Symmetric part preconditioning of the CGM for Stokes type saddle-point systems

by O. Axelsson[1] and J. Karátson[2]

Abstract

A nonsymmetric formulation of saddle-point systems is considered and symmetric part preconditioning of the CGM is applied. Linear and superlinear convergence estimates are derived for the finite element solution of the Stokes problem and of Navier's equations of elasticity.

Key words: conjugate gradient method, preconditioning, symmetric part, regularization, Stokes problem, Navier's equations of elasticity

1 Introduction

Saddle-point problems arise as mathematical models in various applications and have been a subject of intense investigation, e.g. [5, 11, 23, 26]. Besides the widespread Uzawa type methods, an efficient way of solving such problems is the preconditioned conjugate gradient method.

In this paper we consider nonsymmetric formulations of saddle-point systems, following [12]. For nonsymmetric problems, symmetric part preconditioning has proved an efficient tool for the iterative solution (see e.g. [13, 25] and the authors' papers [8, 9]). Hereby Hilbert space background has often been used, in particular, in many PDE settings the convergence results follow from the theory of equivalent operators [14, 15]. We present linear and superlinear convergence estimates of the symmetric part PCG method in the context of saddle-point problems, and apply these results to the finite element solution of the Stokes problem and of Navier's equations of elasticity. The preliminary estimates are given in sections 2 and 3, whereas the applications to the mentioned two models, which form the main part of the paper, are presented in sections 4 and 5.

2 Preliminaries: the GCG-LS method and symmetric part preconditioning

Let us consider a nonsymmetric system of linear algebraic equations

$$\mathbf{L}\mathbf{x} = \mathbf{b} \tag{1}$$

with some given $\mathbf{b} \in \mathbf{R}^n$, where $\mathbf{L} \in \mathbf{R}^{n\times n}$. Here one assumes that

$$\mathbf{L} + \mathbf{L}^T > 0, \tag{2}$$

which ensures the well-posedness of (1). Then, denoting by $\mathbf{x}^*$ the unique solution of (1), we study the error vector $\mathbf{e}_k = \mathbf{x}_k - \mathbf{x}^*$ of the CGM.

[1]Department of Information Technology, Uppsala University, Sweden & Institute of Geonics AS CR, Ostrava, Czech Republic; owea@it.uu.se

[2]Department of Applied Analysis, ELTE University, H-1117 Budapest, Hungary; karatson@cs.elte.hu
The second author was supported by the Hungarian Research Grant OTKA No. T 043765.

2.1 Construction

The generalized conjugate gradient, least square (GCG-LS) method is defined in [2]. Two versions are discussed: the full version which uses all previous search directions, whereas the truncated version uses only a bounded number of search directions. Here we are interested in the so-called truncated GCG-LS(0) method, which requires only a single, namely the current search direction. (The algorithm will be given in preconditioned form below.) As a special case of Theorem 4.1 in [2], it follows that the two algorithms coincide when $\mathbf{L}^T$ is a linear polynomial of $\mathbf{L}$.

'Symmetric part preconditioning' means using the symmetric part

$$\mathbf{M} = (\mathbf{L} + \mathbf{L}^T)/2$$

as preconditioner for the system matrix $\mathbf{L}$. It has been introduced and analysed in [13, 25], see also [2, 8, 24]. Then the $\mathbf{M}$-adjoint $(\mathbf{M}^{-1}\mathbf{L})_M^T$ of the preconditioned matrix $\mathbf{M}^{-1}\mathbf{L}$ satisfies

$$(\mathbf{M}^{-1}\mathbf{L})_M^T = 2\mathbf{I} - \mathbf{M}^{-1}\mathbf{L} \tag{3}$$

(where $\mathbf{I}$ is the identity matrix), hence the above cited result [2, Theorem 4.1] states that the truncated GCG-LS(0) method for the preconditioned equation

$$\mathbf{M}^{-1}\mathbf{L}\mathbf{x} = \mathbf{M}^{-1}\mathbf{b} \tag{4}$$

coincides with the full version. That is, the algorithm is as follows:

$$\begin{cases} (a) & \text{Let } \mathbf{x}_0 \in \mathbf{R}^n \text{ be arbitrary, and let} \\ & \mathbf{r}_0 \text{ be the solution of } \mathbf{M}\mathbf{r}_0 = \mathbf{L}\mathbf{x}_0 - \mathbf{b}; \quad \mathbf{d}_0 = -\mathbf{r}_0; \\ & \quad \text{for any } k \in \mathbf{N}: \text{ when } \mathbf{x}_k, \mathbf{d}_k, \mathbf{r}_k \text{ are obtained, let} \\ (b1) & \mathbf{e}_k \text{ be the solution of } \mathbf{M}\mathbf{e}_k = \mathbf{L}\mathbf{d}_k, \\ & \gamma_k = \langle \mathbf{M}\mathbf{e}_k, \mathbf{e}_k \rangle, \quad \alpha_k = -\frac{1}{\gamma_k}\langle \mathbf{r}_k, \mathbf{M}\mathbf{e}_k \rangle; \\ (b2) & \mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k; \\ (b3) & \mathbf{r}_{k+1} = \mathbf{r}_k + \alpha_k \mathbf{e}_k; \\ (b4) & \beta_k = \frac{1}{\gamma_k}\langle \mathbf{L}\mathbf{r}_{k+1}, \mathbf{e}_k \rangle; \\ (b5) & \mathbf{d}_{k+1} = -\mathbf{r}_{k+1} + \beta_k \mathbf{d}_k. \end{cases} \tag{5}$$

2.2 Convergence

Similarly to the case of symmetric problems [6, 7], the convergence behaviour of the CGM often exhibits a successive linear and a superlinear phase as well for nonsymmetric problems. Such convergence estimates are found e.g. in [2, 3]. In the study of symmetric part preconditioning, we often use the identity $\mathbf{M}^{-1}\mathbf{L} = \mathbf{I} + \mathbf{M}^{-1}\mathbf{N}$, where

$$\mathbf{N} = (\mathbf{L} - \mathbf{L}^T)/2$$

denotes the antisymmetric part, and we consider $\mathbf{M}^{-1}\mathbf{L}$ as a perturbation of the identity matrix. The required convergence results are summarized in the following theorem:

Theorem 2.1 *Let (2) hold. Then the algorithm (5) yields the following estimates, where* $\lambda_i(\mathbf{M}^{-1}\mathbf{N})$ *are the eigenvalues of* $\mathbf{M}^{-1}\mathbf{N}$*, and the norm* $\|\mathbf{x}\|_{\mathbf{M}} = \langle \mathbf{M}\mathbf{x}, \mathbf{x}\rangle^{1/2}$ *is used:*

(1) (linear estimate)

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}}}{\|\mathbf{e}_0\|_{\mathbf{M}}}\right)^{1/k} \le \frac{|\lambda_{max}(\mathbf{M}^{-1}\mathbf{N})|}{\sqrt{1+|\lambda_{max}(\mathbf{M}^{-1}\mathbf{N})|^2}} \qquad (k=1,2,\ldots,n); \tag{6}$$

(2) (superlinear type estimate)

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}}}{\|\mathbf{e}_0\|_{\mathbf{M}}}\right)^{1/k} \le \frac{2}{k}\sum_{i=1}^{k}|\lambda_i(\mathbf{M}^{-1}\mathbf{N})| \qquad (k=1,2,\ldots,n). \tag{7}$$

