Chapter 3

Numerical methods for coupled problem

This chapter is devoted to iterative algorithms for solving the discrete problem (2.18). We propose a few efficient solvers with optimal or quasi-optimal arithmetical complexity. Each of the solvers is based on a block diagonal preconditioner spectrally equivalent (in a sense of Definition 3.1) to the stiffness matrix $\mathcal{A}$ from (2.18). The main framework for deriving block diagonal preconditioners is formulated in Lemma 3.1 which appeared in a number of different forms in [39, 43, 66]. It states that a saddle point problem can be efficiently preconditioned by a block diagonal matrix with as many blocks as there are physical variables. In Section 3.2 we describe a preconditioner for a discrete Lamé operator. Using the Korn inequality we show that a bilinear form associated with the elastic energy generates a norm equivalent to the H^1-norm in a space of displacements. Therefore, a preconditioner for a discrete elliptic operator will be a preconditioner for the discrete Lamé operator. Section 3.3 is the core of this chapter. In Lemmas 3.2–3.4 we describe an abstract way for constructing preconditioners for a discrete diffusion operator given in a mixed form. For the case of topologically cubic meshes, Lemmas 3.5 and 3.7 specify abstract blocks in the preconditioner given by Lemma 3.4. The main result of Chapter 3 is summarized in Theorem 3.1. In Section 3.1 we present results of numerical experiments with a few preconditioners. We shall show that the Dendy multigrid with a Jacobi smoother is not robust enough for problems with severe jumps in coefficients. A simple remedy will be discussed in Chapter 4.

We restrict ourself to MFE discretizations on quasi-uniform partitions $\mathcal{T}_h$. The partition may be characterized by a positive number h such that

$$\frac{1}{c}\, h^3 < |e| < c\, h^3 \qquad \forall e \in \mathcal{T}_h$$

where c is a positive constant independent of h and e. Hereafter, we shall use a generic constant c which may have different values in different places. Contrary to c, constants with lower subscripts (e.g., c_1, c_2, etc.) will have the same values everywhere in this chapter.

To simplify the presentation of the main results, we consider partitions based on a cubic grid with mesh-step size h. In Sections 3.2 and 3.3, the partition $\mathcal{T}_h$ is obtained by splitting each cubic cell into 6 tetrahedra (see Fig. 3.1, left picture). In Section 4.2 we consider the cubic partition $\mathcal{T}_h$. In Section 4.3, the partition $\mathcal{T}_h$ is obtained by splitting each cubic cell into 2 prisms (see Fig. 3.1, right picture). However, we shall always discuss generalizations of the main results to more general partitions.

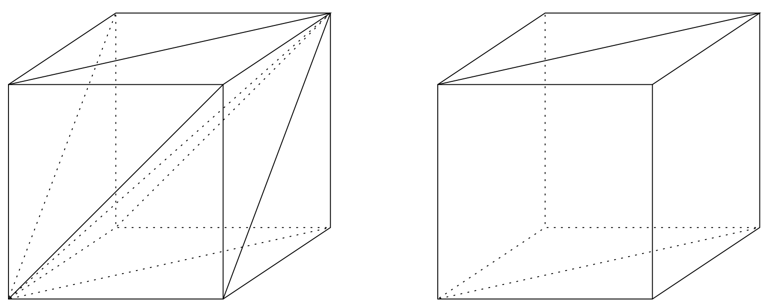


Figure 3.1: The splitting of a cube into 6 tetrahedra and 2 prisms.

3.1 Block diagonal preconditioners

Recall that the algebraic problem $\mathcal{A}x = b$ we are trying to solve reads:

$$\begin{pmatrix} A_{uu} & 0 & A_{pu}^T \\ 0 & A_{ff} & A_{pf}^T \\ A_{pu} & A_{pf} & -c_{pp}A_{pp} \end{pmatrix} \begin{pmatrix} u \\ f \\ p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ b_p \end{pmatrix}. \tag{3.1}$$

Let n_u, n_f and n_p be dimensions of spaces V_h, X_h and Q_h, respectively, and $n = n_u + n_f + n_p$ be the problem size. Then $u \in \Re^{n_u}$, $f \in \Re^{n_f}$ and $p,\ b_p \in \Re^{n_p}$.

Suppose that $\mathcal{B}$ is a positive definite matrix having the following block diagonal structure:

$$\mathcal{B} = \begin{pmatrix} B_u & 0 & 0 \\ 0 & B_f & 0 \\ 0 & 0 & B_p \end{pmatrix}. \tag{3.2}$$

Definition 3.1 *We say that the matrix $\mathcal{B}$ is spectrally equivalent to the saddle point matrix $\mathcal{A}$ ($\mathcal{A} \sim \mathcal{B}$ shortly) if the eigenvalues of the generalized eigenvalue problem*

$$\begin{pmatrix} A_{uu} & 0 & A_{pu}^T \\ 0 & A_{ff} & A_{pf}^T \\ A_{pu} & A_{pf} & -c_{pp}A_{pp} \end{pmatrix} \begin{pmatrix} u \\ f \\ p \end{pmatrix} = \nu \begin{pmatrix} B_u & 0 & 0 \\ 0 & B_f & 0 \\ 0 & 0 & B_p \end{pmatrix} \begin{pmatrix} u \\ f \\ p \end{pmatrix}$$

belong to the union of two segments $[d_1,\ d_2] \cup [d_3,\ d_4]$ with ends $d_1 \le d_2 < 0 < d_3 \le d_4$ independent of the discretization parameter h.

The following lemma is the basis for constructing spectrally equivalent block diagonal preconditioners [43, 66].

