Numerical Linear Algebra with Applications, Vol. 3(5), 427–453 (1996)

Multilevel Preconditioners for Mixed Methods for Second Order Elliptic Problems

Zhangxin Chen

Department of Mathematics, Southern Methodist University, Dallas, TX 75275-0156, USA

and

Richard E. Ewing, Raytcho D. Lazarov and Serguei Maliassov

Institute for Scientific Computation and Department of Mathematics, Texas A&M University, 633 Blocker Bldg., College Station, TX 77843–3404, USA

and

Yuri A. Kuznetsov

Institute of Numerical Mathematics, Russian Academy of Sciences, 32-a Leninski Prospect, 117334 Moscow, Russia

A new approach for constructing algebraic multilevel preconditioners for mixed finite element methods for second order elliptic problems with tensor coefficients on general geometry is proposed. The linear system arising from the mixed methods is first algebraically condensed to a symmetric, positive definite system for Lagrange multipliers, which corresponds to a linear system generated by standard nonconforming finite element methods. Algebraic multilevel preconditioners for this system are then constructed based on a triangulation of the domain into tetrahedral substructures. Explicit estimates of condition numbers and simple computational schemes are established for the constructed preconditioners. Finally, numerical results for the mixed finite element methods are presented to illustrate the present theory.

KEY WORDS mixed method; nonconforming method; multilevel preconditioner; condition number; second order elliptic problem

1. Introduction

We consider the elliptic problem

$$\begin{aligned} \nabla \cdot \sigma = f, \quad \sigma = -a\nabla u \qquad & \text{in } \Omega \\ u = 0 \qquad & \text{on } \partial\Omega \end{aligned} \tag{1.1}$$

CCC 1070–5325/96/050427–27

Received 15 January 1996
Revised 7 May 1996

where $a(x)$ is a uniformly positive definite, bounded, symmetric tensor, $f(x) \in L^2(\Omega)$, and Ω is a bounded domain in $\mathbb{R}^3$ with a polygonal boundary $\partial\Omega$.

Problems of this type arise in various groundwater applications, in particular, in modeling of fluid flows in highly heterogeneous and anisotropic porous media in fairly complex domains Ω. The first equation in (1.1) expresses mass conservation in an arbitrary volume in Ω. This is a fundamental property which is required for any solution method for (1.1) related to groundwater applications. In such applications the transport of various species in the flow is driven by the Darcy velocity σ. Accurate computing of σ is another very desirable feature of a numerical method for solving (1.1). Finite differences with harmonic averaging of the coefficients have been extensively used for such applications. For problems in domains covered by rectangular grids with jumps in $a(x)$, which are aligned with the grid partition, the finite difference method is both accurate and locally conservative. Owing to [32] it is known that this scheme can be obtained from the lowest order mixed finite element method with numerical integration. Now it is accepted that the mixed methods provide a systematic approach for deriving locally conservative and accurate discretizations for problems in general domains with highly inhomogeneous coefficients $a(x)$.

Let $(\cdot, \cdot)_S$ denote the inner product in $L^2(S)$ (we omit S if $S = \Omega$), and let

$$\begin{aligned} V &= H(\mathrm{div}; \Omega) = \left\{ v \in \left(L^2(\Omega)\right)^3 : \nabla \cdot v \in L^2(\Omega) \right\} \\ W &= L^2(\Omega) \end{aligned}$$

Then (1.1) is formulated in the following mixed form for the pair $(\sigma, u) \in V \times W$:

$$\begin{aligned} &(\nabla \cdot \sigma, w) = (f, w), && \forall w \in W \\ &(a^{-1}\sigma, v) - (u, \nabla \cdot v) = 0, && \forall v \in V \end{aligned} \tag{1.2}$$

To define a finite element method, we need a partition $\mathcal{T}_h$ of Ω into elements τ, say, simplices, rectangular parallelepipeds, and/or prisms. We assume that adjacent elements in $\mathcal{T}_h$ completely share their common edge or face; $\partial\mathcal{T}_h$ denotes the set of all *interior* faces e of $\mathcal{T}_h$.

Let $V_h \times W_h \subset V \times W$ be a standard mixed finite element space for second order elliptic problems defined over $\mathcal{T}_h$ (see, e.g., [8,9,10,13,29,30,31]). Then the mixed finite element approximation to (1.1) is to find $(\sigma_h, u_h) \in V_h \times W_h$:

$$\begin{aligned} &(\nabla \cdot \sigma_h, w) = (f, w), && \forall w \in W_h \\ &(a^{-1}\sigma_h, v) - (u_h, \nabla \cdot v) = 0, && \forall v \in V_h \end{aligned} \tag{1.3}$$

Problem (1.2) is a typical saddle-point problem. Its finite element approximation (1.3) represents a linear algebraic system with a symmetric but indefinite matrix. Various preconditioning algorithms for solving saddle-point problems have been proposed and studied in the last decade (see, e.g., [3,5,6,16,18,33,35]). However, their efficiency strongly depends on the geometry of the domain, the coefficient matrix $a(x)$, and the type of finite elements. For example, the method implemented by Ewing and Wheeler [18] eliminates the velocity and reduces system (1.3) to its Schur complement form. This method can be very efficient for rectangular domains and rectangular finite elements. A particular multigrid realization of this method for $a(x)$ being a diagonal tensor is given in [7]. The method of Bramble and Pasciak [5] introduces a new inner product which transforms the indefinite problem (1.3)

to a definite one. This approach combined with preconditioned conjugate gradient iterations can be applied for general equations and finite element spaces. The method studied by Rusten and Winter in [33] (modified and improved in [35]) uses the preconditioned minimal residual iteration method for the indefinite system. The inexact Uzawa method studied in [16,6] is another general iteration method for (1.3). The efficiency of the discussed algorithms in all these papers depends on the choice of the preconditioner. This in turn is a non-trivial problem for highly inhomogeneous media with orthotropy and domains Ω with complex geometry.

The main objective of this paper is construction, study, and testing of optimal (or almost optimal) preconditioners for the algebraic problem (1.3) with symmetric full tensors $a(x)$ and possibly large jumps and anisotropy. Our analysis is restricted to domains which are topologically equivalent to parallelepipeds or are unions of such subdomains. The proposed construction uses the hybridized version of (1.3) and its equivalence to certain modifications of the non-conforming Galerkin approximation to (1.1) (see, e.g., [1,2,11,12,27]). Below we explain our approach in detail.

Since $V_h \subset V$ the vector functions in V_h have normal components which are continuous across the interelement boundaries. Following [2], we relax this constraint on V_h by defining

$$\tilde{V}_h = \{v \in L^2(\Omega) : v|_\tau \in V_h(\tau) \text{ for each } \tau \in \mathcal{T}_h\}$$

where $V_h(\tau) = V_h|_\tau$. Further we shall also use the notation $W_h(\tau) = W_h|_\tau$. In order to impose the continuity of the normal components of the vector function in V_h we use Lagrange multipliers. For this purpose we define the space of the Lagrange multipliers:

$$L_h = \left\{\mu \in L^2\Big(\bigcup_{e \in \partial \mathcal{T}_h} e\Big) : \mu|_e \in V_h \cdot \nu|_e \text{ for each } e \in \partial \mathcal{T}_h\right\}$$

where ν is the unit normal to e. To establish a relationship between the mixed method and the non-conforming Galerkin method, following [11], we introduce the projection of the coefficient tensor $a^{-1}(x)$ on W_h, i.e., $\alpha_h = P_h a^{-1}$, where P_h is the L^2-projection onto W_h. This will lead to a new approximation of (1.2) which we shall show is equivalent to (1.3) with a^{-1} replaced by $P_h a^{-1}$. This modification simplifies the analysis of the method and produces an approximate solution which is $O(h)$ close to the solution of (1.3).

The hybrid form of the mixed finite element method for (1.2) is: find $(\sigma_h, u_h, \lambda_h) \in \tilde{V}_h \times W_h \times L_h$ such that

$$\begin{aligned}
&\sum_{\tau \in \mathcal{T}_h} (\nabla \cdot \sigma_h, w)_\tau = (f, w), && \forall w \in W_h \\
&(\alpha_h \sigma_h, v) - \sum_{\tau \in \mathcal{T}_h} \left[(u_h, \nabla \cdot v)_\tau - (\lambda_h, v \cdot \nu_\tau)_{\partial\tau \setminus \partial\Omega}\right] = 0, && \forall v \in \tilde{V}_h \\
&\sum_{\tau \in \mathcal{T}_h} (\sigma_h \cdot \nu_\tau, \mu)_{\partial\tau \setminus \partial\Omega} = 0, && \forall \mu \in L_h
\end{aligned} \tag{1.4}$$

Note that the last equation in (1.4) enforces the continuity requirement mentioned above, so in fact $\sigma_h \in V_h$. In [2,27] it was shown that in the case of the lowest order Raviart–Thomas elements the solution to (1.4) can be recovered from the Galerkin method which uses non-conforming linear elements augmented with bubble functions. In this paper, following [1,12,14], we show that the linear system generated by (1.4) can be algebraically condensed to a symmetric and positive definite system for the Lagrange multiplier λ_h. It is then shown

that this linear system can be obtained from the Galerkin method for non-conforming linear elements without any bubbles.

It should be pointed out that in many groundwater applications $a(x)$ has large jumps and domain Ω may have fairly complex geometry. A classical general method for preconditioning such problems can be based on incomplete factorization. For theoretical and experimental comparison of a variety of realizations of this approach we refer to [28].

Multilevel/multigrid methods for conforming finite element approximations of various elliptic problems including cases with large jumps and anisotropy (in 2-D) are studied in [4,20,21,25,26,36]. Most of the methods discussed there can be extended to 3-D elliptic problems. Our approach can be viewed as an extension of these methods and algorithms to 3-D mixed and non-conforming finite element approximations.

In this paper the suggested method for solving (1.3) involves two steps: (1) solving the system for the Lagrange multipliers which is shown to be produced by the Galerkin method using linear non-conforming Raviart–Crouzeix finite elements; (2) recovery of the mixed finite element approximations σ_h and u_h from the Lagrange multipliers. Since the second step uses explicit formulas for recovering σ_h and u_h over each element, only the first step is computationally expensive.

The construction of the preconditioner for the Lagrange multipliers system can be summarized as follows. We introduce the partition of the domain into distorted parallelepipeds (called super-elements). This is our coarse level. These super-elements are split into simplices, and the approximation (1.4) is done on this triangulation. The construction of the preconditioner is done in three main stages.

First, we introduce a two-level preconditioner for the system of the Lagrange multipliers which leads to a block 'seven-point' algebraic system with 2×2 blocks on the coarse level. The condition number of the preconditioned matrix is estimated effectively and does not depend on the mesh-size parameter h. The explicit bounds of the spectrum of the preconditioned matrix are obtained using the super-element approach.

On the second stage, introducing a special rotation we reduce the block seven-point algebraic system to a series of plane problems and an exact seven-point-scheme problem with one unknown per parallelepiped. The constructed preconditioners are spectrally equivalent to the original stiffness matrix and their arithmetic cost depends on the method of solving the latter seven-point problem on the coarse level.

In the last step we use an algebraic multigrid method [4,21,25,36] to solve this seven-point problem. It is shown that the application of such solvers for the problems on the coarse level gives the preconditioner with an optimal complexity. Explicit estimates of condition numbers are also obtained for these multilevel preconditioners.

