Programs and Algorithms of Numerical Mathematics 22
J. Chleboun, J. Papež, K. Segeth, J. Šístek, T. Vejchodský (Eds.)
Institute of Mathematics CAS, Prague 2025

COMPARISON OF PRECONDITIONING AND DEFLATION TECHNIQUES OF FETI METHODS FOR PROBLEM OF 2D LINEAR ELASTICITY

Adam Růžička[1,2], David Horák[1,2]

[1] Department of Applied Mathematics, FEECS, VŠB-TUO
17. listopadu 2172/15, Ostrava, Czech Republic
adam.ruzicka.st@vsb.cz, david.horak@vsb.cz
[2] Institute of Geonics, Czech Academy of Sciences
Studentská 1768/9, Ostrava, Czech Republic

Abstract: This paper deals with the basic preconditioning and deflation variants of the FETI-1 and TFETI-1 methods, with (T)FETI-1 with deflation being called (T)FETI-2. It also presents the results of numerical experiments performed on a simple benchmark 2D problem of linear elasticity to compare the computational efficiency of FETI-1 and TFETI-1 and each variant of their preconditioning or deflation in terms of number of executed CG iterations.

Keywords: FETI, domain decomposition, preconditioning, deflation

MSC: 65F08, 65F10

1. Introduction

The Finite element tearing and interconnecting (FETI) methods are probably the most commonly used domain decomposition methods for a parallel numerical solution of PDEs. In Section 2, the mathematical formulations of basic FETI methods: FETI-1 and TFETI-1 are presented. The basic ways of their preconditioning are introduced in Section 3. In Section 4, the principle of the deflated conjugate gradient method is presented. In Section 5, mathematical formulations of applied methods of deflation are introduced. In Section 6, the results of numerical experiments performed on the simple benchmark 2D FEM-discretized problem of linear elasticity are presented.

2. FETI-1 and TFETI-1 methods

In all methods of the FETI-type, the global problem of linear elasticity discretized by FEM, defined on discretized linear elastic domain, is decomposed into several local problems defined on non-overlapping subdomains which are then glued via conditions of displacements continuity across their mutual interfaces, which leads to the constrained minimization problem of quadratic programming [1]:

DOI: 10.21136/panm.2024.13

$$\min\,(1/2)\,u^T K u - u^T f \quad \text{s.t. } Bu = o \tag{1}$$

$$K = \text{diag}\left(K_1 \quad \cdots \quad K_i \quad \cdots \quad K_{N_s}\right), \quad K \in \mathbb{R}^{n\times n} \tag{2}$$

$$u = \left[\, u_1^T \ \cdots \ u_i^T \ \cdots \ u_{N_s}^T \,\right]^T;\ f = \left[\, f_1^T \ \cdots \ f_i^T \ \cdots \ f_{N_s}^T \,\right]^T;\ u \in \mathbb{R}^{n\times 1}; f \in \mathbb{R}^{n\times 1} \tag{3}$$

where blocks K_i, u_i, f_i are blocks associated with ith subdomain denoting its stiffness matrix, vector of deformation parameters of nodes of the subdomain, and vector of external loading concentrated into the subdomains nodes.

Equality conditions: $Bu = o$, $B \in \mathbb{R}^{m\times n}$ ensure the continuity of node displacements by gluing its subdomains on their interfaces. The Dirichlet boundary conditions (BCs) are prescribed by modifying K and f in corresponding columns and rows (FETI-1), or by adding Dirichlet BCs to the problem constraints expressed by $Bu = o$ (TFETI-1). The constraints $Bu = o$ are then enforced by the vector of Lagrange multipliers λ, where $\lambda \in \mathbb{R}^{m\times 1}$.

Problem (1) can be expressed as the following saddle-point problem:

$$\begin{bmatrix} K & B^T \\ B & O \end{bmatrix} \begin{bmatrix} u \\ \lambda \end{bmatrix} = \begin{bmatrix} f \\ o \end{bmatrix}. \tag{4}$$

The vector of solution u can be expressed from the first equation in (4)

$$u = u_{\text{Im}K} + u_{\text{Ker}K} = K^+(f - B^T\lambda) + R\alpha, R \in \mathbb{R}^{n\times r}, \alpha \in \mathbb{R}^{r\times 1} \tag{5}$$

where K^+ is some form of a generalized inverse of K, and R is the matrix whose columns are the basis of Ker$K$, so it should also hold: $R^T(f - B^T\lambda) = o$.

Dualizing this problem and using the standard FETI notation [2]:

$$F = BK^+B^T; d = BK^+f; G = -R^TB^T; e = -R^Tf, \tag{6}$$

the following problem is obtained:

$$\begin{bmatrix} F & G^T \\ G & O \end{bmatrix} \begin{bmatrix} \lambda \\ \alpha \end{bmatrix} = \begin{bmatrix} d \\ e \end{bmatrix}. \tag{7}$$

After homogenizing: $G\lambda = e$ using $\lambda_0 = G^T(GG^T)^{-1}e, \lambda_0 \in \text{Im}G^T$, the remaining part of λ is μ in Ker$G$, and the following minimization problem is obtained [2]:

$$\min\,(1/2)\,\mu^T F\mu - \mu^T(d - F\lambda_0) \text{ s.t. } G\mu = o. \tag{8}$$

The equality constraint can be enforced by dual penalty or more efficiently by the orthogonal projector P onto Ker G, $P \in \mathbb{R}^{m\times m}$ [2]:

$$P = I - G^T(GG^T)^{-1}G, \tag{9}$$

so that the minimization problem is equivalent to the problem of finding the solution $\bar{x}$ of the system of linear equations

$$Ax = b; A = PFP; x = \mu; b = P(d - F\lambda_0), \tag{10}$$

which is solved iteratively, typically by the conjugate gradients (CG).

