AN OPTIMAL PRECONDITIONER FOR A CLASS OF SADDLE POINT PROBLEMS WITH A PENALTY TERM, PART II: GENERAL THEORY

AXEL KLAWONN *

April 1995

Abstract. Iterative methods are considered for saddle point problems with a penalty term. A positive definite preconditioner is constructed and it is proved that the condition number of the preconditioned system can be made independent of the discretization and the penalty parameters. Examples include the pure displacement problem in linear elasticity, the Timoshenko beam, and the Mindlin-Reissner plate.

Key words. saddle point problems, penalty term, nearly incompressible materials, Timoshenko, Mindlin-Reissner, preconditioned conjugate residual method, multilevel, domain decomposition,

AMS(MOS) subject classifications. 65F10,65N22,65N30,65N55,73V05

1. Introduction. In this article, we extend the preconditioning strategy for saddle point problems with a penalty term, discussed in Klawonn [15], to a more general class of problems, including the Timoshenko beam and the Mindlin-Reissner plate. We consider problems of the form

$$a(u,v) + b(v,p) = (f,v)_0 \quad \forall v \in V,$$

 $b(u,q) - t^2 c(p,q) = (g,q)_0 \quad \forall q \in M_c \quad t \in [0,1],$

where V,M and M_c are Hilbert spaces with M_c dense in M. In [15], we required that $a(\cdot,\cdot)$ be V-elliptic and $c(\cdot,\cdot)$ be equivalent to the L_2 -inner product. In this paper, we prove that the condition number of the preconditioned system is bounded from above, independently of the discretization and the penalty parameters, if only the assumptions of the Babuška-Brezzi theory hold, i.e. i) $a(\cdot,\cdot)$ is elliptic on $V_0:=\{v\in V:b(v,q)=0\ \forall q\in M\}$, ii) $b(\cdot,\cdot)$ fulfills an inf-sup condition and iii) $c(\cdot,\cdot)$ is non-negative. From these conditions, one can obtain an inf-sup and a sup-sup condition for the bilinear form

$$\mathcal{A}((u,p),(v,q)) := a(u,v) + b(v,p) + b(u,q) - t^2 c(p,q)$$

defined on the space $X := V \times M_c$. The proof of the bound of the condition number is based mainly on the interpretation of these inf-sup and sup-sup conditions as providing an estimate for the condition number of $\mathcal{A}(\cdot, \cdot)$. Our preconditioning strategy can then be interpreted as introducing a new metric on $V \times M_c$, i.e. performing a change of basis. Let us point out that a unifying multigrid approach for saddle point problems with a penalty term is developed in Brenner [8].

^{*} Westfälische Wilhelms-Universität, Institut für Numerische und instrumentelle Mathematik, Einsteinstraße 62, 48149 Münster, Germany. Electronic mail address: klawonn@math.uni-muenster.de. Dieser Artikel ist Teil einer geplanten Dissertation an der Mathematisch-Naturwissenschaftlichen Fakultät der Westfälischen Wilhelms-Universität Münster. This work was supported in part by the U. S. Department of Energy under contract DE-FG02-92ER25127.

The outline of the remainder of the paper is as follows. In Section 2, we describe an abstract theory for saddle point problems with a penalty term using the bilinear form $\mathcal{A}(\cdot,\cdot)$. As examples, we discuss the Timoshenko beam and the Mindlin-Reissner plate. In Section 3, we analyze the preconditioner discussed in Klawonn [15], using the results from the previous section and give our condition number estimate.

- 2. Saddle point problems with a penalty term. In this section, we first describe an abstract framework for saddle point problems with a penalty term and then give some examples arising in solid mechanics.
- **2.1. The abstract framework.** Let $(V, \|\cdot\|_V)$ and $(M, \|\cdot\|_M)$ be two Hilbert spaces, let M_c be a dense subspace of M, and let

(1)
$$a(\cdot,\cdot): V \times V \to \mathbf{R}, \ b(\cdot,\cdot): V \times M \to \mathbf{R}, \ c(\cdot,\cdot): M_c \times M_c \to \mathbf{R},$$

be three continuous bilinear forms. Additionally, we introduce V_0 , a subspace of V, given by $V_0 := \{v \in V : b(v,q) = 0 \ \forall q \in M\}$. We assume that $a(\cdot,\cdot)$ is V_0 -elliptic and that $c(\cdot,\cdot)$ is M_c -positive semi-definite. We consider the following problem: Find $(u,p) \in V \times M_c$, such that

