Applied Numerical Mathematics 8 (1991) 411–425
North-Holland

Domain decomposition preconditioning for p-version finite elements with high aspect ratios

Jan Mandel *

Computational Mathematics Group, University of Colorado at Denver, Denver, CO 80217-3364, USA

G. Scott Lett

Scientific Software-Intercomp, Denver, CO 80202, USA

Abstract

Mandel, J. and G.S. Lett, Domain decomposition preconditioning for p-version finite elements with high aspect ratios, Applied Numerical Mathematics 8 (1991) 411–425.

A recent domain decomposition type preconditioner for the p-version finite element method in two dimensions treats every element as a subdomain. We show that its condition number deteriorates as the square of the aspect ratio for thin elements, and develop an improved method whose condition number deteriorates only as the first power of the aspect ratio.

Keywords. Domain decomposition, finite elements, aspect ratio, trace inequality, polynomial extension.

1. Introduction

The p-version finite element method uses elements of high order to achieve high accuracy [1,5]. Then the number of degrees of freedom per element is large and domain decomposition methods for the solution of the discrete system can treat every element as a subdomain. This is especially attractive because the finite element data structures can be used also for the domain decomposition. Such a method was developed by Babuška, Craig, Mandel and Pitkäranta [2] for two-dimensional problems and by Mandel [14–17] for three dimensions. For other related methods and further considerations, see Babuška, Griebel, and Pitkäranta [4] and Babuška and Elman [3].

This method was inspired by a substructuring method of Bramble, Pasciak and Schatz [6] for the standard, h-version finite element method. For a recent unifying framework for methods of this type, see Widlund [19]. The advantage of the given subdomain partitioning, however, turns into a disadvantage when the elements have high aspect ratios or are highly distorted, as it is all

* Supported by National Science Foundation under Grant DMS-8704169.

too often the case in practice. The condition number depends upon the square of the aspect ratio. This is due to the separation of functions corresponding to opposite long sides of thin elements. We are thus led to the study of the effect of element distortion on convergence, and to the development of new methods with better properties in the presence of bad aspect ratios. Because we are concerned with self-adjoint, V-elliptic problems, which yield linear systems with symmetric, positive-definite matrices, the basic solution method is the preconditioned conjugate gradient method [10–12] and we concentrate on the design of effective preconditioners. We give a theoretical analysis of condition numbers for high aspect ratios, and present a method with better convergence properties in such a case. The bound on the condition number still depends (weakly) on the aspect ratio, due to estimates of the energy of bilinear interpolation used on thin domains. Very similar problems arise when domain decomposition methods for the standard h-version finite element methods use a given subdomain structure, such as resulting from some mesh generation algorithms. Most of the ideas in this paper carry over to that case, and the results will be presented elsewhere.

The influence of aspect ratios on the classical Schwarz alternating method was studied by Chan et al. [7–9], who have shown that in some cases, the Schwarz method converges independently of aspect ratios. Unfortunately, this result is not directly applicable to our methods.

In Section 2, we present some technical results concerning function spaces on thin domains. In particular, it turns out that the trace theorem for a short side of a rectangle can be formulated so that it is not sensitive to the aspect ratio. In Section 3, we formulate the problem to be solved and the general form of the preconditioner. Section 4 contains an analysis of the preconditioner from [2]. An improved preconditioner is proposed and analyzed in Section 5.

2. Functional spaces on thin domains

In this section, we prove several proposition concerning functional spaces on a thin element, which may be of separate interest.

2.1. Sobolev seminorms

We first define some norms and seminorms. We shall work with these explicitly given seminorms and norms for various domains, rather than with equivalent seminorms and norms as it is usually done. On a domain $\mathcal{O} \subset \mathbb{R}^2$ consider the Sobolev seminorm

$$|u|_{1,\mathcal{O}} = \left(\int\int_{\mathcal{O}} |\nabla u|^2 \, \mathrm{d}x \, \mathrm{d}y \right)^{1/2}, \tag{2.1}$$

and on a segment s the seminorm

$$|u|_{1/2,s} = \left(\int_s \int_s \left(\frac{u(t) - u(r)}{t - r} \right)^2 \mathrm{d}r \, \mathrm{d}t \right)^{1/2}. \tag{2.2}$$