Lemma 3.1 *Suppose that diagonal blocks in the matrix $\mathcal{B}$ are such that*

$$c_1(B_u v,\, v) \le (A_{uu} v,\, v) \le c_2(B_u v,\, v) \qquad \forall v \in \Re^{n_u}, \tag{3.3a}$$

$$c_3(B_f g,\, g) \le (A_{ff} g,\, g) \le c_4(B_f g,\, g) \qquad \forall g \in \Re^{n_f}, \tag{3.3b}$$

$$c_5(B_p q,\, q) \le (S_p q,\, q) \le c_6(B_p q,\, q) \qquad \forall q \in \Re^{n_p}, \tag{3.3c}$$

where

$$S_p \equiv c_{pp} A_{pp} + A_{pf} A_{ff}^{-1} A_{pf}^T + A_{pu} A_{uu}^{-1} A_{pu}^T \tag{3.4}$$

is the Schur complement, and all constants are positive and independent of the discretization parameter h. Then $\mathcal{B} \sim \mathcal{A}$ with

$$d_1 = \frac{1-\sqrt{5}}{2}\theta_2, \quad d_2 = \frac{1-\sqrt{5}}{2}\theta_1, \quad d_3 = \theta_1 \quad \text{and} \quad d_4 = \frac{1+\sqrt{5}}{2}\theta_2$$

where $\theta_1 = \min\{c_1, c_3, c_5\}$ and $\theta_2 = \max\{c_2, c_4, c_6\}$.

The conditions of Lemma 3.1 are verified in the next sections. In order to solve the indefinite symmetric problem (3.1), we can apply the preconditioned Lanczos method of minimal residuals [55]. Let x^0 be an initial guess. Then,

$$\begin{aligned}
x^s &= x^{s-1} - \gamma_s r_s, \quad s \ge 1,\\
\xi^s &= \mathcal{A}x^s - b, \qquad s \ge 0,\\
r_s &= \begin{cases} \mathcal{B}^{-1}\xi^0, & s = 1,\\ \mathcal{B}^{-1}\mathcal{A}r_1 - \alpha_2 r_1, & s = 2,\\ \mathcal{B}^{-1}\mathcal{A}r_{s-1} - \alpha_s r_{s-1} - \beta_s r_{s-2}, & s \ge 3,\end{cases}\\
\alpha_s &= \frac{(\mathcal{A}\mathcal{B}^{-1}\mathcal{A}r_{s-1}, \mathcal{A}r_{s-1})_{\mathcal{B}^{-1}}}{||\mathcal{A}r_{s-1}||^2_{\mathcal{B}^{-1}}},\\
\beta_s &= \frac{||\mathcal{A}r_{s-1}||^2_{\mathcal{B}^{-1}}}{||\mathcal{A}r_{s-2}||^2_{\mathcal{B}^{-1}}}, \qquad \gamma_s = \frac{(\xi^{s-1}, \mathcal{A}r_{s-1})_{\mathcal{B}^{-1}}}{||\mathcal{A}r_s||^2_{\mathcal{B}^{-1}}}.
\end{aligned} \tag{3.5}$$

If the eigenvalues of $\mathcal{B}^{-1}\mathcal{A}$ belong to the union of two segments $[d_1,\ d_2] \cup [d_3,\ d_4]$ with $d_1 \le d_2 < 0 < d_3 \le d_4$, then, the convergence estimate

$$||\xi^{2s}||_{\mathcal{B}^{-1}} \le 2\left(\frac{\kappa-1}{\kappa+1}\right)^s ||\xi^0||_{\mathcal{B}^{-1}}, \qquad \kappa = \frac{\max\{d_4, |d_1|\}}{\min\{d_3, |d_2|\}}$$

holds. It follows that the convergence rate is independent of h when $\mathcal{B}$ is spectrally equivalent to $\mathcal{A}$.

3.2 Preconditioners for the elastic block

In this section we shall prove that A_{uu} is a positive definite matrix and construct an easily invertible preconditioner B_u satisfying (3.3a). The proof is based on the Korn inequality which states that

$$\|\boldsymbol{v}\|_{1,\Omega} \le c\left(\|D(\boldsymbol{v})\|^2 + \|\boldsymbol{v}\|^2\right)^{1/2} \qquad \forall \boldsymbol{v} \in V, \tag{3.6}$$

where c is a constant independent of $\boldsymbol{v}$. The proof of (3.6) can be found in [31]. Applying this inequality to the bilinear form $\mathcal{A}(\boldsymbol{u}, \boldsymbol{v})$, we get that it generates a norm in space V equivalent to the H^1-norm:

$$c_1\|\boldsymbol{v}\|_1^2 \le \mathcal{A}(\boldsymbol{v},\, \boldsymbol{v}) \le c\|\boldsymbol{v}\|_1^2 \qquad \forall \boldsymbol{v} \in V. \tag{3.7}$$

Let us consider a bilinear form

$$\mathcal{B}(\boldsymbol{u},\, \boldsymbol{v}) = \int_\Omega \nabla \boldsymbol{u} : \nabla \boldsymbol{v}\, \mathrm{d}x \qquad \forall \boldsymbol{u}, \boldsymbol{v} \in V$$

which also generates a norm in V equivalent to H^1-norm. Let B_u be a matrix induced by this bilinear form in the discrete space V_h. It is obvious that B_u is symmetric, positive definite and spectrally equivalent to A_{uu}, i.e.,

$$c_1(B_u v,\, v) \le (A_{uu} v,\, v) \le c_2(B_u v,\, v) \qquad \forall v \in \Re^{n_u}. \tag{3.8}$$

Arithmetical complexity of $B_u^{-1} v$

Note that matrix B_u is a block diagonal matrix with three diagonal blocks. In the case of cubic meshes, each block is a separable matrix. Then, a fast direct method based on separation of variables can be used for computing $B_u^{-1}v$, $v \in \Re^{n_u}$. Its arithmetical complexity is estimated as $O(n_u \log^2 n_u)$ or $O(n \log^2 n)$. Moreover, if meshes are uniform, the FFT can be used giving a better estimate for the arithmetical complexity. In the case of a general topologically cubic mesh, the direct computation of $B_u^{-1}v$ is impossible. However, the computational mesh can be mapped onto an orthogonal mesh where a separable matrix $\hat{B}_u$ can be constructed. The classical finite element theory states that condition number of $\hat{B}_u^{-1}B_u$ depends on $C^{1,\infty}$-norm of this mapping [23]. When it is small enough, $\hat{B}_u^{-1}$ may be used in (3.8) instead of B_u.