The primal solution $\bar{u}$ can be reconstructed as follows: [2]

$$\overline{\alpha} = (GG^T)^{-1}G\,(d - F(\lambda_0 + \overline{x}));\ \overline{u} = K^+(f - B^T(\lambda_0 + \overline{x})) + R\overline{\alpha}. \tag{11}$$

3. Preconditiong and preconditioned conjugate gradients (PCG) method

There exist two basic FETI preconditioners for FETI-1 and TFETI-1, both approximating the inverse of the matrix F. To assemble these preconditioners, the stiffness matrix K has to be divided into 4 blocks [3]:

$$K = \begin{bmatrix} K_{ii} & K_{ib} \\ K_{ib}^T & K_{bb} \end{bmatrix}, \tag{12}$$

where K_{ii}, and K_{bb} are composed of elements of K associated with subdomains' internal, respectively boundary nodes, etc. Likewise, the gluing matrix B should be divided into blocks B_i and B_b: $B = \begin{bmatrix} B_i & B_b \end{bmatrix}$. The block B_i, associated with internal nodes of subdomains, is always a zero matrix, since the conditions of equality of the deformation parameters, respectively Dirichlet BCs of the problem, are expressed only between or for boundary nodes of the subdomains.

3.1. Dirichlet preconditioner (DP)

The Dirichlet preconditioner is expressed as follows [3]:

$$F_I^{D^{-1}} = \begin{bmatrix} B_i & B_b \end{bmatrix} \begin{bmatrix} O & O \\ O & S_{bb} \end{bmatrix} \begin{bmatrix} B_i^T \\ B_b^T \end{bmatrix} = B_b S_{bb} B_b^T,\ S_{bb} = K_{bb} - K_{ib}^T K_{ii}^{-1} K_{ib}, \tag{13}$$

where S_{bb} is the Schur complement of the block K_{ii}.

3.2. Lumped preconditioner (LP)

The matrix of the lumped preconditioner is an approximation of the Dirichlet one with only the first term in the relation for computation of S_{bb} used [3]:

$$F_I^{L^{-1}} = \begin{bmatrix} B_i & B_b \end{bmatrix} \begin{bmatrix} O & O \\ O & K_{bb} \end{bmatrix} \begin{bmatrix} B_i^T \\ B_b^T \end{bmatrix} = \begin{bmatrix} B_i & B_b \end{bmatrix} \begin{bmatrix} K_{ii} & K_{ib} \\ K_{ib}^T & K_{bb} \end{bmatrix} \begin{bmatrix} B_i^T \\ B_b^T \end{bmatrix} = BKB^T. \tag{14}$$

The lumped preconditioner is less accurate and less optimal approximation of the inverse of F, so its effect as a preconditioner on improving the spectral properties of the system matrix and reducing the number of PCG iterations is smaller than with the Dirichlet preconditioner, but its computation is significantly cheaper [3].

4. Deflation and deflated conjugate gradient (DCG) method

When solving the system of equations $Ax = b$ using the CG method, the kth approximation x_k of the solution vector is found as the minimizer of quadratic function $f(x) = \frac{1}{2}x^T Ax - x^T b$ over the kth Krylov subspace $\mathcal{K}_k(A, r_0)$.

The basic idea of the DCG method is to enrich the Krylov subspace $\mathcal{K}_k$ by some subspace $\mathcal{W}$, the so-called deflation subspace. If $\mathcal{W}$ is defined conveniently, a faster convergence of the CG method, solving the system, can be anticipated [4], [5].

Let the subspace $\mathcal{W}$ be spanned by column vectors w_j forming the matrix W:

$$W = \begin{bmatrix} w_1 & \dots & w_j & \dots & w_m \end{bmatrix} \tag{15}$$

then the projector P_D on A-conjugate complement of the deflation subspace $\mathcal{W}$ can be formulated as follows [4], [5]:

$$P_D = I - QA = I - W(W^TAW)^{-1}W^TA. \tag{16}$$

In the DCG method, the process of a solution can be split into 2 parts: solution on the deflation subspace $\mathcal{W}$ and solution on its A-conjugate complement. It is achieved using the fact that in a classical CG method it holds that the vector r_k of the residual in the kth iteration is orthogonal to kth Krylov subspace $\mathcal{K}_k(A, r_0)$, over which the quadratic functional $f(x)$ is minimized in the kth iteration [4], [5].