Proof. (1) The preconditioning of (1) by $\mathbf{M}$ in the GCG-LS method means that the residuals are minimized w.r.t. the norm $\|\mathbf{x}\|_{\mathbf{M}} = \langle \mathbf{M}\mathbf{x}, \mathbf{x}\rangle^{1/2}$. Hence Theorem 2.2 in [2] yields

$$\left(\frac{\|\mathbf{e}_k\|^2_{\mathbf{M}}}{\|\mathbf{e}_0\|^2_{\mathbf{M}}}\right)^{1/k} \le 1 - \frac{1}{\left\|\left[\frac{1}{2}(\tilde{\mathbf{L}}+\tilde{\mathbf{L}}^T)\right]^{-1}\right\| \ \left\|\left[\frac{1}{2}(\tilde{\mathbf{L}}^{-1}+\tilde{\mathbf{L}}^{-T})\right]^{-1}\right\|} \tag{8}$$

with $\tilde{\mathbf{L}} = \mathbf{M}^{-1/2}\mathbf{L}\,\mathbf{M}^{-1/2}$. Here $\tilde{\mathbf{L}}+\tilde{\mathbf{L}}^T = \mathbf{I}+\mathbf{M}^{-1/2}\mathbf{N}\,\mathbf{M}^{-1/2}+\mathbf{I}-\mathbf{M}^{-1/2}\mathbf{N}\,\mathbf{M}^{-1/2} = 2\mathbf{I}$, hence the first norm in the denominator of (8) equals 1. Further, by (3), $\mathbf{M}^{-1}\mathbf{L}$ is $\mathbf{M}$-normal and hence it has an $\mathbf{M}$-orthonormal eigensystem $\mathbf{v}_1, ..., \mathbf{v}_n$. Letting for simplicity $\lambda_j = \lambda_j(\mathbf{M}^{-1}\mathbf{L})$, for any vector $\mathbf{y} = \sum_{j=1}^{n} d_j\mathbf{M}^{1/2}\mathbf{v}_j$ we obtain

$$\frac{1}{2}(\tilde{\mathbf{L}}^{-1}+\tilde{\mathbf{L}}^{-T})\mathbf{y} = \sum_{j=1}^{n}\frac{1}{2}(\frac{1}{\lambda_j}+\frac{1}{\overline{\lambda_j}})d_j\mathbf{M}^{1/2}\mathbf{v}_j = \sum_{i=1}^{n}\Big(\mathrm{Re}\frac{1}{\lambda_j}\Big)\,d_j\mathbf{M}^{1/2}\mathbf{v}_j.$$

Similarly, for any $\mathbf{x} = \sum_{j=1}^{n} c_j\mathbf{M}^{1/2}\mathbf{v}_j$ we have

$$\Big[\frac{1}{2}(\tilde{\mathbf{L}}^{-1}+\tilde{\mathbf{L}}^{-T})\Big]^{-1}\mathbf{x} = \sum_{j=1}^{n}\Big(\frac{1}{\mathrm{Re}\frac{1}{\lambda_j}}\Big)\,c_j\mathbf{M}^{1/2}\mathbf{v}_j.$$

Since $\mathbf{M}^{1/2}\mathbf{v}_i$ are orthonormal, we obtain

$$\left\|\left[\frac{1}{2}(\tilde{\mathbf{L}}^{-1}+\tilde{\mathbf{L}}^{-T})\right]^{-1}\right\| \le \max_{j=1,\ldots,n}\Big|\frac{1}{\mathrm{Re}\frac{1}{\lambda_j(\mathbf{M}^{-1}\mathbf{L})}}\Big|.$$

Here $\lambda_j(\mathbf{M}^{-1}\mathbf{L}) = 1+i|\mu_j|$ (where $i^2=-1$) where $\mu_j = \lambda_j(\mathbf{M}^{-1}\mathbf{N})$ since $\mathbf{M}^{-1}\mathbf{N}$ is $\mathbf{M}$-antisymmetric and hence has imaginary eigenvalues. Hence for all j, we have

$$\mathrm{Re}\frac{1}{\lambda_j(\mathbf{M}^{-1}\mathbf{L})} = \frac{1}{1+|\mu_j|^2}, \qquad \left\|\left[\frac{1}{2}(\tilde{\mathbf{L}}^{-1}+\tilde{\mathbf{L}}^{-T})\right]^{-1}\right\| \le \max_{j=1,\ldots,n}(1+|\mu_j|^2),$$

so (8) yields

$$\left(\frac{\|\mathbf{e}_k\|^2_{\mathbf{M}}}{\|\mathbf{e}_0\|^2_{\mathbf{M}}}\right)^{1/k} \le \max_{j=1,\ldots,n}\frac{|\mu_j|^2}{1+|\mu_j|^2},$$

which is the desired estimate.

(2) This is a consequence of [9, Corollary 1]. ∎

Remark 2.1 Similar convergence results for preconditioned Lanczos methods are found in [25].

The case of main interest is when the eigenvalues of $\mathbf{M}^{-1}\mathbf{L}$ cluster around 1. Then the eigenvalues of $\mathbf{M}^{-1}\mathbf{N}$ cluster around 0 and (7) becomes indeed a superlinear convergence estimate. This is the case when $\mathbf{M}$ and $\mathbf{N}$ arise from the discretization of an appropriate infinite-dimensional problem, moreover, one may have mesh independence of the estimate. Namely, let us consider the following setting. For brevity, we will use the following

Definition 2.1 *The numbers* $\lambda_i(K)$ $(i = 1, \ldots, \infty)$ *are the ordered eigenvalues of the operator* K *if each of them is repeated as many times as its multiplicity and* $|\lambda_1(K)| \geq |\lambda_2(K)| \geq \ldots$

Proposition 2.1 *Let* H_M *be a complex separable Hilbert space with inner product* $\langle ., . \rangle_M$ *and let* $K : H_M \to H_M$ *be a bounded linear operator which is antisymmetric, i.e.* $K^* = -K$ *on* H_M*. Let us consider the operator equation*

$$(I + K)u = g \tag{9}$$

with some given $g \in H$*, where* I *is the identity operator, let* $W_n = span\{\varphi_1, \ldots, \varphi_n\} \subset H_M$ *be a given subspace of dimension* n *and*

$$(\mathbf{M}_h)_{i,j} = \langle \varphi_i, \varphi_j \rangle_M \,, \qquad (\mathbf{N}_h)_{i,j} = \langle K\varphi_i, \varphi_j \rangle_M \,. \tag{10}$$

Then for all $k = 1, \ldots, n$

(1) $|\lambda_{max}({\mathbf{M}_h}^{-1}\mathbf{N}_h)| \leq \|K\|$*;*

(2) if K *is also compact then*

$$\sum_{i=1}^{k} |\lambda_i(\mathbf{M}_h^{-1}\mathbf{N}_h)| \leq \sum_{i=1}^{k} |\lambda_i(K)|,$$

where $\lambda_i(K)$ *are the ordered eigenvalues of* K*.*

Proof. (1) Let λ be any eigenvalue of ${\mathbf{M}_h}^{-1}\mathbf{N}_h$ and $\mathbf{c}$ a corresponding eigenvector, $u = \sum_{i=1}^{n} c_i\varphi_i$. Then

$$|\lambda| = \left| \frac{\mathbf{N}_h\, \mathbf{c} \cdot \mathbf{c}}{\mathbf{M}_h\, \mathbf{c} \cdot \mathbf{c}} \right| = \left| \frac{\langle Ku, u \rangle_M}{\langle u, u \rangle_M} \right| \leq \|K\|.$$

(2) The result follows from the proof of [9, Theorem 5].

Remark 2.2 Let $(\mathbf{L}_h)_{i,j} = \langle (I + K)\varphi_i, \varphi_j \rangle_M$. Owing to the antisymmetry of K, the matrices $\mathbf{M}_h$ and $\mathbf{N}_h$ in (10) are the symmetric and antisymmetric parts of $\mathbf{L}_h$, respectively.

Then Theorem 2.1 and Proposition 2.1 yield

Corollary 2.1 *Under the conditions of Proposition 2.1, the algorithm (5) for the matrices* $\mathbf{M}_h$ *and* $\mathbf{L}_h = \mathbf{M}_h + \mathbf{N}_h$ *yields the following estimates (for all* $k = 1, \ldots, n$*), where* K *is the operator in (9):*

(1) there holds

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_h}}{\|\mathbf{e}_0\|_{\mathbf{M}_h}}\right)^{1/k} \leq \frac{\|K\|}{\sqrt{1+\|K\|^2}};$$

(2) if K *is also compact then*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_h}}{\|\mathbf{e}_0\|_{\mathbf{M}_h}}\right)^{1/k} \leq \varepsilon_k, \tag{11}$$

$$\textit{where} \quad \varepsilon_k = \frac{2}{k}\sum_{i=1}^{k}|\lambda_i(K)| \to 0 \qquad (\textit{as } k \to \infty) \tag{12}$$

and ε_k *is a sequence independent of* n *and* W_n*.*

3 Nonsymmetric formulation of saddle-point systems and symmetric part preconditioning

Let us consider a saddle-point system

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & -\mathbf{C} \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \tag{13}$$

where $\mathbf{A}$ and $\mathbf{S} = \mathbf{C} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ are SPD. It has been proposed in [12] to use a nonsymmetric formulation of (13) by changing the sign of the second row,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ -\mathbf{B} & \mathbf{C} \end{pmatrix}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ -\mathbf{b} \end{pmatrix}. \tag{14}$$

Then by suitable preconditioning one can cluster the eigenvalues around 1 in the complex plane [11, 12], and for such nonsymmetric matrices convergence estimates of the CGM are available [3].