The direct solver can be replaced by a multigrid method with the optimal arithmetical complexity. In this case the cost of computing $B_u^{-1}v$ is $O(n_u)$. The powerful multigrids, like algebraic multigrids [7, 15, 70] can be applied for unstructured meshes. Finally, B_u may be replaced by any preconditioner for the Laplace operator. The libraries of such preconditioners are numerous (see e.g., [7, 14, 34, 42] and references therein).

3.3 Preconditioners for the diffusion block

The diffusion part of the coupled system is the most complicated one for preconditioning. The rest of this chapter is dedicated to an analysis of different methods. We begin with a straightforward approach based on Lemma 3.1 where a choice of preconditioners for both the flux and pressure blocks is described. In the next chapter we consider a more promising approach based on an algebraic hybridization of system (3.1).

3.3.1 Preconditioner for matrix A_{ff}

Let B_f be a diagonal matrix spectrally equivalent to A_{ff}. One of the possible choices for constructing B_f is to apply a lumping procedure to A_{ff}. However, we choose B_f

to be the main diagonal of A_{ff}. It is known that for a quasi-uniform mesh and an isotropic tensor $\boldsymbol{K}$, B_f is spectrally equivalent to A_{ff}, i.e.,

$$c_3(B_f g,\, g) \le (A_{ff} g,\, g) \le c_4(B_f g,\, g) \qquad \forall g \in \Re^{n_f} \tag{3.9}$$

with positive constants c_3 and c_4 independent of the discretization parameter h. For the tetrahedral partition shown on Fig. 3.1 (left picture), the element-by-element analysis gives $c_3 = 0.5$ and $c_4 = 2$.

Arithmetical complexity of $B_f^{-1} g$

Since the matrix B_f is diagonal, $B_f^{-1} g$ can be computed with n_f operations, $n_f = O(n)$. Note that (3.9) remains valid for an arbitrary unstructured regular mesh and an isotropic tensor $\boldsymbol{K}$.

3.3.2 Preconditioner for Schur complement S_p

The construction of preconditioner B_p for the Schur complement S_p involves four steps. We shall use notations $S_p^{(k)}$, $k = 1, 2, 3, 4$, for successive approximations to S_p. First, the elastic part $A_{pu} A_{uu}^{-1} A_{pu}^T$ in the Schur complement can be omitted. Let us consider the following matrix

$$S_p^{(1)} = c_{pp} A_{pp} + A_{pf} A_{ff}^{-1} A_{pf}^T. \tag{3.10}$$

Lemma 3.2 *Let either $c_{pp} > 0$ or $|\Gamma| > 0$. Then $S_p^{(1)}$ is spectrally equivalent to S_p.*

<u>Proof</u>. Note that for any pair $(\boldsymbol{v}^h,\, q^h) \in V_h \times Q_h$, the Cauchy-Schwartz inequality gives

$$(q^h,\, \nabla \cdot \boldsymbol{v}^h)^2 \le 2\, \|q^h\|^2\, \|\nabla \boldsymbol{v}^h\|^2 \qquad \forall q^h \in Q_h,\ \forall \boldsymbol{v}_h \in V_h.$$

The algebraic form of the last inequality is

$$(q,\, A_{pu} v)^2 \le 2 (A_{pp} q,\, q)\, (B_u v,\, v) \qquad \forall q \in \Re^{n_p},\ \forall v \in \Re^{n_u}.$$

Applying estimate (3.8) to the last inequality and taking $v = A_{uu}^{-1} A_{pu}^T q$, we get

$$(A_{pu} A_{uu}^{-1} A_{pu}^T q,\, q) \le \frac{2}{c_1} (A_{pp} q,\, q) \qquad \forall q \in \Re^{n_p}.$$

The assertion of the lemma follows from the fact that $c_{pp} > 0$. If $c_{pp} = 0$ but $|\Gamma| > 0$, the discrete LBB condition (2.16) has to be exploited. The algebraic form of (2.16) and definition of a norm in the vector space $\Re^{n_p}$ give:

$$\begin{aligned} \sup_{g \in \Re^{n_f}} \frac{(q,\, A_{pf}\, g)}{(A_{ff} g,\, g)^{1/2}} &= \sup_{w \in \Re^{n_f}} \frac{(q,\, A_{pf} A_{ff}^{-1/2} w)}{\|w\|} \\ &\equiv \|A_{ff}^{-1/2} A_{pf}^T\, q\| \ge \tilde{\beta}_h (A_{pp} q,\, q)^{1/2} \qquad \forall q \in \Re^{n_p}, \end{aligned}$$

where $w = A_{ff}^{-1/2} g$, and $\|\cdot\| \equiv (\cdot,\, \cdot)^{1/2}$ is the Euclidean norm. Since $\tilde{\beta}_h > 0$, the assertion of the lemma follows from the last inequality. $\square$

Second, we replace the matrix $S_p^{(1)}$ by a spectrally equivalent matrix

$$S_p^{(2)} = c_{pp} A_{pp} + A_{pf} B_f^{-1} A_{pf}^T.$$

The third step is logically the most complicated one. It is based on a new technique invented in [44] and described in recent publications [45, 47]. We begin with splitting the matrix

$$L_p = A_{pf} B_f^{-1} A_{pf}^T$$

into substructures. Then, we modify the substructures and assemble them back to get a simplified matrix spectrally equivalent to L_p.