If some arbitrary initial guess x_{-1} is given, then the corresponding vector of residual is $r_{-1} = b - Ax_{-1}$, and the correction x_0 of the initial guess in the deflation subspace $\mathcal{W}$ is then computed as follows [4], [5]:

$$x_0 = x_{-1} + Qr_{-1} = x_{-1} + W(W^TAW)^{-1}W^Tr_{-1}. \tag{17}$$

If the last equation is multiplied by W^TA from the left, then

$$W^TAx_0 = W^TAx_{-1} + W^TAW(W^TAW)^{-1}W^T(b - Ax_{-1}) \tag{18}$$

$$W^Tb - W^TAx_0 = W^Tr_o = o, \tag{19}$$

so that the vector r_0 corresponding to x_0 satisfies the condition of its orthogonality to $\mathcal{W}$, i.e., it has no components in $\mathcal{W}$, and thus x_0 is the exact solution in $\mathcal{W}$.

If columns of the deflation matrix W are exact eigenvectors of the system matrix A computed in exact arithmetics, it holds: $W^TA = \Lambda W^T$, where Λ is a diagonal matrix with eigenvalues of A, with the kth entry corresponding to the kth column of W, i.e., to the kth of the chosen eigenvectors. Thus, in such case also the kth Krylov subspace $\mathcal{K}_k(A, r_0)$ is orthogonal to $\mathcal{W}$ since $W^TA^{k-1} = \Lambda^{k-1}W^T$.

Since the residual r_k in the kth iteration of the CG method belongs to $\mathcal{K}_{k+1}(A, r_0)$, $k = 0, 1, \ldots$, then r_k is orthogonal to and thus has no components in $\mathcal{W}$.

However, since the computations in reality cannot be performed in exact arithmetic and $\mathcal{W}$ generally does not consist of exact eigenvectors of A, the residual r_k is not A-conjugate to $\mathcal{W}$, in general. Thus, conjugate directions p_k are generally not A-conjugate to $\mathcal{W}$ since in standard CG it holds: $p_{k+1} = r_{k+1} + \beta_{k+1}p_k$, which is a problem since the approximations x_k of a solution are searched only on the A-conjugate complement of $\mathcal{W}$. Thus, some sort of correction has to be performed in the iteration process to make the vectors p_k A-conjugate to $\mathcal{W}$, so that the approximations x_k are searched only in the A-conjugate complement of $\mathcal{W}$.

The correction is performed in a way that in relation used to compute the vector p_k in the standard CG method, the vector of residual r_k is projected onto the A-conjugate complement of $\mathcal{W}$, by multiplying it by the projector P_D defined in (16), so the relations for computing p_0 and p_{k+1}, $k = 0, 1, \ldots$ get the following form:

$$p_0 = P_D\, r_0;\; p_{k+1} = P_D\, r_{k+1} + \beta_{k+1}p_k. \tag{20}$$

By ensuring that the approximations x_k of the solution during the iterative process of DCG are searched only on the A-conjugate complement of the deflation subspace $\mathcal{W}$, the required splitting of the solution into components on $\mathcal{W}$ (in the form of correction of the initial guess) and on its A-conjugate complement is achieved. The CG, PCG, and DCG algorithms are presented in Table 1.

CG	PCG	DCG
Input : $A, b, x_0, k = 0$	Input : $A, b, x_0, M^{-1}, k = 0$	Input : $A, b, x_{-1}, W, k = 0$
		$Q = W(W^TAW)^{-1}W^T$
		$P_D = I - QA$
		$r_{-1} = b - Ax_{-1}$
		$x_0 = x_{-1} + Qr_{-1}$
$r_0 = b - Ax_0$	$r_0 = b - Ax_0$	$r_0 = b - Ax_0$
	$z_0 = M^{-1}r_0$	
$p_0 = r_0$	$p_0 = z_0$	$p_0 = P_D\, r_0$
while (some ending criterium)	while (some ending criterium)	while (some ending criterium)
$s = Ap_k$	$s = Ap_k$	$s = Ap_k$
$\alpha_k = (r_k^T r_k)/(s^T p_k)$	$\alpha_k = (r_k^T z_k)/(s^T p_k)$	$\alpha_k = (r_k^T r_k)/(s^T p_k)$
$x_{k+1} = x_k + \alpha_k p_k$	$x_{k+1} = x_k + \alpha_k p_k$	$x_{k+1} = x_k + \alpha_k p_k$
$r_{k+1} = r_k - \alpha_k s$	$r_{k+1} = r_k - \alpha_k s$	$r_{k+1} = r_k - \alpha_k s$
	$z_{k+1} = M^{-1}r_{k+1}$	
$\beta_{k+1} = (r_{k+1}^T r_{k+1})/(r_k^T r_k)$	$\beta_{k+1} = (r_{k+1}^T z_{k+1})/(r_k^T z_k)$	$\beta_{k+1} = (r_{k+1}^T r_{k+1})/(r_k^T r_k)$
$p_{k+1} = r_{k+1} + \beta_{k+1}p_k$	$p_{k+1} = z_{k+1} + \beta_{k+1}p_k$	$p_{k+1} = P_D\, r_{k+1} + \beta_{k+1}p_k$
Output : x_k	Output : x_k	Output : x_k

Table 1: CG, PCG and DCG algorithms

5. (T)FETI-2 – deflated variant of (T)FETI-1

In this section, it is considered that deflation is applied on the CG method solving the final system of equations obtained by decomposition of the FEM-discretized problem of 2D linear elasticity by (T)FETI-1. It is also presumed that both the discretization and decomposition of the analyzed linear elastic domain are conforming.

