigma}(\boldsymbol{u}) = c_1 \operatorname{tr}(\boldsymbol{\epsilon}(\boldsymbol{u}))I + 2c_2\boldsymbol{\epsilon}(\boldsymbol{u}) \quad \text{in} \quad \omega,$$

where "tr" denotes the trace of matrices, $I \in \mathbb{R}^{2\times 2}$ is the identity matrix and $c_1, c_2 > 0$ are the Lame constants.

Denote $V_{\boldsymbol{g}}(\omega) = \{\boldsymbol{v} \in (H^1(\omega))^2 |\ \boldsymbol{v} = \boldsymbol{g} \text{ on } \gamma_u\}$. The weak formulation of Eq. (32) reads as follows:

$$\left.\begin{aligned} &\textit{Find } \boldsymbol{u} \in V_{\boldsymbol{g}}(\omega) \textit{ such that} \\ &a_\omega(\boldsymbol{u},\boldsymbol{v}) = f_\omega(\boldsymbol{v}) + (\boldsymbol{p},\boldsymbol{v})_{\gamma_p} \quad \forall \boldsymbol{v} \in \mathbb{V}_{\boldsymbol{0}}(\omega), \end{aligned}\right\} \tag{33}$$

where

$$a_\omega(\boldsymbol{u},\boldsymbol{v}) = \int_\omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{v})\,\mathrm{d}\boldsymbol{x}, \quad f_\omega(\boldsymbol{v}) = \int_\omega \boldsymbol{f}\cdot\boldsymbol{v}\,\mathrm{d}\boldsymbol{x},$$

and $(\cdot,\cdot)_{\gamma_p}$ is the scalar product in $(L^2(\gamma_p))^2$.

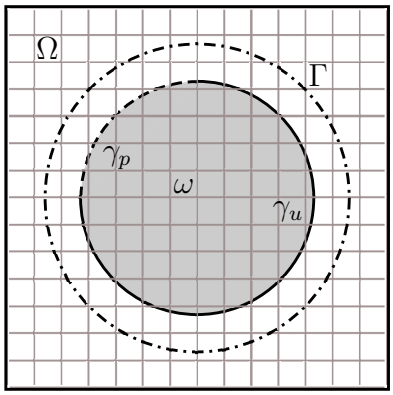

Fig. 1: FD method.

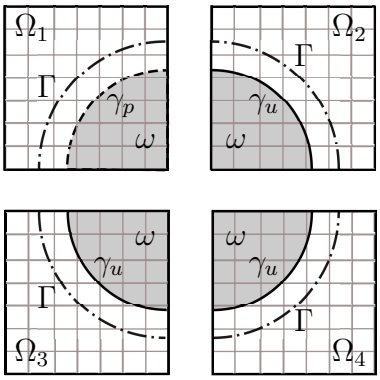

Fig. 2: FD-FETI method.

Let us take a box Ω such that $\overline{\omega} \subset \Omega$ and construct a closed curve Γ surrounding ω; see Fig. 1. Instead of Eq. (33), we propose to solve the following fictitious domain (FD) formulation of Eq. (32) in Ω:

$$\begin{aligned} &\textit{Find } (\hat{\boldsymbol{u}}, \boldsymbol{\lambda}_\Gamma) \in (H^1(\Omega))^2 \times (H^{-1/2}(\Gamma))^2 \quad \textit{such that} \\ &a_\Omega(\hat{\boldsymbol{u}},\boldsymbol{v}) + b_\Gamma(\boldsymbol{\lambda}_\Gamma,\boldsymbol{v}) = f_\Omega(\boldsymbol{v})\ \forall \boldsymbol{v} \in (H^1(\Omega))^2, \\ &b_{\gamma_u}(\boldsymbol{\mu}_{\gamma_u},\hat{\boldsymbol{u}}) = b_{\gamma_u}(\boldsymbol{\mu}_{\gamma_u},\boldsymbol{g})\ \forall \boldsymbol{\mu}_{\gamma_u} \in (H^{-1/2}(\gamma_u))^2, \\ &b_{\gamma_p}(\boldsymbol{\mu}_{\gamma_p},\boldsymbol{\sigma}(\hat{\boldsymbol{u}})\boldsymbol{\nu}) = b_{\gamma_p}(\boldsymbol{\mu}_{\gamma_p},\boldsymbol{p})\ \forall \boldsymbol{\mu}_{\gamma_p} \in (H^{1/2}(\gamma_p))^2, \end{aligned}$$

where b_Γ, b_{γ_u}, b_{γ_p} are duality pairings between the respective trace spaces and their duals. It is readily seen that $\hat{\boldsymbol{u}}_{|\omega}$ solves Eq. (33). Notice that $\boldsymbol{\lambda}_\Gamma$ plays the role of the control variable defined on Γ enforcing the

boundary conditions on γ to be satisfied. This new reformulation of Eq. (33) enables us to use (non-fitted) uniform meshes on Ω and, in addition, it improves the convergence rate of finite element approximations; see [16], [15].