Let us represent the matrix L_p in the form

$$L_p = \sum_{k=1}^{\hat{m}} \mathcal{N}_k L_k \mathcal{N}_k^T \tag{3.11}$$

with local "stiffness" matrices $L_k \in \Re^{n_k \times n_k}$ and the corresponding assembling matrices $\mathcal{N}_k$. The sparsity structure of L_p is shown in Fig. 3.2. Each cubic cell contains 6 degrees of freedom for the pressure function p. They are associated with the vertices of a right six-sided polygon (hexagon). The non-zero off-diagonal entices of L_p are associated with both hexagon edges and dashed lines connecting vertices of neighboring hexagons.

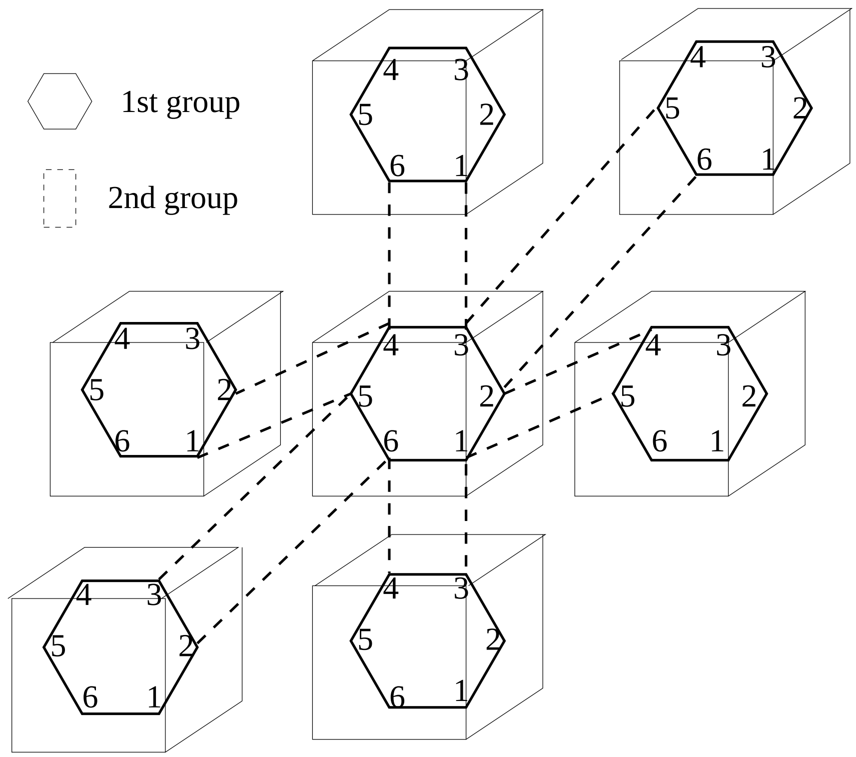


Figure 3.2: The sparsity structure of matrix L_p.

For the sake of simplicity of further analysis of L_p, henceforth we assume that the diffusion tensor $\boldsymbol{K}$ is the identity matrix, Ω is the unit cube, and $\Gamma = \emptyset$. In this case the set of pairs $\{L_k; \mathcal{N}_k\}$ can be partitioned into two groups. The first group consists of $\hat{m}_1 = h^{-3}$ pairs $\{L_k \in \Re^{6\times6}; \mathcal{N}_k \in \Re^{n_p\times6}\}$, and the second group consists of $\hat{m}_2 = 3(h^{-1} - 1)^3$ pairs $\{L_k \in \Re^{4\times4}; \mathcal{N}_k \in \Re^{n_p\times4}\}$. Thus, $\hat{m} = \hat{m}_1 + \hat{m}_2$. It can be easily shown that for the second group we always have

$$L_k = \frac{h}{24}\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}. \tag{3.12}$$

If we denote the pairs in the first group by $\{L_{1,k}; \mathcal{N}_{1,k}\}$, and in the second group by $\{L_{2,k}; \mathcal{N}_{2,k}\}$, then, instead of (3.11) we can use

$$L_p = \sum_{k=1}^{\hat{m}_1} \mathcal{N}_{1,k} L_{1,k} \mathcal{N}_{1,k}^T + \sum_{k=1}^{\hat{m}_2} \mathcal{N}_{2,k} L_{2,k} \mathcal{N}_{2,k}^T. \tag{3.13}$$

Let index k be associated with a cubic cell ω_k and s_k be the number of sides of ω_k belonging to $\partial\Omega$, $0 \le s_k \le 3$. Then

$$L_{1,k} = \sum_{t=1}^{6} \beta_{kt} \mathcal{N}_{kt} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \mathcal{N}_{kt}^T$$

where

$$\beta_{kt} = h/12, \qquad 1 \le t \le s_k,$$

$$\beta_{kt} = h/24, \qquad s_k + 1 \le t \le 6,$$

and $\mathcal{N}_{kt} \in \Re^{6\times 2}$ are the corresponding assembling matrices. If ω_k is an interior cell, all β_{kt} are equal to $h/24$, and we have

$$L_{1,k} = \frac{h}{24}\begin{pmatrix} 2 & -1 & 0 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}.$$

The matrices $L_{2,k}$ are given by (3.12).

Let us denote the identity $s \times s$ matrix by I_s and define the local "mass" matrices

$$\begin{aligned} M_{1,k} &= h^2 \sum_{t=1}^{6} \beta_{kt} \mathcal{N}_{kt} \mathcal{N}_{kt}^T, \qquad && 1 \le k \le \hat{m}_1, \\ M_{2,k} &= \frac{h^3}{24} I_4, && 1 \le k \le \hat{m}_2. \end{aligned} \tag{3.14}$$

The triples $\{L_{1,k};\, M_{1,k};\, \mathcal{N}_{1,k}\}$ and $\{L_{2,k};\, M_{2,k};\, \mathcal{N}_{2,k}\}$ are called *substructures* associated with the matrix L_p. In general, the substructures are introduced algebraically (e.g., (3.11) and (3.14)) or formed from FE stiffness and mass matrices. It is easy to check that

$$M_p = \sum_{k=1}^{\hat{m}_1} \mathcal{N}_{1,k} M_{1,k} \mathcal{N}_{1,k}^T + \sum_{k=1}^{\hat{m}_2} \mathcal{N}_{2,k} M_{2,k} \mathcal{N}_{2,k}^T = \frac{h^3}{6} I_{n_p}.$$

We remark, that $A_{pp} = M_p$.