The corresponding norms are given by

$$\| u \|_{1,\mathcal{O}}^2 = \| u \|_{L^2(\mathcal{O})}^2 + |u|_{1,\mathcal{O}}^2,$$

$$\| u \|_{1/2,s}^2 = \| u \|_{L^2(s)}^2 + |u|_{1/2,s}^2.$$

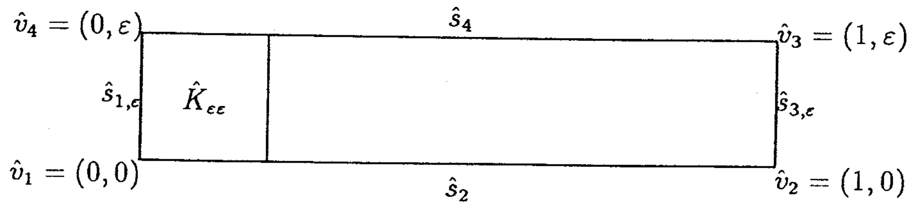

Fig. 1. Thin domain $\hat{K}_\varepsilon$.

We further denote

$$\| u \|_{k,\infty,\mathcal{O}} = \sum_{|\alpha| \leqslant k} \operatorname*{ess\,sup}_{\mathcal{O}} | D^\alpha u |.$$

2.2. Invariance and equivalence of seminorms

Consider the family of domains $\hat{K}_\varepsilon = (0, 1) \times (0, \varepsilon)$, $0 < \varepsilon \leqslant 1$ (see Fig. 1). We shall use the associated family of bijections

$$\hat{K}_1 \to \hat{K}_\varepsilon, \qquad F_\varepsilon(x, y) = (x, \varepsilon y).$$

We will also occasionally use the domains $\hat{K}_{\varepsilon\varepsilon} = (0, \varepsilon) \times (0, \varepsilon)$ and the corresponding mappings

$$\hat{K}_1 \to \hat{K}_{\varepsilon\varepsilon}, \qquad F_{\varepsilon\varepsilon}(x, y) = (\varepsilon x, \varepsilon y).$$

Let $\hat{s}_{1,\varepsilon} = \{0\} \times (0, \varepsilon)$ be one short side of $\hat{K}_\varepsilon$ as in Fig. 1. By a simple computation, we have for any $u \in H^1(\hat{K}_1)$ the invariance of seminorms,

$$|u \circ F_{\varepsilon\varepsilon}^{-1}|_{1,\hat{K}_{\varepsilon\varepsilon}} = |u|_{1,\hat{K}_1}, \tag{2.3}$$

$$|u \circ F_{\varepsilon\varepsilon}^{-1}|_{1/2,\hat{s}_{1,\varepsilon}} = |u \circ F_\varepsilon^{-1}|_{1/2,\hat{s}_{1,\varepsilon}} = |u|_{1/2,\hat{s}_{1,1}}. \tag{2.4}$$

Because

$$\int\int_{\hat{K}_\varepsilon} |\nabla(u \circ F_\varepsilon^{-1})|^2 = \varepsilon \int\int_{\hat{K}_1} \left(\frac{\partial u}{\partial x}\right)^2 + \frac{1}{\varepsilon^2}\left(\frac{\partial u}{\partial y}\right)^2,$$

we have the equivalence of seminorms,

$$\varepsilon^{1/2} |u|_{1,\hat{K}_1} \leqslant |u \circ F_\varepsilon^{-1}|_{1,\hat{K}_\varepsilon} \leqslant \varepsilon^{-1/2} |u|_{1,\hat{K}_1}. \tag{2.5}$$

It is easy to see that the constants in (2.5) are sharp by considering the functions $u(x, y) = x$ and $u(x, y) = y$. Finally note that

$$\int\int_{\hat{K}_1} u = \frac{1}{\varepsilon} \int\int_{\hat{K}_\varepsilon} u \circ F_\varepsilon^{-1} = \frac{1}{\varepsilon^2} \int\int_{\hat{K}_{\varepsilon\varepsilon}} u \circ F_{\varepsilon\varepsilon}^{-1}. \tag{2.6}$$

We shall often take a known theorem for the fixed domain $\hat{K}_1$ and then use the scaling properties of the seminorms given by (2.1)–(2.2) to obtain a result for $\hat{K}_\varepsilon$.