To increase the computational efficiency, we also apply the FETI method. We decompose Ω into subboxes Ω_i, $i = 1, \dots, s$ such that $\overline{\Omega} = \bigcup_{i=1}^{s} \overline{\Omega}_i$; see Fig. 2. To ensure the global continuity of the solution, we add the "gluing" condition on interfaces of the subboxes. A common interface between Ω_i and Ω_j, $i \neq j$, is defined by $\Delta_{ij} = \partial\Omega_i \cap \partial\Omega_j$, $|\Delta_{ij}| > 0$. Before giving the FD-FETI formulation of Eq. (32), we introduce notation:

$$\mathbb{V}(\Omega) = \prod_{i=1}^{s} (H^1(\Omega_i))^2$$

for $\boldsymbol{v} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_s) \in \mathbb{V}(\Omega)$, $\boldsymbol{v}_i \in (H^1(\Omega_i))^2$;

$$\tilde{a}_\Omega(\boldsymbol{v}, \boldsymbol{w}) = \sum_{i=1}^{s} a_{\Omega_i}(\boldsymbol{v}_i, \boldsymbol{w}_i), \;\; \tilde{f}_\Omega(\boldsymbol{v}) = \sum_{i=1}^{s} f_{\Omega_i}(\boldsymbol{v}_i)$$

for $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{V}(\Omega)$;

$$\mathbb{X}(\Delta) = \prod_{\Delta_{ij} \in \Delta} (H^{-1/2}(\Delta_{ij}))^2$$

with $\Delta = \bigcup_{i \neq j} \Delta_{ij}$; and

$$b_\Delta(\boldsymbol{\mu}, \boldsymbol{v}) = \sum_{\Delta_{ij} \in \Delta} b_{\Delta_{ij}}(\boldsymbol{\mu}_{|_{\Delta_{ij}}}, [\boldsymbol{v}]_{ij})$$

for $\boldsymbol{\mu} \in \mathbb{X}(\Delta)$, $\boldsymbol{v} \in \mathbb{V}(\Omega)$, where

$$[\boldsymbol{v}]_{ij} = \boldsymbol{v}_{i|_{\Delta_{ij}}} - \boldsymbol{v}_{j|_{\Delta_{ij}}}$$

denotes the jump of $\boldsymbol{v}$ on Δ_{ij}. The decomposition of the trace spaces on $\pi \in \{\Gamma, \gamma_u, \gamma_p\}$ is defined as follows:

$$\mathbb{X}(\pi) = \prod_{i: \pi \cap \overline{\Omega}_i \neq \emptyset} (H^\epsilon(\pi \cap \overline{\Omega}_i))^2,$$

where $\epsilon = -1/2$ for $\pi = \Gamma, \gamma_u$ and $\epsilon = 1/2$, if $\pi = \gamma_p$; the respective duality pairings are given by

$$\tilde{b}_\pi(\boldsymbol{\mu}, \boldsymbol{v}) = \sum_{i: \pi \cap \overline{\Omega}_i \neq \emptyset} b_{\pi,i}(\boldsymbol{\mu}_{|_{\pi \cap \overline{\Omega}_i}}, \boldsymbol{v}_i)$$

for $\boldsymbol{\mu} \in \mathbb{X}(\pi)$, $\boldsymbol{v} \in \mathbb{V}(\Omega)$, where $b_{\pi,i}$ are duality pairings on subspaces and their duals.

The FD-FETI formulation of Eq. (32) reads as follows:

Find $(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}_\Gamma, \tilde{\boldsymbol{\lambda}}_\Delta) \in \mathbb{V}(\Omega) \times \mathbb{X}(\Gamma) \times \mathbb{X}(\Delta)$ *such that*

$$\tilde{a}_\Omega(\tilde{\boldsymbol{u}}, \boldsymbol{v}) + \tilde{b}_\Gamma(\tilde{\boldsymbol{\lambda}}_\Gamma, \boldsymbol{v}) + b_\Delta(\tilde{\boldsymbol{\lambda}}_\Delta, \boldsymbol{v}) = \tilde{f}_\Omega(\boldsymbol{v}) \;\; \forall \boldsymbol{v} \in \mathbb{V}(\Omega),$$