Let us consider generalized eigenvalue problems

$$L_k w = \nu M_k w$$

with the pairs $\{L_k;\, M_k\}$ equal to either $\{L_{1,k};\, M_{1,k}\}$ or $\{L_{2,k};\, M_{2,k}\}$. Denote by W_k the matrix of M_k-orthogonal eigenvectors, $W_k = [w_k^{(1)}, \ldots, w_k^{(n_k)}]$, $(w_k^{(i)})^T M_k w_k^{(j)} = \delta_{ij}$, $1 \le i, j \le n_k$, and by Λ_k the matrix of eigenvalues:

$$\Lambda_k = \mathrm{diag}\{\nu_k^{(1)}, \ldots, \nu_k^{(n_k)}\}, \qquad 0 = \nu_k^{(1)} < \nu_k^{(2)} \le \ldots \le \nu_k^{(n_k)}.$$

We can prove that

$$\frac{4}{h^2} \sin^2(\frac{\pi}{6}) \le \nu_k^{(2)} \quad \text{and} \quad \nu_k^{(n_k)} \le \frac{4}{h^2}. \tag{3.15}$$

Using the spectral decomposition of matrix L_k, $L_k = M_k W_k \Lambda_k W_k^T M_k$, and estimates (3.15), we can easily prove that L_k is spectrally equivalent to the following matrix [45]:

$$S_k^{(3)} = \frac{4}{h^2} M_k W_k \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} W_k^T M_k = \frac{4}{h^2} \left[M_k - M_k w_k^{(1)} (w_k^{(1)})^T M_k \right].$$

Lemma 3.3 *The matrix $S_p^{(3)}$ defined by*

$$S_p^{(3)} = c_{pp}A_{pp} + \sum_{k=1}^{\hat{m}} \mathcal{N}_k S_k^{(3)} \mathcal{N}_k^T$$

is spectrally equivalent to the matrix $S_p^{(2)}$, i.e.,

$$\frac{1}{c}(S_p^{(3)}q,\, q) \le (S_p^{(2)}q,\, q) \le c(S_p^{(3)}q,\, q) \qquad \forall q \in \Re^{n_p}, \tag{3.16}$$

where c is independent of the discretization parameter h.

Proof. The standard steps in the proof of similar propositions are as follows. First, simple algebra gives:

$$\frac{(S_p^{(2)}q,\, q)}{(S_p^{(3)}q,\, q)} \ge \min_{1\le k\le \hat{m}} \frac{(L_k\mathcal{N}_k^T q,\, \mathcal{N}_k^T q)}{(S_k^{(3)}\mathcal{N}_k^T q,\, \mathcal{N}_k^T q)}$$

Second, the spectral decompositions of L_k and $S_k^{(3)}$ imply that

$$(L_k q_k,\, q_k) \ge \sin^2(\frac{\pi}{6})\,(S_k^{(3)} q_k,\, q_k), \qquad q_k = \mathcal{N}_k^T q.$$

This gives the lower estimate in (3.16). The upper estimate is proved similarly. $\square$

Let us consider a solution algorithm for a system

$$S_p^{(3)} q = b, \qquad b \in \Re^{n_p}, \tag{3.17}$$

with the matrix $S_p^{(3)}$ from Lemma 3.3. We shall reduce equation (3.17) to a more simple equation $\hat{S}\hat{q} = \hat{b}$. Hereafter, hats will denote transformed matrices and vectors. Since $A_{pp} = M_p$, the definition of matrices $S_k^{(3)}$ yields:

$$S_p^{(3)} = dA_{pp} - \frac{4}{h^2}\sum_{k=1}^{\hat{m}} \mathcal{N}_k M_k w_k^{(1)} (w_k^{(1)})^T M_k \mathcal{N}_k^T, \qquad d = c_{pp} + 4/h^2.$$

Since A_{pp} is the diagonal matrix, the solution vector q can be computed as follows:

$$q = \frac{1}{d}\, A_{pp}^{-1} b + \frac{4}{d\, h^2} \sum_{k=1}^{\hat{m}} \hat{q}_k A_{pp}^{-1} \mathcal{N}_k M_k w_k^{(1)} \tag{3.18}$$

where $\hat{q}_k$ is the mean value of q corresponding to the k-th substructure:

$$\hat{q}_k = (w_k^{(1)})^T M_k \mathcal{N}_k^T q = (M_k w_k^{(1)}, \mathcal{N}_k^T q), \qquad 1 \le k \le \hat{m}.$$

The following trick is used for computing the mean values. Multiplying both sides of (3.18) by $(w_l^{(1)})^T M_l \mathcal{N}_l^T$, $1 \le l \le \hat{m}$, and using the definition of $\hat{q}_l$, we get that the vector $\hat{q} = (\hat{q}_1, \ldots, \hat{q}_{\hat{m}})^T$ is the solution of the system

$$\hat{S}\hat{q} = \hat{b}, \qquad \hat{b} \in \Re^{\hat{m}}, \tag{3.19}$$

where $\hat{S}$ is a symmetric positive definite matrix with entries

$$\hat{s}_{lk} = \delta_{lk} - \frac{4}{d\,h^2}(w_l^{(1)})^T M_l \mathcal{N}_l^T A_{pp}^{-1} \mathcal{N}_k M_k w_k^{(1)}, \qquad 1 \le k, l \le \hat{m}.$$

Recall that δ_{lk} denotes the Kronecker symbol. The components of the right hand side vector $\hat{b}$ are given by

$$\hat{b}_l = d^{-1}(w_l^{(1)})^T M_l \mathcal{N}_l^T A_{pp}^{-1} b = d^{-1}(M_l w_l^{(1)}, \mathcal{N}_l^T A_{pp}^{-1} b), \qquad 1 \le l \le \hat{m}.$$