We study the case when $\mathbf{A}$ and $\mathbf{C}$ are SPD, and we are interested in the superlinear convergence of the preconditioned CGM under symmetric part preconditioning. This convergence result is readily derived from Section 2.

Namely, let $\mathbf{L}$ denote the system matrix of (14). Its symmetric and antisymmetric parts are

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \qquad \text{and} \qquad \mathbf{N} = \begin{pmatrix} \mathbf{0} & \mathbf{B}^T \\ -\mathbf{B} & \mathbf{0} \end{pmatrix}, \tag{15}$$

respectively, therefore the identity $\mathbf{M}^{-1}\mathbf{L} = \mathbf{I} + \mathbf{M}^{-1}\mathbf{N}$ implies that the eigenvalues of $\mathbf{M}^{-1}\mathbf{L}$ are the perturbations of 1 in the complex plane by the eigenvalues of

$$\mathbf{M}^{-1}\mathbf{N} = \begin{pmatrix} \mathbf{0} & \mathbf{A}^{-1}\mathbf{B}^T \\ -\mathbf{C}^{-1}\mathbf{B} & \mathbf{0} \end{pmatrix}. \tag{16}$$

Then the assumptions $\mathbf{A}, \mathbf{C} > 0$ imply $\mathbf{L} + \mathbf{L}^T = 2\mathbf{M} > 0$, i.e. (2) holds and hence Theorem 2.1 applies for system (14) involving the eigenvalues of (16). Further, if the matrices (15) are stiffness matrices of the form (10), then Proposition 2.1 and Corollary 2.1 are valid.

We note that in various applications $\mathbf{C} = \mathbf{0}$ in (13), i.e. the assumption $\mathbf{C} > 0$ fails. However, for such systems one sometimes uses some regularization procedure which just yields that a matrix $\mathbf{C} = \sigma\mathbf{G}$ appears instead of $\mathbf{0}$ in the bottom right-hand side block, where $\mathbf{G}$ is some SPD matrix and σ is a positive, small parameter. Next we consider such a regularization for the Stokes problem.

4 A superlinear PCG algorithm for a regularized Stokes problem

4.1 The regularized Stokes problem

The presentation in this subsection is based on [5].

Let us consider the classical Stokes problem of the form

$$\begin{cases} -\Delta \mathbf{u} + \nabla p = \mathbf{f} \\ \operatorname{div} \mathbf{u} = 0 \\ u_{|\partial\Omega} = 0 \end{cases} \tag{17}$$

in a bounded domain $\Omega \subset \mathbf{R}^N$ ($N = 2$ or 3) with a Lipschitz boundary. Here $\mathbf{f} \in L^2(\Omega)^N$.

Since p in (17) is determined up to an additive constant only, for uniqueness one introduces the space

$$L_0^2(\Omega) := \{p \in L^2(\Omega) : \int_\Omega p = 0\}. \tag{18}$$

The standard weak formulation then reads as follows, where the involved Sobolev spaces are defined to consist of real-valued functions. Find $(\mathbf{u}, p) \in H_0^1(\Omega)^N \times L_0^2(\Omega)$ satisfying

$$\begin{cases} \int_\Omega \nabla \mathbf{u} \cdot \nabla \mathbf{v} - \int_\Omega p\,(\operatorname{div} \mathbf{v}) = \int_\Omega \mathbf{f} \cdot \mathbf{v} & (\forall \mathbf{v} \in H_0^1(\Omega)^N) \\ \int_\Omega q\,(\operatorname{div} \mathbf{u}) = 0 & (\forall q \in L_0^2(\Omega)), \end{cases} \tag{19}$$

where the notation $\nabla \mathbf{u} \cdot \nabla \mathbf{v} := \sum_{i=1}^N \nabla u_i \cdot \nabla v_i$ is used. Then problem (19) has a unique weak solution. (Moreover [19, 23], if $\partial\Omega$ is sufficiently regular then $(\mathbf{u}, p) \in H^2(\Omega)^N \times H^1(\Omega)$.)

For the finite element solution of (19) one chooses suitable FEM subspaces $V_h \subset H_0^1(\Omega)^N$ and $P_h \subset L_0^2(\Omega)$, and wants to find $(\mathbf{u}_h, p_h) \in V_h \times P_h$ satisfying

$$\begin{cases} \int_\Omega \nabla \mathbf{u}_h \cdot \nabla \mathbf{v}_h - \int_\Omega p_h\,(\operatorname{div} \mathbf{v}_h) = \int_\Omega \mathbf{f} \cdot \mathbf{v}_h & (\forall \mathbf{v}_h \in V_h) \\ \int_\Omega q_h\,(\operatorname{div} \mathbf{u}_h) = 0 & (\forall q_h \in P_h). \end{cases} \tag{20}$$

Here a crucial issue with the choice of V_h and P_h is to satisfy the LBB-condition, which is not always straightforward (see [5] for a discussion). Therefore an important effort has

been done to circumvent this problem and define regularized versions of (20), which are consistent with the original problem but allow equal-order approximation (i.e. both the velocity and pressure are looked for in H^1).

We study the following regularized version of the Stokes problem, formulated for the discrete case (20). It is taken from [5] with the modification that the real Sobolev spaces are replaced by complex ones (this will be required to apply the theory of normal operators from [9] to our problem). Let the FEM subspaces

$$V_h \subset H_0^1(\Omega)^N, \qquad P_h \subset \dot{H}^1(\Omega) := H^1(\Omega) \cap L_0^2(\Omega)$$

consist of piecewise linear functions. We fix a parameter $\sigma > 0$, and want to find $(\mathbf{u}_h, p_h) \in V_h \times P_h$ satisfying

$$\left\{ \begin{array}{lll} \displaystyle\int_\Omega \nabla \mathbf{u}_h \cdot \nabla \overline{\mathbf{v}}_h \;-\; \int_\Omega p_h \,(\operatorname{div} \overline{\mathbf{v}}_h) = \int_\Omega \mathbf{f} \cdot \overline{\mathbf{v}}_h & \qquad (\forall \mathbf{v}_h \in V_h) \\ \displaystyle\int_\Omega (\operatorname{div} \mathbf{u}_h)\, \overline{q}_h + \sigma \int_\Omega \nabla p_h \cdot \nabla \overline{q}_h = \sigma \int_\Omega \mathbf{f} \cdot \nabla \overline{q}_h & \qquad (\forall q_h \in P_h). \end{array} \right. \tag{21}$$

Problem (21) is the special case of (3.5)-(3.6) in [5] with the choice (3.8) there (see [17, 19] and the other references in [5]).

Remark 4.1 Formally, in terms of the strong form (17) and assuming sufficient regularity, the regularization in the second row comes from adding the relation $-\sigma\Delta p = -\sigma \operatorname{div} \mathbf{f}$, which follows from taking the divergence of $-\Delta \mathbf{u} + \nabla p = \mathbf{f}$ and using $\operatorname{div} \mathbf{u} = 0$.

4.2 The superlinear PCG algorithm

Our goal is to verify the superlinear convergence of the CGM when a symmetric part preconditioning is used for problem (21), based on sections 2 and 3. First we rewrite problem (21) in order to balance the parameter σ in the diagonal. Letting

$$s_h := \sigma^{1/2}\, p_h \,,$$

problem (21) is equivalent to

$$\left\{ \begin{array}{lll} \displaystyle\int_\Omega \nabla \mathbf{u}_h \cdot \nabla \overline{\mathbf{v}}_h \;-\; \sigma^{-1/2} \int_\Omega s_h \,(\operatorname{div} \overline{\mathbf{v}}_h) = \int_\Omega \mathbf{f} \cdot \overline{\mathbf{v}}_h & \qquad (\forall \mathbf{v}_h \in V_h) \\ \displaystyle\sigma^{-1/2} \int_\Omega (\operatorname{div} \mathbf{u}_h)\, \overline{q}_h + \int_\Omega \nabla s_h \cdot \nabla \overline{q}_h = \sigma^{1/2} \int_\Omega \mathbf{f} \cdot \nabla \overline{q}_h & \qquad (\forall q_h \in P_h). \end{array} \right. \tag{22}$$