$$\tilde{b}_{\gamma_u}(\boldsymbol{\mu}_{\gamma_u}, \tilde{\boldsymbol{u}}) = \tilde{b}_{\gamma_u}(\boldsymbol{\mu}_{\gamma_u}, \boldsymbol{g}) \quad \forall \boldsymbol{\mu}_{\gamma_u} \in \mathbb{X}(\gamma_u),$$

$$\tilde{b}_{\gamma_p}(\boldsymbol{\mu}_{\gamma_p}, \boldsymbol{\sigma}(\tilde{\boldsymbol{u}})\boldsymbol{\nu}) = \tilde{b}_{\gamma_p}(\boldsymbol{\mu}_{\gamma_p}, \boldsymbol{p}) \quad \forall \boldsymbol{\mu}_{\gamma_p} \in \mathbb{X}(\gamma_p),$$

$$b_\Delta(\boldsymbol{\mu}_\Delta, \tilde{\boldsymbol{u}}) = 0 \quad \forall \boldsymbol{\mu}_\Delta \in \mathbb{X}(\Delta).$$

Note that $\hat{\boldsymbol{u}}$ satisfies $\hat{\boldsymbol{u}}_{|_{\Omega_i}} = \tilde{\boldsymbol{u}}_i$, $i = 1, \dots, s$, since the last equation ensures the zero jump of $\tilde{\boldsymbol{u}}$ across all interfaces Δ_{ij}. The mixed finite element approximation leads to the two-by-two block linear system Eq. (1), in which A is symmetric, positive semi-definite with the block diagonal structure, $A = diag(A_1, \dots, A_s)$, where $A_i \in \mathbb{R}^{n_i \times n_i}$, $n = \sum_{i=1}^{s} n_i$, and $C = 0$. The off-diagonal blocks B_1, B_2 and g in Eq. (1) are given by

$$B_1 = \begin{pmatrix} B_\Gamma \\ B_\Delta \end{pmatrix}, \; B_2 = \begin{pmatrix} B_{\gamma_u} \\ B_{\gamma_p} \\ B_\Delta \end{pmatrix}, \; g = \begin{pmatrix} g_{\gamma_u} \\ g_{\gamma_p} \\ 0 \end{pmatrix},$$

where $B_\Gamma \in \mathbb{R}^{m_\Gamma \times n}$, $B_\Delta \in \mathbb{R}^{m_\Delta \times n}$, $B_{\gamma_u} \in \mathbb{R}^{m_{\gamma_u} \times n}$, $B_{\gamma_p} \in \mathbb{R}^{m_{\gamma_p} \times n}$, $g_{\gamma_u} \in \mathbb{R}^{m_{\gamma_u}}$, and $g_{\gamma_p} \in \mathbb{R}^{m_{\gamma_p}}$ with $m := m_\Gamma + m_\Delta = m_{\gamma_u} + m_{\gamma_p} + m_\Delta$. Here, B_Γ and B_{γ_u} are the Dirichlet trace matrices on Γ and γ_u, respectively, and B_{γ_p} is the Neumann trace matrix (representing the trace of $\boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{\nu}$) on γ_p. If the Ladyzhenskaya-Babuska-Brezzi condition [3] is satisfied, we get the linear system Eq. (1) with $\mathcal{A}$ nonsingular.

5.2. Model Example

Let ω be the interior of the circle:

$$\omega = \{(x, y) \in \mathbb{R}^2 |\; (x - 0.5)^2 + (y - 0.5)^2 < 0.3^2\},$$

and $\Omega = (0, 1) \times (0, 1)$. The data in Eq. (32) are chosen as follows: $\boldsymbol{f} = -\mathrm{div}\,\boldsymbol{\sigma}(\boldsymbol{u})$, $\boldsymbol{g} = \boldsymbol{u}_{|_{\gamma_u}}$, and $\boldsymbol{p} = \boldsymbol{\sigma}(\boldsymbol{u}_{|_{\gamma_p}})\boldsymbol{\nu}$, where $\boldsymbol{u}(x, y) = (0.1xy, 0.1xy)$, $(x, y) \in \mathbb{R}^2$ and γ_p is the (northwest) quarter of the circle $\gamma := \partial\omega$ as in Fig. 2. The decomposition of Ω into Ω_i is given by dividing the sides of Ω into n_Ω equidistant segments of length H so that $s = n_\Omega^2$ is the total number of the subboxes. The mesh in each Ω_i is constructed analogously: we divide the sides of Ω_i into n_{Ω_i} equidistant segments of size h so that $(n_{\Omega_i}+1)^2$ is the total number of the finite element nodes in $\overline{\Omega}_i$. The finite element discretization of each $H^1(\Omega_i)$ is given by piecewise bilinear functions so that $n = 2n_\Omega^2(n_{\Omega_i} + 1)^2$. Note that the used meshes are conforming. The auxiliary curve Γ is constructed by shifting γ three h units in the direction of the outward normal vector to γ. To define finite element approximations of $\mathbb{X}(\Gamma)$, $\mathbb{X}(\gamma_u)$, and $\mathbb{X}(\gamma_p)$ we proceed as follows. First we find intersection points of Γ, γ_u, and γ_p with the finite element meshes on Ω_i, $i = 1, \dots, s$, that define polygonal approximations of Γ, γ_u, and γ_p, respectively, with non-equidistant straight segments. Then we construct partitions of these approximations $\tilde{\Gamma}, \tilde{\gamma}_u$, and $\tilde{\gamma}_p$ given by polygonal macrosegments, i.e., unions of few following straight segments. We assume that $\tilde{\Gamma}, \tilde{\gamma}_u$, and $\tilde{\gamma}_p$ contain m_Γ, m_{γ_u}, and m_{γ_p} almost equidistant macrosegments for which $m_\Gamma = m_{\gamma_u} + m_{\gamma_p}$. The spaces $\mathbb{X}(\Gamma)$, $\mathbb{X}(\gamma_u)$, and