The rest of the paper is organized as follows. In Section 2 we consider an elimination procedure for (1.4). Then, in Section 3 we develop multilevel preconditioners for the resulting linear system. There we analyze the case where $a(x)$ is a diagonal tensor and Ω is a regular parallelepiped. In Section 4, we extend these results to the case in which $a(x)$ is a full tensor and Ω is a rather general domain. Finally, in Section 5 extensive numerical results are presented for both regular and logical parallelepiped domains to illustrate the present theory.

2. The mixed finite element method

We now consider the most useful partition $\mathcal{T}_h$ of Ω into tetrahedra. The lowest-order Raviart–Thomas–Nedelec space [29,31] defined over $\tau \in \mathcal{T}_h$ is given by

$$\begin{aligned} V_h(\tau) &= \big(P_0(\tau)\big)^3 \oplus \big((x, y, z)P_0(\tau)\big) \\ W_h(\tau) &= P_0(\tau) \\ L_h(e) &= P_0(e) \end{aligned}$$

where $P_i(\tau)$ is the restriction of the set of all polynomials of total degree not bigger than $i \geq 0$ to the set $\tau \in \mathcal{T}_h$. For each τ in $\mathcal{T}_h$, let

$$\overline{f}_\tau = \frac{1}{|\tau|}(f, 1)_\tau$$

where $|\tau|$ denotes the volume of τ. Also, set $\alpha_h = (\alpha_{ij})$ and $\sigma_h|_\tau = (\sigma_{\tau 1}, \sigma_{\tau 2}, \sigma_{\tau 3}) = (q_\tau^1 + t_\tau x, q_\tau^2 + t_\tau y, q_\tau^3 + t_\tau z)$. Then, by the first equation of (1.4), it follows that

$$t_\tau = \overline{f}_\tau / 3 \tag{2.1}$$

Now, take $v = (1, 0, 0)$ in τ and $v = 0$ elsewhere, $v = (0, 1, 0)$ in τ and $v = 0$ elsewhere, and $v = (0, 0, 1)$ in τ and $v = 0$ elsewhere, respectively, in the second equation of (1.4) to obtain

$$\left(\sum_{i=1}^{3} \alpha_{ji}\sigma_{\tau i}, 1\right)_\tau + \sum_{i=1}^{4} |e_\tau^i|\, v_\tau^{i(j)}\, \lambda_h|_{e_\tau^i} = 0, \qquad j = 1, 2, 3 \tag{2.2}$$

where $|e_\tau^i|$ is the area of the face e_τ^i, and $v_\tau^i = (v_\tau^{i(1)}, v_\tau^{i(2)}, v_\tau^{i(3)})$. Let $\beta = (\beta_{ij}) = ((\alpha_{ij}, 1)_\tau)^{-1}$. Then (2.2) can be solved for q_τ^j:

$$\begin{aligned} q_\tau^j = \; & -\sum_{i=1}^{4} |e_\tau^i|\, (\beta_{j1} v_\tau^{i(1)} + \beta_{j2} v_\tau^{i(2)} + \beta_{j3} v_\tau^{i(3)}) \cdot \lambda_h|_{e_\tau^i} \\ & -\frac{\overline{f}_\tau}{3}\left(\sum_{i=1}^{3} \beta_{ji}\, (\alpha_{i1}x + \alpha_{i2}y + \alpha_{i3}z),\ 1\right)_\tau, \qquad j = 1, 2, 3 \end{aligned} \tag{2.3}$$

Let the basis in L_h be chosen as usual. Namely, take $\mu = 1$ on one face and $\mu = 0$ elsewhere in the last equation of (1.4). Then, apply (2.1) and (2.3) to see that the contributions of the tetrahedron τ to the stiffness matrix and the right-hand side are:

$$A_{\tau,ij} = \overline{v}_\tau^i\, \beta\, \overline{v}_\tau^j, \qquad F_{\tau,i} = -\frac{(J_\tau^f, \overline{v}_\tau^i)_\tau}{|\tau|} + (J_\tau^f, v_\tau^i)_{e_\tau^i}, \qquad \tau \in \mathcal{T}_h$$

where $\overline{v}_\tau^i = |e_\tau^i|\, v_\tau^i$ and $J_\tau^f = \overline{f}_\tau(x, y, z)/3$. Hence we obtain the system for λ_h:

$$A\lambda = F \tag{2.4}$$

After the computation of λ_h, we can recover σ_h via (2.1) and (2.3). Also, if u_h is required, it follows from the second equation of (1.4) that

$$u_\tau = \frac{1}{3|\tau|}\left((\alpha_h\sigma_h, (x, y, z))_\tau + \sum_{i=1}^{4} \lambda_h|_{e_\tau^i}((x, y, z), \nu_\tau^i)_{e_\tau^i}\right), \qquad \tau \in \mathcal{T}_h$$

The above result is summarized in the following lemma.

Lemma 2.1. *Let*

$$A_h(\chi, \mu) = \sum_{\tau\in\mathcal{T}_h} (\chi, \nu_\tau)_{\partial\tau}\ \beta\ (\mu, \nu_\tau)_{\partial\tau}, \qquad \chi, \mu \in L_h$$

$$F_h(\mu) = -\sum_{\tau\in\mathcal{T}_h} \frac{1}{|\tau|}(J^f, 1)_\tau \cdot (\mu, \nu_\tau)_{\partial\tau} + \sum_{\tau\in\mathcal{T}_h} (\mu J^f, \nu_\tau)_{\partial\tau}, \qquad \mu \in L_h$$

where J^f is such that $J^f|_\tau = J_\tau^f$. Then $\lambda_h \in \mathcal{L}_h$ satisfies

$$A_h(\lambda_h, \mu) = F_h(\mu), \qquad \forall \mu \in \mathcal{L}_h \tag{2.5}$$

where

$$\mathcal{L}_h = \{\mu \in L_h : \mu|_e = 0 \text{ for each } e \subset \partial\Omega\}$$

Note that there are at most seven non-zero entries per row in the stiffness matrix A. Also, it is easy to see that matrix A is symmetric and positive definite; moreover, if the angles of every τ in $\mathcal{T}_h$ are not bigger than $\pi/2$, then it is an M-matrix. Finally, (2.4) corresponds to the P_1 non-conforming finite element method system, as described below. This equivalence is used to construct our multilevel preconditioners later.

Let

$$\begin{aligned}\mathcal{N}_h = \{v \in L^2(\Omega) :\ & v|_\tau \in P_1(\tau),\ \forall \tau \in \mathcal{T}_h;\ v \text{ is continuous at} \\ & \text{the barycenters of interior faces and} \\ & \text{vanishes at the barycenters of faces on } \partial\Omega\}\end{aligned} \tag{2.6}$$

Proposition 2.1. *Let $f_h = P_h f$. Then* (2.4) *corresponds to the linear system produced by the problem: find $\psi_h \in \mathcal{N}_h$ such that*

$$a_h(\psi_h, \varphi) = (f_h, \varphi), \quad \forall \varphi \in \mathcal{N}_h \tag{2.7}$$

where $a_h(\psi_h, \varphi) = \sum_{\tau\in\mathcal{T}_h}(\alpha_h^{-1}\nabla\psi_h, \nabla\varphi)_\tau$.

Proof From the definition of the basis $\{\psi_i^h\}$ of $\mathcal{N}_h$, for each $\tau \in \mathcal{T}_h$ we have

$$\psi_i^h|_\tau = \frac{1}{|\tau|}\overline{\nu}_\tau^i \cdot ((x, y, z) - p_l), \qquad i \neq l$$

where p_l is the barycenter of face l. Then, we see that

$$(\alpha_h^{-1}\nabla\psi_i^h, \nabla\psi_j^h)_\tau = \overline{\nu}_\tau^i\ \beta\ \overline{\nu}_\tau^j$$

which is $A_{\tau,ij}$. Also, note that for any linear functions ψ and ϕ on a tetrahedron $\tau \in \mathcal{T}_h$

$$(\psi, \phi)_\tau = \frac{1}{4}|\tau| \sum_{i=1}^{4} \psi(p_i)\phi(p_i) \tag{2.8}$$

where the p_i's are the barycenters of the faces of τ, so that

$$F_{\tau,i} = -\frac{(J_\tau^f, \overline{v}_\tau^i)_\tau}{|\tau|} + (J_\tau^f, v_\tau^i)_{e_\tau^i} = -\frac{\overline{f}_\tau}{3}(1, \psi_i^h)_\tau + \frac{|\tau|\overline{f}_\tau}{3|e_\tau^i|}(\psi_i^h, 1)_{e_\tau^i} = \overline{f}_\tau(1, \psi_i^h)_\tau$$

which is $(f_h, \psi_i^h)_\tau$. ∎

3. Multilevel preconditioners over a cube

In this section we consider multilevel preconditioners for (2.4) based on partitioning regular parallelepipeds into tetrahedral substructures, following the ideas in [17,23]. Here we treat the case where Ω is a unit cube and $a(x)$ is a diagonal tensor. A general case is considered in the next section.

3.1. *Two-level preconditioners*

Let $\mathcal{C}_h = \{C^{(i,j,k)}\}$ be a partition of Ω into uniform cubes of length $h = 1/n$, where (x_i, y_j, z_k) is the right back upper corner of the cube $C^{(i,j,k)}$. Next, each cube $C^{(i,j,k)}$ is divided into two prisms $P_1 = P_1^{(i,j,k)}$ and $P_2 = P_2^{(i,j,k)}$ as shown in Figure 1. The resulting partition of Ω is denoted by $\mathcal{P}_h$. Finally, we divide each prism into three tetrahedra as illustrated in Figure 1 and denote this partition of Ω into tetrahedra by $\mathcal{T}_h$. Clearly, there are various ways to subdivide a cube into two prisms. In our approach we partition each cube by plane parallel to the 'z-direction'. The importance of such partitioning will be explained later.

Let $W_{c,h}$ be the space of piecewise constants associated with $\mathcal{C}_h$, and $P_{c,h}$ be the L^2-projection onto $W_{c,h}$. To define our preconditioners, we introduce $\alpha_h = P_{c,h}a^{-1}$ in the hybrid form (1.4) instead of $\alpha_h = P_h a^{-1}$. Obviously, Lemma 2.1 and Proposition 2.1 are still valid for this modification since $\mathcal{T}_h$ is a refinement of $\mathcal{C}_h$. With this modification, α_h^{-1} is a constant on each cube. For notational convenience, we drop the subscript h and simply write $\alpha_h^{-1} = \text{diag}(a_1, a_2, a_3)$.

Let $\mathcal{N}_h$ be the non-conforming finite element space associated with $\mathcal{T}_h$ as defined in (2.6), and let its dimension be N. All the unknowns on the faces of $\partial\Omega$ are excluded. For any function $v_h \in \mathcal{N}_h$, we denote by $\mathbf{v} \in \mathbb{R}^N$ the corresponding vector of its degrees of freedom. Introduce the inner product

$$(\mathbf{u}, \mathbf{v})_N = \sum_{p_i \in \partial\mathcal{T}_h} u_h(p_i)\, v_h(p_i), \qquad u_h, v_h \in \mathcal{N}_h \tag{3.1}$$

where the p_i's are the barycenters of the interior faces. By (2.8) the norm induced by (3.1) is equivalent to the L^2-norm on Ω.