Thus, the solution vector q is computed by formulae (3.19) and (3.18). For theoretical reasons, one can formally write the matrix $S_p^{(3)}$ in an implicit form:

$$\left[S_p^{(3)}\right]^{-1} = \frac{1}{d} A_{pp}^{-1} + \frac{4}{d^2\,h^2} \sum_{k,l=1}^{\hat{m}} \hat{t}_{kl} A_{pp}^{-1} \mathcal{N}_k M_k w_k^{(1)} (w_l^{(1)})^T M_l \mathcal{N}_l^T A_{pp}^{-1} \tag{3.20}$$

where $\hat{t}_{kl}$, $1 \le k, l \le \hat{m}$, are entries of the matrix $\hat{T} = \hat{S}^{-1}$.

Lemma 3.4 *Let the matrix $\tilde{T} = \left\{\tilde{t}_{kl}\right\}_{k,l=1}^{\hat{m}}$ be spectrally equivalent to the matrix $\hat{T}$ and*

$$\left[S_p^{(4)}\right]^{-1} = \frac{1}{d} A_{pp}^{-1} + \frac{4}{d^2\,h^2} \sum_{k,l=1}^{\hat{m}} \tilde{t}_{kl} A_{pp}^{-1} \mathcal{N}_k M_k w_k^{(1)} (w_l^{(1)})^T M_l \mathcal{N}_l^T A_{pp}^{-1}.$$

Then the matrices $S_p^{(4)}$ and $S_p^{(3)}$ are spectrally equivalent.

<u>Proof</u>. It is sufficient to note that

$$\left[S_p^{(3)}\right]^{-1} = d^{-1} A_{pp}^{-1} + P^T \hat{T} P$$

where $P\colon \Re^{n_p} \to \Re^{\hat{m}}$ is a linear operator (a matrix) implicitly given by (3.20). $\square$

Remark 3.1 *Lemmas 3.2 – 3.4 show an abstract way for constructing preconditioners for the Schur complement S_p; namely, the design of preconditioner $B_p = S_p^{(4)}$ consists in choosing substructures $\{L_k;\ M_k;\ \mathcal{N}_k\}$, $k = 1, \ldots, \hat{m}$, and finding an appropriate matrix $\tilde{S}$ spectrally equivalent to $\hat{S}$ (or a matrix $\tilde{T}$ spectrally equivalent to $\hat{T}$).*

Hereafter we consider one of the possible ways for constructing the matrix $\tilde{S}$. Let us rewrite matrix $\hat{S}$ in a 2x2 block form:

$$\hat{S} = \begin{pmatrix} \hat{S}_{11} & \hat{S}_{12} \\ \hat{S}_{21} & \hat{S}_{22} \end{pmatrix}$$

where the first and the second diagonal blocks correspond to the first and the second groups of substructures in (3.13), respectively. It is obvious that the matrix $\hat{S}_{11}$ is diagonal.

Lemma 3.5 *Let $\tilde{S}_{22}$ be the diagonal lumping of $\hat{S}_{22}$ and*

$$\tilde{S}^{(1)} = \begin{pmatrix} \hat{S}_{11} & \hat{S}_{12} \\ \hat{S}_{21} & \tilde{S}_{22} \end{pmatrix}.$$

Then the matrices $\tilde{S}^{(1)}$ and $\hat{S}$ are spectrally equivalent.

The proof is based on the superelement analysis of blocks of $\hat{S}$. The basis idea for lumping of $\hat{S}_{22}$ is discussed in [51] and [60, Chapter 6]. It is quite technical and we omit it. Instead, we shall study numerically the spectral properties of the resulting preconditioner in the next section. Let us introduce a Schur complement

$$\tilde{R}_{11}^{(1)} = \hat{S}_{11} - \hat{S}_{12}\tilde{S}_{22}^{-1}\hat{S}_{21}.$$

Lemma 3.6 *Let matrix $\tilde{R}_{11}^{(2)}$ be spectrally equivalent to $\tilde{R}_{11}^{(1)}$ and*

$$\tilde{S}^{(2)} = \begin{pmatrix} \tilde{R}_{11}^{(2)} + \hat{S}_{12}\tilde{S}_{22}^{-1}\hat{S}_{21} & \hat{S}_{12} \\ \hat{S}_{21} & \tilde{S}_{22} \end{pmatrix}. \tag{3.21}$$

Then matrices $\tilde{S}^{(2)}$ and $\tilde{S}^{(1)}$ are spectrally equivalent.

Proof. It is sufficient to show that the eigenvalues of the generalized eigenvalue problem

$$\tilde{S}^{(2)}\hat{q} = \nu \tilde{S}^{(1)}\hat{q}$$

are bounded independently of the discretization parameter h. Let $\nu \neq 1$. Then, the eigenvalue problem is equivalent to

$$\tilde{R}_{11}^{(2)}\hat{q}_1 = \nu \tilde{R}_{11}^{(1)}\hat{q}_1$$

where $\hat{q} = (\hat{q}_1,\, \hat{q}_2)^T$. The assertion of the lemma follows from spectral equivalence of matrices $\tilde{R}_{11}^{(2)}$ and $\tilde{R}_{11}^{(1)}$. $\square$

Let us consider a solution algorithm for system $\tilde{S}^{(2)}\hat{q} = \hat{b}$. Since $\tilde{S}_{22}$ is the diagonal matrix, $\hat{q}_2$ may be eliminated from the system resulting in a system with matrix $\tilde{R}_{11}^{(1)}$.