Note that problem (22) leads to an algebraic system of the form (14) with SPD blocks in the diagonal. To formulate this algebraic system, the following obvious notations will be used for the discrete operators arising from the four integrals on the left-hand side of (22). Let ∇_h, div_h, Δ_h^0 and Δ_h^ν denote the Gram matrices of the operators ∇, div, Δ with Dirichlet boundary conditions and Δ with Neumann boundary conditions, respectively, in the considered subspaces. Further, let $diag_N(-\Delta_h^0)$ denote the block diagonal matrix with $-\Delta_h^0$ blocks repeated N times. Then the algebraic system corresponding to (22) can be written as follows:

$$\begin{pmatrix} diag_N(-\Delta_h^0) & \sigma^{-1/2}\, \nabla_h \\ \sigma^{-1/2} \operatorname{div}_h & -\Delta_h^\nu \end{pmatrix} \begin{pmatrix} \xi_h \\ \eta_h \end{pmatrix} = \begin{pmatrix} \mathbf{f}_h \\ \sigma^{-1/2} \operatorname{div} \mathbf{f}_h \end{pmatrix}, \tag{23}$$

where ξ_h and η_h are the coefficient vectors of $\mathbf{u}_h$ and p_h in the given basis of V_h and P_h, respectively. Following the notations of Section 3, we let

$$\mathbf{L}_h = \begin{pmatrix} diag_N(-\Delta_h^0) & \sigma^{-1/2}\,\nabla_h \\ \sigma^{-1/2}\,\mathrm{div}_h & -\Delta_h^\nu \end{pmatrix}, \tag{24}$$

and for its symmetric and antisymmetric parts

$$\mathbf{M}_h = \begin{pmatrix} diag_N(-\Delta_h^0) & \mathbf{0} \\ \mathbf{0} & -\Delta_h^\nu \end{pmatrix} \qquad \text{and} \qquad \mathbf{N}_h = \begin{pmatrix} \mathbf{0} & \sigma^{-1/2}\,\nabla_h \\ \sigma^{-1/2}\,\mathrm{div}_h & \mathbf{0} \end{pmatrix}. \tag{25}$$

Here $\mathbf{L}_h \in \mathbf{R}^{n\times n}$ where $n = dim(V_h) + dim(P_h)$.

Theorem 4.1 *There exists a sequence*

$$\varepsilon_k \to 0 \qquad (as\ \ k \to \infty),$$

depending on σ and Ω, such that the PCG algorithm (5) for the matrices $\mathbf{M}_h$ and $\mathbf{L}_h$ in (24)-(25) yields

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_h}}{\|\mathbf{e}_0\|_{\mathbf{M}_h}}\right)^{1/k} \le\ \varepsilon_k \qquad (k = 1, \ldots, n). \tag{26}$$

In particular, there holds

$$\varepsilon_k = \ \frac{2}{\sigma^{1/2}k}\ \sum_{i=1}^{k} |\lambda_i(Q)| \tag{27}$$

where Q is a compact linear operator defined below in (29).

Proof. We check the conditions of Proposition 2.1, statement (2). Let us consider the complex separable Hilbert space

$$H_M := H_0^1(\Omega)^N \times \dot{H}^1(\Omega)$$

with the inner product

$$\left\langle \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M := \int_\Omega \nabla \mathbf{u}_h \cdot \nabla \overline{\mathbf{v}}_h \ + \int_\Omega \nabla s_h \cdot \nabla \overline{q}_h\,. \tag{28}$$

We now verify that the following relation defines a compact linear operator $Q : H_M \to H_M$ which is antisymmetric:

$$\left\langle Q \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M = -\int_\Omega s\,(\mathrm{div}\,\overline{\mathbf{v}}) + \int_\Omega (\mathrm{div}\,\mathbf{u})\,\overline{q} \qquad \left(\forall\ \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \in H_M\right). \tag{29}$$

Namely, here $\mathbf{u} \in H_0^1(\Omega)^N$ and the divergence theorem implies

$$\int_\Omega (\mathrm{div}\,\mathbf{u})\,\overline{q}\ = -\int_\Omega \mathbf{u} \cdot \nabla \overline{q}\,,$$

hence the r.h.s. of (29) satisfies

$$\left| -\int_\Omega s\,(\mathrm{div}\,\overline{\mathbf{v}}) + \int_\Omega (\mathrm{div}\,\mathbf{u})\,\overline{q} \right| \le \|s\|_{L^2(\Omega)} \|\nabla \mathbf{v}\|_{L^2(\Omega)^N} + \|\mathbf{u}\|_{L^2(\Omega)^N} \|\nabla q\|_{L^2(\Omega)^N}$$

$$\leq \left\| \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} \right\|_{L^2(\Omega)^{N+1}} \left\| \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\|_M \tag{30}$$

where the product spaces $L^2(\Omega)^N$ and $L^2(\Omega)^{N+1}$ are endowed with the usual Euclidean-like norms. Let us fix $\begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} \in H_M$. Then (30) yields that the r.h.s. of (29) defines a bounded linear functional on H_M in the variable $\begin{pmatrix} \mathbf{v} \\ q \end{pmatrix}$, hence the Riesz theorem ensures the existence of the vector $Q\begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} \in H_M$ satisfying (29). Clearly $Q : H_M \to H_M$ is linear. Further, (29)-(30) also shows that one can define Q on $L^2(\Omega)^{N+1}$ and that $Q : L^2(\Omega)^{N+1} \to H_M$ is bounded. Since the embedding of $H^1(\Omega)$ into $L^2(\Omega)$ is compact [1] and the norm of H_M is equivalent to the standard $H^1(\Omega)^N$-norm, therefore the embedding $E : H_M \to L^2(\Omega)^{N+1}$ is also compact. That is, $Q : H_M \to H_M$ is the composition of a compact and a bounded operator and hence it is also compact. Finally, relation (29) shows directly that

$$\left\langle Q\begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M = -\left\langle \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, Q\begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M ,$$

i.e. Q_M is is antisymmetric.

A similar argument with the Riesz theorem ensures the existence of the vector $\begin{pmatrix} \mathbf{g} \\ r \end{pmatrix} \in H_M$ satisfying

$$\left\langle \begin{pmatrix} \mathbf{g} \\ r \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M = \int_\Omega \mathbf{f} \cdot \overline{\mathbf{v}} + \sigma^{1/2} \int_\Omega \mathbf{f} \cdot \nabla \overline{q} \qquad \left(\forall \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \in H_M \right).$$

Now let

$$K = \sigma^{-1/2} Q. \tag{31}$$

Then the operator equation

$$(I + K)\begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} = \begin{pmatrix} \mathbf{g} \\ r \end{pmatrix} \tag{32}$$

(where I is the identity operator on H_M) can be written as

$$\left\langle \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M + \sigma^{-1/2} \left\langle Q\begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M = \left\langle \begin{pmatrix} \mathbf{g} \\ r \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M \qquad \left(\forall \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \in H_M \right),$$

that is,

$$\int_\Omega \left(\nabla \mathbf{u} \cdot \nabla \overline{\mathbf{v}} + \nabla s \cdot \nabla \overline{q} \right) + \sigma^{-1/2} \int_\Omega \left(-s\,(\operatorname{div} \overline{\mathbf{v}}) + (\operatorname{div} \mathbf{u})\,\overline{q} \right) = \int_\Omega \left(\mathbf{f} \cdot \overline{\mathbf{v}} + \sigma^{1/2}\, \mathbf{f} \cdot \nabla \overline{q} \right) \tag{33}$$

$$(\forall \mathbf{v} \in H_0^1(\Omega)^N,\ \forall q \in \dot{H}^1(\Omega)).$$

This is equivalent to the system

$$\begin{cases} \displaystyle\int_\Omega \nabla \mathbf{u} \cdot \nabla \overline{\mathbf{v}} \ - \ \sigma^{-1/2} \int_\Omega s\,(\operatorname{div} \overline{\mathbf{v}}) = \int_\Omega \mathbf{f} \cdot \overline{\mathbf{v}} & (\forall \mathbf{v} \in H_0^1(\Omega)^N) \\ \sigma^{-1/2} \displaystyle\int_\Omega (\operatorname{div} \mathbf{u})\,\overline{q} + \int_\Omega \nabla s \cdot \nabla \overline{q} \ = \sigma^{1/2} \int_\Omega \mathbf{f} \cdot \nabla \overline{q} & (\forall q \in \dot{H}^1(\Omega)) \end{cases} \tag{34}$$

since here $\mathbf{v}$ and q are varied independently. That is, system (22) is the FEM discretization of the operator equation (32) in the subspace $V_h \times P_h$.