$\mathbb{X}(\gamma_p)$ are discretized by piecewise constant functions over $\tilde{\Gamma}, \tilde{\gamma}_u$, and $\tilde{\gamma}_p$, respectively. Next we shall suppose that $h_\gamma/h = \log_2 h$, where h_γ denotes the maximal length of the macrosegments on $\tilde{\gamma} = \tilde{\gamma} \cup \gamma_p$; see [16]. The entries of the matrices B_Γ, B_{γ_u}, B_{γ_p} are given by: $(B_\Gamma)_{ij} = \tilde{b}_\Gamma(\boldsymbol{\xi}_i, \boldsymbol{\phi}_j)$, $i = 1, \ldots, m_\Gamma$, $j = 1, \ldots, n$; $(B_{\gamma_u})_{ij} = \tilde{b}_{\gamma_u}(\boldsymbol{\psi}_i, \boldsymbol{\phi}_j)$, $i = 1, \ldots, m_{\gamma_u}$, $j = 1, \ldots, n$; $(B_{\gamma_p})_{ij} = \tilde{b}_{\gamma_p}(\boldsymbol{\psi}_{i+m_{\gamma_u}}, \boldsymbol{\sigma}(\boldsymbol{\phi}_j)\boldsymbol{\nu})$, $i = 1, \ldots, m_{\gamma_p}$, $j = 1, \ldots, n$; where $\boldsymbol{\xi}_i$, $\boldsymbol{\psi}_i$, and $\boldsymbol{\phi}_j$ are the basis functions of the finite element spaces on $\tilde{\Gamma}$, $\tilde{\gamma} = \tilde{\gamma}_u \cup \tilde{\gamma}_p$, and Ω, respectively. As the duality pairings are represented by integrals over $\tilde{\Gamma}$, $\tilde{\gamma}_u$, and $\tilde{\gamma}_p$, we evaluate their values by the composite trapezoidal rule [22]. Since γ and Γ are not far apart, the satisfaction of Eq. (21) is guaranteed. The space $\mathbb{X}(\Delta)$ is approximated by the Dirac distributions defined at the matching nodes on the common interfaces Δ_{ij}. The discrete "gluing" condition at the matching node $\boldsymbol{x}_k$ on Ω_i and $\boldsymbol{x}_l$ on Ω_j reads as $\boldsymbol{v}_i(\boldsymbol{x}_k) - \boldsymbol{v}_j(\boldsymbol{x}_l) = 0$. Thus each row of B_Δ contains only two nonzero entries 1 and -1 in appropriate positions.

The model example is computed with $h = 4 \cdot 10^{-3}$ and $H = 2 \cdot 10^{-1}$. Fig. 3 and Fig. 4 show the deformation and the stress field, respectively. The pointwise errors between the exact and computed solutions are drawn in Fig. 5 and Fig. 6. One can observe that the most significant error is concentrated in a vicinity of γ, especially on the part γ_p.

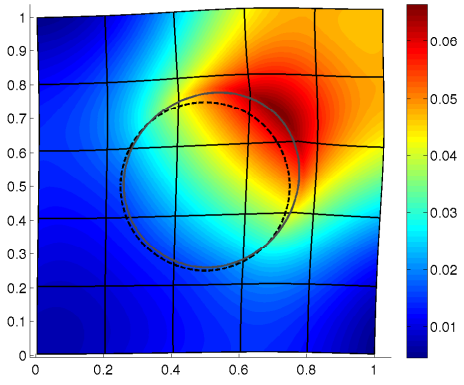

Fig. 3: Deformation field.

