UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

# Preconditioning Saddle-Point Problems Arising from Mixed Finite Element Discretization of Elliptic Equations

Panayot S. Vassilevski
Raytcho D. Lazarov

November 1992

CAM Report 92-46

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

# PRECONDITIONING SADDLE-POINT PROBLEMS ARISING FROM MIXED FINITE ELEMENT DISCRETIZATION OF ELLIPTIC EQUATIONS

PANAYOT S. VASSILEVSKI AND RAYTCHO D. LAZAROV

November 1992

ABSTRACT. We consider saddle-point problems that typically arise from the mixed finite element discretization of second order elliptic problems. By a proper equivalent algebraic operations the considered saddle-point problem is transformed to another saddle–point problem. The resulting problem can be then efficiently preconditioned by a block-diagonal matrix (with blocks corresponding to velocity and pressure, respectively) with first block on the diagonal corresponding to the bilinear form $\int_\Omega \left[a^{-1}\underline{\chi}\cdot\underline{\theta}+\frac{1}{\delta}\nabla\cdot\underline{\chi}\;\nabla\cdot\underline{\theta}\right]$ ($\delta$ is a positive parameter) and the second block equals to a constant times the identity operator. We derive uniform bounds for the negative and positive eigenvalues of the preconditioned operator. Then any known preconditioner for the above bilinear form can be applied. We also show some numerical experiments that illustrate the convergence properties of the proposed technique.

**1. Introduction.** We consider the following saddle-point problem in operator form

$$\begin{bmatrix} M & N^\star \\ N & 0 \end{bmatrix}\begin{bmatrix} \mathbf{u}_h \\ p_h \end{bmatrix} = \begin{bmatrix} 0 \\ -\phi \end{bmatrix}, \tag{1}$$

typically arising from mixed finite element discretizations of second order elliptic problems.

This is a saddle-point problem, i.e., the corresponding matrix is indefinite. This makes it hard for preconditioning and thus using efficient iterative procedure for solving (1) not as easy. There are several known techniques for solving saddle-point problems, cf., e.g., Bramble and Pasciak [8], Bank, Welfert, and Yserentant [6],

---

1991 *Mathematics Subject Classification.* 65F10, 65N20, 65N30.

*Key words and phrases.* preconditioning saddle-point problems, eigenvalue estimation, mixed finite element method, minimum residual method, second order elliptic problems.

The first author while performing this research was on leave from the Center of Informatics and Computer Technology, "Acad. G. Bontchev" str., Bl.25 A, 1113 Sofia, Bulgaria

This work was partially supported by the National Science Foundation grant INT-8914472. The first author was also supported in part by the Army Research Office under contract ARO DAAL03-91-G-150, and by the National Science Foundation under grant FDP NSF ASC 9003002. A funding through the Institute of Scientific Computation, Texas A&M University for a visit there is also acknowledged.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Axelsson and Vassilevski [4], Axelsson [2], Axelsson [3], Mathew [21], Rusten and Winther [24], Ewing, Lazarov, Lu, and Vassilevski [15], and Cai, Goldstein and Pasciak [12], Ewing and Wang [17], [18], Vassilevski and Wang [25], Ewing, Pasciak and Vassilevski [16]. We can distinguish several approaches; namely in Bramble and Pasciak [8] the original problem (1) by equivalent algebraic transformations is transformed to a positive definite problem in a certain inner product. Another approach, exploiting inner – outer iteration was studied in Bank, Welfert and Yserentant [6] and Axelsson and Vassilevski [4]. In Axelsson [3] a general treatment of iterative solution of saddle–point problems was presented. In Mathew [21] and Ewing and Wang [17], [18] the original problem is reduced to a symmetric and positive definite one by working in a certain subspace (in their application this is the space of divergence–free fluxes). In Axelsson [2], Cai, Goldstein and Pasciak [12], and Ewing, Pasciak and Vassilevski [16], and Vassilevski and Wang [25] methods based on introducing penalty term were studied. The latter approach as a discretization procedure was first introduced by Bercovier [7]. Rusten and Winther [24] used proper block–diagonal preconditioning of the original saddle–point operator in the MINRES iterative method (cf., Paige and Saunders [22], or Chandra [13]). Our method is similar to that of Rusten and Winther [24], see also Ewing, Lazarov, Lu, and Vassilevski [15], but we first transform the original saddle–point problem to a more suitable again saddle–point problem. We also introduce a parameter $\delta > 0$ which need not necessarily be small. This transformation was essentially described (in different context and assuming $\delta \to 0$) in Axelsson [3]. In our case, by choosing $\delta \to 0$ we can control the length of the intervals that contain the eigenvalues of the preconditioned problem thus improving (to a certain extend) the conditioning of the preconditioned operator. In this respect our method has much in common with the penalty approach studied in some of the above papers. But there is a principal difference. Even in the case $\delta \to 0$ we precondition a problem that is equivalent to the original one, whereas in the penalty approach one first approximates the original problem with a slightly perturbed one (hence only close to the original one and not exactly equivalent to it).

The preconditioner we propose is

$$\mathcal{D} \equiv \begin{bmatrix} M + \frac{1}{\delta} N^\star N & 0 \\ 0 & \delta_1 I \end{bmatrix}. \tag{2}$$

Here $I$ is the identity operator and $\delta_1$ is just a scaling parameter. Possible choices could be $\delta_1 = 1$ or $\delta_1 = O(\delta)$.