5.1. Deflation by equality of displacements in corner nodes (CE)

Equation $B_C u = o$ expresses the equality conditions of the corresponding displacement components of mutually corresponding corner nodes on the interfaces of neighbouring subdomains in two perpendicular directions x and y.

Since conditions $B_C u = o$ are already included in conditions $Bu = o$ using the matrix B, the matrix B_C can be obtained by splitting B into two parts as follows $B = \begin{bmatrix} B_C^T & B_R^T \end{bmatrix}^T$, with B_R expressing the equality conditions of displacements of the remaining nodes by $B_R u = o$, which are not in corners of subdomains.

The deflation matrix W is: $W = BB_C^T = \begin{bmatrix} B_C^T & B_R^T \end{bmatrix}^T B_C^T = \begin{bmatrix} B_C B_C^T & B_C B_R^T \end{bmatrix}^T = \begin{bmatrix} B_C B_C^T & O^T \end{bmatrix}^T$, where in case of orthonormal rows of B it holds: $W = \begin{bmatrix} I & O^T \end{bmatrix}^T$.

5.2. Deflation by equality of the displacement averages and by moment equilibrium of gluing forces on subdomains' interfaces

In this method of defining the deflation subspace $\mathcal{W}$, at first the matrix B_A has to be defined. This matrix will be divided into two vertical blocks B_{A_A} and B_{A_M}, i.e. $B_A = \begin{bmatrix} B_{A\_A}^T & B_{A\_M}^T \end{bmatrix}^T$ for purposes of following formulations.

The block B_{A_A} in the relation $B_{A\_A}u = o$ expresses the conditions enforcing the equality of the averages of values of displacement components of each node along opposite sides of each corresponding interface of 2 subdomains.

The block B_{A_M} is used to express the conditions of moment equilibrium of the force system of solitary forces of contributions of the notional total gluing force, acting on the interfaces of two subdomains, distributed continuously and uniformly along their length, as contributions concentrated in each corresponding node on either side of the interface, with the total force distributed uniformly (averaged) among the contributions, in terms of their magnitude and direction. The moments of the forces of contributions concentrated in corresponding nodes are all related to the same reference point, here always the point with the global coordinates [0,0].

The moment of the corresponding component, in direction of axis x, or y, of the solitary force of corresponding contribution of the total gluing force, concentrated into the node on interface, denoted as 12, of two subdomains: 1 and 2, related to the point [0,0] is equal to the corresponding element of the vector of the product of transpose of the corresponding row of matrix B_{A_M}, with the corresponding element, denoted e.g. as $\lambda_{A\_M_{12}}$ of the vector λ_{A_M}, where sum of all the elements of the product equals zero, i.e., the moment equilibrium of the force system is ensured.

The deflation matrix W is computed as $W = BB_A^T$.

5.2.1. Conditions of averages equality (AE)

The formulation of the conditions of equality of averages of displacement components in two directions x and y, of nodes lying along opposite sides of the interface, denoted as 12, of two subdomains: 1 and 2, of decomposed domain, is following:

$$x : \frac{1}{n}\sum_{k=1}^{n} u_{1_{k\_x}} - u_{2_{k\_x}} = 0, y : \frac{1}{n}\sum_{k=1}^{n} u_{1_{k\_y}} - u_{2_{k\_y}} = 0, \tag{21}$$

where n is the number of nodes on side of the interface and $u_{1(2)_{k\_x(y)}}$ is the displacement component of the k-th node in the $x(y)$ direction. The structure of the two corresponding rows of the B_{A_A} matrix expressing these two conditions is then:

$$\frac{1}{n}\begin{bmatrix} & 1_{1\_x} & 1_{1\_y} & \cdots & 1_{k\_x} & 1_{k\_y} & \cdots & 1_{n\_x} & 1_{n\_y} & & 2_{1\_x} & 2_{1\_y} & \cdots & 2_{k\_x} & 2_{k\_y} & \cdots & 2_{n\_x} & 2_{n\_y} & \\ O & 1 & 0 & \cdots & 1 & 0 & \cdots & 1 & 0 & O & -1 & 0 & \cdots & -1 & 0 & \cdots & -1 & 0 & O \\ O & 0 & 1 & \cdots & 0 & 1 & \cdots & 0 & 1 & O & 0 & -1 & \cdots & 0 & -1 & \cdots & 0 & -1 & O \end{bmatrix} \tag{22}$$

The conditions of equality of displacement averages on the interface are implicitly also the conditions of equilibrium of the force system of solitary forces of discrete contributions of the total gluing force into corresponding nodes on that interface.