The conditions of Proposition 2.1 are thus satisfied for the operator equation (32). In virtue of Remark 2.2, the symmetric and antisymmetric parts $\mathbf{M}_h$ and $\mathbf{N}_h$ of $\mathbf{L}_h$ in (25) arise in the form of (10) for this operator equation. Consequently, using also (31), Corollary 2.1 yields that (26) is valid with

$$\varepsilon_k = \frac{2}{k} \sum_{i=1}^{k} |\lambda_i(K)| = \frac{2}{\sigma^{1/2} k} \sum_{i=1}^{k} |\lambda_i(Q)| \,. \tag{35}$$

Here the eigenvalues $\lambda_i(Q)$ depend only on Ω, hence ε_k depends on σ and Ω. The compactness of Q implies that $\varepsilon_k \to 0$. ∎

Remark 4.2 *(Efficiency).* The crucial part of the preconditioned CG algorithm (5) for problem (23) is step (b1), i.e., the solution of the auxiliary problems $\mathbf{M}_h \mathbf{e}_k = \mathbf{L}_h \mathbf{d}_k$ with system matrix (25). This requires the FEM solution of N(=2 or 3) Poisson equations with homogeneous Dirichlet b.c. and one additional Poisson equation with homogeneous Neumann b.c. Note the following:

(i) These auxiliary problems are standard ones for which various fast solvers are available (like fast Fourier transform, cyclic reduction, multigrid or multilevel, see e.g. [3, 16, 20, 22])

(ii) The auxiliary problems are uncoupled and can be solved in parallel.

(iii) The work added by the regularization is the extra Neumann problem. The Poisson equations are present in the standard linearly convergent Uzawa algorithms (see e.g. [26]), and also extra Neumann problems are present in some preconditioned versions [18].

Remark 4.3 The preceding result can be extended in an obvious way to the so-called *generalized Stokes problem*, which is defined by adding a term $\alpha \mathbf{u}$ (with $\alpha > 0$ constant) to the first equation of (17):

$$\begin{cases} -\Delta \mathbf{u} + \alpha \mathbf{u} + \nabla p = \mathbf{f} \\ \operatorname{div} \mathbf{u} = 0 \\ u_{|\partial\Omega} = 0 \,, \end{cases} \tag{36}$$

see e.g. [18]. Then the same regularization can be used as for (17), since we can repeat the argument as shown by Remark 4.1: now the identity $-\sigma \Delta p = -\sigma \operatorname{div} \mathbf{f}$ follows from the relation $\operatorname{div}(-\Delta \mathbf{u} + \alpha \mathbf{u} + \nabla p) = -\Delta(\operatorname{div} \mathbf{u}) + \alpha \operatorname{div} \mathbf{u} + \Delta p = \Delta p$ and the first row of (36). The symmetric part matrix in (25) is now changed to

$$\mathbf{M}_h = \begin{pmatrix} diag_N\big((-\Delta + \alpha I)_h^0\big) & \mathbf{0} \\ \mathbf{0} & -\Delta_h^\nu \end{pmatrix},$$

and for this case Theorem 4.1 can be proved in the same way.

5 PCG algorithms for Navier's equations of elasticity

We consider a mixed formulation of an elasticity model from [10]. The displacement $\mathbf{u}$ of an isotropic elastic body Ω subject to a body force $\mathbf{f}$ in the case of pure displacement is described by

$$\begin{cases} -\mu\,\Delta\mathbf{u} - (\lambda+\mu)\,\nabla\,\mathrm{div}\,\mathbf{u} = \mathbf{f} \\ \mathbf{u}_{|\partial\Omega} = 0 \end{cases}$$

where λ, μ are the Lamé coefficients, which are here assumed to be constant. The pressure p satisfies

$$\mathrm{div}\,\mathbf{u} = -(1-2\nu)p,$$

and using the relation $(1-2\nu)(\lambda+\mu) = \mu$, we obtain the mixed formulation

$$\begin{cases} -\Delta\mathbf{u} + \nabla p = \dfrac{1}{\mu}\,\mathbf{f} \\ \mathrm{div}\,\mathbf{u} + (1-2\nu)p = 0 \\ \mathbf{u}_{|\partial\Omega} = 0\,. \end{cases} \tag{37}$$

The Poisson ratio ν satisfies

$$0 < \nu < \frac{1}{2}.$$

System (37) is closely related to the Stokes problem (17), and can be studied in the same spaces $H_0^1(\Omega)^N$ and $L_0^2(\Omega) = \{p \in L^2(\Omega) : \int_\Omega p = 0\}$ as in Section 4. We note that the condition $p \in L_0^2(\Omega)$, prescribed artificially for (17) to have uniqueness, is automatically satisfied now:

$$\int_\Omega p = -(1-2\nu)^{-1}\int_\Omega \mathrm{div}\,\mathbf{u} = -(1-2\nu)^{-1}\int_{\partial\Omega}\mathbf{u}\cdot\nu = 0.$$

Similarly to (19), problem (37) has a unique weak solution.

Following Section 4, for the finite element solution of (37) one chooses suitable FEM subspaces

$$V_h \subset H_0^1(\Omega)^N, \qquad P_h \subset L_0^2(\Omega)$$

where the Sobolev spaces are taken to be complex (this will be required to apply the theory of normal operators from [9] to our problem). One then looks for $(\mathbf{u}_h, p_h) \in V_h \times P_h$ satisfying

$$\begin{cases} \displaystyle\int_\Omega \nabla\mathbf{u}_h\cdot\nabla\overline{\mathbf{v}}_h - \int_\Omega p_h\,(\mathrm{div}\,\overline{\mathbf{v}}_h) = \int_\Omega \frac{1}{\mu}\,\mathbf{f}\cdot\overline{\mathbf{v}}_h & (\forall \mathbf{v}_h \in V_h) \\ \displaystyle\int_\Omega (\mathrm{div}\,\mathbf{u}_h)\,\overline{q}_h + (1-2\nu)\int_\Omega p_h\,\overline{q}_h = 0 & (\forall q_h \in P_h). \end{cases} \tag{38}$$

Further, by letting

$$s_h := (1-2\nu)^{1/2}\,p_h$$

we can rewrite problem (38) in the equivalent form

$$\begin{cases} \displaystyle\int_\Omega \nabla\mathbf{u}_h\cdot\nabla\overline{\mathbf{v}}_h \;-\; (1-2\nu)^{-1/2}\int_\Omega s_h\,(\mathrm{div}\,\overline{\mathbf{v}}_h) = \int_\Omega \frac{1}{\mu}\,\mathbf{f}\cdot\overline{\mathbf{v}}_h & (\forall \mathbf{v}_h \in V_h) \\ \displaystyle(1-2\nu)^{-1/2}\int_\Omega (\mathrm{div}\,\mathbf{u}_h)\,\overline{q}_h \quad + \quad \int_\Omega s_h\,\overline{q}_h \;=\; 0 & (\forall q_h \in P_h). \end{cases} \tag{39}$$

The presence of the term $(1-2\nu)p$ in (37) (compared to (17)) enables us to verify the convergence of the symmetric part PCG without regularization in subsection 5.1, then we develop the analogue of subsection 4.2 for the similarly regularized problem in subsection 5.2. The two methods exhibit different dependence on parameters: the convergence of the algorithm without regularization depends on ν, whereas in the regularized case (which involves the solution of an extra Neumann problem) the convergence depends on the regularization parameter σ but is uniform w.r.t. ν. Therefore the first method is only advisable for ν away from $1/2$, whereas in the opposite case the regularized algorithm is more efficient in spite of the required extra job.

5.1 Symmetric part preconditioning without regularization

Problem (38) leads to an algebraic system of the form (14) with SPD blocks in the diagonal. To formulate this algebraic system, we follow the obvious notations of subsection 4.2 for the matrices arising from the four integrals on the left-hand side of (38). Namely, let ∇_h, div_h and Δ_h^0 denote the Gram matrices of the operators ∇, div, Δ with Dirichlet b.c., respectively, in the considered subspaces, further, let I_h denote the mass matrix corresponding to the subspace P_h. Finally, let $diag_N(-\Delta_h^0)$ denote the block diagonal matrix with $-\Delta_h^0$ blocks repeated N times. Then the algebraic system corresponding to (38) can be written as follows:

$$\begin{pmatrix} diag_N(-\Delta_h^0) & (1-2\nu)^{-1/2}\,\nabla_h \\ (1-2\nu)^{-1/2}\,\mathrm{div}_h & I_h \end{pmatrix} \begin{pmatrix} \xi_h \\ \eta_h \end{pmatrix} = \begin{pmatrix} \frac{1}{\mu}\,\mathbf{f}_h \\ 0 \end{pmatrix}, \tag{40}$$

where ξ_h and η_h are the coefficient vectors of $\mathbf{u}_h$ and p_h in the given basis of V_h and P_h, respectively. Following the notations of Section 3, we let

$$\mathbf{L}_h = \begin{pmatrix} diag_N(-\Delta_h^0) & (1-2\nu)^{-1/2}\,\nabla_h \\ (1-2\nu)^{-1/2}\,\mathrm{div}_h & I_h \end{pmatrix}, \tag{41}$$

and for its symmetric and antisymmetric parts

$$\mathbf{M}_h = \begin{pmatrix} diag_N(-\Delta_h^0) & \mathbf{0} \\ \mathbf{0} & I_h \end{pmatrix} \quad \text{and} \quad \mathbf{N}_h = \begin{pmatrix} \mathbf{0} & (1-2\nu)^{-1/2}\,\nabla_h \\ (1-2\nu)^{-1/2}\,\mathrm{div}_h & \mathbf{0} \end{pmatrix}.$$

Here $\mathbf{L}_h \in \mathbf{R}^{n\times n}$ where $n = dim(V_h) + dim(P_h)$.