Since this is block–diagonal symmetric positive definite operator it allows (it will be clear from the following analysis) that the first block $M + \frac{1}{\delta} N^\star N$ can be replaced with any available preconditioner $B_\delta$. As a result of our approach the preconditioning of the original saddle–point problem is reduced to the preconditioning of the operator

$$M_\delta \equiv M + \frac{1}{\delta} N^\star N. \tag{3}$$

For the case of mixed finite element discretization of second order elliptic problems

$M_\delta$ corresponds to the following bilinear form

$$\int_\Omega \left( a^{-1} \underline{\chi} \cdot \underline{\theta} + \frac{1}{\delta} \nabla \cdot \underline{\chi} \, \nabla \cdot \underline{\theta} \right). \tag{4}$$

This bilinear form is defined for $\underline{\chi}$ and $\underline{\theta}$ which are vector-functions from the corresponding finite element spaces for the velocity unknown used in the mixed finite element discretization of the second order elliptic problem. The coefficient $a$ is the coefficient matrix in the considered second order elliptic operator. It is supposed to be symmetric and uniformly in $x \in \overline{\Omega}$ ($\Omega$ is the problem domain) positive definite. The bilinear form (4) is the same that appears in the papers by Cai, Goldstein, and Pasciak [12], Ewing, Pasciak, and Vassilevski [16] and in Vassilevski and Wang [25]. In the latter two papers optimal order hierarchical basis methods and multigrid methods of almost optimal order were derived, respectively. All these methods have convergence independent of the penalty parameter $\delta \to 0$, and the hierarchical method is also shown to be insensitive with respect to possible jumps in the coefficient matrix $a$ as long as these occur only across edges from the elements from the initial (coarse) triangulation of the problem domain $\Omega$. One can apply the above mentioned hierarchical basis approach for the bilinear form (4) and thus to get an optimal order multilevel preconditioner for the transformed problem. Note that here we have to perform iterations simultaneously for the velocity $\mathbf{u}_h$ and the pressure $p_h$ in a preconditioned conjugate gradient type method (namely, the MINRES or a least squares version of it, or more generally, polynomially preconditioned CG methods as studied in Ashby, Manteuffel and Saylor [1]).

Our objective in this paper is to derive uniform eigenvalue estimates (for the negative and positive eigenvalues) of the preconditioned operator $\mathcal{D}^{-1}\mathcal{A}_\delta$, where

$$\mathcal{A}_\delta \equiv \begin{bmatrix} M_\delta & N^\star \\ N & 0 \end{bmatrix}, \tag{5}$$

is the operator that is obtained from (1) using equivalent algebraic transformation.

The paper is organized as follows. In §2 we present the mixed finite element discretization and then we introduce the transformation of the problem to a problem with coefficient operator $\mathcal{A}_\delta$. In §3 we derive the eigenvalue estimates. Finally in §4 some numerical results are presented that illustrate the convergence properties of the studied preconditioning technique.

**2. The Problem.** In this section we consider the second order elliptic problem: Given $f \in L^2(\Omega)$ for a polygonal domain $\Omega \subset \mathbb{R}^2$, find $p \in H^1(\Omega)$ such that

$$-\nabla \cdot a \nabla p = f(x,y), \quad (x,y) \in \Omega,$$

with Neumann boundary conditions

$$-a\nabla p \cdot \mathbf{n} = 0 \text{ on } \partial\Omega\cdot$$

As commonly used $\mathbf{n}$ denotes the outward unit vector normal to $\partial\Omega$. Here $a = a(x,y)$ is symmetric and uniformly in $(x,y) \in \overline{\Omega}$ positive definite coefficient matrix. The right–hand side $f$ satisfies the following compatibility condition,

$$\int_\Omega f(x,y)dxdy = 0.$$

Note that the solution $p$ is determined up to an additive constant.

By introducing the velocity (flux) unknown (a vector function)

$$\mathbf{u} = -a\nabla p,$$

we can rewrite the above equation for $p$ as a system of first order equations,

$$\begin{array}{lcl} a^{-1}\mathbf{u} + \nabla p & = & 0, \\ \nabla \cdot \mathbf{u} & = & f, \end{array}$$

with boundary conditions

$$\mathbf{u} \cdot \mathbf{n} = 0 \text{ on } \partial\Omega\cdot$$

In a variational form the above system can be rewritten as follows:

$$\begin{array}{lcl} \int_\Omega a^{-1}\mathbf{u} \cdot \mathbf{v} - \int_\Omega \nabla \cdot \mathbf{v} p & = & 0, \\ \int_\Omega \nabla \cdot \mathbf{u}\, w & = & \int_\Omega f w, \end{array} \tag{6}$$

for all test functions $\mathbf{v}$ from the space

$$H_0(div; \Omega) \equiv \left\{\mathbf{v} \in L^2(\Omega)^2 : \quad \nabla \cdot \mathbf{v} \in L^2(\Omega), \quad \mathbf{v} \cdot \mathbf{n} = 0 \text{ on } \partial\Omega\right\}$$

and all $w \in L^2(\Omega) \setminus Span\{1\}$, i.e., $w$ is in the quotient space $L^2(\Omega)$ over the space spanned by the constants.

The mixed finite element solution of problem (6) is obtained by restricting the test functions to finite dimensional subspaces

$$\begin{array}{ll} \mathcal{V} = & \mathcal{V}_h \subset H_0(div; \Omega), \\ \mathcal{W} = & \mathcal{W}_h \subset L^2(\Omega) \setminus Span\{1\}, \end{array}$$

of piecewise polynomial functions that satisfy certain (weak) continuity conditions. Namely, for the velocity unknown the space $\mathcal{V}$ has to satisfy continuity of the normal component of the test functions across the element boundaries. I.e., assuming that the domain $\Omega$ is partitioned into a set of elements (e. g., triangles or rectangles) so that the functions from $\mathcal{V}$ and $\mathcal{W}$ are piecewise polynomials. Suitable pairs of finite element spaces $\mathcal{V}$ and $\mathcal{W}$ that provide a stable discretization of the system of first order equations (6) are proposed by many authors, cf., e.g, Raviart and Thomas [23], Brezzi, Douglas and Marini [10], Douglas and Wang [14], etc. The stability requirement (i.e., the Babuška – Brezzi condition, cf. Babuška [5], Brezzi [9], or the text by Brezzi and Fortin [11]) reads

$$\beta \|\psi\|_{L^2(\Omega)} \leq \sup_{\underline{\chi} \in \mathcal{V}} \frac{(\psi, \nabla \cdot \underline{\chi})}{\|\underline{\chi}\|_{\mathcal{H}_0(\mathrm{div},\Omega)}}, \quad \text{for all } \psi \in \mathcal{W}, \tag{7}$$

where $\beta > 0$ is a constant independent of the mesh size $h \to 0$.