For each prism $P = P^{(i,j,k)} \in \mathcal{P}_h$, denote by $\mathcal{N}_h^P$ the subspace of the restriction of the functions in $\mathcal{N}_h$ onto P. For each $\mathbf{v} \in \mathcal{N}_h^P$, we indicate by $\mathbf{v}_P$ its corresponding vector.

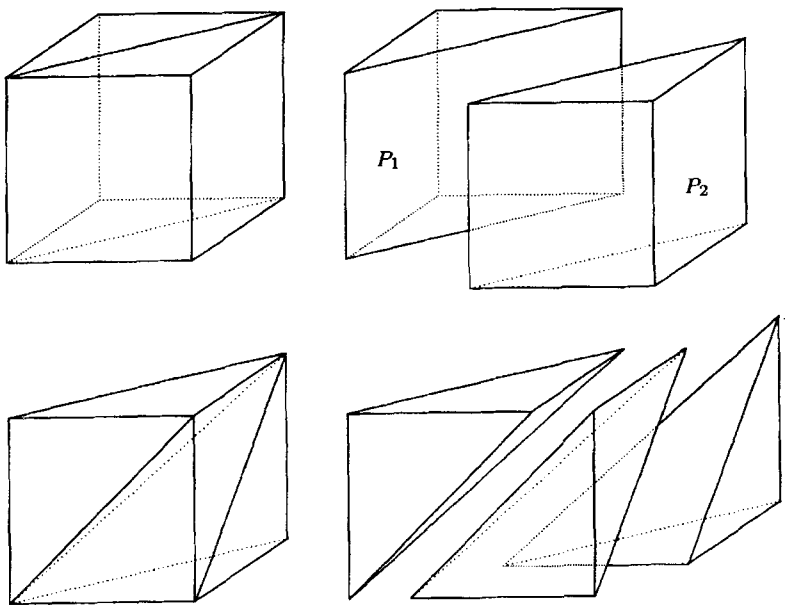


Figure 1. The partition of a cube into prisms and tetrahedra

The dimension of $\mathcal{N}_h^P$ is denoted by N^P. Obviously, for a prism without faces on $\partial\Omega$ its dimension is $N^P = 10$.

The local stiffness matrix A^P on prism $P \in \mathcal{P}_h$ is given by

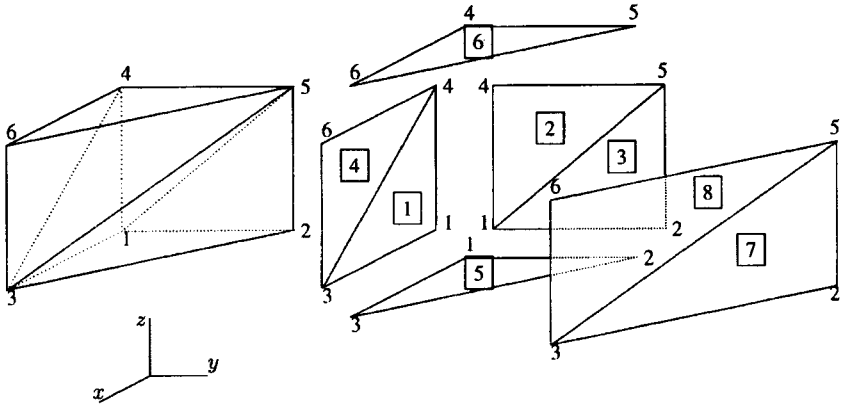
$$(A^P \mathbf{u}_P, \mathbf{v}_P)_{N^P} = \sum_{\tau \subset P} (\alpha_h^{-1} \nabla u_h, \nabla v_h)_\tau \tag{3.2}$$

Then the global stiffness matrix is determined by assembling the local stiffness matrices:

$$(A\mathbf{u}, \mathbf{v})_N = \sum_{P \in \mathcal{P}_h} (A^P \mathbf{u}_P, \mathbf{v}_P)_{N^P} \tag{3.3}$$

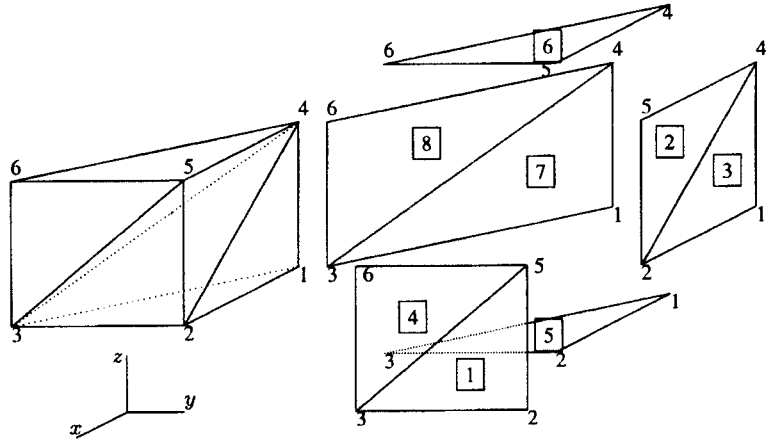
Now we consider a prism P of a cube that has no face on the boundary $\partial\Omega$ and enumerate the faces s_j, $j = 1, \ldots, 10$ of the tetrahedra in this prism as shown in Figure 2. Then the local

$$A^P = \frac{3h}{2} \left[\begin{array}{cccccc|cccc}
a_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -a_2 & 0 \\
0 & a_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -a_1 \\
0 & 0 & a_1 & 0 & 0 & 0 & -a_1 & 0 & 0 & 0 \\
0 & 0 & 0 & a_2 & 0 & 0 & 0 & -a_2 & 0 & 0 \\
0 & 0 & 0 & 0 & a_3 & 0 & 0 & 0 & -a_3 & 0 \\
0 & 0 & 0 & 0 & 0 & a_3 & 0 & 0 & 0 & -a_3 \\
\hline
0 & 0 & -a_1 & 0 & 0 & 0 & a_1+a_2 & 0 & -a_2 & 0 \\
0 & 0 & 0 & -a_2 & 0 & 0 & 0 & a_1+a_2 & 0 & -a_1 \\
-a_2 & 0 & 0 & 0 & -a_3 & 0 & -a_2 & 0 & 2(a_2+a_3) & -a_3 \\
0 & -a_1 & 0 & 0 & 0 & -a_3 & 0 & -a_1 & -a_3 & 2(a_1+a_3)
\end{array}\right]$$



(a) Prism P_1

$s_1 = (1, 4, 3)$ $s_3 = (1, 2, 5)$ $s_5 = (1, 2, 3)$ $s_7 = (2, 5, 3)$ $s_9 = (1, 5, 3)$
$s_2 = (1, 4, 5)$ $s_4 = (3, 4, 6)$ $s_6 = (4, 5, 6)$ $s_8 = (3, 5, 6)$ $s_{10} = (3, 4, 5)$



(b) Prism P_2

$s_1 = (2, 3, 5)$ $s_3 = (1, 2, 4)$ $s_5 = (1, 2, 3)$ $s_7 = (1, 3, 4)$ $s_9 = (2, 3, 4)$
$s_2 = (2, 4, 5)$ $s_4 = (3, 5, 6)$ $s_6 = (4, 5, 6)$ $s_8 = (3, 4, 6)$ $s_{10} = (3, 4, 5)$

Figure 2. Local enumeration of faces in prisms

which we write as

$$A^P = \frac{3h}{2}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{3.4}$$

Along with matrix A^P we also introduce the matrix B^P. The purpose of introducing B^P is to simplify the graph of connectedness in the local stiffness matrix in such a way that the kernel is preserved and the elimination of the unknowns internal for the prism leads to a simpler Schur complement. Matrix B^P is defined on the same space $\mathcal{N}_h^P$ by

$$B^P = \frac{3h}{2}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & B_{22} \end{bmatrix} \tag{3.5}$$

where

$$B_{22} = \begin{bmatrix} a_1 + a_2 + b & -b & -a_2 & 0 \\ -b & a_1 + a_2 + b & 0 & -a_1 \\ -a_2 & 0 & 2a_2 + a_3 & 0 \\ 0 & -a_1 & 0 & 2a_1 + a_3 \end{bmatrix}$$

with some parameter b. This parameter will be chosen in such a way that matrix B^P is spectrally equivalent to A^P (with respect to the kernel) with a smallest possible relative condition number.

Proposition 3.1. *Assume that $b > 0$. Then it holds that* $\ker A^P = \ker B^P$.

Proof It is easy to see from the definitions of A^P and B^P that

$$\ker A^P = \ker B^P = \{\mathbf{v} = (v_1, v_2, \cdots, v_{10})^T \in \mathbb{R}^{10} : v_i = v_1,\ i = 2, \cdots, 10\}$$

■

Remark 1. If the prism $P \in \mathcal{P}_h$ has a face on $\partial\Omega$, then the matrix A^P does not have the rows and columns which correspond to the nodes on that face, and the modification of B_{22} is obvious.

Now we define the $N \times N$ matrix B by the following equality:

$$(B\mathbf{u}, \mathbf{v})_N = \sum_{P \in \mathcal{P}_h} (B^P \mathbf{u}_P, \mathbf{v}_P)_{N^P}, \qquad \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^N$$

Since matrix B is used for preconditioning the original problem (2.4), it is important to estimate the condition number of $B^{-1}A$. Thus, we consider an eigenvalue problem:

$$A\,\mathbf{u} = \mu B\,\mathbf{u} \tag{3.6}$$

Lemma 3.1. *Let $\mu_P \neq 0$ satisfy the equality*

$$A^P \mathbf{u}_P = \mu_P B^P \mathbf{u}_P, \qquad P \in \mathcal{P}_h,\ \mathbf{u}_P \neq \mathbf{0} \tag{3.7}$$

Then we have

$$\max_{(B\mathbf{u},\mathbf{u})_N \neq 0} \frac{(A\mathbf{u}, \mathbf{u})_N}{(B\mathbf{u}, \mathbf{u})_N} \leq \max_{P \in \mathcal{P}_h} \mu_P \quad \textit{and} \quad \min_{(B\mathbf{u},\mathbf{u})_N \neq 0} \frac{(A\mathbf{u}, \mathbf{u})_N}{(B\mathbf{u}, \mathbf{u})_N} \geq \min_{P \in \mathcal{P}_h} \mu_P \tag{3.8}$$

Proof For each $P \in \mathcal{P}_h$, it follows from (3.7) that

$$(A^P \mathbf{u}_P, \mathbf{u}_P)_{N^P} = \mu_P (B^P \mathbf{u}_P, \mathbf{u}_P)_{N^P}$$

Then, from the fact that all the local stiffness matrices are non-negative it follows that

$$\begin{aligned} \sum_{P \in \mathcal{P}_h} (A^P \mathbf{u}_P, \mathbf{u}_P)_{N^P} &= \sum_{P \in \mathcal{P}_h} \mu_P (B^P \mathbf{u}_P, \mathbf{u}_P)_{N^P} \\ &\leq \max_{P \in \mathcal{P}_h} \mu_P \sum_{P \in \mathcal{P}_h} (B^P \mathbf{u}_P, \mathbf{u}_P)_{N^P} \end{aligned}$$

Hence from the definitions of A and B we see that

$$(A\mathbf{u}, \mathbf{u})_N \leq \max_{P \in \mathcal{P}_h} \mu_P (B\mathbf{u}, \mathbf{u})_N$$

Consequently, the first inequality in (3.8) is true. The same argument can be used to verify the second inequality. ∎

From Lemma 3.1, we see that, to estimate the condition number of $B^{-1}A$, it suffices to consider the local problems (3.7). Using a super-element analysis [22] to estimate $\max_{P \in \mathcal{P}_h} \mu_P$ and $\min_{P \in \mathcal{P}_h} \mu_P$, it suffices to treat the worst case where the prism $P \in \mathcal{P}_h$ has no face on the boundary $\partial\Omega$. From (3.4) and (3.5), direct calculations show that the eigenvalues μ_P are within the interval $[\mu_P^-, \mu_P^+]$, where

$$\mu_P^{\pm} = \frac{1}{2}\left(1 + \frac{a_3}{a_1} + \frac{a_3}{a_2} + \frac{a_3}{b}\right)\left(1 \pm \sqrt{1 - \frac{4a_3/b}{(1 + a_3/a_1 + a_3/a_2 + a_3/b)^2}}\right) \tag{3.9}$$

Obviously, $\mu_P^{\pm}$ depends on the parameter b. We shall choose b to minimize the ratio μ_P^+/μ_P^-, which then gives an upper bound for the condition number $\mathrm{Cond}(B^{-1}A)$.