Theorem 5.1 *There holds*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_h}}{\|\mathbf{e}_0\|_{\mathbf{M}_h}}\right)^{1/k} \le \frac{1}{\sqrt{2(1-\nu)}} \qquad (k=1,\dots,n). \tag{42}$$

Proof. We proceed similarly to Theorem 4.1, now checking the conditions of statement (1) of Proposition 2.1. We now consider the complex separable Hilbert space

$$H_M := H_0^1(\Omega)^N \times L_0^2(\Omega)$$

with the inner product

$$\left\langle \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_M := \int_\Omega \nabla\mathbf{u}_h \cdot \nabla\overline{\mathbf{v}}_h \; + \int_\Omega s_h \overline{q}_h \,.$$

With the present definition of H_M, the expression (29):

$$\left\langle Q\begin{pmatrix}\mathbf{u}\\ s\end{pmatrix},\begin{pmatrix}\mathbf{v}\\ q\end{pmatrix}\right\rangle_M = -\int_\Omega s\,(\operatorname{div}\overline{\mathbf{v}}) + \int_\Omega (\operatorname{div}\mathbf{u})\,\overline{q} \qquad \left(\forall\ \begin{pmatrix}\mathbf{u}\\ s\end{pmatrix},\begin{pmatrix}\mathbf{v}\\ q\end{pmatrix}\in H_M\right) \tag{43}$$

now defines a bounded linear operator $Q : H_M \to H_M$, since

$$\begin{aligned}\left|-\int_\Omega \big(s\,(\operatorname{div}\overline{\mathbf{v}}) + (\operatorname{div}\mathbf{u})\,\overline{q}\,\big)\right|^2 &\le \left|\int_\Omega \sqrt{|s|^2+|\operatorname{div}\mathbf{u}|^2}\,\sqrt{|q|^2+|\operatorname{div}\mathbf{v}|^2}\ \right|^2\\ &\le \int_\Omega \big(|s|^2+|\operatorname{div}\mathbf{u}|^2\big)\ \int_\Omega \big(|q|^2+|\operatorname{div}\mathbf{v}|^2\big)\\ &\le C\int_\Omega \big(|s|^2+|\nabla\mathbf{u}|^2\big)\ \int_\Omega \big(|q|^2+|\nabla\mathbf{v}|^2\big) = C\left\|\begin{pmatrix}\mathbf{u}\\ s\end{pmatrix}\right\|_M\ \left\|\begin{pmatrix}\mathbf{v}\\ q\end{pmatrix}\right\|_M .\end{aligned} \tag{44}$$

Moreover (see e.g. [4]), the boundary condition $\mathbf{u}_{|\partial\Omega} = 0$ implies the relation

$$\int_\Omega \partial_i u_i\ \partial_j\overline{u}_j = \int_\Omega \partial_j u_i\ \partial_i\overline{u}_j$$

for all i, j (where $\partial_j u_i \equiv \partial u_i/\partial x_j$), hence

$$\begin{aligned}\int_\Omega |\operatorname{div}\mathbf{u}|^2 &= \int_\Omega \Big(\sum_i |\partial_i u_i|^2 + 2\sum_{i<j}\partial_i u_i\ \partial_j\overline{u}_j\Big) = \int_\Omega \Big(\sum_i |\partial_i u_i|^2 + 2\sum_{i<j}\partial_j u_i\ \partial_i\overline{u}_j\Big)\\ &\le \int_\Omega \Big(\sum_i |\partial_i u_i|^2 + \sum_{i<j}(|\partial_j u_i|^2+|\partial_i u_j|^2\,)\Big) = \int_\Omega \sum_{i,j}|\partial_j u_i|^2 = \int_\Omega |\nabla\mathbf{u}|^2\ ,\end{aligned}$$

hence in (44) we have $C = 1$, which implies

$$\|Q\| \le 1.$$

Now we can proceed similarly to Theorem 4.1. Here Q is antisymmetric, and letting

$$K = (1-2\nu)^{-1/2}Q \tag{45}$$

and with a suitable vector $\begin{pmatrix}\mathbf{g}\\ r\end{pmatrix}\in H_M$, the operator equation

$$(I+K)\begin{pmatrix}\mathbf{u}\\ s\end{pmatrix} = \begin{pmatrix}\mathbf{g}\\ r\end{pmatrix} \tag{46}$$

can be written as

$$\begin{cases}\displaystyle\int_\Omega \nabla\mathbf{u}\cdot\nabla\overline{\mathbf{v}}\ -\ (1-2\nu)^{-1/2}\int_\Omega s\,(\operatorname{div}\overline{\mathbf{v}}) = \frac{1}{\mu}\int_\Omega \mathbf{f}\cdot\overline{\mathbf{v}} & (\forall\mathbf{v}\in H_0^1(\Omega)^N)\\ \displaystyle(1-2\nu)^{-1/2}\int_\Omega (\operatorname{div}\mathbf{u})\,\overline{q} + \int_\Omega s\,\overline{q} = 0 & (\forall q\in L_0^2(\Omega)),\end{cases} \tag{47}$$

that is, system (39) is the FEM discretization of the operator equation (47) in the subspace $V_h\times P_h$. The conditions of Proposition 2.1 are thus satisfied for the operator equation (46). In virtue of $\|K\|\le(1-2\nu)^{-1/2}$ we have

$$\frac{\|K\|}{\sqrt{1+\|K\|^2}} \le \frac{1}{\sqrt{2(1-\nu)}},$$

i.e. statement (1) of Corollary 2.1 yields the required estimate. ∎

5.2 Superlinear PCGM for the regularized system

Now we study a regularized version of the elasticity system (38), using the same way of regularization as in (21). Let us consider (38) in which now we define the FEM subspaces

$$V_h \subset H_0^1(\Omega)^N, \qquad P_h \subset \dot{H}^1(\Omega) := H^1(\Omega) \cap L_0^2(\Omega)$$

to consist of piecewise linear functions. Let us fix a parameter $\sigma > 0$. The regularized version reads as follows: find $(\mathbf{u}_h, p_h) \in V_h \times P_h$ satisfying

$$\begin{cases} \displaystyle\int_\Omega \nabla \mathbf{u}_h \cdot \nabla \overline{\mathbf{v}}_h - \int_\Omega p_h \,(\operatorname{div} \overline{\mathbf{v}}_h) = \int_\Omega \frac{1}{\mu}\, \mathbf{f} \cdot \overline{\mathbf{v}}_h \qquad (\forall \mathbf{v}_h \in V_h) \\ \displaystyle\int_\Omega (\operatorname{div} \mathbf{u}_h)\, \overline{q}_h + \, 2\sigma(1-\nu) \int_\Omega \nabla p_h \cdot \nabla \overline{q}_h \, + (1-2\nu) \int_\Omega p_h\, \overline{q}_h = \frac{\sigma}{\mu} \int_\Omega \mathbf{f} \cdot \nabla \overline{q}_h \qquad (\forall q_h \in P_h). \end{cases} \tag{48}$$

Following Remark 4.1, in terms of the strong form (37) and assuming sufficient regularity, the regularization in the second row of (48) formally comes from adding the relation $-2\sigma(1-\nu)\Delta p = -\frac{\sigma}{\mu}\operatorname{div} \mathbf{f}$, which follows by taking the divergence of $-\Delta \mathbf{u} + \nabla p = \frac{1}{\mu}\mathbf{f}$ and using $\operatorname{div} \mathbf{u} = -(1-2\nu)p$.

The superlinear convergence result can be established much similarly to subsection 4.2 if first a parameter-dependent norm is used; then we can pass to a parameter-independent norm using a simple estimate.