5.3. Comparison of the Algorithms

We compare the efficiency of the PSCM with respect to different solvers used in the inner level. By ProjGMRES(P_1F) and ProjGMRES(P_1) we denote our projected GMRES variants with $M_P = M$ and $M_P = P_1$, respectively. Note that the former leads to the equation with the symmetric, positive definite operator so that the projected CGM may be also used in this case. We refer to this method as

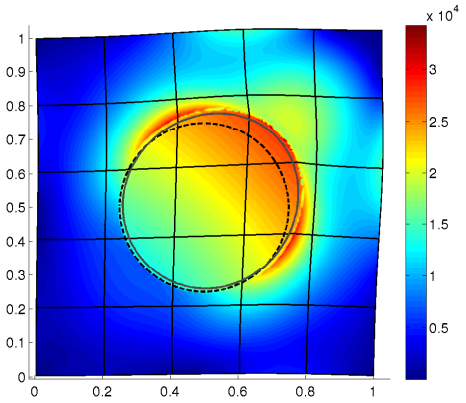

Fig. 4: Stress field.

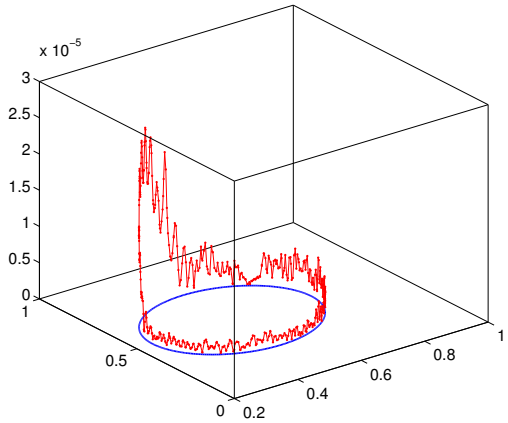

Fig. 5: Pointwise error on γ.

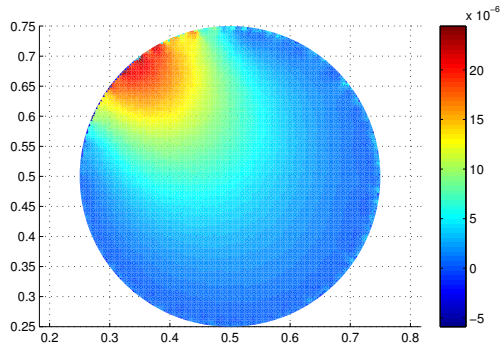

Fig. 6: Pointwise error in ω

ProjCGM(P_1F). Its implementation was done with the reorthogonalization of the Lanczos vectors [11]. The projected BiCGSTAB algorithm is denoted here by ProjBiCGSTAB(P_1). It was proposed in [16] starting from the standard BiCGSTAB [26] by analogy with the projected CGM [8] but only for the *Variant 2*. Ac-

tions of generalized inverses of A to vectors are realized by combining the Cholesky factorization with the singular value decomposition [4]. The choice of the generalized inverse has no influnce on convergence of the projected methods as follows from Theorem 2 (see [20], [19] for numerical experiments). All computations were performed by using 32 cores with 2GB memory per core of the HP Blade system, model BLc7000.

Figure 7 and Fig. 8 show convergence of the relative residual norm (i.e. err/β in the ProjGMRES) typical for small and large problems, respectively. Table 1 reports the number of iterations (**iter**) and the computational time in seconds (CPU_time). One can observe

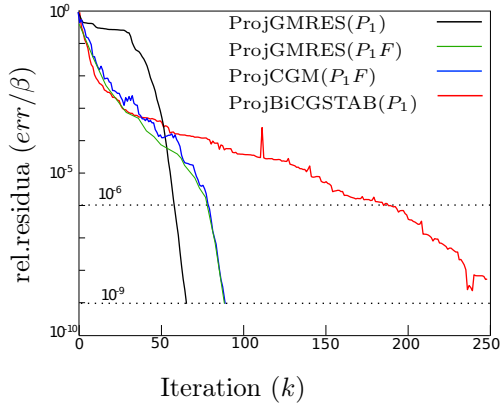

Fig. 7: $n = 3528$, $m = 186$.

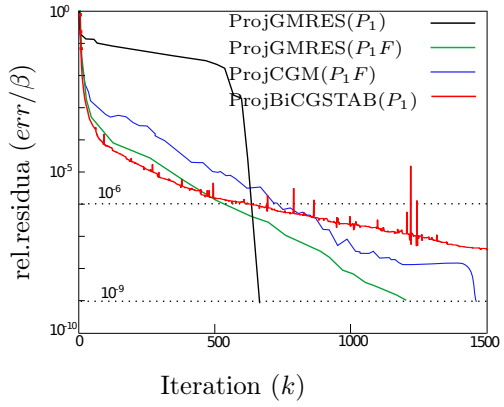

Fig. 8: $n = 520200$, $m = 18372$.

that ProjGMRES(P_1) is the most efficient method, if a high accuracy of the computed solution is required. The progress is more expressive in CPU_time, since the only one action of the generalized inverse of A per iteration is needed.