We denote by $X := V \times M_c$ the product space and by

$$A(x,y) := a(u,v) + b(u,q) + b(v,p) - t^2 c(p,q),$$

$$x = (u,p) \in X, \ y = (v,q) \in X,$$

the bilinear form of problem (2) on X. With the additional definition

$$\mathcal{F}(y) := \langle f, v \rangle + \langle g, q \rangle,$$

we obtain an equivalent formulation of problem (2)

(3)
$$\mathcal{A}(x,y) = \mathcal{F}(y) \quad \forall y \in X.$$

We equip X with a new norm. We assume that we have an additional norm on M_c , i.e. $|||\cdot|||_M$, and introduce the new norm on X by

$$|||x||| := ||u||_V + |||q|||_M$$
 for $x = (u, p) \in X$.

Remark 1. If the bilinear form $c(\cdot,\cdot)$ is continuous on $M\times M$, we define $|||p|||_M:=||p||_M$. Otherwise, $|||p|||_M$ is defined by $||p||_M+t|p|_c$, where $|p|_c:=\sqrt{c(p,p)}$ is a semi-norm on M_c .

According to the well-known theory of Babuška; see [5], we have to verify a supsup and an inf-sup condition to guarantee the well-posedness of the problem. From the assumptions, we can conclude that $\mathcal{A}(\cdot,\cdot)$ is a continuous bilinear form on X, i.e.

(4)
$$\sup_{y \in X} \sup_{x \in X} \frac{\mathcal{A}(x, y)}{|||x||| |||y|||} \leq \gamma_1,$$

where $\gamma_1 > 0$ is independent of $t \in [0,1]$. Additionally, $\mathcal{A}(\cdot,\cdot)$ has to fulfill an inf-sup condition,

(5)
$$\inf_{y \in X} \sup_{x \in X} \frac{\mathcal{A}(x, y)}{|||x||| \, |||y|||} \ge \gamma_0 > 0,$$

where γ_0 is independent of $t \in [0, 1]$.

Theorem 1. Let the following three assumptions be satisfied:

(i) The continuous bilinear form $a(\cdot,\cdot)$ is V_0 -elliptic, i.e.

$$\exists \alpha_0 > 0$$
, such that $a(v, v) \geq \alpha_0 ||v||_V^2 \quad \forall v \in V_0$,

(ii) The continuous bilinear form $b(\cdot,\cdot)$ fulfills an inf-sup condition, i.e.

$$\exists \beta_0 > 0, \ such \ that \ \inf_{q \in M_c} \sup_{v \in V} \frac{b(v,q)}{\|v\|_V \|q\|_M} \ge \beta_0,$$

(iii) The continuous bilinear form $c(\cdot,\cdot)$ is M_c -positive semi-definite, i.e.

$$c(q,q) \geq 0 \quad \forall q \in M_c.$$

Then, the inf-sup condition (5) holds if in addition one of the following conditions is satisfied:

- 1) The bilinear form $c(\cdot, \cdot)$ is continuous on $M \times M$.
- 2) The bilinear form $a(\cdot, \cdot)$ is V-elliptic.

Proof: Let us first assume that condition 1) holds. Then, we define the norm $|||\cdot|||_M$ by $|||p|||_M := ||p||_M$; see Remark 1. The proof that the inf-sup condition (5) holds, can be found in Braess and Blömer [7].

Now, let us assume that 2) is fulfilled. In this case, we define $|||p|||_M := ||p||_M + t|p|_c$, see Remark 1. For a special formulation of the Mindlin-Reissner plate a proof for (5) can be found in Huang [14], Lemma 3.1. The arguments given in that proof immediately carry over to the abstract setting used in this section.

All these results are also valid for suitable finite element spaces; see 2.2. We then require, additionally, that the constants in Theorem 1 are independent of h. The continuity assumptions turn into uniform boundedness with respect to h; see e.g. Braess [6].