First, letting

$$\varrho := 2\sigma(1-\nu), \quad \hat{\varrho} := \varrho^{-1}(1-2\nu) \quad \text{and} \quad s_h := \varrho^{1/2}\, p_h\,,$$

we rewrite problem (48) as

$$\begin{cases} \displaystyle\int_\Omega \nabla \mathbf{u}_h \cdot \nabla \overline{\mathbf{v}}_h - \varrho^{-1/2} \int_\Omega s_h \,(\operatorname{div} \overline{\mathbf{v}}_h) = \int_\Omega \frac{1}{\mu}\, \mathbf{f} \cdot \overline{\mathbf{v}}_h \qquad (\forall \mathbf{v}_h \in V_h) \\ \displaystyle \varrho^{-1/2} \int_\Omega (\operatorname{div} \mathbf{u}_h)\, \overline{q}_h + \, \int_\Omega \Big(\nabla s_h \cdot \nabla \overline{q}_h \, + \hat{\varrho}\, s_h\, \overline{q}_h \Big) = \frac{\varrho^{-1/2}\sigma}{\mu} \int_\Omega \mathbf{f} \cdot \nabla \overline{q}_h \qquad (\forall q_h \in P_h). \end{cases} \tag{49}$$

Following the notations of subsection 4.2, the algebraic system corresponding to (49) has the system matrix

$$\mathbf{L}_{h,\hat{\varrho}} := \begin{pmatrix} diag_N(-\Delta_h^0) & \varrho^{-1/2}\, \nabla_h \\ \varrho^{-1/2} \operatorname{div}_h & (-\Delta + \hat{\varrho} I)_h^\nu \end{pmatrix}, \tag{50}$$

with symmetric and antisymmetric parts

$$\mathbf{M}_{h,\hat{\varrho}} = \begin{pmatrix} diag_N(-\Delta_h^0) & \mathbf{0} \\ \mathbf{0} & (-\Delta + \hat{\varrho} I)_h^\nu \end{pmatrix} \qquad \text{and} \qquad \mathbf{N}_{h,\hat{\varrho}} = \begin{pmatrix} \mathbf{0} & \varrho^{-1/2}\, \nabla_h \\ \varrho^{-1/2} \operatorname{div}_h & \mathbf{0} \end{pmatrix}, \tag{51}$$

respectively.

Proposition 5.1 *There exists a sequence $\varepsilon_k^{(\hat{\varrho})} \to 0$ (as $k \to \infty$), depending on $\hat{\varrho}$, σ and Ω, such that the PCG algorithm (5) for the matrices $\mathbf{M}_{h,\hat{\varrho}}$ and $\mathbf{L}_{h,\hat{\varrho}}$ in (50)-(51) yields*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_{h,\hat{\varrho}}}}{\|\mathbf{e}_0\|_{\mathbf{M}_{h,\hat{\varrho}}}} \right)^{1/k} \le \; \varepsilon_k^{(\hat{\varrho})} \qquad (k = 1, \dots, n). \tag{52}$$

PROOF. Let

$$\varepsilon_k^{(\hat{\varrho})} = \frac{2}{\varrho^{1/2}k} \sum_{i=1}^{k} |\lambda_i(Q_{\hat{\varrho}})| \tag{53}$$

where $Q_{\hat{\varrho}}$ is the compact linear operator defined in (29) in which the inner product (28) on the space $H_M := H_0^1(\Omega)^N \times \dot{H}^1(\Omega)$ is replaced by

$$\left\langle \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{v} \\ q \end{pmatrix} \right\rangle_{M,\hat{\varrho}} := \int_\Omega \nabla \mathbf{u} \cdot \nabla \overline{\mathbf{v}} \; + \int_\Omega \Big(\nabla s \cdot \nabla \overline{q} \; + \hat{\varrho}\, s\, \overline{q} \Big) . \tag{54}$$

Then the proof goes exactly as for Theorem 4.1 if $\mathbf{M}_h$, $\mathbf{L}_h$, $\langle .,. \rangle_M$ and σ are replaced by $\mathbf{M}_{h,\hat{\varrho}}$, $\mathbf{L}_{h,\hat{\varrho}}$, $\mathbf{L}_h$, $\langle .,. \rangle_{M,\hat{\varrho}}$ and ϱ, respectively. ∎

One can get rid of the parameter $\hat{\varrho}$, by which the result becomes independent of ν:

Theorem 5.2 *There exists a sequence $\varepsilon_k \to 0$ (as $k \to \infty$), depending on σ and Ω but independent of ν, such that the PCG algorithm (5) for the matrices $\mathbf{M}_{h,\hat{\varrho}}$ and $\mathbf{L}_{h,\hat{\varrho}}$ in (50)-(51) yields*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_{h,\hat{\varrho}}}}{\|\mathbf{e}_0\|_{\mathbf{M}_{h,\hat{\varrho}}}} \right)^{1/k} \leq \; \varepsilon_k \qquad (k = 1, \ldots, n). \tag{55}$$

PROOF. First, the fact $0 < \nu < 1/2$ and the definition of ϱ and $\hat{\varrho}$ imply

$$\varrho \geq \sigma \quad \text{and} \quad 0 \leq \hat{\varrho} \leq \frac{1}{\sigma}. \tag{56}$$

We define the functions $f_k : [0, \frac{1}{\sigma}] \to \mathbf{R}^+$,

$$f_k(\hat{\varrho}) := \frac{2}{\sigma^{1/2}k} \sum_{i=1}^{k} |\lambda_i(Q_{\hat{\varrho}})|,$$

then $\varepsilon_k^{(\hat{\varrho})} \leq f_k(\hat{\varrho})$. The continuous dependence of the eigenvalues of the bounded operators $Q_{\hat{\varrho}}$ on the parameter $\hat{\varrho}$ implies that the functions f_k are continuous. Further, the ordering of eigenvalues by Definition 2.1 implies that $f_k(\hat{\varrho}) \to 0$ monotonically for each fixed $\hat{\varrho}$. Then, by Dini's theorem [21], $f_k \to 0$ uniformly on the interval $[0, \frac{1}{\sigma}]$. Letting $\varepsilon_k := max\{ f_k(\hat{\varrho}) : \; \hat{\varrho} \in [0, \frac{1}{\sigma}]$, we obtain that $\varepsilon_k \to 0$ and that $\varepsilon_k^{(\hat{\varrho})} \leq \varepsilon_k$, which by (52) yields the required result. ∎

In addition, one can avoid the parameter $\hat{\varrho}$ in the $\mathbf{M}_h$-norm too. For this we observe that the norms $\|.\|_{\mathbf{M}_h}$ and $\|.\|_{\mathbf{M}_{h,\hat{\varrho}}}$ are equivalent uniformly in ν. Namely, using (56) and the Sobolev inequality

$$\|\nabla s\|^2_{L^2(\Omega)} \geq c_0 \|s\|^2_{L^2(\Omega)} \qquad (s \in \dot{H}^1(\Omega))$$

with $c_0 > 0$, the norms corresponding to the inner products (28) and (54) are related via

$$\left\| \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} \right\|^2_M \leq \left\| \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} \right\|^2_{M,\hat{\varrho}} \leq \Big(1 + \frac{1}{c_0 \sigma} \Big) \left\| \begin{pmatrix} \mathbf{u} \\ s \end{pmatrix} \right\|^2_M \tag{57}$$

on $H_0^1(\Omega)^N \times \dot{H}^1(\Omega)$. By (10), the norms $\|.\|_{\mathbf{M}_h}$ and $\|.\|_{\mathbf{M}_{h,\hat{\varrho}}}$ are the traces of the above Sobolev norms in the the FEM subspace $V_h \times P_h$, hence they inherit the above estimate:

$$\|\xi\|_{\mathbf{M}_{h,\hat{\varrho}}} \leq \|\xi\|_{\mathbf{M}_h} \leq \Big(1 + \frac{1}{c_0 \sigma} \Big) \|\xi\|_{\mathbf{M}_{h,\hat{\varrho}}} \qquad (\xi \in \mathbf{R}^n). \tag{58}$$

Corollary 5.1 *There exists a sequence $\tilde{\varepsilon}_k \to 0$ (as $k \to \infty$), depending on σ and Ω but independent of ν, such that the PCG algorithm (5) for the matrices $\mathbf{M}_{h,\hat{\varrho}}$ and $\mathbf{L}_{h,\hat{\varrho}}$ in (50)-(51) yields*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_h}}{\|\mathbf{e}_0\|_{\mathbf{M}_h}}\right)^{1/k} \le \tilde{\varepsilon}_k \qquad (k = 1, \ldots, n).$$

PROOF. Relations (58) and (55) imply

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_h}}{\|\mathbf{e}_0\|_{\mathbf{M}_h}}\right)^{1/k} \le \left(1 + \frac{1}{c_0\sigma}\right)^{1/k} \left(\frac{\|\mathbf{e}_k\|_{\mathbf{M}_{h,\hat{\varrho}}}}{\|\mathbf{e}_0\|_{\mathbf{M}_{h,\hat{\varrho}}}}\right)^{1/k} \le \left(1 + \frac{1}{c_0\sigma}\right)^{1/k} \varepsilon_k \le \left(1 + \frac{1}{c_0\sigma}\right) \varepsilon_k =: \tilde{\varepsilon}_k$$

for $k = 1, \ldots, n$. ∎

Remark 5.1 The preconditioned algorithms in Sections 4-5 reduce the original problem to the solution of Poisson equations (cf. Remark 4.2). In terms of Section 3, equation (14), problems of the saddle-point form

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ -\mathbf{B} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \tag{59}$$

are thus reduced to problems with system matrix $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}$, where $\mathbf{A}$ corresponds to the discrete Laplacian.