Finally, Fig. 9 displays the condition number

$$\kappa(P_2F^\top P_1F|\mathcal{N}(G_2)) = \frac{\sigma_{\max}(P_2F^\top P_1F|\mathcal{N}(G_2))}{\sigma_{\min}(P_2F^\top P_1F|\mathcal{N}(G_2))}$$

for the different number of the subboxes provided that the ratio H/h is fixed. The obtained results indicate

Tab. 1: Complexity of computations: **iter** and CPU_time (in brackets) for two terminating tolerances ε.

subboxes (s)	25	100	225
primal DOFs (n)	130050	520200	1170450
dual DOFs (m)	4148	18372	42692
nullity of A (l)	75	300	675
$\varepsilon = 10^{-6}$			
ProjGMRES(P_1)	**322**	**633**	**1001**
	(34.7)	(289.4)	(2474.1)
ProjGMRES(P_1F)	**289**	**543**	**743**
	(48.5)	(396.7)	(2164.3)
ProjCGM(P_1F)	**378**	**732**	**1046**
	(63.3)	(588.4)	(4339.8)
ProjBiCGSTAB(P_1)	**538**	**628**	**672**
	(163.6)	(793.3)	(1849.8)
$\varepsilon = 10^{-9}$			
ProjGMRES(P_1)	**342**	**665**	**1040**
	(35.9)	(327.5)	(2817.1)
ProjGMRES(P_1F)	**592**	**1210**	**1678**
	(122.6)	(1100.4)	(7462.2)
ProjCGM(P_1F)	**686**	**1462**	**2198**
	(126.8)	(1243.0)	(9666.7)
ProjBiCGSTAB(P_1)	>2500	>2500	>2500

*if >2500, the default maximum of iterations is achieved

the boundedness of $\kappa = \kappa(P_2F^\top P_1F|\mathcal{N}(G_2))$ that corresponds to the scalability of the FETI method [25].

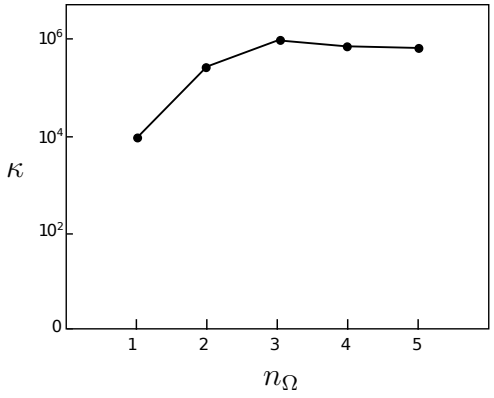

Fig. 9: Experimental scalability of the FD-FETI method, $H/h = 50$.

6. Conclusions

The paper deals with the solution of large non-symmetric two-by-two block linear systems with a singular leading submatrix by the PSCM algorithm. This algorithm extends the ideas used in the FETI domain decomposition methods. It results from the necessity to solve the linear equation given by the invertible operator M between two different subspaces $\mathbb{V}_2$ and $\mathbb{V}_1$ in $\mathbb{R}^m$. We propose the general technique that enables us to derive from the standard Krylov methods for non-symmetric matrices their projected analogies ge-

nerating iterations on $\mathbb{V}_2$. Two variants how to choose the auxiliary operator M_P that maps the basis of $\mathbb{V}_2$ into $\mathbb{V}_1$ are analyzed. The first variant with $M_P = M$ guarantees convergence for each input data. Unfortunately, two expensive actions of M and $M^\top$ are necessary when multiplying. In addition, the condition number is too high so that the convergence rate may be slow. The second variant with $M_P = P_1$ uses the orthogonal projector P_1 onto $\mathbb{V}_1$. In this case, the action of $M^\top$ is replaced by the action of P_1 that is considerably cheaper in computations. However, convergence is guaranteed only when the angle θ between $\mathbb{V}_2$ and $\mathbb{V}_1^\perp$ is non-zero. Moreover, if θ is small, then the problem is equivalent to the system of linear equations whose matrix is close to the singular one.

The second part of the paper is devoted to numerical experiments illustrated by examples arising from the FD-FETI method. We compare two variants of the projected GMRES derived in this paper, namely with the projected CGM (for $M_P = M$) and the projected BiCGSTAB (for $M_P = P_1$). We may conclude that the GMRES approach with $M_P = P_1$ leads to the best performance of the PSCM algorithm in almost all experiments, especially, when the high accuracy of the computed solution is needed.