The discrete problem reads:

Find a pair of functions $(\mathbf{u}_h, p_h)$ belonging to $\mathcal{V} \times \mathcal{W}$ such that

$$\begin{array}{lcl} \int_\Omega a^{-1}\mathbf{u}_h \cdot \underline{\chi} - \int_\Omega \nabla \cdot \underline{\chi}\, p_h & = & 0, \\ \int_\Omega \nabla \cdot \mathbf{u}_h \psi & = & \int_\Omega f \psi, \end{array} \tag{8}$$

for all $\underline{\chi} \in \mathcal{V}$ and all $\psi \in \mathcal{W}$. In matrix–operator form we have

$$\begin{bmatrix} M & N^\star \\ N & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_h \\ p_h \end{bmatrix} = \begin{bmatrix} 0 \\ -\phi \end{bmatrix}. \tag{9}$$

Here $\phi$ is the $L^2$ – projection of $f$ onto $\mathcal{W}$. The operators $M: \mathcal{V} \to \mathcal{V}$, $N: \mathcal{V} \to \mathcal{W}$, and $N^\star: \mathcal{W} \to \mathcal{V}$ are defined as follows:

$$(M\underline{\chi}, \underline{\theta}) = \int_\Omega a^{-1}\underline{\chi} \cdot \underline{\theta}, \quad \text{for all } \underline{\chi} \text{ and } \underline{\theta} \in \mathcal{V},$$

$$(N\underline{\chi}, \psi) = -\int_\Omega \nabla \cdot \underline{\chi}\, \psi, \quad \text{for all } \underline{\chi} \in \mathcal{V} \text{ and all } \psi \in \mathcal{W},$$

$$(N^\star\psi, \underline{\theta}) = -\int_\Omega \nabla \cdot \underline{\theta}\, \psi, \quad \text{for all } \psi \in \mathcal{W} \text{ and } \underline{\theta} \in \mathcal{V}\cdot$$

Here and in what follows $(.,.)$ denotes the standard $L^2(\Omega)$ or $L^2(\Omega)^2$ – inner product.

Next we describe the transformation of (8) to an equivalent problem with a matrix-operator that is again of saddle–point type. We use the equivalent variational form of (9); namely, fo a given parameter $\delta > 0$ we first set in the second equation of (8) $\psi = -\frac{1}{\delta}\nabla \cdot \underline{\chi}$ and add it to the first one. We get

$$\begin{array}{lcl} \int_\Omega \left[a^{-1}\mathbf{u}_h \cdot \underline{\chi} + \frac{1}{\delta}\nabla \cdot \mathbf{u}_h \nabla \cdot \underline{\chi}\right] - \int_\Omega \nabla \cdot \underline{\chi}\, p_h & = & \frac{1}{\delta}\int_\Omega f \nabla \cdot \underline{\chi}, \\ -\int_\Omega \nabla \cdot \mathbf{u}_h \psi & = & -\int_\Omega f\psi\cdot \end{array} \tag{10}$$

Note that we have substantially used the fact that for any $\underline{\chi} \in \mathcal{V}$ its divergence $\nabla \cdot \underline{\chi}$ belongs to $\mathcal{W}$.

In matrix-operator form (10) reads

$$\begin{bmatrix} M + \frac{1}{\delta}N^\star N & N^\star \\ N & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_h \\ p_h \end{bmatrix} = -\begin{bmatrix} \frac{1}{\delta}N^\star\phi \\ \phi \end{bmatrix}. \tag{11}$$

We recall that $\phi$ was the $L^2$ - projection of $f$ onto the finite element space $\mathcal{W}$.

In what follows we study the matrix-operator of (11) $\mathcal{A}_\delta$. Note that $\mathcal{A}_\delta$ is again a saddle–point operator. We however will show that this transformed operator is spectrally equivalent to a block–diagonal symmetric and positive definite one; namely the operator $\mathcal{D}$, (2) for any choice of the parameter $\delta > 0$.

**3. Eigenvalue Estimation.** In this section we prove a basic lemma concerning estimation of the minimal eigenvalue of the operator $NM_\delta^{-1}N^\star$ in terms of the parameter $\delta > 0$ and the minimal eigenvalue of $NM^{-1}N^\star$. We recall that we use the standard $L^2(\Omega)$ or $L^2(\Omega)^2$ – inner product denoted by $(.,.)$. This estimate then is used to derive bounds for the spectrum (the negative and positive eigenvalues) of the preconditioned operator $\mathcal{D}^{-1}\mathcal{A}_\delta$.

**Lemma 1.** *Let $\lambda_0$ be the minimal eigenvalue of $NM^{-1}N^\star$. Then the following spectral bounds for the operator $NM_\delta^{-1}N^\star$ hold*

$$\frac{\delta\lambda_0}{\delta+\lambda_0} \leq \lambda < \delta,$$

*for any eigenvalue $\lambda = \lambda\left[NM_\delta^{-1}N^\star\right]$.*

*Proof.* Consider $\hat{N} = NM^{-\frac{1}{2}}$. Then

$$NM_\delta^{-1}N^\star = N\left(M + \frac{1}{\delta}N^\star N\right)^{-1} N^\star = \hat{N}\left(I + \frac{1}{\delta}\hat{N}^\star\hat{N}\right)^{-1}\hat{N}^\star.$$

Consider the following eigenvalue problem in $\mathcal{W} = \mathcal{W}_h$,

$$\hat{N}\hat{N}^\star q_k = \mu_k q_k, \quad k = 1, 2, \ldots, n = O(h^{-2}).$$

Here $n = O(h^{-2})$ is the dimension of $\mathcal{W} = \mathcal{W}_h$.