2.3. Short side trace theorem

With the definitions (2.1) and (2.2), we can prove a version of the trace theorem for the family of domains $\hat{K}_\varepsilon$.

Theorem 2.1. *There is a constant C independent of* ε *such that for all* $\varepsilon \in (0, 1]$,

$$|u|_{1/2,\hat{s}_{1,\varepsilon}} \leqslant C|u|_{1,\hat{K}_\varepsilon}, \quad \forall u \in H^1(\hat{K}_\varepsilon).$$

Proof. Let $u \in H^1(\hat{K}_\varepsilon)$. Since $|u+c|_{1,\hat{K}_\varepsilon} = |u|_{1,\hat{K}_\varepsilon}$ for any constant function c, we may assume that $\iint_{\hat{K},\varepsilon} u = 0$. Put $v = u \circ F_\varepsilon$. Then $v \in \hat{K}_1$, and by (2.6), $\iint_{\hat{K}_1} v = 0$. From the trace theorem (see, e.g., [18]) we have:

$$|v|_{1/2,\hat{s}_{1,1}} \leqslant \|v\|_{1/2,\hat{s}_{1,1}} \leqslant C_1 \|v\|_{1,\hat{K}_1} \leqslant C|v|_{1,\hat{K}_1},$$

and using (2.3) and (2.4) we get

$$|u|_{1/2,\hat{s}_{1,\varepsilon}} \leqslant C|u|_{1,\hat{K}_{\varepsilon\varepsilon}}, \quad \forall u \in H^1(\hat{K}_{\varepsilon\varepsilon}).$$

It remains to note that $|u|_{1,\hat{K}_{\varepsilon\varepsilon}} \leqslant |u|_{1,\hat{K}_\varepsilon}$. □

2.4. Polynomial spaces on $\hat{K}_\varepsilon$

Let $P_p(\hat{K}_\varepsilon)$ be the space of all real tensor product polynomials of degree up to p on $\hat{K}_\varepsilon$, that is, $u \in P_p(\hat{K}_\varepsilon)$ if and only if

$$u(x, y) = \sum_{i,j=0}^{p} c_{ij} x^i y^j, \quad c_{ij} \in \mathbb{R}.$$

Our first proposition is concerned with bounding the $|\cdot|_{1,\hat{K}_\varepsilon}$ seminorm of a bilinear function from its vertex values.

Lemma 2.2. *Let* $u_0 \in P_1(\hat{K}_\varepsilon)$, *and* $\hat{v}_1, \hat{v}_2, \hat{v}_3, \hat{v}_4$ *be the vertices of* $\hat{K}_\varepsilon$ *as in Fig. 1. Then*

$$|u_0|_{1,\hat{K}_\varepsilon} \leqslant C\varepsilon^{-1/2}\left(|u_0(\hat{v}_1) - u_0(\hat{v}_4)| + |u_0(\hat{v}_2) - u_0(\hat{v}_3)|\right) + \varepsilon^{1/2} \sum_{i=1}^{4} |u_0(\hat{v}_i)|,$$

where C does not depend on u_0 *or* $\varepsilon \in (0, 1]$.

Proof. First, let $u_0(\hat{v}_1) = u_0(\hat{v}_4) = 0$. Then for any $(x_0, y_0) \in \hat{K}_\varepsilon$,

$$\frac{\partial u_0}{\partial x}(x_0, y_0) = \left(1 - \frac{y_0}{\varepsilon}\right) u_0(\hat{v}_2) + \frac{y_0}{\varepsilon} u_0(\hat{v}_3),$$

$$\frac{\partial u_0}{\partial y}(x_0, y_0) = x_0 \frac{u_0(\hat{v}_3) - u_0(\hat{v}_2)}{\varepsilon},$$