5.2.2. Conditions of moment equilibrium (ME)

The condition of moment equilibrium of the discrete contributions of the total gluing force acting along entire length of the interface 12 of two subdomains: 1 and 2, concentrated and uniformly distributed in each node on either corresponding side of the interface, with moments all related to the reference point [0,0], is enforced using the corresponding row of B_{A_M} of the following structure:

$$\begin{bmatrix} & 1_{1\_x} & 1_{1\_y} & \cdots & 1_{k\_x} & 1_{k\_y} & \cdots & 1_{n\_x} & 1_{n\_y} & & 2_{1\_x} & 2_{1\_y} & \cdots & 2_{k\_x} & 2_{k\_y} & \cdots & 2_{n\_x} & 2_{n\_y} & \\ O & -y_1 & x_1 & \cdots & -y_k & x_k & \cdots & -y_n & x_n & O & y_1 & -x_1 & \cdots & y_k & -x_k & \cdots & y_n & -x_n & O \end{bmatrix}, \quad (23)$$

where, $x(y)_1(k,n)$ is the $x(y)$-coordinate, in the global coordinate system, belonging to the first (k-th, last) node on each side of the interface.

5.3. Deflation by the eigenvectors of the system matrix (EIG)

As it was mentioned in Section 4 concerning the DCG method, the deflation matrix W should be in an ideal case composed of the (exact) eigenvectors of the system matrix A. If the columns W are exact eigenvectors of the system matrix A computed in exact arithmetic, then $W^T A = \Lambda_M W^T$, where Λ_M is a diagonal matrix with eigenvalues of A, where the kth diagonal entry (kth eigenvalue) corresponds to the kth column of W.

To achieve the desired effect of deflation in significantly improving the spectral properties of the spectral operator $P_D A$ in the iterative process of the DCG method, and thus speeding up convergence of the iterative process, the eigenvectors of the matrix A that slow down convergence the most should be deflated, which are usually those corresponding to the extremal, usually the lowest, eigenvalues of A. If the eigenvectors of A, which form W, are favourably selected, then the desired effect of deflation can be reached for a relatively small number of eigenvectors of A, which leads to a small matrix $W^T AW$ of the coarse problem (CP) in the DCG method and thus to a computationally cheap solution of CP [4], [5].

However, the process of obtaining the eigenvalues and eigenvectors is generally very costly, and thus the solution of the system of linear algebraic equations $Ax = b$ by the CG method with a good preconditioner is often faster, in terms of the total time needed for the assembly of the preconditioner, or the deflation matrix, and the subsequent solution of the system using the PCG, or DCG method.

5.4. Deflation by discrete wavelet transform (DWT)

In the following text, it is considered that the discrete Haar wavelet, as the structurally simplest, is applied during the DWT. The Haar wavelet has two filters, the "low-frequency" and "high-frequency", which are used to obtain the components of some signal corresponding to the low/high frequencies.

The process of splitting 1D signal, represented by vector x, into its low- and high-frequency components, is in kth level of forward DWT represented by decomposition of the vector a_{k-1}, with $a_0 = x$, lying in so-called $(k-1)$th discretized scaling subspace V_{k-1}, on its so-called approximation coefficients a_k (corresponding to the lower frequencies), lying in kth discretized scaling subspace V_k, and detailed coefficients d_k (higher frequencies), in so-called kth discretized wavelet subspace W_k as orthogonal complement of V_k in V_{k-1}, is carried out by the gradual application of corresponding orthogonal projectors H_k (from V_{k-1} onto V_k) and G_k (from V_{k-1} onto W_k).

This means that in the k-th level, where $k = 1, \ldots, M$, with M being the given chosen total number of levels of DWT performed, it holds:

$$a_k = H_k a_{k-1},\ d_k = G_1 a_{k-1},\ \begin{bmatrix} a_k \\ d_k \end{bmatrix} = \begin{bmatrix} H_k \\ G_k \end{bmatrix} a_{k-1}. \tag{24}$$

Thus the vector a_M obtained after M compressions (M levels of forward DWT) applied on the vector x, of length equal to $N/2^M$, or to its closest higher or lower integer to $N/2^M$, of original signal x (of length N) can be expressed using the matrix H of the total projector from the space $V_0 = l^2(N)$ to the V_M in the following way:

$$a_M = H_M H_{M-1} \dots H_k \dots H_2 H_1 x = Hx. \tag{25}$$

The inverse DWT of a_M is then given by multiplication of a_M by transpose of H, i.e. by H^T, where the vector obtained by applying M levels of inverse DWT using H^T on the vector obtained by M levels of DWT on x using matrix H, only the trend part of the signal represented by x is preserved.

If DWT is applied to square matrices A in order to obtain only its components corresponding to its lower eigenvalues, the projector H is applied to the columns and its transpose H^T on transformed rows, so that the matrix obtained by 2D FDWT of A is a matrix $A_T = HAH^T$. The deflation matrix W is obtained as $W = H^T$.

The matrix H_k of the orthogonal projector from V_{k-1} onto V_k has structure:

$$H_k = \frac{1}{\sqrt{2}} \begin{bmatrix} \ddots & 1 & 1 & 0 & 0 & \\ & 0 & 0 & 1 & 1 & \ddots \end{bmatrix}. \tag{26}$$

The vector (matrix) on which the projector H_k, without modification, is applied at the kth level of DWT must have the length (dimensions) divisible by 2. If it does not hold, then some adjustment of the structure of H_k has to be performed; see [6].