We can choose $\{q_k\}$ to be a complete and $L^2(\Omega)$ – orthogonal system in $\mathcal{W}$. This is possible since $\hat{N}\hat{N}^\star$ is symmetric and since $\mathcal{W}$ is finite dimensional. Then we can expand any $\psi \in \mathcal{W}$ in terms of $\{q_k\}$,

$$\psi = \sum c_k q_k.$$

We also have

$$\hat{N}^\star\hat{N}(\hat{N}^\star q_k) = \mu_k(\hat{N}^\star q_k).$$

Multiplying this equation by $\frac{1}{\delta}$ and adding it with the trivial identity $\hat{N}^\star q_k = \hat{N}^\star q_k$ leads to

$$\left(I + \frac{1}{\delta}\hat{N}^\star\hat{N}\right)\hat{N}^\star q_k = \left(1 + \frac{\mu_k}{\delta}\right)\hat{N}^\star q_k.$$

Or equivalently

$$\left(I + \frac{1}{\delta}\hat{N}^\star\hat{N}\right)^{-1}\hat{N}^\star q_k = \left(1 + \frac{\mu_k}{\delta}\right)^{-1}\hat{N}^\star q_k.$$

Finally, multiplying the last equation by $\hat{N}$ and using the fact that $\{q_k\}$ are eigenfunctions of $\hat{N}\hat{N}^\star$, we get

$$\begin{aligned}
\hat{N}\left(I + \tfrac{1}{\delta}\hat{N}^\star\hat{N}\right)^{-1}\hat{N}^\star q_k &= \left(1 + \tfrac{\mu_k}{\delta}\right)^{-1}\hat{N}\hat{N}^\star q_k \\
&= \left(1 + \tfrac{\mu_k}{\delta}\right)^{-1}\mu_k q_k \\
&= \tfrac{\mu_k\delta}{\delta+\mu_k} q_k.
\end{aligned}$$

We also have

$$\frac{(\psi, NM_\delta^{-1}N^\star\psi)}{\|\psi\|_0^2} = \frac{\sum c_k^2 \frac{\mu_k\delta}{\delta+\mu_k}\|q_k\|_0^2}{\sum c_k^2\|q_k\|_0^2}.$$

($\|.\|_0$ stands for $L^2$–norm.)

This shows that the spectrum of $NM_\delta^{-1}N^\star$ is contained in the interval

$$\left[\min_k \frac{\mu_k\delta}{\delta+\mu_k}, \max_k \frac{\mu_k\delta}{\delta+\mu_k}\right].$$

Since the function $g(t)=\frac{\delta t}{\delta+t}$ is increasing we get the following interval that contains the spectrum of $NM_\delta^{-1}N^\star$,

$$\left[\frac{\delta\lambda_0}{\delta+\lambda_0}, \delta\right].$$

□

*Remark 1.* Lemma 1 is related to Lemma 3.2 of Axelsson [3]. The difference is that in [3] $N^\star N$ was assumed invertible. Also our result gives more explicit eigenvalue bounds.

In the case of mixed finite element discretization of second order elliptic equations the following lower bound for $\lambda_0$, the minimal eigenvalue of $NM^{-1}N^\star$, can be derived. We remark that similar estimate can be derived in the discretization of the stationary Stokes equations but we shall not pursue this issue further in the present paper.

**Lemma 2.** *The minimal eigenvalue of $NM^{-1}N^\star$ is uniformly in $h \to 0$ bounded below. We have*

$$\lambda_0 \geq \beta^2 \inf_{(x,y)\in\Omega} \lambda_{\min}[a],$$

*where $\beta$ is the constant from the Babuška–Brezzi stability condition (7) and $a$ is the coefficient matrix.*

*Proof.* Recalling the Babuška–Brezzi stability condition (7) we get

$$\begin{aligned}\beta\|q\|_0 &\leq \sup_{\underline{\chi}\in H(\mathrm{div};\Omega)} \frac{(\nabla\cdot\underline{\chi},q)}{\|\underline{\chi}\|_{H(\mathrm{div};\Omega)}}\\ &= \sup_{\underline{\chi}\in H(\mathrm{div};\Omega)} \frac{(N^\star q,\underline{\chi})}{\|\underline{\chi}\|_{H(\mathrm{div};\Omega)}}\\ &\leq \|N^\star q\|_0.\end{aligned}$$

In other words, we have

$$\lambda_{\min}[NN^\star] \geq \beta^2.$$

Then, since the operator $M$ is spectrally equivalent to the identity operator (it differs from $I$ by the coefficient matrix $a^{-1}$ ) we see that

$$\lambda_0 \equiv \lambda_{\min}\left[NM^{-1}N^\star\right] \geq \beta^2 \inf_{(x,y)\in\Omega} \lambda_{\min}[a] = O(1).$$

□

*Corollary 1.* Lemma 1 and Lemma 2 imply that $NM_\delta^{-1}N^\star$ is a well–conditioned operator.

We are now ready to prove our first main result.

**Theorem 1.** *The eigenvalues of the generalized eigenvalue problem*

$$\begin{bmatrix} M_\delta & N^\star \\ N & 0 \end{bmatrix} \begin{bmatrix} \underline{\chi} \\ \psi \end{bmatrix} = \lambda \begin{bmatrix} M_\delta & 0 \\ 0 & \delta_1 I \end{bmatrix} \begin{bmatrix} \underline{\chi} \\ \psi \end{bmatrix}$$

*are real and are contained in the following union of intervals* $\Delta^- \cup \Delta^+$, *where*

$$\Delta^- = \left[ \frac{1}{2} - \frac{1}{2}\sqrt{1 + 4\frac{\delta}{\delta_1}}, \frac{1}{2} - \frac{1}{2}\sqrt{1 + 4\frac{\lambda_0}{\delta + \lambda_0}\frac{\delta}{\delta_1}} \right]$$

*contains the negative eigenvalues, and*

$$\Delta^+ = \{1\} \cup \left[ \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\frac{\lambda_0}{\delta + \lambda_0}\frac{\delta}{\delta_1}},\ \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\frac{\delta}{\delta_1}} \right]$$

*contains the positive part of the spectrum. The eigenvalue 1 is of large multiplicity; it corresponds to the divergence–free space* $N\underline{\chi} = 0$ *(and* $\psi = 0$*). These estimates in particular show, since* $\delta > 0$, $\delta_1 > 0$ *and* $\lambda_0 = O(1)$ *are constants independent of the mesh parameter* $h \to 0$, *that* $\mathcal{D} = \begin{bmatrix} M_\delta & 0 \\ 0 & \delta_1 I \end{bmatrix}$ *is a symmetric and positive definite operator spectrally equivalent to the transformed saddle–point operator* $\mathcal{A}_\delta = \begin{bmatrix} M_\delta & N^\star \\ N & 0 \end{bmatrix}$.