so

$$\left(\frac{\partial u_0}{\partial x}\right)^2 + \left(\frac{\partial u_0}{\partial y}\right)^2 \leqslant (u_0(\hat{v}_2) + u_0(\hat{v}_3))^2 + \frac{1}{\varepsilon^2}(u_0(\hat{v}_3) - u_0(\hat{v}_2))^2,$$

and the lemma follows for this case. The case when $u_0(\hat{v}_2)=u_0(\hat{v}_3)=0$ is analogous. In the general case, write

$$u_0=u_1+u_2,$$

where

$$u_1,\ u_2\in P_1(\hat{k}_\varepsilon),\qquad u_1(\hat{v}_1)=u_1(\hat{v}_4)=u_2(\hat{v}_2)=u_2(\hat{v}_3)=0,$$

and use the result already proved for u_1 and u_2 separately along with the triangle inequality. □

In [2, Corollary 6.3], it was proved that

$$\|v\|_{0,\infty,\partial\hat{K}_1}\leqslant C(1+\log^{1/2}p)\,\|v\|_{1,\hat{K}_1},\quad \forall v\in P_p(\hat{K}_1). \tag{2.7}$$

Here we extend this result to a family of thin domains.

Theorem 2.3. *There is a constant C independent of ε such that*

$$\|u\|_{0,\infty,\partial\hat{K}_\varepsilon}\leqslant C\varepsilon^{-1/2}(1+\log^{1/2}p)\,|u|_{1,\hat{K}_\varepsilon}$$

for all $u\in P_p(\hat{K}_\varepsilon)$ *such that* $\iint_{\hat{K}_\varepsilon}u=0$, *and all* $\varepsilon\in(0,1]$.

Proof. Write $u=v\circ F_\varepsilon^{-1}$, and use (2.5), (2.6), and (2.7). □

On the short sides $\hat{s}_{1,\varepsilon}$ and $\hat{s}_{3,\varepsilon}$ we can get a bound independent of ε.

Theorem 2.4. *There is a constant C independent of ε such that*

$$\max_{r,s\in\hat{s}_{1,\varepsilon}}|u(r)-u(s)|\leqslant C(1+\log^{1/2}p)\,|u|_{1,\hat{K}_\varepsilon},\quad \forall u\in P_p(\hat{K}_\varepsilon). \tag{2.8}$$

for all $\varepsilon\in(0,1]$ *and all p.*

Proof. Let $\varepsilon=1$. Because the inequality (2.8) is invariant to adding a constant function to u, suppose without loss of generality that $\iint_{\hat{K}_1}u=0$; then (2.8) follows from (2.7). For a general $\varepsilon\in(0,1]$, (2.3) and (2.6) give

$$\max_{r,s\in\hat{s}_{1,\varepsilon}}|u(r)-u(s)|\leqslant C(1+\log^{1/2}p)\,|u|_{1,\hat{K}_{\varepsilon\varepsilon}},$$

and it remains to note that $|u|_{\hat{K}_{\varepsilon\varepsilon}}\leqslant|u|_{\hat{K}_\varepsilon}$. □

We will also need the following extension theorem.

Theorem 2.5. *Let* $u\in P_p(\hat{K}_\varepsilon)$ *such that* $u(\hat{v}_1)=u(\hat{v}_4)=0$. *Then there exist a function* $v\in P_p(\hat{K}_\varepsilon)$ *such that* $v=u$ *on* $\hat{s}_{1,\varepsilon}$, $v=0$ *on* $\partial\hat{K}_\varepsilon\backslash\hat{s}_{1,\varepsilon}$, *and*

$$|v|_{1,\hat{K}_\varepsilon}\leqslant C\varepsilon^{-1/2}\left(|u|_{1/2,\hat{s}_{1,\varepsilon}}+(1+\log^{1/2}p)\,\|u\|_{0,\infty,\hat{s}_{1,\varepsilon}}\right) \tag{2.9}$$

for all $\varepsilon\in(0,1]$, *and the constant C does not depend on ε or u.*

Proof. Let $\varepsilon = 1$. It was proved in [2, Theorem 6.6] that

$$\| u \|_{H_{00}^{1/2}(\hat{s}_{1,1})} \leqslant C\left(\| u \|_{1/2,\hat{s}_{1,1}} + (1 + \log^{1/2} p) \| u \|_{0,\infty,\hat{s}_{1,1}} \right)$$