In numerical experiments, the case where the structure of matrix of orthogonal projector H_k was adjusted with regard to the fact that 2D decomposed discretized problem of linear elasticity is solved in a following way (27), was also tested:

$$H_{k,2D} = \frac{1}{\sqrt{2}} \begin{bmatrix} \ddots & 1 & 0 & 1 & 0 & \\ & 0 & 1 & 0 & 1 & \ddots \end{bmatrix}, \tag{27}$$

so that $H_{2D} = H_{M,2D} H_{M-1,2D} \dots H_{1,2D}$ and $W_{2D} = H_{2D}^T$.

5.5. Deflation by discrete Fourier transform (DFT)

DFT of some vector $x \in l^2(N)$ is in fact the computation of its coordinate vector c in complex orthonormal discrete Fourier basis of the vector space $l^2(N)$. The kth component c_k, of the vector c as the DFT of the vector x can be computed as the complex inner product of x, with the kth Fourier basis vector having the structure:

$$F_k = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & (e^{2\pi i/N})^k & \cdots & (e^{2\pi i/N})^{nk} & \cdots & (e^{2\pi i/N})^{(N-1)k} \end{bmatrix}^T. \tag{28}$$

The deflation matrix W is then composed of the first M vectors F_k, of a discrete Fourier basis, where $k = 0, 1, \dots, M-1$ as follows:

$$W = \begin{bmatrix} F_0 & F_1 & \cdots & F_k & \cdots & F_{M-1} \end{bmatrix}. \tag{29}$$

The deflation matrix W can again have its structure adjusted with regard to solution of 2D elasticity problem, where the block $F_{k,2D}$ of the deflation matrix W, replacing the kth discrete Fourier basis vector, is constructed using kth discrete Fourier basis vector of the vector space $l^2(N/2)$, and it has the following structure:

$$F_{k,2D} = \frac{2}{\sqrt{N}} \begin{bmatrix} 1 & 0 & \cdots & e^{\frac{2\pi i(nk)}{N/2}} & 0 & \cdots & e^{\frac{2\pi i(N/2-1)k}{N/2}} & 0 \\ 0 & 1 & \cdots & 0 & e^{\frac{2\pi i(nk)}{N/2}} & \cdots & 0 & e^{\frac{2\pi i(N/2-1)k}{N/2}} \end{bmatrix}^T . \quad (30)$$

The deflation matrix, with structure adjusted with regard to solving the 2D decomposed discretized problem of linear elasticity, denoted as W_{2D}, is defined as follows:

$$W_{2D} = \left[\begin{array}{cccccc} F_{0,2D} & F_{1,2D} & \cdots & F_{k,2D} & \cdots & F_{M-1,2D} \end{array} \right]. \quad (31)$$

5.6. Deflation by discrete cosine transform (DCT)

DCT works on similar principle as DFT, only the complex discrete Fourier basis is replaced by real discrete cosine basis, whose kth vector has following structure:

$$C_k = \sqrt{\frac{2-\delta_{k,0}}{N}} \left[\cos\frac{(1/2)k\pi}{N} \ \cdots \ \cos\frac{(n+1/2)k\pi}{N} \ \cdots \ \cos\frac{(N-1+1/2)k\pi}{N} \right]^T, \quad (32)$$

and the deflation matrix W is then composed of the first M vectors of this discrete cosine basis C_k, $k = 0, 1, ..., M-1$:

$$W = \left[\begin{array}{cccccc} C_0 & C_1 & \cdots & C_k & \cdots & C_{M-1} \end{array} \right]. \quad (33)$$

The structure of the deflation matrix W, respectively of the vectors C_k as the columns of W can be again adjusted with regard to solving 2D problem of elasticity

$$C_{k,2D} = \sqrt{\frac{2-\delta_{k,0}}{N/2}} \begin{bmatrix} \cdots & \cos\frac{(n+1/2)k\pi}{N/2} & 0 & \cdots \\ \cdots & 0 & \cos\frac{(n+1/2)k\pi}{N/2} & \cdots \end{bmatrix}^T, \ n = 0, 1, \ldots, \frac{N}{2}-1, \quad (34)$$

resulting in W_{2D} with the structure:

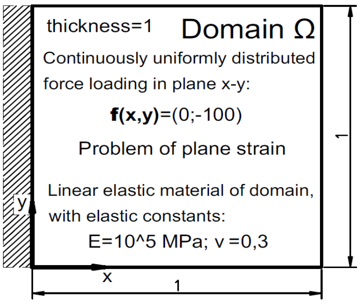
$$W_{2D} = \left[\begin{array}{cccccc} C_{0,2D} & C_{1,2D} & \cdots & C_{k,2D} & \cdots & C_{M-1,2D} \end{array} \right]. \quad (35)$$

6. Numerical experiments

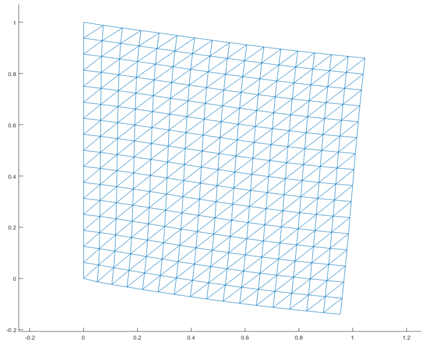
As the benchmark problem used in numerical experiments, model 2D linear elasticity problem, see Fig. 1, defined on 2D linear elastic domain, decomposed into $(1/H)^2$ identical square non-overlapping subdomains with edge lengths $H = 1/2, 1/4, 1/8, 1/16$, was chosen. The subdomains were discretized by 2D identical finite elements of shape of isoscleses right-angled triangle with length h. There are presented the results of numerical experiments only for value of the ratio $H/h = 8$.