*Proof.* Since $\mathcal{A}_\delta$ and $\mathcal{D}$ are symmetric and $\mathcal{D}$ is positive definite it is clear that the eigenvalues $\lambda$ are real. The eigenvalue problem is equivalent to the following system

$$\begin{aligned} (1-\lambda)M_\delta\underline{\chi} \quad &+N^\star\psi &&= 0, \\ N\underline{\chi} \quad &-\delta_1\lambda\psi &&= 0\cdot \end{aligned} \tag{12}$$

Consider first the case $\lambda > 0$. Taking inner product of the first equation with $\underline{\chi}$ and inner product of the second equation with $-\psi$ and adding the results we get the identity

$$(1-\lambda)(M_\delta\underline{\chi}, \underline{\chi}) + \delta_1\lambda\|\psi\|_0^2 = 0\cdot$$

If $\lambda = 1$ it follows that $\psi = 0$ and hence $N\underline{\chi} = 0$ (see the second equation of (12)). If $\lambda \neq 1$ then since $\lambda > 0$ we get that $\lambda > 1$. I.e., all positive eigenvalues that are different from one are to the right of one. Next, we estimate those positive eigenvalues and the negative part of the spectrum.

We solve the first equation in (12) for $\underline{\chi}$ and substitute it in the second one. We get

$$NM_\delta^{-1}N^\star\psi = -\delta_1\lambda(1-\lambda)\ \psi\cdot$$

Using now Lemma 1 that provides $O(1)$ bound $\frac{\delta\lambda_0}{\delta+\lambda_0}$ for the minimal eigenvalue of $NM_\delta^{-1}N^\star$ (because of Lemma 2) we get the inequality

$$-\delta_1\lambda(1-\lambda) \geq \frac{\delta\lambda_0}{\delta + \lambda_0},$$

or

$$\lambda^2 - \lambda - \delta_1^{-1} \frac{\delta \lambda_0}{\delta + \lambda_0} \geq 0,$$

which gives us the following negative upper bound for $\Delta^-$

$$\lambda \leq \frac{1}{2} - \frac{1}{2}\sqrt{1 + 4\frac{\lambda_0}{\delta + \lambda_0}\frac{\delta}{\delta_1}},$$

and the following positive lower bound of $\Delta^+ \setminus \{1\}$

$$\lambda \geq \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\frac{\lambda_0}{\delta + \lambda_0}\frac{\delta}{\delta_1}}.$$

Similarly, using the upper bound $\delta$ for the eigenvalues of $NM_\delta^{-1}N^\star$ provided by Lemma 1 we get the inequality

$$-\delta_1 \lambda(1 - \lambda) \leq \delta,$$

or

$$\lambda^2 - \lambda - \delta_1^{-1}\delta \leq 0,$$

which gives us the following lower bound for $\Delta^-$

$$\lambda \geq \frac{1}{2}\left(1 - \sqrt{1 + 4\frac{\delta}{\delta_1}}\right),$$

and the following upper bound for $\Delta^+$

$$\lambda \leq \frac{1}{2}\left(1 + \sqrt{1 + 4\frac{\delta}{\delta_1}}\right).$$

This completes the proof of the theorem. □

**Theorem 2.** *If $M_\delta$ is replaced by some spectrally equivalent preconditioner $B_\delta$, then the modified block–diagonal preconditioner $\mathcal{D} = \begin{bmatrix} B_\delta & 0 \\ 0 & \delta_1 I \end{bmatrix}$ will still be spectrally equivalent to the transformed saddle–point operator $\mathcal{A}_\delta$. More precisely, assume that for some positive constants $\gamma_1$ and $\gamma_2$ the following spectral equivalence relations hold*

$$\gamma_1(M_\delta \underline{\chi}, \underline{\chi}) \leq (B_\delta \underline{\chi}, \underline{\chi}) \leq \gamma_2(M_\delta \underline{\chi}, \underline{\chi}) \quad \text{for all } \underline{\chi} \in \mathcal{V}. \tag{15}$$

*Then the eigenvalues of $\mathcal{D}^{-1}\mathcal{A}_\delta$ are contained in the following union of intervals $\Delta^- \cup \Delta^+$ where*

$$\Delta^- = \left[\frac{1}{2\gamma_1} - \frac{1}{2\gamma_1}\sqrt{1 + 4\frac{\gamma_1}{\delta_1}\delta},\ \frac{1}{2\gamma_2} - \frac{1}{2\gamma_2}\sqrt{1 + 4\frac{\lambda_0}{\delta + \lambda_0}\frac{\gamma_2}{\delta_1}\delta}\right] \tag{16}$$

*contains the negative eigenvalues, and*

$$(17)\quad \Delta^+ = [\gamma_2^{-1}, \gamma_1^{-1}] \cup \left[ \frac{1}{2\gamma_2} + \frac{1}{2\gamma_2}\sqrt{1 + 4\frac{\lambda_0}{\delta+\lambda_0}\frac{\gamma_2}{\delta_1}\delta}\ ,\ \frac{1}{2\gamma_1} + \frac{1}{2\gamma_1}\sqrt{1+4\frac{\gamma_1}{\delta_1}\delta} \right]$$

*contains the positive eigenvalues. All constants ($\lambda_0$, $\delta_1$ and $\delta$) are as in Theorem 1.*

*Proof.* The eigenvalue problem is equivalent to the following system

$$(18)\qquad \begin{array}{lll} (M_\delta - \lambda B_\delta)\underline{\chi} & +N^\star\psi & = 0, \\ N\underline{\chi} & -\delta_1\lambda\psi & = 0\cdot \end{array}$$