Until the end of the section we shall use the following assumption.

Assumption 31.. *Assume that the matrix coefficient of equation* (1.1) *is a diagonal tensor* $a(x) = diag\ \{a_1, a_2, a_3\}$, *where* a_i, $i = 1, 2, 3$, *are constants on each prism* $P \in \mathcal{P}_h$, *and there exists a parameter* κ *such that*

$$\max_{P \in \mathcal{P}_h} \left\{ \frac{a_3}{a_1}, \frac{a_3}{a_2} \right\} \leq \kappa \tag{3.10}$$

Remark 2. Generally speaking, we need only the assumption that the coefficient a_* in some direction multiplied by some fixed parameter $1/\kappa$ is not greater than the coefficients in the other directions. For the sake of simplicity we assume that this is the 'z-direction'.

The optimal choice of b is given in the following theorem.

Theorem 3.1. *The eigenvalues of problem* (3.7) *with the parameter* $b^{-1} = a_1^{-1} + a_2^{-1} + a_3^{-1}$ *belong to the interval*

$$\left[(1 + 2\kappa)\left(1 - \sqrt{\frac{2\kappa}{1 + 2\kappa}}\right),\ (1 + 2\kappa)\left(1 + \sqrt{\frac{2\kappa}{1 + 2\kappa}}\right)\right]$$

and the condition number is then estimated as follows:

$$Cond\ (B^{-1}A) \leq 3 + 8\kappa$$

Proof With the choice $b^{-1} = a_1^{-1} + a_2^{-1} + a_3^{-1}$, the expression $\mu_P^{\pm}$ can be written as

$$\mu_P^{\pm} = \left(1 + \frac{a_3}{a_1} + \frac{a_3}{a_2}\right)\left(1 \pm \sqrt{1 - \frac{1}{1 + \frac{a_3}{a_1} + \frac{a_3}{a_2}}}\right)$$

Then we consider the functions

$$f_{\pm}(x) = x\left(1 \pm \sqrt{1 - \frac{1}{x}}\right), \quad x \geq 1$$

Note that f_+ is a non-decreasing function and f_- is a non-increasing function. Hence, the desired result follows from the definition of κ. ∎

Remark 3. If the parameter b is chosen by the simple relation $b = a_3$, then the eigenvalues of problem (3.7) belong to the interval

$$\left[1 + \kappa - \sqrt{\kappa^2 + 2\kappa},\ 1 + \kappa + \sqrt{\kappa^2 + 2\kappa}\right]$$

and the condition number is thus estimated by

$$\text{Cond}\ (B^{-1}A) \leq 3 + 8\kappa + 4\kappa^2$$

We stress that the condition number of the matrix $B^{-1}A$ is bounded by a constant independent of the step size of the mesh h. Since we introduced a two level subdivision, matrix B can be referred to as a two-level preconditioner.

Remark 4. Because the condition number of matrix $B^{-1}A$ depends on the value of the parameter κ it is very important to choose the 'z-direction' in the proper way. Note that we can always rearrange the co-ordinate axes (make a change of co-ordinates) to ensure Assumption 3.1.

3.2. Three-level preconditioners

While preconditioner B has good properties, it is not economical to invert it. In this subsection we propose a modification of matrix B and consider its properties and computational scheme. Toward that end, in this section we divide all unknowns in the system into two groups:

1. The first group consists of all unknowns corresponding to the faces of the prisms in partition $\mathcal{P}_h$, excluding the faces on $\partial\Omega$ (see Figure 2).
2. The second group consists of the unknowns corresponding to the faces of the tetrahedra that are internal for each prism (these are faces s_9 and s_{10} in Figure 2).

This splitting of the space $\mathbb{R}^N$ induces the presentation of the vectors: $\mathbf{v} = (\mathbf{v}_1^T, \mathbf{v}_2^T)^T$, where $\mathbf{v}_1 \in \mathbb{R}^{N_1}$ and $\mathbf{v}_2 \in \mathbb{R}^{N_2}$. Obviously, $N_1 = N - 4n^3$. Then matrix B can be presented

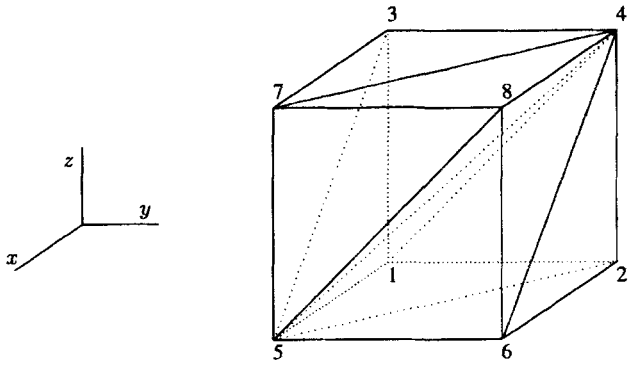


Figure 3. Enumeration of the vertices of a cube $C^{(i,j,k)}$

in the following block form:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \qquad \dim B_{11} = N_1 \tag{3.11}$$

Denote now by $\hat{B}_{11} = B_{11} - B_{12}B_{22}^{-1}B_{21}$ the Schur complement of B obtained by elimination of vector $\mathbf{v}_2$. Then $B_{11} = \hat{B}_{11} + B_{12}B_{22}^{-1}B_{21}$, and hence matrix B has the form:

$$B = \begin{bmatrix} \hat{B}_{11} + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{3.12}$$

Note that for each prism $P \in \mathcal{P}_h$ the unknowns on faces s_9 and s_{10} (see Figure 2) are connected only with the unknowns associated with this prism and therefore can be eliminated locally; that is, matrix B_{22} is diagonal. Thus matrix $\hat{B}_{11}$ is easily computable. The proposed modification of matrix B in (3.12) is of the form

$$\tilde{B} = \begin{bmatrix} \tilde{B}_0 + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where $\tilde{B}_0$ is to be defined later.

3.2.1. Group partitioning of grid points For the sake of simplicity of representation of matrices and computational schemes we introduce the partitioning of all nodes in $\partial\mathcal{T}_h$ into the following three groups. Denote by $s_{r,l,m}^{(i,j,k)}$ the tetrahedral face in the cube $C^{(i,j,k)}$ with vertices r, l, m (see Figure 3).

1. First, we group the nodes on the faces

$$s_{2,4,5}^{(i,j,k)} \quad \text{and} \quad s_{4,5,7}^{(i,j,k)}, \qquad i, j, k = \overline{1, n}$$

we denote the unknowns at these nodes by $VI_\ell^{(i,j,k)}$, $\ell = 1, 2$, $i, j, k = \overline{1, n}$.

2. Second, we number the nodes on the faces perpendicular to x, y, and z axes:

$$\text{(i)}\quad s_{1,2,4}^{(i,j,k)},\quad s_{1,3,4}^{(i,j,k)},\quad i=\overline{2,n},\quad j,k=\overline{1,n}$$

we denote the unknowns at these nodes by $Vx_{\ell}^{(i,j,k)}$, $\ell=1,2$, $i=\overline{2,n}$, $j,k=\overline{1,n}$.

$$\text{(ii)}\quad s_{1,3,5}^{(i,j,k)},\quad s_{5,3,7}^{(i,j,k)},\quad j=\overline{2,n},\quad i,k=\overline{1,n}$$

we denote the unknowns at these nodes by $Vy_{\ell}^{(i,j,k)}$, $\ell=1,2$, $j=\overline{2,n}$, $i,k=\overline{1,n}$.

$$\text{(iii)}\quad s_{1,2,5}^{(i,j,k)},\quad s_{2,5,6}^{(i,j,k)},\quad i,j=\overline{1,n},\quad k=\overline{2,n}$$

we denote the unknowns at these nodes by $Vz_{\ell}^{(i,j,k)}$, $\ell=1,2$, $i,j=\overline{1,n}$, $k=\overline{2,n}$.

3. Finally, we number the remaining nodes on the faces

$$s_{1,4,5}^{(i,j,k)},\quad s_{3,4,5}^{(i,j,k)},\quad s_{4,5,6}^{(i,j,k)},\quad s_{4,5,8}^{(i,j,k)},\quad i,j,k=\overline{1,n}$$

we denote the unknowns at these nodes by $VA_{\ell}^{(i,j,k)}$, $\ell=\overline{1,4}$, $i,j,k=\overline{1,n}$.

3.2.2. Three-level preconditioners We partition each cube $C^{(i,j,k)}$ into left and right prisms $P_p^{(i,j,k)}$, $p=1,2$ (see Figure 1). Below we ignore the indices '(i,j,k)' and the superscript 'P' when no ambiguity occurs.

In the local numeration (see Figure 2) matrices B_1 and B_2, corresponding to the left and right prisms have the form (3.5). We rewrite these matrices in the above group partitioning:

$$B_1=\frac{3h}{2}\left[\begin{array}{cc|cccccc|cc}
a_1+a_2+b & -b & -a_1 & 0 & 0 & 0 & 0 & 0 & -a_2 & 0 \\
-b & a_1+a_2+b & 0 & 0 & 0 & -a_2 & 0 & 0 & 0 & -a_1 \\
\hline
-a_1 & 0 & a_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_1 & 0 & 0 & 0 & 0 & 0 & -a_1 \\
0 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & -a_2 & 0 \\
0 & -a_2 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_3 & 0 & -a_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & 0 & -a_3 \\
\hline
-a_2 & 0 & 0 & 0 & -a_2 & 0 & -a_3 & 0 & 2a_2+a_3 & 0 \\
0 & -a_1 & 0 & -a_1 & 0 & 0 & 0 & -a_3 & 0 & 2a_1+a_3
\end{array}\right]$$

$$B_2=\frac{3h}{2}\left[\begin{array}{cc|cccccc|cc}
a_1+a_2+b & -b & 0 & 0 & -a_2 & 0 & 0 & 0 & -a_1 & 0 \\
-b & a_1+a_2+b & 0 & -a_1 & 0 & 0 & 0 & 0 & 0 & -a_2 \\
\hline
0 & 0 & a_1 & 0 & 0 & 0 & 0 & 0 & -a_1 & 0 \\
0 & -a_1 & 0 & a_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-a_2 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 & -a_2 \\
0 & 0 & 0 & 0 & 0 & 0 & a_3 & 0 & -a_3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & 0 & -a_3 \\
\hline
-a_1 & 0 & -a_1 & 0 & 0 & 0 & -a_3 & 0 & 2a_1+a_3 & 0 \\
0 & -a_2 & 0 & 0 & 0 & -a_2 & 0 & -a_3 & 0 & 2a_2+a_3
\end{array}\right]$$

The partitioning of nodes into the above three groups induces the following block forms of matrices B_p, $p = 1, 2$:

$$B_p = \begin{bmatrix} B_{11,p} & B_{12,p} \\ B_{21,p} & B_{22,p} \end{bmatrix}, \qquad p = 1, 2 \tag{3.13}$$

where blocks $B_{22,p}$ correspond to the unknowns of the last group and blocks $B_{11,p}$ correspond to the unknowns of the first and second groups.