Alternatively, one can reduce problem (59) to a block-diagonal matrix containing the Schur complement $\mathbf{S} = \mathbf{C} + \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ via the transformation

$$\begin{pmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{B}\mathbf{A}^{-1} & \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ -\mathbf{B} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{I}_1 & -\mathbf{A}^{-1}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \tag{60}$$

(where $\mathbf{I}_1$, $\mathbf{I}_2$ are the identity matrices of the proper size). This is particularly useful if one does not use the exact matrix $\mathbf{A}$ in the preconditioner but a readily solved approximation $\tilde{\mathbf{A}}$ of it. Namely, the left-hand side of (60) is then replaced by the matrix

$$\tilde{\mathcal{A}} := \begin{pmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{B}\tilde{\mathbf{A}}^{-1} & \mathbf{I}_2 \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ -\mathbf{B} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{I}_1 & -\tilde{\mathbf{A}}^{-1}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I}_2 \end{pmatrix}.$$

Then, letting $\tilde{\mathbf{S}} = \mathbf{C} + \mathbf{B}\tilde{\mathbf{A}}^{-1}\mathbf{B}^T$, a direct calculation shows

$$\tilde{\mathcal{A}} = \begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{S}} \end{pmatrix} + \begin{pmatrix} \mathbf{A} - \tilde{\mathbf{A}} & -(\mathbf{A} - \tilde{\mathbf{A}})\tilde{\mathbf{A}}^{-1}\mathbf{B}^T \\ \mathbf{B}\tilde{\mathbf{A}}^{-1}(\mathbf{A} - \tilde{\mathbf{A}}) & \mathbf{B}\tilde{\mathbf{A}}^{-1}(\mathbf{A} - \tilde{\mathbf{A}})\tilde{\mathbf{A}}^{-1}\mathbf{B}^T \end{pmatrix}. \tag{61}$$

Consequently, the matrix $\begin{pmatrix} \tilde{\mathbf{A}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{S}} \end{pmatrix}$ is an efficient preconditioner of $\tilde{\mathcal{A}}$ since its inverse transforms (61) into identity plus a perturbation matrix whose norm is asymptotically bounded by *const.*$\cdot \|\mathbf{A} - \tilde{\mathbf{A}}\|$, that is, the nonsymmetric part dimishes when $\tilde{\mathbf{A}}$ approaches $\mathbf{A}$.

References

[1] Adams, R.A., *Sobolev Spaces*, Academic Press, 1975.

[2] Axelsson, O., A generalized conjugate gradient least square method, *Numer. Math.* 51 (1987), 209-227.

[3] Axelsson, O., *Iterative Solution Methods,* Cambridge University Press, 1994.

[4] Axelsson, O., On iterative solvers in structural mechanics; separate displacement orderings and mixed variable methods, *Math. Comput. Simulation* 50 (1999), no. 1-4, 11–30.

[5] Axelsson, O., Barker, V. A., Neytcheva, M., Polman, B., Solving the Stokes problem on a massively parallel computer, *Math. Model. Anal.* 6 (2001), no. 1, 7–27.

[6] Axelsson, O., Kaporin, I., On the sublinear and superlinear rate of convergence of conjugate gradient methods. Mathematical journey through analysis, matrix theory and scientific computation (Kent, OH, 1999), *Numer. Algorithms* 25 (2000), no. 1-4, 1–22.

[7] Axelsson, O., Karátson J., On the rate of convergence of the conjugate gradient method for linear operators in Hilbert space, *Numer. Funct. Anal.* **23** (2002), No. 3-4, 285-302.

[8] Axelsson, O., Karátson J., Symmetric part preconditioning for the conjugate gradient method in Hilbert space, *Numer. Funct. Anal.* 24 (2003), No. 5-6, 455-474.

[9] Axelsson, O., Karátson J., Superlinearly convergent CG methods via equivalent preconditioning for nonsymmetric elliptic operators, *Numer. Math.* 99 (2004), No. 2, 197-223.

[10] Axelsson, O., Neytcheva, M., Scalable algorithms for the solution of Navier's equations of elasticity, *J. Comput. Appl. Math.* 63 (1995), no. 1-3, 149–178.

[11] Axelsson, O., Neytcheva, M., Preconditioning methods for linear systems arising in constrained optimization problems, *Numer. Linear Algebra Appl.* 10 (2003), no. 1-2, 3–31.

[12] Axelsson, O., Neytcheva, M., Eigenvalue estimates for preconditioned saddle-point matrices, Technical Report of the Department of Information Technology 2004-019, Uppsala University. To appear in *Numer. Linear Algebra Appl.*

[13] Concus, P., Golub, G.H., A generalized conjugate method for non-symmetric systems of linear equations, in: *Lect. Notes Math. Syst.* 134 (eds. Glowinski, R., Lions, J.-L.), pp. 56-65, Springer, 1976.

[14] Faber, V., Manteuffel, T., Parter, S.V., On the theory of equivalent operators and application to the numerical solution of uniformly elliptic partial differential equations, *Adv. in Appl. Math.,* 11 (1990), 109-163.

[15] Goldstein, C. I., Manteuffel, T. A., Parter, S. V., Preconditioning and boundary conditions without H_2 estimates: L_2 condition numbers and the distribution of the singular values, *SIAM J. Numer. Anal.* 30 (1993), no. 2, 343–376.

[16] Hackbusch, W., *Multigrid methods and applications*, Springer Series in Computational Mathematics 4, Springer, Berlin, 1985.

[17] Hughes, T. J. R., Franca, L. P., Balestra, M., A new finite element formulation for computational fluid dynamics V. Circumventing the Babuška-Brezzi condition: a stable Petrov-Galerkin formulation of the Stokes problem accommodating equal-order interpolations, *Comput. Methods Appl. Mech. Engrg.* 59 (1986), no. 1, 85–99.

[18] Kobelkov, G. M., Olshanskii, M. A., Effective preconditioning of Uzawa type schemes for a generalized Stokes problem, *Numer. Math.* 86 (2000), no. 3, 443–470.

[19] Pierre, R., Regularization procedures of mixed finite element approximations of the Stokes problem, *Numer. Methods Partial Diff. Eqns.* 5 (1989), no. 3, 241–258.

[20] Rossi, T., Toivanen, J., A parallel fast direct solver for block tridiagonal systems with separable matrices of arbitrary dimension, *SIAM J. Sci. Comput.* 20 (1999), no. 5, 1778–1796 (electronic).

[21] Rudin, W., *Principles of mathematical analysis.* Third edition, International Series in Pure and Applied Mathematics, McGraw-Hill, 1976.

[22] Swarztrauber, P. N., The methods of cyclic reduction, Fourier analysis and the FACR algorithm for the discrete solution of Poisson's equation on a rectangle, *SIAM Rev.* 19 (1977), no. 3, 490–501.

[23] Temam, R., *Navier-Stokes equations. Theory and numerical analysis*, Studies in Mathematics and its Applications, Vol. 2. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977.

[24] van der Vorst, H. A., Iterative solution methods for certain sparse linear systems with a nonsymmetric matrix arising from PDE-problems, *J. Comput. Phys.* 44 (1981), no. 1, 1–19.

[25] Widlund, O., A Lanczos method for a class of non-symmetric systems of linear equations, *SIAM J. Numer. Anal.,* 15 (1978), 801-812.

[26] Zulehner, W., Analysis of iterative methods for saddle point problems: a unified approach, *Math. Comp.* 71 (2002), no. 238, 479–505.