Let $\lambda > 0$ do not lie in $[\gamma_2^{-1}, \gamma_1^{-1}]$. Hence $M_\delta - \lambda B_\delta$ is definite (negative or positive). From the identity

$$\big((M_\delta - \lambda B_\delta)\underline{\chi}, \underline{\chi}\big) + \delta_1\lambda\|\psi\|_0^2 = 0,$$

we see that $M_\delta - \lambda B_\delta$ is negative definite. Hence $\lambda > \gamma_1^{-1}$. I.e., all positive eigenvalues (outside the interval $[\gamma_2^{-1}, \gamma_1^{-1}]$) are greater than $\gamma_1^{-1}$. We next estimate this part of the spectrum. i.e., all $\lambda > \gamma_1^{-1}$. In this case $M_\delta - \lambda B_\delta$ is negative definite. We solve the first equation in (18) for $\underline{\chi}$ and substitute it in the second one. We get

$$(19)\qquad N(M_\delta - \lambda B_\delta)^{-1}N^\star\psi = -\delta_1\lambda\psi\cdot$$

Noting now that the spectrum of $M_\delta(M_\delta - \lambda B_\delta)^{-1}$ is contained in $[\frac{1}{1-\lambda\gamma_1}, \frac{1}{1-\lambda\gamma_2}]$. Using now Lemma 1 we get the following inequalities

$$-\delta_1\lambda(1 - \lambda\gamma_1) \le \delta,$$

and

$$-\delta_1\lambda(1-\lambda\gamma_2) \ge \frac{\delta\lambda_0}{\delta+\lambda_0}\cdot$$

Solving these inequalities for $\lambda > 0$ we obtain the desired bounds for the second interval in $\Delta^+$.

The negative eigenvalues are estimated similarly as in Theorem 1. In this case $M_\delta - \lambda B_\delta$ is positive definite and the spectrum of $M_\delta(M_\delta - \lambda B_\delta)^{-1}$ is contained in $[\frac{1}{1-\lambda\gamma_2}, \frac{1}{1-\lambda\gamma_1}]$. Using them and Lemma 1 in (19) we obtain the inequalities

$$-\delta_1\lambda(1 - \lambda\gamma_1) \le \delta,$$

and

$$-\delta_1\lambda(1-\lambda\gamma_2) \ge \frac{\delta\lambda_0}{\delta+\lambda_0}\cdot$$

The bounds for $\Delta^-$ are obtained by solving the above inequalities for $\lambda < 0$. This completes the proof of the theorem. □

*Remark 2.* Note that Theorem 2 when $\gamma_2 = \gamma_1 = 1$ reduces to Theorem 1.

Theorem 2 implies that the problem of constructing spectrally equivalent preconditioners of the saddle–point operator is reduced to the construction of preconditioners for the symmetric and positive operator $M + \frac{1}{\delta}NN^\star$ that corresponds to the bilinear form

$$\int_\Omega \left( a^{-1}\underline{\chi} \cdot \underline{\theta} + \frac{1}{\delta}\nabla \cdot \underline{\chi}\ \nabla \cdot \underline{\theta} \right) \quad \text{for all } \underline{\chi},\ \underline{\theta} \in \mathcal{V}\cdot$$

We finally note that $\delta > 0$ need not necessarily be a small parameter. However, if $\delta \to 0$ (and $\delta_1 = O(1)$) the lengths of both intervals $\Delta^-$ and $\Delta^+$ decrease (see (15) and (16)), which improves (to a certain extend) the condition number of the preconditioned operator $\mathcal{D}^{-1}\mathcal{A}_\delta$. Note also that both ends of $\Delta^-$ approach zero as $\delta \to 0$ and $\delta_1 = O(1)$. In such a limit situation the convergence behavior of the iterative methods applied for solving systems of equations with the preconditioned matrix is not clear. So there is no obvious reason to choose $\delta \to 0$.

For preconditioning $M_\delta$ (including the case $\delta \to 0$) one possible choice can be the multilevel preconditioners constructed on the basis of the hierarchical decomposition of the space $\mathcal{V} = \mathcal{V}_h$ for the velocity unknown (see, Cai, Goldstein and Pasciak [12] and its optimal stabilization the hybrid multilevel preconditioner from Ewing, Pasciak and Vassilevski [16]). Another possible choice (only for two – dimensional domains) can be the multigrid methods from Vassilevski and Wang [25].

**4. Iterative Methods and Numerical Experiments.** In this section we mention some iterative techniques for solving systems of equations with symmetric and indefinite matrices. We also show some experiments that illustrate the convergence of the MINRES method with a preconditioner $\mathcal{D}$ for $\mathcal{A}_\delta$ for various choices of $\delta$ and $\delta_1$.

We first remark that the preconditioned operator $\mathcal{D}^{-1}\mathcal{A}_\delta$ is symmetric in the inner product generated by $\mathcal{D}$, i.e., defined by

$$(B_\delta \underline{\chi}, \underline{\theta}) + \delta_1(\psi, \phi) \quad \text{for all } (\underline{\chi}, \psi) \text{ and } (\underline{\theta}, \phi) \in \mathcal{V} \times \mathcal{W}\cdot$$

Therefore we can consider the iterative methods for the symmetric indefinite operator $\mathcal{C} \equiv \mathcal{D}^{-1}\mathcal{A}_\delta$ in the special inner product $(.,.)_\mathcal{D}$ defined above. The MINRES method (first presented in Paige and Saunders [22]) applied to the system $\mathcal{A}_\delta \mathbf{x} = \mathbf{b}$, generates at $k$th iteration of the method new search vector $\mathbf{d}_k$ which is $(\mathcal{C}\cdot, \mathcal{C}\cdot)_\mathcal{D}$ orthogonal to the previous $k$ search vectors $\{\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_{k-1}\}$. Then the $k$th iterate $\mathbf{x}_k$ is computed such that the residual norm $\|\mathbf{r}_k\|_{\mathcal{D}^{-1}} \equiv (\mathbf{r}_k, \mathbf{r}_k)^{\frac{1}{2}}_{\mathcal{D}^{-1}}$ $(\mathbf{r}_k = \mathbf{b} - \mathcal{A}_\delta \mathbf{x}_k)$ is minimized over the shifted space

$$\mathbf{x}_0 + \operatorname{Span}\left\{\mathbf{d}_0, \mathcal{C}\mathbf{d}_0, \dots, \mathcal{C}^{k-1}\mathbf{d}_0\right\}\cdot$$

Here, $\mathbf{x}_0$ is the initial iterate, $\mathbf{r}_0 = \mathbf{b} - \mathcal{A}_\delta \mathbf{x}_0$ is the initial residual and $\mathbf{d}_0 = \mathcal{D}^{-1}\mathbf{r}_0$ is the initial search vector.