We eliminate the unknowns of the last group from each matrix B_p, $p = 1, 2$, which is done locally on each prism. Then we get the matrices

$$\hat{B}_{11,p} = B_{11,p} - B_{12,p} B_{22,p}^{-1} B_{21,p}, \qquad p = 1, 2$$

where

$$B_{12,1} B_{22,1}^{-1} B_{21,1} = \frac{3h}{2} \left[\begin{array}{cc|cc|cc|cc} \frac{a_2^2}{2a_2+a_3} & 0 & 0 & 0 & \frac{a_2^2}{2a_2+a_3} & 0 & \frac{a_2a_3}{2a_2+a_3} & 0 \\ 0 & \frac{a_1^2}{2a_1+a_3} & 0 & \frac{a_1^2}{2a_1+a_3} & 0 & 0 & 0 & \frac{a_1a_3}{2a_1+a_3} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{a_1^2}{2a_1+a_3} & 0 & \frac{a_1^2}{2a_1+a_3} & 0 & 0 & 0 & \frac{a_1a_3}{2a_1+a_3} \\ \hline \frac{a_2^2}{2a_2+a_3} & 0 & 0 & 0 & \frac{a_2^2}{2a_2+a_3} & 0 & \frac{a_2a_3}{2a_2+a_3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \frac{a_2a_3}{2a_2+a_3} & 0 & 0 & 0 & \frac{a_2a_3}{2a_2+a_3} & 0 & \frac{a_3^2}{2a_2+a_3} & 0 \\ 0 & \frac{a_1a_3}{2a_1+a_3} & 0 & \frac{a_1a_3}{2a_1+a_3} & 0 & 0 & 0 & \frac{a_3^2}{2a_1+a_3} \end{array}\right]$$

and a similar form holds for $B_{12,2} B_{22,2}^{-1} B_{21,2}$.

Following [23], we introduce on each prism a modification of matrices $\hat{B}_{11,p}$:

$$B_0 = \frac{3h}{2} \left[\begin{array}{cc|cc|cc|cc} a_1+a_2+b+s_2 & -b & -a_1 & 0 & -a_2 & 0 & -s_2/2 & -s_2/2 \\ -b & a_1+a_2+b+s_1 & 0 & -a_1 & 0 & -a_2 & -s_1/2 & -s_1/2 \\ \hline -a_1 & 0 & a_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -a_1 & 0 & a_1 & 0 & 0 & 0 & 0 \\ \hline -a_2 & 0 & 0 & 0 & a_2 & 0 & 0 & 0 \\ 0 & -a_2 & 0 & 0 & 0 & a_2 & 0 & 0 \\ \hline -s_2/2 & -s_1/2 & 0 & 0 & 0 & 0 & \frac{s_1+s_2}{2} & 0 \\ -s_2/2 & -s_1/2 & 0 & 0 & 0 & 0 & 0 & \frac{s_1+s_2}{2} \end{array}\right]$$

with some parameters s_1 and s_2.

Proposition 3.2. *Assume that $s_1 > 0$ and $s_2 > 0$. Then matrices $\hat{B}_{11,p}$, $p = 1, 2$, and B_0 have the same kernel, i.e.* $\ker \hat{B}_{11,p} = \ker B_0$.

Proof It can be easily checked that $\ker \hat{B}_{11,p} = \ker B_0 = \{\mathbf{v} = (v_1, v_2, \cdots, v_8)^T \in \mathbb{R}^8 : v_i = v_1,\ i = 2, \cdots, 8\}$, $\quad p = 1, 2.$ ∎

We now consider the eigenvalue problem

$$\hat{B}_{11,p}\mathbf{u} = \mu B_0 \mathbf{u}, \quad \mathbf{u} \in \mathbb{R}^8 \setminus \ker B_0, \qquad p = 1, 2 \tag{3.14}$$

with the following choices of s_1 and s_2.

Proposition 3.3. *For the case of* $s_i = 2a_i a_3/(2a_i + a_3)$, $i = 1, 2$, *the eigenvalues of problem* (3.14) *belong to the interval*

$$\left[\frac{3+2\kappa}{4+2\kappa}\left(1-\frac{1}{\sqrt{3}}\right), \frac{3+2\kappa}{4+2\kappa}\left(1+\frac{1}{\sqrt{3}}\right)\right]$$

If we choose $s_i = a_3$, $i = 1, 2$, *the eigenvalues of problem* (3.14) *are within the interval*

$$\left[\frac{3+\kappa}{4+2\kappa}\left(1-\frac{1}{\sqrt{3}}\right), \frac{3+\kappa}{4+2\kappa}\left(1+\frac{1}{\sqrt{3}}\right)\right]$$

Both cases have the same estimate of the condition number

$$Cond\,(B_0^{-1}\hat{B}_{11,p}) \leq 2+\sqrt{3}$$

where the condition number is defined as the ratio of the biggest and the smallest non-zero eigenvalues of problem (3.14).

Proof A direct calculation shows that $\mu \in [\mu^-, \mu^+]$ where

$$\mu^- = \min_{i=1,2}\left\{\frac{a_i}{4a_i+2a_3}\left(1+\frac{a_3}{a_i}+\frac{2a_3}{s_i}\right)\left(1-\sqrt{1-\frac{a_3^2/(a_i s_i)+2a_3/s_i}{1+a_3/a_i+2a_3/s_i}}\right)\right\}$$

and

$$\mu^+ = \max_{i=1,2}\left\{\frac{a_i}{4a_i+2a_3}\left(1+\frac{a_3}{a_i}+\frac{2a_3}{s_i}\right)\left(1+\sqrt{1-\frac{a_3^2/(a_i s_i)+2a_3/s_i}{1+a_3/a_i+2a_3/s_i}}\right)\right\}$$

With $s_i = 2a_i a_3/(2a_i + a_3)$, $i = 1, 2$, and the definition of κ, it can be seen as in Theorem 3.1 that

$$\mu_- \geq \frac{3+2\kappa}{4+2\kappa}\left(1-\sqrt{1-\frac{2+3\kappa/2+\kappa^2/2}{3+2\kappa}}\right)$$

and

$$\mu_+ \geq \frac{3+2\kappa}{4+2\kappa}\left(1+\sqrt{1-\frac{2+3\kappa/2+\kappa^2/2}{3+2\kappa}}\right)$$

Note that

$$1-\frac{2+3\kappa/2+\kappa^2/2}{3+2\kappa} \leq \frac{1}{3}$$

so that the first case follows. The same argument applies to the second case. ∎

Now we define a new matrix on each prism:

$$\tilde{B}_p = \begin{bmatrix} B_0 + B_{12,p}B_{22,p}^{-1}B_{21,p} & B_{12,p} \\ B_{21,p} & B_{22,p} \end{bmatrix}, \qquad p = 1, 2 \tag{3.15}$$

As we noted in Remark 3.1, when cube C has a non-empty intersection with $\partial\Omega$, matrices B_0, $B_{12,p}$, and $B_{21,p}$, $p = 1, 2$, do not have rows and columns corresponding to the nodes on the boundary.

For each prism $P \in \mathcal{P}_h$ we now consider the eigenvalue problem

$$B^P\mathbf{u} = \mu\tilde{B}^P\mathbf{u} \tag{3.16}$$

where $B^P = B_p^P$ is defined in (3.13) and $\tilde{B}^P = \tilde{B}_p^P$ in (3.15), $p = 1, 2$. Below we consider only the simplest choice: $s_i = a_3$, $i = 1, 2$.

Proposition 3.4. *The eigenvalues of problem* (3.16) *belong to the interval*

$$\left[\frac{3+\kappa}{4+2\kappa}\left(1-\frac{1}{\sqrt{3}}\right), \frac{3+\kappa}{4+2\kappa}\left(1+\frac{1}{\sqrt{3}}\right)\right]$$

Moreover, on each prism $P \in \mathcal{P}_h$ the eigenvalues of the problem

$$A^P\mathbf{u} = \mu\tilde{B}^P\mathbf{u} \tag{3.17}$$

are within the interval $[\mu_-, \mu_+]$, where

$$\mu_\pm = (1+2\kappa)\left(1 \pm \sqrt{\frac{2\kappa}{1+2\kappa}}\right)\frac{3+\kappa}{4+2\kappa}\left(1 \pm \frac{1}{\sqrt{3}}\right)$$

Proof The first statement follows directly from Proposition 3.3, and the second one then follows from Theorem 3.1. ∎

Now we define the symmetric positive-definite $N_1 \times N_1$ matrix $\tilde{B}_0$ by

$$(\tilde{B}_0\mathbf{u}_1, \mathbf{v}_1) = \sum_{P\in\mathcal{P}_h} (B_0\mathbf{u}_{1,P}, \mathbf{v}_{1,P})$$

where $\mathbf{v}_1, \mathbf{u}_1 \in \mathbb{R}^{N_1}$, and $\mathbf{u}_{1,P}$ and $\mathbf{v}_{1,P}$ are their respective restrictions on prism P. As in (3.12), we introduce the matrix

$$\tilde{B} = \begin{bmatrix} \tilde{B}_0 + B_{12}B_{22}^{-1}B_{21} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \tag{3.18}$$

Using Proposition 3.2 and the same proof as in Theorem 3.1, we have the following theorem.

Theorem 3.2. *Matrix $\tilde{B}$ defined in* (3.18) *is spectrally equivalent to matrix A, i.e.*

$$\mu_*\tilde{B} \le A \le \mu^*\tilde{B}$$

Moreover,

$$Cond\,(\tilde{B}^{-1}A) \le \overline{\mu} \equiv \mu^*/\mu_* \le (3+8\kappa)(2+\sqrt{3}) \tag{3.19}$$

Instead of matrix B in the form (3.12) we take matrix $\tilde{B}$ from (3.18) as a preconditioner for matrix A. Because we have introduced a two-level subdivision of matrix $\tilde{B}_0$, matrix $\tilde{B}$ can be considered a three-level preconditioner.