2.2. Examples. We now discuss some problems from solid mechanics that can be treated within this abstract framework. We denote the finite element spaces approximating V and M (resp. M_c) with V^h and M^h (resp. M_c^h). Since we have already treated the pure displacement problem for nearly incompressible materials in Klawonn [15], we refer to that article for that case.

The notations that are used for the operators and finite element spaces, are collected in Section 4.

3

2.2.1. The Timoshenko beam problem. Let $\Omega := I \subset \mathbf{R}$ be a finite interval and $f \in L_2(I)$. The mixed Ansatz for the Timoshenko beam is given by

(6)
$$(\theta', \psi')_0 + (v' - \psi, \gamma)_0 = (f, v)_0 \ \forall (v, \psi) \in V,$$

(7)
$$(w'-\theta,\eta)_0 - t^2(\gamma,\eta)_0 = 0 \qquad \forall \eta \in M,$$

with $V := (H_0^1(I))^2$ and $M := L_2(I)$. Equations (6) and (7) represent a saddle point problem with a penalty term, in the sense of Theorem 1, with

$$a((w,\theta),(v,\psi)) := (\theta',\psi')_0,$$

 $b((v,\psi),\eta) := (v'-\psi,\eta)_0$
 $c(\gamma,\eta) := (\gamma,\eta)_0.$

We obtain the finite element formulation by replacing $V \times M := (H_0^1(I))^2 \times L_2(I)$ with $V^h \times M^h := (\mathcal{M}_{0,0}^k(\mathcal{T}_h))^2 \times \mathcal{M}^{k-1}(\mathcal{T}_h), \ k \geq 2$; see Arnold [1], Braess [6], p.278, and Braess and Blömer [7]. The proof that the assumptions of Theorem 1 hold, can also be found in these references.

2.2.2. The Mindlin-Reissner plate problem. Let $\Omega \subset \mathbf{R}^2$ be a polygonal domain and $f \in L_2(\Omega)$. The mixed Ansatz for the Mindlin-Reissner plate that we consider is part of the Brezzi-Fortin formulation of this problem; see Brezzi and Fortin [10]. The problem considered is:

Find $(\psi, q) \in (H_0^1(\Omega))^2 \times H^1(\Omega)/\mathbf{R}$, such that

(8)
$$a(\theta,\psi) + (rot(\psi),p)_0 = (f,\psi)_0 \quad \forall \psi \in \left(H_0^1(\Omega)\right)^2,$$

(9)
$$(rot(\theta), q)_0 - t^2(curl(p), curl(q))_0 = 0 \qquad \forall q \in H^1(\Omega)/\mathbf{R},$$

with $a(\theta,\psi):=\int_{\Omega}\left\{2\mu\epsilon(\theta):\epsilon(\psi)+\frac{\lambda}{2}\frac{2\mu}{\lambda+2\mu}div(\theta)\,div(\psi)\right\}dx$. The constants μ and λ denote the Lamé parameters and t represents the thickness of the plate.

By setting $x^{\perp} := (-x_2, x_1)$ for $x \in \mathbb{R}^2$, we have

$$rot(\psi)=div(\psi^{\perp}).$$

By using the definition of curl(p), we obtain an equivalent Stokes problem with a penalty term from (8),(9)

(10)
$$a(\theta^{\perp}, \psi^{\perp}) + (div(\psi^{\perp}), p)_{0} = (f, \psi^{\perp})_{0} \quad \forall \psi^{\perp} \in (H_{0}^{1}(\Omega))^{2},$$

(11)
$$(\operatorname{div}(\theta^{\perp}), q)_{0} - t^{2}(p, q)_{1} = 0 \qquad \forall q \in H^{1}(\Omega)/\mathbf{R}.$$

The finite element formulation is obtained by replacing $V \times M := (H_0^1(\Omega))^2 \times H^1(\Omega)/\mathbf{R}$ with $V^h \times M^h := (\mathcal{N}_{0,0}(\mathcal{T}_h))^2 \times \mathcal{M}_0^1(\mathcal{T}_h)/\mathbf{R}$, where $\mathcal{N}_{0,0}(\mathcal{T}) := \mathcal{M}_{0,0}^1(\mathcal{T}_h) \oplus \mathcal{B}^3(\mathcal{T}_h)$; see Arnold, Brezzi, and Fortin [3], or Arnold and Falk [4]. These finite element spaces correspond to the MINI-element of Arnold, Brezzi, and Fortin; see [3]. Since the problem considered is similar to the Stokes problem, it is also possible to use other elements that are common in fluid dynamics, e.g. the Taylor-Hood element; see Huang [14] and also Klawonn [15].