Lemma 3.7 *The matrix $\tilde{R}_{11}^{(1)}$ is spectrally equivalent to the matrix*

$$\tilde{R}_{11}^{(2)} = (K_0 + c_{pp}M_0) \otimes M_0 \otimes M_0 + M_0 \otimes K_0 \otimes M_0 + M_0 \otimes M_0 \otimes K_0 \tag{3.22}$$

where $K_0 \in \Re^{n\times n}$ is the tridiagonal matrix,

$$K_0 = \frac{1}{h}\begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}, \qquad M_0 = h \ \mathrm{diag}\{0.5, 1, \ldots, 1, 0.5\},$$

and $n = 1/h$.

The matrix $\tilde{R}_{11}^{(2)}$ represents the well known seven-point stencil discretization of the Laplace operator in the unit cube with Neumann boundary conditions. Another choice for $\tilde{R}_{11}^{(2)}$ is a multigrid preconditioner for the matrix $\tilde{R}_{11}^{(1)}$. Finally, the preconditioner $B_p = S_p^{(4)}$ is defined by Lemmas 3.2 – 3.7.

Arithmetical complexity of $B_p^{-1}q$

The most time-consuming step in computing $B_p^{-1}q$ is the solution of an algebraic problem with matrix $\tilde{S}^{(2)}$. The other steps require just $O(n_p)$ operations. Let us consider a problem $\tilde{S}^{(2)}\hat{q} = \hat{b}$. It turns out that it is reduced to an algebraic problem with a separable matrix $\tilde{R}_{11}^{(2)}$. Therefore the arithmetical complexity of computing $B_p^{-1}q$ is estimated as $O(n_p \log^2 n_p)$ or $O(n \log^2 n)$. The direct solver can be again replaced by a multigrid method with the optimal arithmetical complexity. In this case the arithmetical cost of $B_p^{-1}q$ is $O(n_p)$.

Theorem 3.1 *Under the assumptions made above, the block diagonal preconditioner $\mathcal{B}$ is spectrally equivalent to the matrix $\mathcal{A}$.*

The results of this subsection are easily extended to other types of boundary conditions as well as to other shapes of the domain Ω. The latter usually means that $\mathcal{T}_h$ is an unstructured quasi-uniform partition and $\tilde{R}_{11}^{(2)}$ is a multigrid preconditioner. Generally speaking, for the case of arbitrary meshes, an algebraic multigrid [7, 15, 70] may be applied directly to a system with matrix $S_p^{(2)}$.

Remark 3.2 *One of the technical problems with block diagonal preconditioners is the adequate scaling of diagonal blocks. In other words, the constants c_1, c_3 and c_5 in (3.3) must be of the same order of magnitude (coincide in the ideal case). Practically, a choice of scaling parameters requires one inversion of $\mathcal{B}$ and elemental analysis of block weights.*

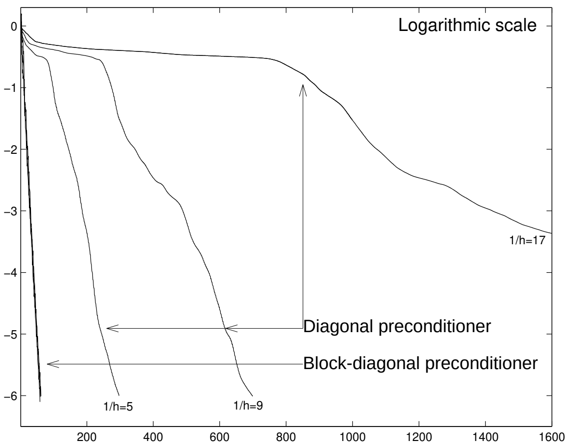
3.4 Numerical experiments

Let Ω be the unit cube, $c_{pp} = 1$, and Γ be the top side of $\partial\Omega$, i.e., $\Gamma = \{(x,\ y,\ 1) \in \partial\Omega\}$. Define the right hand side in (2.15) as follows:

$$S_f(x,y,z) = \begin{cases} 1, & \text{if } 0.25 \le x \le 0.75, \quad 0.25 \le y \le 0.75, \quad z \le 0.5, \\ 0, & \text{otherwise.} \end{cases}$$

In all numerical experiments the Lanczos iterations (3.5) are stopped when the Euclidean norm of the initial residual is decreased by a factor of 10^6. The time values correspond to DEC Alpha workstation (500 MHz) and are measured in seconds.

In the first experiment we take $\boldsymbol{K} = \boldsymbol{I}$ and $\lambda = \mu = 1$. A method of separation of variables is used for solving algebraic problems with matrix $\tilde{R}_{11}^{(2)}$ from Lemma 3.7. As it is predicted by the theory, no dependence on the discretization parameter h is observed (see Fig. 3.3). The number of Lanczos iterations (#itr) is between 61 and 67 for all discretizations. Since we used direct methods for computing $B_u^{-1}v$ and $\left[\tilde{R}_{11}^{(2)}\right]^{-1}\hat{q}_1$, the time $t_{\mathcal{B}^{-1}}$ needed to compute $\mathcal{B}^{-1}\xi$ grows slightly faster than the problem size n.



1/h	n	#itr	$t_{\mathcal{B}^{-1}}$,s
5	3990	61	0.0006
9	19374	62	0.005
17	116958	64	0.05
33	805182	66	1.2
49	2580894	67	3.9

Figure 3.3: Comparison with a diagonal preconditioner.

The comparison of the preconditioner $\mathcal{B}$ with a diagonal preconditioner is shown on Fig. 3.3. The absolute values of elements lying on the main diagonal of the stiffness

$1/h$	n	$\mu = \lambda = 1$	$\mu = 100,\ \lambda = 1$	$\mu = 1,\ \lambda = 100$
5	3990	72	71	71
9	19374	86	86	86
17	116958	116	114	114
25	357390	137	138	138

Table 3.1: The number of Lanczos iterations with the Dendy algorithm.

matrix $\mathcal{A}$ form the non-zero entries of the diagonal preconditioner. Obviously such a simple preconditioner can be used only for very rough discretizations.