As we noted earlier, matrix B_{22} is block-diagonal and can be inverted locally on prisms. So we concentrate on the linear system

$$\tilde{B}_0 \mathbf{u} = \mathbf{G} \tag{3.20}$$

In terms of the group partitioning in Section 3.2.1, matrix $\tilde{B}_0$ has the block form

$$\tilde{B}_0 = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{3.21}$$

where block C_{22} corresponds to the nodes from the second group, which are on the faces of tetrahedra perpendicular to the co-ordinate axes. From the definition of B_0, it can be seen that matrix C_{22} is diagonal. In the above partitioning, we present $\mathbf{u}$ and $\mathbf{G}$ in (3.20) in the form

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \qquad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}$$

Then, after elimination of the second group of unknowns:

$$\mathbf{u}_2 = C_{22}^{-1}(\mathbf{G}_2 - C_{21}\mathbf{u}_1)$$

we get the system of linear equations

$$(C_{11} - C_{12}C_{22}^{-1}C_{21})\mathbf{u}_1 = \mathbf{G}_1 - C_{12}C_{22}^{-1}\mathbf{G}_2 \equiv \tilde{\mathbf{G}}_1$$

where vector $\mathbf{u}_1$ and block C_{11} correspond to the unknowns from the first group, which have only two unknowns per cube. The dimension of vectors $\mathbf{u}_1$ and $\mathbf{G}_1$ is equal to dim $(\mathbf{u}_1) = 2n^3$. The above simplification of (3.20) is carried out in detail in the next subsection.

Remark 5. We note that all the estimates in this section depend on parameter κ introduced in Assumption 3.1. Hence, it is very important to arrange the co-ordinate axes in such a way that parameter κ has the smallest value.

Remark 6. Note that the estimate of the condition number of the preconditioned matrix (3.19) is proportional to the value of parameter κ. In some sense we benefit from anisotropy. The smaller the coefficient a_3 of matrix a (the coefficient in the 'z-direction') the better the preconditioner B.

3.2.3. Computational scheme We now consider the computational scheme for (3.20). For simplicity of presentation we consider here the case of diagonal constant tensor coefficient $a(x) \equiv \text{diag}\,\{a_1, a_2, a_3\}$ in the whole domain Ω. In terms of the unknowns introduced

in Section 3.2.1:

$$
\begin{aligned}
&uI_\ell^{(i,j,k)}, \quad GI_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad i,j,k = \overline{1,n} \\
&ux_\ell^{(i,j,k)}, \quad Gx_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad i = \overline{2,n}, \qquad j,k = \overline{1,n} \\
&uy_\ell^{(i,j,k)}, \quad Gy_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad j = \overline{2,n}, \qquad i,k = \overline{1,n} \\
&uz_\ell^{(i,j,k)}, \quad Gz_\ell^{(i,j,k)}, \quad \ell = 1,2, \quad k = \overline{2,n}, \qquad i,j = \overline{1,n}
\end{aligned}
$$

the system (3.20) with $s_i = a_3$, $i = 1,2$, can be written as

$$
\begin{aligned}
2\begin{bmatrix} a_1+a_2+a_3+b & -b \\ -b & a_1+a_2+a_3+b \end{bmatrix} &\mathbf{u}I^{(i,j,k)} \\
-a_1\left((1-\delta_{i1})\mathbf{u}x^{(i-1,j,k)} + (1-\delta_{in})\mathbf{u}x^{(i,j,k)}\right) & \\
-a_2\left((1-\delta_{j1})\mathbf{u}y^{(i,j-1,k)} + (1-\delta_{jn})\mathbf{u}y^{(i,j,k)}\right) & \\
-\frac{a_3}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left((1-\delta_{k1})\mathbf{u}z^{(i,j,k-1)} + (1-\delta_{kn})\mathbf{u}z^{(i,j,k)}\right) & \\
= \frac{2}{3h}\mathbf{G}I^{(i,j,k)}, \qquad i,j,k = \overline{1,n} &
\end{aligned}
\tag{3.22}
$$

and

$$
a_1\left(2\mathbf{u}x^{(i,j,k)} - \mathbf{u}I^{(i+1,j,k)} - \mathbf{u}I^{(i,j,k)}\right) = \frac{2}{3h}\mathbf{G}x^{(i,j,k)}, \quad i = \overline{1,n-1},\ j,k = \overline{1,n}
$$

$$
a_2\left(2\mathbf{u}y^{(i,j,k)} - \mathbf{u}I^{(i,j+1,k)} - \mathbf{u}I^{(i,j,k)}\right) = \frac{2}{3h}\mathbf{G}y^{(i,j,k)}, \quad j = \overline{1,n-1},\ i,k = \overline{1,n}
$$

$$
\begin{aligned}
a_3\left(2\mathbf{u}z^{(i,j,k)} - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left(\mathbf{u}I^{(i,j,k+1)} + \mathbf{u}I^{(i,j,k)}\right)\right) &= \frac{2}{3h}\,\mathbf{G}z^{(i,j,k)} \\
k = \overline{1,n-1},\ i,j = \overline{1,n} &
\end{aligned}
\tag{3.23}
$$

where δ_{ij} (the Kronecker symbol) is introduced to take into account the Dirichlet boundary conditions. Eliminating unknowns $ux_\ell^{(i,j,k)}$, $uy_\ell^{(i,j,k)}$, $uz_\ell^{(i,j,k)}$, $\ell = 1,2$, from equations (3.22), we obtain the block 'seven-point' scheme with 2×2-blocks for the unknowns

$uI_\ell^{(i,j,k)}$, $\ell = 1, 2$, $i, j, k = \overline{1, n}$. From (3.23) we have

$$\mathbf{u}x^{(i,j,k)} = \frac{1}{3ha_1}\mathbf{G}x^{(i,j,k)} + \frac{1}{2}\left(\mathbf{u}I^{(i+1,j,k)} + \mathbf{u}I^{(i,j,k)}\right), \quad i = \overline{1, n-1},\ j, k = \overline{1, n}$$

$$\mathbf{u}y^{(i,j,k)} = \frac{1}{3ha_2}\mathbf{G}y^{(i,j,k)} + \frac{1}{2}\left(\mathbf{u}I^{(i,j+1,k)} + \mathbf{u}I^{(i,j,k)}\right), \quad j = \overline{1, n-1},\ i, k = \overline{1, n}$$

$$\mathbf{u}z^{(i,j,k)} = \frac{1}{3ha_3}\mathbf{G}z^{(i,j,k)} + \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left(\mathbf{u}I^{(i,j,k+1)} + \mathbf{u}I^{(i,j,k)}\right)$$
$$k = \overline{1, n-1},\ i, j = \overline{1, n} \tag{3.24}$$

Substituting (3.24) into (3.22), we see that

$$\begin{aligned} &2\begin{bmatrix} a_1+a_2+a_3+b & -b \\ -b & a_1+a_2+a_3+b \end{bmatrix}\mathbf{u}I^{(i,j,k)} \\ &\quad -\frac{a_1}{2}\left((1-\delta_{i1})\left(\mathbf{u}I^{(i-1,j,k)} + \mathbf{u}I^{(i,j,k)}\right) + (1-\delta_{in})\left(\mathbf{u}I^{(i+1,j,k)} + \mathbf{u}I^{(i,j,k)}\right)\right) \\ &\quad -\frac{a_2}{2}\left((1-\delta_{j1})\left(\mathbf{u}I^{(i,j-1,k)} + \mathbf{u}I^{(i,j,k)}\right) + (1-\delta_{jn})\left(\mathbf{u}I^{(i,j+1,k)} + \mathbf{u}I^{(i,j,k)}\right)\right) \\ &-\frac{a_3}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left((1-\delta_{k1})\left(\mathbf{u}I^{(i,j,k-1)} + \mathbf{u}I^{(i,j,k)}\right) + (1-\delta_{kn})\left(\mathbf{u}I^{(i,j,k+1)} + \mathbf{u}I^{(i,j,k)}\right)\right) \\ &\qquad = \mathbf{g}^{(i,j,k)}, \quad i, j, k = \overline{1, n} \end{aligned} \tag{3.25}$$

where

$$\begin{aligned} \mathbf{g}^{(i,j,k)} = \frac{2}{3h}\Bigg\{ &\mathbf{G}I^{(i,j,k)} + \frac{1}{2a_1}\left((1-\delta_{i1})\mathbf{G}x^{(i-1,j,k)} + (1-\delta_{in})\mathbf{G}x^{(i,j,k)}\right) \\ &+\frac{1}{2a_2}\left((1-\delta_{j1})\mathbf{G}y^{(i,j-1,k)} + (1-\delta_{jn})\mathbf{G}y^{(i,j,k)}\right) \\ &+\frac{1}{4a_3}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\left((1-\delta_{k1})\mathbf{G}z^{(i,j,k-1)} + (1-\delta_{kn})\mathbf{G}z^{(i,j,k)}\right)\Bigg\} \end{aligned} \tag{3.26}$$

To solve system (3.25) we introduce the rotation matrix

$$Q = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

and new vectors $\mathbf{v}^{(i,j,k)} = (v_1^{(i,j,k)}, v_2^{(i,j,k)})^T$, $i, j, k = \overline{1, n}$, given by

$$\mathbf{v}^{(i,j,k)} = Q \cdot \mathbf{u}I^{(i,j,k)}, \qquad i, j, k = \overline{1, n} \tag{3.27}$$

Then multiplying both sides of the matrix equation (3.25) by matrix Q and using the

relation

$$\mathbf{u}I^{(i,j,k)} = Q^T \cdot \mathbf{v}^{(i,j,k)}, \qquad i, j, k = \overline{1, n} \tag{3.28}$$

we obtain the following problem for the unknowns $\mathbf{v}^{(i,j,k)}$:

$$\begin{aligned}
&\begin{bmatrix} 2(a_1+a_2+a_3) & 0 \\ 0 & 2(a_1+a_2+a_3+b) \end{bmatrix} \mathbf{v}^{(i,j,k)} \\
&\quad -\frac{a_1}{2}\left((1-\delta_{i1})\left(\mathbf{v}^{(i-1,j,k)}+\mathbf{v}^{(i,j,k)}\right)+(1-\delta_{in})\left(\mathbf{v}^{(i+1,j,k)}+\mathbf{v}^{(i,j,k)}\right)\right) \\
&\quad -\frac{a_2}{2}\left((1-\delta_{j1})\left(\mathbf{v}^{(i,j-1,k)}+\mathbf{v}^{(i,j,k)}\right)+(1-\delta_{jn})\left(\mathbf{v}^{(i,j+1,k)}+\mathbf{v}^{(i,j,k)}\right)\right) \\
&-\frac{a_3}{2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\left((1-\delta_{k1})\left(\mathbf{v}^{(i,j,k-1)}+\mathbf{v}^{(i,j,k)}\right)+(1-\delta_{kn})\left(\mathbf{v}^{(i,j,k+1)}+\mathbf{v}^{(i,j,k)}\right)\right) \\
&\qquad = Q\cdot\mathbf{g}^{(i,j,k)} \equiv \tilde{\mathbf{g}}^{(i,j,k)}, \qquad i, j, k = \overline{1, n}
\end{aligned} \tag{3.29}$$