The following convergence estimate holds

$$\|\mathbf{r}_k\|_{\mathcal{D}^{-1}} \le E_k \|\mathbf{r}_0\|_{\mathcal{D}^{-1}},$$

where

$$E_k = \min_{p \in \Pi_k, p(0)=1} \max_{\lambda \in \Delta^- \cup \Delta^+} |p(\lambda)|.$$

Here $\Pi_k$ is the set of all polynomials of degree $\leq k$; $\Delta^+$ and $\Delta^-$ contain the positive and negative part of the spectrum of $C$. Explicit expression for $E_n$ is known only in some special cases, for example when length $(\Delta^+) =$ length $(\Delta^-)$. Explicit formulas (involving elliptic functions) have been derived in Freund [19] for the asymptotic convergence factor of the method defined by $\lim_{k \to \infty} E_k^{\frac{1}{k}}$ for general intervals $\Delta^-$ and $\Delta^+$. Since in our application we are interested in a few iterations (i.e., $k$ small) the asymptotic convergence factors might not give good information about the actual convergence rate of the method. At any rate, however, if the intervals $\Delta^+$ and $\Delta^-$ are fixed (i.e., independent of the mesh size $h \to 0$) the convergence of MINRES will be independent of the mesh parameter since $E_k$ is independent of $h \to 0$.

Another way of solving systems of equations with indefinite symmetric matrices (in the complex case with indefinite hermitian matrices) was studied in Ashby, Manteuffel and Saylor [1]. The idea is to use a polynomial $S(\lambda)$ such that the transformed problem

$$S(\mathcal{C})\mathcal{C}\mathbf{x} = S(\mathcal{C})\hat{\mathbf{b}}, \quad (\hat{\mathbf{b}} = \mathcal{D}^{-1}\mathbf{b})$$

has a positive definite matrix $S(\mathcal{C})\mathcal{C}$ (in the corresponding inner product $(.,.)_{\mathcal{D}}$). This technique is known as polynomial preconditioning. Since now the preconditioned matrix $S(\mathcal{C})\mathcal{C}$ is positive definite then the standard CG (conjugate gradient) method can be applied. A simple choice is $S(\mathcal{C}) = \mathcal{C}$ which leads to the least squares preconditioned CG method. This technique needs a priori eigenvalue estimation in order to construct polynomial $S(\lambda)$. As shown in Ashby, Manteuffel and Saylor [1] this can be done adaptively.

At the end we present some numerical experiments that show the number of iterations in the MINRES method applied to solve the system (9). It corresponded to the discretization of (6) in $\Omega = (0,1)^2$ using the lowest order Raviart–Thomas spaces on square elements of size $h = \frac{1}{n}$, $n > 1$ a given integer. The test problem corresponded to coefficient $a^{-1} = 1 + 10(x^2 + y^2)$ and exact solution $p = x(1-x)^2y(1-y)^2$. The stiffness matrix corresponding to the operator $M$ was obtained by using numerical integration based on tensor product Simpson rule in each element.

The preconditioner $B_\delta$ for $M_\delta$ was defined by

$$B_\delta = D + \frac{1}{\delta} N N^\star,$$

where $D$ was computed from the bilinear form that corresponds to the operator $M$ using quadrature formula that gave rise to a diagonal form od $D$. Then the action of $B_\delta^{-1}$ was computed using the Sherman–Morrison–Woodbury formula (cf., e.g., Golub and van Loan, [20], p. 51)

$$B_\delta^{-1} = D^{-1} - D^{-1}N(\delta I + N^\star D^{-1} N)^{-1} N^\star D^{-1}.$$

Note now that $\delta I + N^\star D^{-1}N$ is very close to a cell-centered finite difference approximation of the operator $-\nabla \cdot a\nabla p + \delta p$. Since the goal of the present paper was to study the behavior of the global preconditioner $\mathcal{D}$, in our test the actions of $(\delta I + N^\star D^{-1}N)^{-1}$ were computed exactly based on Choleski factorization.

We used the preconditioned MINRES method with a stopping criterion when the $\mathcal{D}^{-1}$–norm of the residual was less than $10^{-9}$. In the tables below we show the number of iterations *iter* and the average reduction factor $\rho = \left(\frac{\|\text{last residual}\|_{\mathcal{D}^{-1}}}{\|\text{initial residual}\|_{\mathcal{D}^{-1}}}\right)^{\frac{1}{iter}}$ for various choices of $\delta > 0$ and $\delta_1 = 1$ or $\delta_1 = \delta$.

TABLE 1. Number of iterations, *iter* and average reduction factors $\rho$; $\delta = \delta_1$

| $h^{-1}$ | $\delta = 1$ | $\delta = 0.1$ | $\delta = 10^{-2}$ | $\delta = 10^{-3}$ | $\delta = 10^{-4}$ |
|---|---|---|---|---|---|
| 16 | 27<br>0.4516 | 23<br>0.3610 | 19<br>0.2831 | 19<br>0.2621 | 17<br>0.2133 |
| 32 | 26<br>0.4359 | 22<br>0.3451 | 18<br>0.26 | 18<br>0.24 | 16<br>0.1913 |
| 64 | 25<br>0.4191 | 19<br>0.3032 | 17<br>0.2417 | 17<br>0.2187 | 15<br>0.1697 |

TABLE 2. Number of iterations, *iter* and average reduction factors $\rho$; $\delta_1 = 1$

| $h^{-1}$ | $\delta = 1$ | $\delta = 0.1$ | $\delta = 10^{-2}$ | $\delta = 10^{-3}$ | $\delta = 10^{-4}$ |
|---|---|---|---|---|---|
| 16 | 27<br>0.4516 | 24<br>0.3999 | 23<br>0.3850 | 28<br>0.4548 | $\star$ |
| 32 | 26<br>0.4359 | 23<br>0.3857 | 22<br>0.3691 | 27<br>0.4420 | $\star$ |
| 64 | 25<br>0.4191 | 22<br>0.3694 | 21<br>0.3511 | 26<br>0.4272 | $\star$ |

"$\star$" stands for no convergence.