All algorithms were implemented and numerical experiments were performed in Matlab, as the stopping criteria was used: $||r_k|| \leq \epsilon||b||$, where always: $\epsilon = 10^{-6}$, and the initial guess x_0 was always zero vector.

Table 2 presents the corresponding dimensions of the problems for given H using FETI-1/TFETI-1. The dimensions of the deflation subspace $\mathcal{W}$ for each tested variant of deflation are presented in Table 3 .



(a) Definition of the problem



(b) Solution of the problem

Figure 1: The benchmark: 2D problem of linear elasticity

H	1/2	1/4	1/8	1/16
Primal dimension	648	2592	10368	41472
Dual dimension (dim(A))	70/104	414/480	1918/2048	8190/8448
dim(Ker$K$) = (dim(Ker$A$)	6/12	36/48	168/192	720/768
rank(A) = dim(A) − dim(Ker$A$)	64/92	378/432	1750/1856	7470/7680

Table 2: Problems dimensions ($A = PFP$, $H/h = 8$)

	H	1/2	1/4	1/8	1/16
FETI-2	CE1/CE2	6/14	54/78	294/350	1350/1470
	AE/AE+ME	8/12	48/72	224/336	960/1440
	EIG	1/2/4/.../32/64	1/2/.../256/378	1/2/.../1024/1750	1/2/.../2048/4096
	DWT1	36/18/10/6	208/104/52/26	960/480/240/120	4096/2048/1024/512
	DWT2	35/18/9/5	207/104/52/26	959/480/240/120	4095/2048/1024
	DFT1	2/4/8/.../32/64	2/4/.../256/378	2/4/.../1024/1734	2/4/.../2048/4096
	DFT2	1/2/4/.../32/64	1/2/.../256/376	1/2/.../1024/1728	1/2/.../2048/4096
	DCT1	2/4/8/.../32/64	2/4/.../256/378	2/4/.../1024/1738	2/4/.../2048/4096
	DCT2	1/2/4/.../32/64	1/2/.../256/376	1/2/.../1024/1730	1/2/.../2048/4096
TFETI-2	CE1/CE2/CE3	6/14/16	54/78/84	294/350/364	1350/1470/1500
	AE/AE+ME	8/12	48/72	224/336	960/1440
	EIG	1/2/4/.../64/92	1/2/.../256/432	1/2/.../1024/1856	1/2/.../2048/4096
	DWT1	52/26/14/8	240/120/60/30	1024/512/256/128	4224/2112/1056/528
	DWT2	52/26/13/7	240/120/60/30	1024/512/256/128	4224/2112/1056/528
	DFT1	2/4/8/.../64/92	2/4/.../256/432	2/4/.../1024/1854	2/4/.../2048/4096
	DFT2	1/2/4/.../64/92	1/2/.../256/432	1/2/.../1024/1854	1/2/.../2048/4096
	DCT1	2/4/8/.../64/92	2/4/.../256/432	2/4/.../1024/1854	2/4/.../2048/4096
	DCT2	1/2/4/.../64/92	1/2/.../256/432	1/2/.../1024/1855	1/2/.../2048/4096

Table 3: Dimensions of the deflation subspaces

The numbers of performed iterations of (P)CG method with no, lumped and Dirichlet preconditioners, and of the DCG method for each tested variant of deflation, solving the system of equations obtained by (T)FETI-1, are depicted in Table 4.