Acknowledgment

This work was supported by the grant GA CR P201/12/0671 (RK, JH), by the European Regional Development Fund in the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 (RK, TK, AM, JH), by the project SPOMECH, reg. no. CZ.1.07/2.3.00/20.0070 (TK), and by the project Opportunity for young researchers, reg. no. CZ.1.07/2.3.00/30.0016 (AM) within Operational Programme 'Education for competitiveness' funded by Structural Funds of the European Union, co-financed by the European Social Fund and the state budget of the Czech Republic.

References

[1] BEN–ISRAEL, A. and T. GREVILLE. *Generalized inverses: theory and applications*. New York: Springer, 2003. ISBN 978-0-387-21634-8.

[2] BENZI, M., G. H. GOLUB and J. LIESEN. Numerical solution of saddle point systems. *Acta Numerica*. 2005, vol. 14, iss. 1, pp. 1–137. ISSN 0962-4929. DOI: 10.1017/S0962492904000212.

[3] BREZZI, F. and M. FORTIN. *Mixed and hybrid finite elements methods*. New York: Springer-Verlag, 1991. ISBN 978-1-4612-7824-5.

[4] BRZOBOHATY, T., Z. DOSTAL, T. KOZUBEK, A. MARKOPOULOS and P. KOVAR. Cholesky decomposition with fixing nodes to stable evaluation of a generalized inverse of the stiffness matrix of a floating structure. *International Journal for Numerical Methods in Engineering*. 2011, vol. 88, iss. 5, pp. 1384–1405. ISSN 1097-0207. DOI: 10.1002/nme.3187.

[5] DOSTAL, Z. *Optimal quadratic programming algorithms: with applications to variational inequalities*. New York: Springer, 2009. ISBN 978-0-387-84806-8.

[6] DOSTAL, Z., D. HORAK and R. KUCERA. Total FETI - an easier implementable variant of the FETI method for numerical solution of elliptic PDE. *Communications in Numerical Methods in Engineering*. 2006, vol. 22, iss. 12, pp. 1155–1162. ISSN 2040-7947. DOI: 10.1002/cnm.881.

[7] DOSTAL, Z., T. KOZUBEK, A. MARKOPOULOS and M. MENSIK. Cholesky decomposition and a generalized inverse of the stiffness matrix of a floating structure with known null space. *Applied Mathematics and Computation*. 2011, vol. 217, iss. 13, pp. 6067–6077. ISSN 0096-3003. DOI: 10.1016/j.amc.2010.12.069.

[8] FARHAT, C., J. MANDEL and F. X. ROUX. Optimal convergence properties of the FETI domain decomposition method. *Computer Methods in Applied Mechanics and Engineering*. 1994, vol. 115, iss. 3–4, pp. 365–385. ISSN 0045-7825. DOI: 10.1016/0045-7825(94)90068-X.

[9] GIRAULT, V. and R. GLOWINSKI. Error analysis of a fictitious domain method applied to a Dirichlet problem. *Japan Journal of Industrial and Applied Mathematics*. 1995, vol. 12, iss. 3, pp. 487–514. ISSN 0916-7005. DOI: 10.1007/BF03167240.

[10] GLOWINSKI, R., T. PAN and J. PERIAUX. A fictitious domain method for Dirichlet problem and applications. *Computer Methods in Applied Mechanics and Engineering*. 1994, vol. 111, iss. 3–4, pp. 283–303. ISSN 0045-7825. DOI: 10.1016/0045-7825(94)90135-X.

[11] GOLUB, G. H. and C. F. VAN LOAN. *Matrix computations*. 3rd ed. Baltimore: The Johns Hopkins University Press, 1996. ISBN 978-0801854149.

[12] GOULD, N., M. HRIBAR and J. NOCEDAL. On the solution of equality constrained quadratic programming problems arising in optimization. *SIAM Journal on Scientific Computing*. 2000, vol. 23, iss. 4, pp. 1376–1395. ISSN 1064-8275. DOI: 10.1137/S1064827598345667.

[13] HASLINGER, J. and A. KLARBRING. Fictitious domain/mixed finite element approach for a class of optimal shape design problems. *ESAIM: Mathematical Modelling and Numerical Analysis*. 1995, vol. 29, iss. 4, pp. 435–450. ISSN 0764-583X.

[14] HASLINGER, J. and T. KOZUBEK. *A fictitious domain approach for a class of Neumann boundary value problems with applications in shape optimization*. Jyvaskyla: University of Jyvaskyla, 2000. ISBN 978-9513905934.