The assumptions of Theorem 1 hold with $V := (H_0^1(\Omega))^2$, $M := L_2(\Omega)/\mathbf{R}$, $M_c := H^1(\Omega)/\mathbf{R}$. A proof can be found in the references just provided.

It is also possible to use the MITCn-elements, introduced by Brezzi, Bathe, and Fortin in [9]; see also Peisker and Braess [16].

Finally, we would like to point out that the approach considered in this section is not the only one possible. There are direct approaches that do not use the Helmholtz decomposition; see Arnold and Brezzi [2].

3. The preconditioner. To construct the preconditioner, we work with the matrix representation of the saddle point problem,

(12)
$$\mathcal{A} := \begin{pmatrix} A & B^t \\ B & -t^2C \end{pmatrix} \in \mathbf{R}^{n+m} \times \mathbf{R}^{n+m},$$

and give the preconditioner the form

(13)
$$\hat{\mathcal{B}} := \begin{pmatrix} \hat{A} & O \\ O & \hat{C} \end{pmatrix} \in \mathbf{R}^{n+m} \times \mathbf{R}^{n+m}.$$

Here \hat{A} and \hat{C} satisfy certain ellipticity conditions, i.e. there exist positive constants a_0, a_1 and c_0, c_1 , such that

$$\begin{aligned} a_0^2 \, ||u||_V^2 & \leq u^t \hat{A} u \leq a_1^2 \, ||u||_V^2, \\ c_0^2 \, |||p|||_M^2 & \leq p^t \hat{C} p \leq c_1^2 \, |||p|||_M^2. \end{aligned}$$

The next lemma shows that $\hat{\mathcal{B}}$ is positive definite and defines a norm on X, which is equivalent to $|||\cdot|||$.

LEMMA 1. There exist positive constants b_0, b_1 , such that

$$|b_0|||x||| \le ||\hat{\mathcal{B}}^{1/2}x||_2 \le |b_1|||x|||.$$

Proof: Using the ellipticity of \hat{A} and \hat{C} , we obtain

$$\begin{split} \|\hat{\mathcal{B}}^{1/2}x\|_2^2 &= x^t \hat{\mathcal{B}}x \\ &= u^t \hat{A}u + p^t \hat{C}p \\ &\leq a_1^2 \|u\|_V^2 + c_1^2 \||p|||_M^2 \\ &\leq \max\{a_1^2, c_1^2\} (\|u\|_V^2 + \||p|||_M^2) \\ &\leq \max\{a_1^2, c_1^2\} (\|u\|_V + \||p|||_M)^2 \\ &= \max\{a_1^2, c_1^2\} \||x|||^2. \end{split}$$

Analogously, we get

$$\begin{split} |||x|||^2 &= (||u||_V + |||p|||_M)^2 \\ &\leq 2 ||u||_V^2 + 2|||p|||_M^2 \\ &\leq 2 \max\{\frac{1}{a_0^2}, \frac{1}{c_0^2}\} ||\hat{\mathcal{B}}^{1/2}x||_2^2 \\ &= 2 \left(\min\{a_0^2, c_0^2\}\right)^{-1} ||\hat{\mathcal{B}}^{1/2}x||_2^2. \end{split}$$

Examples for very fast and efficient methods that fulfill the ellipticity requirements, are given by domain decomposition and multigrid methods or, more generally, by Schwarz methods; see e.g. Dryja, Smith, and Widlund [12] or Chan and Mathew [11].