From both Tables 1 and 2 we see that the preconditioner is of optimal order; the number of iterations *iter* and the average reduction factors $\rho$ are bounded independently of the mesh size. The preconditioner works also for small (but not too small) $\delta$ and $\delta_1$. Note that in that case $M_\delta$ is very ill–conditioned. We also see that there is some optimum of how small $\delta$ and $\delta_1$ should be and when $\delta$ becomes less than that value (see Table 2) the number of iterations began to rapidly increase.

This was not observed in Table 1 ($\delta_1 = \delta$). Note that in that case with $\delta \to 0$ the intervals $\Delta^-$ and $\Delta^+$ (see Theorem 2) approach limit values and this was also observed in our test; i.e., the number of iterations *iter* tended to stabilize.

The convergence behavior in Table 2 ($\delta_1 = 1$) when $\delta \to 0$ is explained as follows. Note that in that case $\Delta^-$ tends to the origin and $M_\delta$ tends to a singular matrix. This is a possible explanation of the increase of the number of iterations for $\delta \leq 10^{-3}$.

## References

1. S.F. Ashby, T.A. Manteuffel and P.E. Saylor, *Adaptive polynomial preconditioning for hermitian indefinite systems*, BIT **29** (1989), 583-609.
2. O. Axelsson, *Preconditioning of indefinite problems by regularization*, SIAM J. Numer. Anal. **18** (1979), 58-69.
3. O. Axelsson, *Numerical algorithms for indefinite problems*, Elliptic Problems Solvers, II, (G. Birkhoff and A. Schoenstadt, eds.), Academic Press, Orlando, 1984, pp. 219-232.
4. O. Axelsson and P.S. Vassilevski, *A black box generalized conjugate gradient solver with inner iterations and variable–step preconditioning*, SIAM J. Matrix Anal. Appl. **12** (1991), 625-644.
5. I. Babuška, *The finite element method with Lagrangian multipliers*, Numer. Math. **20** (1973), 179-192.
6. R.E. Bank, B.D. Welfert, and H. Yserentant, *A class of iterative methods for solving saddle–point problems*, Numer. Math. **56** (1990), 645-666.
7. M. Bercovier, *Perturbation of a mixed variational problem, applications to mixed finite element methods*, R.A.I.R.O. Anal. Numer. **12** (1978), 211-236.
8. J.H. Bramble and J.E. Pasciak, *A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems*, Math. Comp. **50** (1988), 1-17.
9. F. Brezzi, *On the existence, uniqueness, and approximation of saddle point problems arising from Lagrangian multipliers*, R.A.I.R.O. Anal. Numer. **2** (1974), 129-151.
10. F. Brezzi, J. Douglas, and L.D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math. **47** (1985), 217-235.
11. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer–Verlag, New York-Berlin-Heidelberg, 1991.
12. Z. Cai, C.I. Goldstein, and J.E. Pasciak, *Multilevel iteration for mixed finite element systems with penalty*, SIAM J. Sci. Stat. Comput., to appear.
13. R. Chandra, *Conjugate gradient methods for partial differential equations*, Research Report # 119, Department of Computer Science, Yale University, 1978.
14. J.Douglas and J. Wang, *A new family of mixed finite elements*, R.A.I.R.O. Anal. Numer., to appear.
15. R.E. Ewing, R.D. Lazarov, Peng Lu, and P.S. Vassilevski, *Preconditioning indefinite systems arising from mixed finite element discretization of second–order elliptic problems*, Lecture Notes in Math. # 1457, Springer, Berlin, 1990, pp. 28-43.
16. R.E. Ewing, J.E. Pasciak and P.S. Vassilevski, *Hybrid hierarchical methods for mixed finite element systems with penalty*, E.O.R.I., preprint (1991).
17. R.E. Ewing and J. Wang, *Analysis of Schwarz algorithm for mixed finite element methods*, R.A.I.R.O. Math. Modelling and Numer. Anal. **26** (1992), 739-756.
18. R.E. Ewing and J. Wang, *Analysis of multilevel decomposition iterative methods for mixed finite element methods*, R.A.I.R.O. Math. Modelling and Numer. Anal., to appear.
19. R.W. Freund, *On polynomial preconditioning and asymptotic convergence factors for indefinite Hermitian matrices*, Linear Algebra Appl. **154–156** (1991), 259–288.
20. G.H. Golub and C.F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London, 1989.
21. T.P. Mathew, *Domain Decomposition and Iterative Refinement Methods for Mixed Finite Element Discretizations of Elliptic Problems*, Ph.D. Thesis, Courant Institute, New York (1989).

22. C.C. Paige and M.A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. **12** (1975), 617-629.
23. P.A. Raviart and J.M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, Mathematical Aspects of Finite Element Methods, Lect. Notes in Math., vol. 606, Springer–Verlag, New York, 1977, pp. 292-315.
24. T. Rusten and R. Winther, *A preconditioned iterative method for saddle–point problems*, SIAM J. Matrix Anal. Appl. **13** (1992), 887-904.
25. P.S. Vassilevski and J. Wang, *Multilevel iterative methods for mixed finite element discretizations of elliptic problems*, Numer. Math. **63** (1992), 503-520.

Department of Mathematics, University of California at Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90024-1555, U.S.A.
*E-mail address*: panayot@math.ucla.edu

Department of Mathematics, Texas A&M University, College Station, TX 77843-3388, U.S.A.
*E-mail address*: lazarov@math.tamu.edu