[15] HASLINGER, J., T. KOZUBEK and R. KUCERA. Fictitious domain formulation of unilateral problems: analysis and algorithms. *Computing*. 2009, vol. 84, iss. 1–2, pp. 69–96. ISSN 0010-485X. DOI: 10.1007/s00607-009-0026-y.

[16] HASLINGER, J., T. KOZUBEK, R. KUCERA and G. PEICHL. Projected Schur complement method for solving non-symmetric saddle-point systems arising from fictitious domain approach. *Numerical Linear Algebra with Applications*. 2007, vol. 14, iss. 9, pp. 713–739. ISSN 1099-1506. DOI: 10.1002/nla.550.

[17] ISPEN, I. C. F. and C. D. MEYER. The angle between complementary subspaces. *The American Mathematical Monthly*. 1995, vol. 102, no. 10, pp. 904–911. ISSN 0002-9890. DOI: 10.2307/2975268.

[18] KOZUBEK, T., A. MARKOPOULOS, T. BRZOBOHATY, R. KUCERA, V. VONDRAK and Z. DOSTAL. MatSol–MATLAB efficient solvers for problems in engineering. In: *IT4Innovations* [online]. 2009. Available at: `http://matsol.vsb.cz/`.

[19] KUCERA, R., T. KOZUBEK and A. MARKOPOULOS. On large-scale generalized inverses in solving two-by-two block linear systems. *Linear Algebra and Its Applications*. 2013, vol. 438, iss. 7, pp. 3011–3029. ISSN 0024-3795. DOI: 10.1016/j.laa.2012.09.027.

[20] KUCERA, R., T. KOZUBEK, A. MARKOPOULOS and J. MACHALOVA. On the Moore-Penrose inverse in solving saddle-point systems with singular diagonal blocks. *Numerical Linear Algebra with Applications*. 2012, vol. 19, iss. 4, pp. 677–699. ISSN 1099-1506. DOI: 10.1002/nla.798.

[21] MEURANT, G. *Computer solution of large linear systems*. Amsterdam: Elsevier, 1999. ISBN 978-0444501691.

[22] MOCEK, L. *Numerical solution of partial differential equations using a fictitious domain method*. Ostrava, 2012. Disseration thesis. VSB–Technical University of Ostrava.

[23] D. Orban. Projected Krylov methods for unsymmetric augmented systems. In: *Sparse Days Meeting 2008 at CERFACS* [online]. 2008. Available at: `http://www.cerfacs.fr/algor/PastWorkshops/SparseDays2008/Abstracts/orban.pdf`.

[24] SAAD, Y. and M. H. SCHULTZ. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*. 1986, vol. 7, iss. 3, pp. 856–869. ISSN 1064-8275. DOI: 10.1137/0907058.

[25] TOSELLI, A. and O. B. Widlund. *Domain decomposition methods–algorithms and theory*. New York: Springer, 2005. ISBN 978-3-540-26662-4.

[26] VAN DER VORST, H. A. BiCGSTAB: a fast and smoothly converging variant of BiCG for solution of nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*. 1992, vol. 13, iss. 2, pp. 631–644. ISSN 1064-8275. DOI: 10.1137/0913035.

About Authors

Radek KUCERA was born in 1968 in Opava, Czech Republic. Since 2001, he is employed as the associated professor at the VSB–Technical University of Ostrava and since 2012 he is Head of Department of Mathematics and Descriptive Geometry. The research specialization includes numerical linear algebra, optimization, domain decomposition methods, or contact problems of mechanics.

Tomas KOZUBEK was born in 1975 in Karvina, Czech Republic. Since 2011, he is Head of the Research Programme Supercomputing for Industry at the National Supercomputing Center IT4Innovations and since 2013, Professor at the VSB–Technical University of Ostrava. The research specialization includes numerical linear algebra, optimization, domain decomposition and fictitious domain methods, and contact problems of mechanics.

Alexandros MARKOPOULOS was born in 1981 in Krnov, Czech Republic. Since 2009, he is employed as the junior researcher at the VSB–Technical University of Ostrava. The research specialization includes applied mechanics, numerical linear algebra, domain decomposition methods.

Jaroslav HASLINGER was born in 1946 in

Olomouc, Czech Republic, is the member of the Department of Numerical Mathematics at the Faculty of Mathematics and Physics, Charles University, Prague. His main research areas cover: structural optimization, approximation and numerical realization of variational inequalities, in particular contact problems with Coulomb friction and fictitious domain methods for solving elliptic problems.

Lukas MOCEK was born in 1983 in Ostrava. Graduated Ph.D. studies in 2012. His research specializations cover fictitious domain and domain decompositions methods.