In view of Theorem 2 in Klawonn [15], see also Hackbusch [13], p. 270, our goal is to give an estimate of the condition number $\kappa(\hat{\mathcal{B}}^{-1}\mathcal{A}) := \rho(\hat{\mathcal{B}}^{-1}\mathcal{A})\rho((\hat{\mathcal{B}}^{-1}\mathcal{A})^{-1})$. Since $\hat{\mathcal{B}}$ is positive definite, $\hat{\mathcal{B}}^{-1}\mathcal{A}$ and $\hat{\mathcal{B}}^{-1/2}\mathcal{A}\hat{\mathcal{B}}^{-1/2}$ have the same eigenvalues and we obtain $\rho(\hat{\mathcal{B}}^{-1}\mathcal{A}) = \|\hat{\mathcal{B}}^{-1/2}\mathcal{A}\hat{\mathcal{B}}^{-1/2}\|_2$. Here, we use that $\hat{\mathcal{B}}^{-1/2}\mathcal{A}\hat{\mathcal{B}}^{-1/2}$ is normal. The same argument applies to $\rho\left((\hat{\mathcal{B}}^{-1}\mathcal{A})^{-1}\right)$. Thus, we only have to provide upper bounds for $\|\hat{\mathcal{B}}^{-1/2}\mathcal{A}\hat{\mathcal{B}}^{-1/2}\|_2$ and $\|(\hat{\mathcal{B}}^{-1/2}\mathcal{A}\hat{\mathcal{B}}^{-1/2})^{-1}\|_2$.

The next two Lemmata are well known and are given here for the sake of completeness only.

Lemma 2. Let L be a $(n+m) \times (n+m)$ - Matrix. Then, the following three inequalities are equivalent:

$$\inf_{y \in X} \sup_{x \in X} \frac{x^t L y}{\|x\|_2 \|y\|_2} \ge \alpha,$$
$$\|L y\|_2 \ge \alpha \|y\|_2 \quad \forall y \in X,$$
$$\|L^{-1}\|_2 \le \frac{1}{\alpha}.$$

Lemma 3. Let L be a $(n+m) \times (n+m)$ -Matrix. Then, the following three inequalities are equivalent:

$$\sup_{y \in X} \sup_{x \in X} \frac{x^t L y}{\|x\|_2 \|y\|_2} \le C,$$
$$\|L y\|_2 \le C \|y\|_2 \quad \forall y \in X,$$
$$\|L\|_2 \le C.$$

In the following lemma, we prove a lower (resp. upper) bound for the inf-sup (resp. sup-sup) of $\hat{\mathcal{B}}^{-1/2}\mathcal{A}\hat{\mathcal{B}}^{-1/2}$.

Lemma 4. There exist positive constants C_0, C_1 , such that

$$C_0 \le \inf_{y \in X} \sup_{x \in X} \frac{x^t \hat{\mathcal{B}}^{-1/2} \mathcal{A} \hat{\mathcal{B}}^{-1/2} y}{\|x\|_2 \|y\|_2},$$

$$C_1 \ge \sup_{y \in X} \sup_{x \in X} \frac{x^t \hat{\mathcal{B}}^{-1/2} \mathcal{A} \hat{\mathcal{B}}^{-1/2} y}{\|x\|_2 \|y\|_2}.$$

Proof: The Lemma follows immediately from (4), (5) by changing basis and by applying Lemma 1.

The next theorem follows from the definition of the condition number in combination with Lemmata 2, 3 and 4.

Theorem 2. The condition number of $\hat{\mathcal{B}}^{-1}\mathcal{A}$ is bounded independently of the discretization and the penalty parameters, i.e.

$$\kappa(\hat{\mathcal{B}}^{-1}\mathcal{A}) \le \frac{C_1}{C_0}.$$

4. List of Notations. We introduce the following product for matrices

$$\sigma: \tau := \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij} \tau_{ij},$$

where $\sigma, \tau \in Mat(d,d)$.

The **Sobolev spaces** are defined by:

$$\begin{array}{lll} H^1_\Gamma(\Omega) &:=& \{v\in H^1(\Omega): v_{|\Gamma_0}=0\}, \\ H^1_0(\Omega) &:=& H^1_\Gamma(\Omega) \text{ with } \Gamma=\partial\Omega. \end{array}$$