It is easy to see that problem (3.29) can be decomposed into the following two independent problems:

$$\begin{aligned}
&2(a_1+a_2+a_3)v_1^{(i,j,k)} \\
&-\frac{a_1}{2}\left((1-\delta_{i1})\left(v_1^{(i-1,j,k)}+v_1^{(i,j,k)}\right)+(1-\delta_{in})\left(v_1^{(i+1,j,k)}+v_1^{(i,j,k)}\right)\right) \\
&-\frac{a_2}{2}\left((1-\delta_{j1})\left(v_1^{(i,j-1,k)}+v_1^{(i,j,k)}\right)+(1-\delta_{jn})\left(v_1^{(i,j+1,k)}+v_1^{(i,j,k)}\right)\right) \\
&-\frac{a_3}{2}\left((1-\delta_{k1})\left(v_1^{(i,j,k-1)}+v_1^{(i,j,k)}\right)+(1-\delta_{kn})\left(v_1^{(i,j,k+1)}+v_1^{(i,j,k)}\right)\right) = \tilde{g}_1^{(i,j,k)} \\
&\qquad\qquad i, j, k = \overline{1, n}
\end{aligned} \tag{3.30}$$

and

$$\begin{aligned}
&2(a_1+a_2+a_3+b)v_2^{(i,j,k)} \\
&-\frac{a_1}{2}\left((1-\delta_{i1})\left(v_2^{(i-1,j,k)}+v_2^{(i,j,k)}\right)+(1-\delta_{in})\left(v_2^{(i+1,j,k)}+v_2^{(i,j,k)}\right)\right) \\
&-\frac{a_2}{2}\left((1-\delta_{j1})\left(v_2^{(i,j-1,k)}+v_2^{(i,j,k)}\right)+(1-\delta_{jn})\left(v_2^{(i,j+1,k)}+v_2^{(i,j,k)}\right)\right) = \tilde{g}_2^{(i,j,k)} \\
&\qquad\qquad i, j = \overline{1, n}, \quad \forall\, k = \overline{1, n}
\end{aligned} \tag{3.31}$$

Hence, we reduced the linear system (3.29) of dimension $(2n^3)$ to one linear system of equations (3.30) of dimension n^3 and n linear systems of equations (3.31) of dimension n^2.

For all these problems we can use either the method of separation of variables [34] or an algebraic multigrid method [4,21,25]. An implementation cost of the first method is estimated by $O(h^{-3}\ln(h^{-1}))$. The AMG methods have the optimal order of arithmetic complexity $O(h^{-3})$.

After we find the solution of problems (3.30) and (3.31) we easily retrieve vectors $\mathbf{u}I^{(i,j,k)}$ by using relations (3.28).

3.2.4. Preconditioned conjugate gradient method The underlying method to solve system (2.4) is a preconditioned iterative method. The choice of a particular method within a certain class is not essential, but for the purpose of this exposition we may think of the well-known preconditioned conjugate gradient method [19,24], which is often used in practice.

Proposition 3.5. *The number of operations for solving system* (2.4) *by a preconditioned conjugate method with preconditioner* $\tilde{B}$ *defined in* (3.16) *and with accuracy* ϵ *in the sense*

$$\|\lambda^{k_\epsilon+1} - \lambda^*\|_A \leq \epsilon \|\lambda^0 - \lambda^*\|_A, \qquad 0 < \epsilon \ll 1 \tag{3.32}$$

is estimated by $cN \ln\left(\frac{2}{\epsilon}\right)$, *where* $\lambda^* = A^{-1}F$, $\lambda^0 \in \mathbb{R}^N$ *is any initial vector, and the constant* $c > 0$ *does not depend on* N *and the coefficients of the matrix function* $a(x)$.

4. Preconditioners for a general case

In this section we consider the case when the coefficient a is a full symmetric tensor and the domain Ω satisfies the following assumptions:

(a) There is an orientation-preserving smooth map $\mathcal{L}$ of the unit cube $\hat{\Omega}$ onto Ω and there are positive constants d and C (see [15]) such that

$$\begin{aligned} d^{-1}\|\mathcal{J}(x)\| &\leq C, \qquad \forall x \in \hat{\Omega} \\ d\,\|\mathcal{J}^{-1}(x)\| &\leq C, \qquad \forall x \in \Omega \end{aligned} \tag{4.1}$$

where $\mathcal{J}(x)$ is the Jacobian matrix of $\mathcal{L}$ at x and $\|\cdot\|$ denotes a matrix norm.

(b) The transformed tensor $\hat{a}(x) = \frac{1}{|\det(\mathcal{J})|}\mathcal{J}^T a(x)\mathcal{J}$, $x \in \hat{\Omega}$ is a smooth matrix function which is a small perturbation of a diagonal constant matrix in the entire domain. It means that there exist a diagonal constant matrix $\tilde{a} = \text{diag}\,\{a_1, a_2, a_3\}$, and some positive constants $\check{c}$, $\hat{c}$, such that

$$\check{c}\,\tilde{a} \leq a(x) \leq \hat{c}\,\tilde{a}, \qquad \forall x \in \Omega \tag{4.2}$$

The definition of the non-conforming finite element space for the domains satisfying (4.1) is given below. Let $\mathcal{C}_{\hat{h}}$ and $\mathcal{T}_{\hat{h}}$ be the partitions of $\hat{\Omega}$ into cubes and tetrahedra, respectively, which are associated with the mesh-size $\hat{h} = 1/n$, and let $\mathcal{N}_{\hat{h}}$ be the P_1-non-conforming space associated with $\mathcal{T}_{\hat{h}}$, as given in (2.6). Set $h = r \cdot \hat{h}$ and define

$$\mathcal{N}_h(\Omega) = \left\{\varphi = \psi \circ \mathcal{L}^{-1} : \psi \in V_{\hat{h}}(\hat{\Omega})\right\}$$

We also introduce the map $\mathcal{J} : V_h(\Omega) \to V_{\hat{h}}(\hat{\Omega})$ defined by $\mathcal{J}v = v \circ \mathcal{L}$.

Now we define the stiffness matrix A on domain Ω by

$$(A\mathbf{u}, \mathbf{v})_N = a_h(u, v), \qquad \forall u, v \in \mathcal{N}_h(\Omega) \tag{4.3}$$

where

$$\begin{aligned} a_h(u,v) &= \sum_{\tau\in\mathcal{T}_h} \int_\tau a(x)\,\nabla u\cdot\nabla v\,dx \\ &= \sum_{\hat\tau\in\mathcal{T}_{\hat h}} \int_{\hat\tau} \tfrac{1}{|\det(\mathcal{J})|}\mathcal{J}^T a(x)\mathcal{J}\,\nabla\mathcal{J}u\cdot\nabla\mathcal{J}v\,dx \end{aligned} \tag{4.4}$$

and $|\det(\mathcal{J})|$ is the determinant of the Jacobian $\mathcal{J}(x)$.

Note that taking into account (4.4), we can treat the bilinear form (4.3) as a form generated by some elliptic positive definite operator with a piecewise smooth 3×3 symmetric matrix-valued function $a(x)$ on the cube $\hat\Omega$. This function satisfies the uniform positive definiteness condition. For this reason, below without loss of generality we suppose that $\Omega \equiv \hat\Omega$ is a parallelepiped with the partition into cubes $\mathcal{C}_h$ and into tetrahedra $\mathcal{T}_h$.

For each cube $C \in \mathcal{C}_{\hat h}$, we introduce diagonal matrix $\mathcal{A}_C = \mathrm{diag}\{a_{1,C}, a_{2,C}, a_{3,C}\}$ with some as yet unspecified constants $a_{i,C}$, $i = 1, 2, 3$. Then on the reference parallelepiped $\hat\Omega$ we define a bilinear form

$$b_h(u,v) = \sum_{C\in\mathcal{C}_h} \delta_C \left(\sum_{\tau\in C} \int_\tau \mathcal{A}_C\nabla u\cdot\nabla v\,dx \right), \qquad \forall u, v \in \mathcal{N}_{\hat h} \tag{4.5}$$

where the constants δ_C are scaling factors. One reasonable choice is to take $\delta_C = (\lambda_{1,C} + \lambda_{0,C})/2$, where $\lambda_{1,C}$ and $\lambda_{0,C}$ are the largest and smallest eigenvalues of the eigenvalue problem $\hat a(x_0)\mathbf{v} = \lambda_C\mathcal{A}_C\mathbf{v}$, $\mathbf{v}\in\mathbb{R}^3$, where $\hat a(x) = \frac{1}{|\det(\mathcal{J})|}\mathcal{J}^T a(x)\mathcal{J}$ and $x_0 \in \mathcal{L}(C) \subset \Omega$ is some point.

We assume that the matrix function defined above, $\delta_C\mathcal{A}_C$, is a small perturbation of a diagonal constant matrix in the entire cube $\hat\Omega$.

Note that assumptions (4.1) imply that there are two constants c_0 and c_1 independent of r and $\hat h$ such that

$$c_0 a_h(u,u) \le r\cdot b_h(\mathcal{J}u, \mathcal{J}u) \le c_1 a_h(u,u), \qquad \forall u \in \mathcal{N}_h \tag{4.6}$$

We choose matrices $\mathcal{A}_C$ in the form: $\mathcal{A}_C = \mathrm{diag}\{\hat a(x_0)\}$, $\forall C \in \mathcal{C}_h$, i.e. the matrix $\mathcal{A}_C$ is the diagonal part of $\hat a(x_0)$ at some point $x_0 \in \mathcal{L}(C)$. In this case constants c_0 and c_1 in (4.6) depend only on the local variation of the coefficients $\{(\hat a)_{kl}\}$. Hence the problem of defining a preconditioner for $a_h(\cdot,\cdot)$ is reduced to the problem of finding a preconditioner for $r\cdot b_h(\cdot,\cdot)$, which has a diagonal coefficient tensor and is defined on the unit cube $\hat\Omega$. Therefore, all the analysis of Section 3 can be carried out here.

5. Results of the numerical experiments

In this section the method of preconditioning being presented is tested on the model problem

$$\begin{aligned} -\sum_{i=1}^{3} \frac{\partial}{\partial x_i}\left(a_i\frac{\partial u}{\partial x_i}\right) &= f \qquad \text{in } \Omega \\ u &= 0 \qquad \text{on } \partial\Omega \end{aligned} \tag{5.1}$$

We present three numerical examples. In the first example, the domain Ω is the unit cube: $\Omega = [0,1]^3$. The domain is divided into n^3 cubes (n in each direction) and each cube

Table 1. The results of experiments for anisotropic problem

			$20 \times 20 \times 20$ $N = 93\,600$		$30 \times 30 \times 30$ $N = 318\,600$	
a_1	a_2	a_3	Iter	Cond	Iter	Cond
1	1	1	18	7.5	17	7.6
1	1	0.1	13	3.7	13	3.8
1	1	0.01	10	2.8	11	3.0
10	1	1	16	6	16	6.2
1	10	1				
100	1	1	14	4.7	14	5.2
1	100	1				
1	1	10	34	41	34	42
1	1	100	75	315	80	328
0.1	1	1	32	30	31	29
1	0.1	1				
0.01	1	1	68	198	72	203
1	0.01	1				

is partitioned into six tetrahedra. The dimension of the original algebraic system is $N = 12n^3 - 6n^2$. The right hand side is generated randomly, and the accuracy parameter is taken as $\epsilon = 10^{-6}$. The condition number of the matrix $B^{-1}A$ is calculated by the relation between the conjugate gradient and Lanczos algorithms [19]. In this experiment we have considered the dependency of the condition number on the coefficients of the problem. The coefficients a_i, $i = 1, 2, 3$, are constants on each cube. The results are summarized in Table 1, where 'Iter' and 'Cond' denote the iteration number and condition number, respectively.