	H	1/2	1/4	1/8	1/16
FETI-1 (NO/LP/DP)		23/14/8	37/20/13	45/24/17	56/29/25
TFETI-1 (NO/LP/DP)		25/14/8	34/16/8	34/16/11	33/16/11
FETI-2	CE1/CE2	18/15	22/20	25/24	26/26
	AE/AE+ME	16/14	24/20	25/21	26/22
	EIG	28/23/19/ 16/12/7/0	46/46/47/32/27/ 22/15/9/6/0	56/56/56/58/44/29/ 27/22/15/10/6/0	72/70/70/70/62/52/ 29/28/27/22/16/11/7
	DWT1	25/21/13/9	53/36/19/11	67/49/27/15	89/61/34/16
	DWT2	22/18/15/–	38/29/26/–	49/36/32/–	62/48/35/–
	DFT1	27/26/19/ 15/13/0	62/5954/2262/1079 /215/49/23/16/0	74/7186/–/8522/9267 /–/2285/148/46/23/5	95/–/–/–/–/–/–/ –/–/231/61/33
	DFT2	28/1029/640 /57/24/17/0	58/5183/–/6069/5471 /3946/335/54/14/2	67/7443/7904/8719/9251/ 9190/8404/–/1193/134/25/7	84/7788/9757/7437/9456/ 9979/–/–/–/–/2799/308/33
	DCT1	29/23/19/ 13/8/0	59/80/102/66/ 44/29/16/8/0	74/98/213/221/236/ 238/116/50/25/13/4	93/140/231/242/238/280 /289/336/343/70/33/15
	DCT2	27/34/32/ 23/19/15/0	57/67/114/103/110 /75/49/30/19/2	66/85/148/148/156/155 /165/191/102/47/24/7	87/108/155/165/165/182/ 198/211/231/261/169/69/30
TFETI-2	CE1/CE2/CE3	23/19/18	26/22/20	28/24/20	28/25/20
	AE/AE+ME	22/21	30/27	31/28	31/29
	EIG	25/24/22/19 /16/11/7/0	33/33/32/31/29/ 24/17/12/8/0	34/34/33/33/33/32/ 30/25/17/12/8/0	33/33/33/33/33/33/ 33/32/30/25/18/13/8
	DWT1	22/19/15/10	31/25/17/11	32/26/18/11	32/28/18/11
	DWT2	22/19/17/–	29/26/25/–	31/28/27/–	31/29/27/–
	DFT1	25/25/21/ 19/16/13/0	33/34/33/33/ 31/23/21/19/0	34/34/34/34/34/ 34/33/24/22/20/2	34/34/34/34/34/34/ 34/34/33/28/23/22
	DFT2	25/24/23/22 /21/18/10/0	33/33/33/33/32/ 29/29/27/16/0	34/34/34/34/34/34/ 33/30/30/30/20/2	33/33/33/33/33/33/ 33/33/33/30/30/30/25
	DCT1	25/23/21/ 18/12/7/0	33/33/33/32/ 30/22/14/9/0	34/34/34/33/33/ 33/32/25/15/10/2	34/33/34/34/34/34/ 33/33/33/29/18/10
	DCT2	25/24/23/21 /18/17/12/0	33/33/33/33/31/ 28/25/24/18/0	34/34/34/34/33/33/ 33/30/27/25/22/1	33/33/33/33/33/33/ 33/33/33/31/29/26/22

Table 4: Number of performed iterations of the (P)CG and DCG methods

In numerical experiments the following variants of deflation were tested:

- CE1 (CE with displacement equality conditions between corner nodes on the boundary of the domain not included in W), CE2 (CE with conditions between corner nodes on the domain boundary included), CE3 (only TFETI-2 – CE2 + Dirichlet BCs assigned in corner nodes on the domain boundary),
- AE (with displacement components of no corner nodes on the interface included into the computation of averages on the interface), AE+ME ((T)FETI-2 – with no corner nodes of interface included into averages, solitary forces of contributions of total gluing force concentrated into the corner nodes on the domain boundary in case of FETI-2, inside the domain in case of TFETI-2, not included)
- EIG1 (deflation by a given number of eigenvectors of the system matrix A),
- DWT1 and DWT2 (4/3/2/1 levels of 2D DWT applied on A with and without the modification of W with regard to solving 2D problem of elastictity),

- DFT1, DFT2 (deflation by first M vectors of discrete Fourier basis with and without the modification of W with regard to solving 2D problem of elastictity),
- DCT1, DCT2 (deflation by first M vectors of discrete cosine basis with and without the modification of W with regard to solving 2D problem of elastictity),

7. Conclusion

This paper provides experimental evidence of an effect of standard FETI-1 and TFETI-1 preconditioners and various types of deflation resulting in FETI-2 and TFETI-2 variants for a model 2D linear elasticity problem. This effect considering the numbers of iterations should always be taken into account with its costs. A detailed analysis in parallel environment is work in progress.

It should be mentioned that the benchmark 2D plane strain linear elasticity problem discretized by FEM on which the numerical experiments were performed was well-conditioned and thus the effect of the deflation was not that significant. If deflation were applied, for example, to a decomposed problem of linear elasticity with plates or shells, or to a decomposed problem without dualization, the effect of the deflation would be even more considerable. A more significant effect of deflation could also appear in the case of nonconforming and irregular subdomains' meshes resulting in a worse conditioned system matrix.

Acknowledgements

This work was supported by the SGS grant No. SP2024/067 of VSB-Technical University of Ostrava.

References

[1] Dostál, Z., Horák, D. and Kučera, R.: Total FETI - an easier implementable variant of the FETI method for numerical solution of elliptic PDE Commun. Numer. Meth. Engng **22** (2006), 1155–1162.

[2] Dostál, Z.: Rozděl a slep aneb jak řešit soustavu s bilionem lineárních rovnic. Pokroky matematiky, fyziky a astronomie **63** (2018), 28–40.

[3] Farhat, C., Mandel, J., and Roux, F.X.: Optimal convergence properties of the FETI domain decomposition method. Comput. Methods Appl. Mech. Engrg. **115** (1994), 365–385.

[4] Kružík, J.: *Implementation of the deflated variants of the conjugate gradient method*. Diploma thesis, VSB - Technical University of Ostrava, 2018.

[5] Saad, Y., Yeung, M., Erhel, J., and Guyomarc'H, F.: A Deflated Version of the Conjugate Gradient Algorithm. SIAM J. Sci. Comput., **21** (2000), 1909–1926.

[6] Taswell, C., McGill, K.C., Algorithm 735: Wavelet transform algorithms for finite duration discrete-time signals ACM Trans. Math. Softw. **20** (1994), 398–412.