The finite element spaces by:

$$\begin{split} \mathcal{M}^k(\mathcal{T}_h) &:= \{v \in L_2(\Omega) : v_{|T} \in \mathcal{P}^k \text{ for all } T \in \mathcal{T}_h\} \\ \mathcal{M}^k_0(\mathcal{T}_h) &:= \mathcal{M}^k(\mathcal{T}_h) \cap H^1(\Omega), \\ \mathcal{M}^k_{0,0}(\mathcal{T}_h) &:= \mathcal{M}^k(\mathcal{T}_h) \cap H^1_0(\Omega), \\ \mathcal{B}^3(\mathcal{T}_h) &:= \{v \in \mathcal{M}^3_0(\mathcal{T}_h) : v \text{ vanishes on the boundary of every element }\}, \\ \mathcal{N}^1_0(\mathcal{T}_h) &:= \mathcal{M}^1_0(\mathcal{T}_h) \oplus \mathcal{B}^3(\mathcal{T}_h), \end{split}$$

where \mathcal{T}_h is a triangulation of Ω . Here, \mathcal{P}^k is the space of polynomials of degree $\leq k$. The **differential operators** are defined by:

$$div(v) := \sum_{i=1}^{d} \frac{\partial v_i}{\partial x_i},$$

$$rot(v) := -\frac{\partial v_1}{\partial x_2} + \frac{\partial v_2}{\partial x_1},$$

$$curl(q) := (-\frac{\partial q}{\partial x_2}, \frac{\partial q}{\partial x_1})^t,$$

$$\nabla v := (\nabla v_i)_{i=1,\dots,d},$$

$$\epsilon(v) := \frac{1}{2} \left(\nabla v + (\nabla v)^t\right).$$

All of these operators can be defined element by element on the space $\mathcal{M}^k(\mathcal{T}_h)$. The resulting discrete operators are marked by a subscript h.

Acknowledgements. This work was in part carried out while visiting the Courant Institute in New York in March 1995. I would like to thank Olof Widlund for making this visit possible and for many interesting discussions.

REFERENCES

- [1] Douglas N. Arnold. Discretization by finite elements of a model parameter dependent problem. Numer. Math., 37:405-421, 1981.
- [2] Douglas N. Arnold and Franco Brezzi. Some new elements for the Reissner-Mindlin plate model. In Boundary Value Problems for Partial Differential Equations and Applications, pages 287–292. Masson, 1993.
- [3] Douglas N. Arnold, Franco Brezzi, and Michel Fortin. A stable finite element for the Stokes equations. Calcolo, 21:337-344, 1984.
- [4] Douglas N. Arnold and Richard S. Falk. A uniformly accurate finite element method for the Reissner- Mindlin plate. SIAM J. Numer. Anal., 26(6):1276-1290, 1989.
- [5] Ivo Babuška and A.K. Aziz. Survey lectures on the mathematical foundations of the finite element method. In A.K. Aziz, editor, The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, pages 3-359. Academic Press, 1972.
- [6] Dietrich Braess. Finite Elemente. Springer-Verlag, 1992.
- [7] Dietrich Braess and C. Blömer. A multigrid method for a parameter dependent problem in solid mechanics. Numer. Math., 57:747-761, 1990.
- [8] Susanne C. Brenner. Multigrid methods for parameter dependent problems. Technical report, University of South Carolina, Columbia, 1993.
- [9] Franco Brezzi, Klaus-Jürgen Bathe, and Michel Fortin. Mixed-interpolated elements for Reissner-Mindlin plates. Int. J. Num. Meth. Eng. 28, 28:1787-1801, 1989.
- [10] Franco Brezzi and Michel Fortin. Numerical approximation of Mindlin-Reissner plates. Math. Comp., 47:151-158, 1986.
- [11] Tony F. Chan and Tarek P. Mathew. Domain decomposition algorithms. Acta Numerica, 1994.
- [12] Maksymilian Dryja, Barry F. Smith, and Olof B. Widlund. Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions. SIAM J. Numer. Anal., 31(6), December 1994.
- [13] Wolfgang Hackbusch. Iterative Lösung großer schwachbesetzter Gleichungssysteme. Teubner-Verlag, Stuttgart, 1991.
- [14] Ziping Huang. A Multi-Grid Algorithm for Mixed Problems with Penalty. Numer. Math., 57:227-247, 1990.
- [15] Axel Klawonn. An optimal preconditioner for a class of saddle point problems with a penalty term. Technical Report 19/94, Westfälische Wilhelms-Universität Münster, Germany, November 1994. http://www.math.uni-muenster.de/math/users/klawonn/.
- [16] Petra Peisker and Dietrich Braess. Uniform convergence of mixed interpolated elements for Reissner-Mindlin plates. M²AN, 26:557-574, 1992.