From Table 1 we see that the condition number depends on the maximal ratio $\kappa = \max_{C \in \mathcal{C}_h} \left\{ \frac{a_3}{a_1}, \frac{a_3}{a_2} \right\}$. The case of $\kappa < 1$ has a better convergence than the case of the Poisson equation (i.e., $a_1 = a_2 = a_3 = 1$) as is predicted by the theory (see estimate (3.19)).

We note that the condition numbers in all experiments depend on parameter κ introduced in Assumption 3.1. Namely, the estimate of the condition number of the preconditioned matrix (3.19) is proportional to the value of parameter κ. Obviously, it is important to arrange the co-ordinate axis in such a way that parameter κ has the smallest value. In some sense we can benefit from anisotropy. The smaller coefficient a_3 (the coefficient in the 'z-direction') leads to a better preconditioner B.

In the second example we have considered the dependency of the condition number on the jump of the coefficients. Again, the computational domain is the unit cube, which is subdivided into n^3 cubes and each cube is partitioned into six tetrahedra. The coefficient $a(x)$ is piecewise constant and is defined to be

$$a_1 = a_2 = a_3 = \begin{cases} a, & (x, y, z) \in [0.5, 1] \times [0.5, 1] \times [0.5, 1] \\ 1, & \text{elsewhere} \end{cases} \tag{5.2}$$

The results are summarized in Table 2.

In the third example we treat the Poisson equation on the domain Ω as shown in Figure 4. The domain is subdivided into $90 \times 90 \times 10$ cubes and the number of unknowns is then $N = 955\,440$. This problem is solved with accuracy $\epsilon = 10^{-6}$. Twenty iterations are needed

Table 2. The results of experiments with jump of the coefficient

	$20 \times 20 \times 20$ $N = 77\,600$		$30 \times 30 \times 30$ $N = 264\,600$		$40 \times 40 \times 40$ $N = 630\,400$	
a	Iter	Cond	Iter	Cond	Iter	Cond
1	18	7.50	17	7.61	15	6.03
10	17	7.59	17	7.79	17	7.59
100	18	7.86	18	7.85	17	7.84
1 000	18	7.90	18	7.90	17	7.88
10^4	18	7.90	18	7.90	17	7.88
0.1	16	7.11	16	7.10	16	7.10
0.01	16	7.11	16	7.10	16	7.10
0.001	16	7.11	16	7.10	16	7.10
10^{-4}	16	7.11	16	7.10	16	7.10

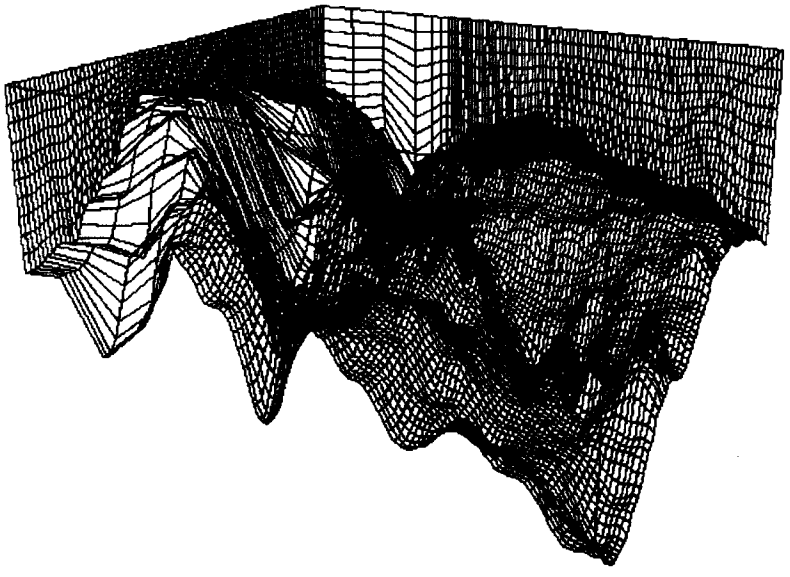


Figure 4. An example of the grid domain Ω

to achieve the desired accuracy; the computed condition number of matrix $B^{-1}A$ is equal to 10. All experiments were carried out on a Sun Workstation.

Acknowledgements

The authors are grateful to the referees for their helpful remarks and to Prof. Panayot Vassilevski for his valuable comments and suggestions.

The results of this paper are part of the Partnership in Computational Sciences (PICS) Project on Groundwater Contaminant Transport sponsored by the US Department of Energy under Grant #DE-FG05-92ER25143. This research together with the results published in [1,6,7,14,17,23,35] have given the theoretical foundation to the solution methods which have been implemented in the PICS codes. A detailed description of procedures to construct preconditioners using non-overlapping domain decomposition and domain decomposition on non-matching grids is given in [23].

REFERENCES

1. T. Arbogast and Z. Chen. On the implementation of mixed methods as nonconforming methods for second order elliptic problems. *Math. Comp.*, 64(211), 943–972, 1995.
2. D. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. *RAIRO Model. Math. Anal. Numer.*, 19, 7–32, 1985.
3. K. Arrow, L. Hurwicz and H. Uzawa. *Studies in Nonlinear Programming*. Stanford University Press, Stanford, CA, 1958.
4. O. Axelsson and P. Vassilevski. Algebraic multilevel preconditioning methods, I. *Numer. Math.*, 56, 157–177, 1989.
5. J. Bramble and J. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comp.*, 50, 1–18, 1990.
6. J. Bramble, J. Pasciak, and A. Vassilev. Analysis of the inexact Uzawa algorithm for saddle-point problems. Technical Report ISC-94-09-MATH, Institute for Scientific Computation, Texas A&M University, College Station, TX, 1994.
7. J. Bramble, R.Ewing, J. Pasciak, and J. Shen. Analysis of the multigrid algorithms for cell-centered finite difference approximations. *Advances in Comp. Mathematics*, 1996. (to appear).
8. F. Brezzi, J. J. Douglas, R. Duràn, and M. Fortin. Mixed finite elements for second order elliptic problems in three variables. *Numer. Math.*, 51, 237–250, 1987.
9. F. Brezzi, J. J. Douglas, M. Fortin and L. D. Marini. Efficient rectangular mixed finite elements in two and three space variables. *RAIRO Math. Model. Numer. Anal.*, 21, 581–604, 1987.
10. F. Brezzi, J. J. Douglas and L. Marini. Two families of mixed elements for second order elliptic problems. *Numer. Math.*, 88, 217–235, 1985.
11. Z. Chen. Analysis of mixed methods using conforming and nonconforming finite element methods. *RAIRO Math. Model. Numer. Anal.*, 27, 9–34, 1993.
12. Z. Chen. Equivalence between multigrid algorithms for mixed and nonconforming methods for second order elliptic problems. *East–West J. Numer. Math.*, 4, 1–33, 1996.
13. Z. Chen and J. Douglas. Prismatic mixed finite elements for second order elliptic problems. *Calcolo*, 26, 135–148, 1989.
14. Z. Chen, R. Ewing, and R. Lazarov. Domain decomposition algorithms for mixed methods for second order elliptic problems. *Math. Comp.*, 65, 467–490, 1996.
15. E. D'yakonov. On triangulations in the finite element method and efficient iterative methods. In I. Miller, editor, *Topics in Numerical Analysis, III*, pages 103–124. Academic Press, London, 1977.
16. H. Elman and G. Golub. Inexact and preconditioned Uzawa algorithms for saddle-point problems. *SIAM J. Numer. Anal.*, 31, 1645–1661, 1994.
17. R. Ewing, S. Maliassov, Y. Kuznetsov, and R. Lazarov. Substructure reconditioning for porous flow problems. In G. Garey, editor, *Finite Element Modeling of Environmental Problems*, pages 303–332, John Wiley & Sons Inc., New York, 1995.
18. R. Ewing and M. Wheeler. Computational aspects of mixed finite element methods. In

R. Stepleman, editor, *Numerical Methods for Scientific Computing*, pages 163–172, North-Holland, New York, 1983.
19. G. Golub and C. V. Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, PA, 1989.
20. Y. Hakopian and Y. Kuznetsov. Algebraic multigrid/substructuring preconditioners on triangular grids. *Sov. J. Numer. Anal. and Math Modeling*, 6, 453–484, 1991.
21. Y. Kuznetsov. Algebraic multigrid domain decomposition methods. *Soviet J. Numer. Anal. and Math. Modeling*, 4, 351–380, 1989.
22. Y. Kuznetsov. Multigrid domain decomposition methods for elliptic problems. *Comput. Meth. Appl. Mech. and Eng.*, 75, 185–193, 1989.
23. S. Maliassov. Optimal order preconditioners for mixed and nonconforming finite element approximations of elliptic problems with anisotropy. PhD thesis, Department of Mathematics, Texas A&M University, College Station, TX, March 1996.
24. G. Marchuk and Y. Kuznetsov. On optimal iteration processes. *Soviet Math. Dokl.*, 9(4), 1041–1045, 1968.
25. S. Margenov and P. Vassilevski. Algebraic multilevel preconditioning of anisotropic elliptic problems. *SIAM J. Scientific Computing*, 15(5), 1026–1037, 1994.
26. S. Margenov, L. Xanthis and L. Zikatanov. On the optimality of the semi-coarsening amli algorithm. In S. Margenov and P. Vassilevski, editors, *Second IMACS Int. Symp. on Iterative Methods in Linear Algebra*, pages 270–280, IMACS, Piscataway, NJ, 1995.
27. L. Marini. An inexpensive method for the evaluation of the solution of the lowest order Raviart-Thomas mixed method. *SIAM J. Numer. Anal.*, 22, 493–496, 1985.
28. M. Monga-Made and B. Polman. Theoretical comparison of pintwise, linewise and planewise preconditioners for solving three-dimensional problems. Preprint, 1996.
29. J. Nedelec. Mixed finite elements in $\mathbf{R}^3$. *Numer. Math.*, 35, 315–341, 1980.
30. J. Nedelec. A new family of mixed finite elements in $\mathbf{R}^3$. *Numer. Math.*, 50, 57–81, 1980.
31. P. Raviart and J. Thomas. A mixed finite element method for 2-nd order elliptic problems. In I. Galligani and E. Magenes, editors, *Mathematical Aspects of Finite Element Methods*, Lecture Notes in Mathematics, 606, pages 292–315, Springer-Verlag, New York, 1977.
32. T. Russell and M. Wheeler. Finite element and finite difference methods for continuous flows in porous media. In R. Ewing, editor, *The Mathematics of Reservoir Simulation*, pages 35–106. SIAM, Philadelphia, PA, 1983.
33. T. Rusten and R. Winter. A preconditioned iterative method for saddle point problems. *SIAM J. Matrix Anal.*, 13, 887–904, 1992.
34. A. Samarski and E. Nikolaev. *Méthodes de Résolution des Équations de Mailles*. Mir, Moscow, Russia, 1981.
35. P. Vassilevski and R. Lazarov. Preconditioning saddle-point problems arising from mixed finite element discretization of elliptic equations. *Numer. Lin. Alg. Appl.*, 3(1), 1–20, 1996.
36. P. S. Vassilevski. Hybrid V-cycle algebraic multilevel preconditioners. *Math. Comp.*, 58(198), 489–512, 1992.