In real applications, the diffusion tensor $\boldsymbol{K}$ is usually a heterogeneous tensor with strongly varying entries. If so, the matrix $\tilde{R}_{11}^{(1)}$ can not be optimally preconditioned by a separable matrix. Therefore, we replace the exact computation of $\left[\tilde{R}_{11}^{(1)}\right]^{-1} \hat{q}_1$ by a multigrid method. It turns out that simple multigrid schemes do not result in matrices spectrally equivalent to $\tilde{R}_{11}^{(1)}$. The main reason is that $\tilde{R}_{11}^{(1)}$ corresponds to a mesh approximation where the surfaces of tensor discontinuity do not coincide with the mesh surfaces.

For the second experiment, we realized a more sophisticated algebraic multigrid method known in the literature as the Dendy algorithm [1]. Like the classical multigrids, it uses geometric ideas for constructing a sequence of coarse meshes but matrix-dependent interpolation and restriction operators. The method behaves similarly for all choices of the Lamé coefficients; however, a moderate dependence on the discretization parameter h is observed (see Table 3.1). There are two possible reasons for that. First, the weights of diagonal blocks are not optimal. Second, the Jacobi relaxation steps fail to provide a robust convergence rate (see [34]).

In the third experiment, the diffusion tensor is isotropic, $\boldsymbol{K} = K\boldsymbol{I}$, but piecewise constant, and $\lambda = \mu = 1$. Suppose that the initial cubic grid is decomposed into $s \times s \times s$ big cubes and K is piecewise constant in each of these cubes.

Let $s = 3$ (resp., $s = 5$) and K have jump discontinuities along z-axes taking values 100, 1 and 0.01 (resp., 100, 0.01, 0.1, 1 and 10) as z runs from 0 to 1. The results of the numerical experiment are shown in Table 3.2 and Fig. 3.4. The number of Lanczos iterations (#itr) is close to the similar number in Table 3.1, especially for $s = 3$. This means that the preconditioner $\mathcal{B}$ is slightly sensitive to jumps in coefficients. However, when $s = 5$, the unstable behavior of the Dendy algorithm slows the convergence rate. Analysis of the convergence histories shown on Fig. 3.4 implies that the Dendy algorithm does not provide a spectrally equivalent preconditioner for the matrix $\tilde{R}_{11}^{(1)}$ for both values of s.

$1/h$	n	#itr	$t_{\mathcal{B}^{-1}}, s$
9	19374	91	0.030
15	81690	111	0.165
21	215286	135	0.617

$1/h$	n	#itr	$t_{\mathcal{B}^{-1}}, s$
5	3990	89	0.004
15	81690	147	0.166
25	357390	192	1.064

Table 3.2: The number of Lanczos iterations for $s = 3$ (on the left) and $s = 5$.

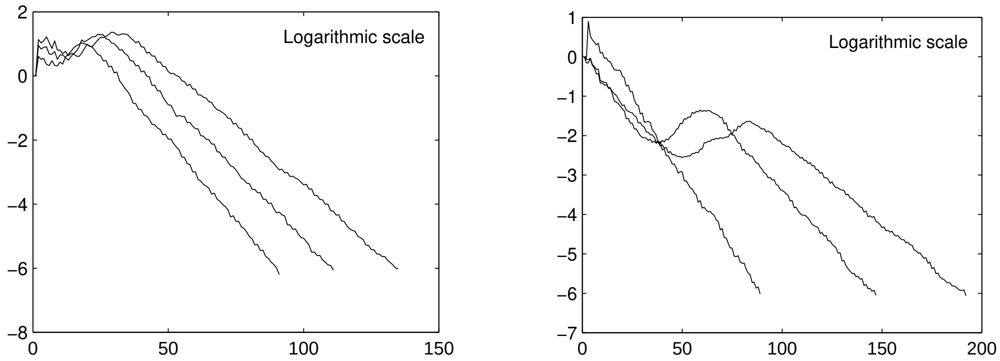


Figure 3.4: Convergence histories for the Euclidean norm of residual: $s = 3$ (on the left) and $s = 5$.

Let $s = 3$ and the diffusion coefficient K have jump discontinuities along both z- and x-directions, i.e., $K = K(x)K(z)$. Furthermore, let the first factor take values 1,

10^{α}, and 1 as x runs from 0 to 1, and the second factor take values 100, 1 and 0.01 as z runs from 0 to 1. Comparing with the previous experiment, we have about 18% (resp., 43%) more iterations for $\alpha = -3$ (resp., $\alpha = 3$) (see Tables 3.2 and 3.3).

$1/h$	n	#itr, $\alpha = -3$	#itr, $\alpha = 3$	$t_{\mathcal{B}^{-1}}, s$
9	19374	108	127	0.030
15	81690	128	159	0.165
21	215286	157	194	0.616

Table 3.3: The number of Lanczos iterations for $s = 3$ and $K = K(x)K(z)$.

The theory described in the previous sections is not applicable for anisotropic diffusion tensors because neither B_f nor B_p satisfy Lemma 3.1. Therefore, we run only one experiment to study the method behavior. Let $s = 3$, $K = K(z)\text{diag}\{1, 1, 10^{\alpha}\}$ and $K(z)$ take values 100, 1 and 0.01 as z runs from 0 to 1. The results of numerical experiments for a few different values of α are shown in Table 3.4. It seems that the preconditioner, oriented to isotropic problems, can not be applied for strongly anisotropic problems.

$1/h$	n	#itr, $\alpha = 0$	#itr, $\alpha = 1$	#itr, $\alpha = 2$	$t_{\mathcal{B}^{-1}}, s$
9	19374	91	254	1538	0.030
15	81690	111	285	1704	0.165
21	215286	135	311	1801	0.617

Table 3.4: The number of Lanczos iterations for an anisotropic diffusion tensor.

Our experiments show that the convergence rate of Lanczos iterations strongly depends on preconditioning properties of B_p. In the next chapter we shall try to overcome this problem by modifying the original system of algebraic equations.