Noting that

$$\| u \|_{1/2,\hat{s}_{1,1}}^2 = | u |_{1/2,\hat{s}_{1,1}}^2 + \| u \|_{L^2(\hat{s}_{1,1})}^2 \leqslant | u |_{1/2,\hat{s}_{1,1}}^2 + \| u \|_{0,\infty,\hat{s}_{1,1}}^2,$$

we put $v = u$ on $\hat{s}_{1,1}$, $v = 0$ on $\partial\hat{K}_1 \setminus \hat{s}_{1,1}$, and extend by [2, Theorem 7.5] to a function $v \in P_p(\hat{K}_1)$ to get (2.9) for $\varepsilon = 1$. The general case then follows using (2.3), (2.4), and (2.5). □

3. General preconditioning by local space decomposition

3.1. Preliminaries

In this paper, we are interested in the following model problem:

$$u \in V: \qquad a(u, v) = f(v), \quad \forall v \in V, \tag{3.1}$$

where $V = H_0^1(\Omega)$, Ω is a bounded domain with piecewise smooth boundary in $\mathbb{R}^2$, $f \in V'$, and

$$a(u, v) = \int\int_\Omega (\nabla u)^{\mathrm{T}} A \,\nabla v, \tag{3.2}$$

$A = (a_{ij}(x, y))_{i,j=1}^2$ is a symmetric coefficient matrix, $a_{ij} = a_{ji}$, such that

$$a_{ij} \in L^\infty(\Omega), \tag{3.3}$$

and A is uniformly positive-definite: there is a constant $\alpha > 0$ such that almost everywhere in Ω,

$$Z^{\mathrm{T}} A(x, y) Z \geqslant \alpha Z^{\mathrm{T}} Z, \quad \forall Z \in \mathbb{R}^2. \tag{3.4}$$

We consider solution of (3.1) by the usual finite element method [20] except that we use a family of reference elements defined by $\{\hat{K}_\varepsilon\}$, $\varepsilon \in (0, 1]$, rather than a single reference element. Let $\mathscr{E} = \{K\}$ be a partition of Ω into elements K, where each K is the image of one reference element $\hat{K}_\varepsilon$,

$$\bar{\Omega} = \bigcup_{K \in \mathscr{E}} \bar{K}, \qquad K = G_K\left(\hat{K}_{\varepsilon_K}\right), \quad \varepsilon_K \in (0, 1],$$

where $G_k \in P_i(\hat{K}_{\varepsilon_K})$ are bijections such that

$$\| G_K \|_{1,\infty,\hat{K}_{\varepsilon_K}} \leqslant C h_K, \qquad \left\| G_K^{-1} \right\|_{1,\infty,\hat{K}_{\varepsilon_K}} \leqslant C h_K^{-1},$$

and

$$\left\| J_{G_K} \right\|_{0,\infty,\hat{K}_{\varepsilon_K}} \leqslant C h_K^2, \qquad \left\| J_{G_K}^{-1} \right\|_{0,\infty,\hat{K}_{\varepsilon_K}} \leqslant C h_K^{-2},$$

J_{G_K} is the Jacobian of G_K, $h_K = \mathrm{diam}(K)$. We then define the finite element space

$$V_{\mathscr{E}} = \left\{ u \in V\colon u|_K \circ G_K \in P_p\left(\hat{K}_{\varepsilon_K}\right), \forall K \in \mathscr{E} \right\},$$

and the discretization of (3.1) is

$$u \in V_{\mathscr{E}}: \qquad a(u, v) = f(v), \quad \forall v \in V_{\mathscr{E}}. \tag{3.5}$$

We will be concerned with the numerical solution of the problem (3.5). After selecting a suitable basis of the space $V_{\mathscr{E}}$, (3.5) gives rise to a system of linear equations. We will study the solution of that system by the preconditioned conjugate gradients method. To this end, we need to construct another symmetric bilinear form $c(u, v)$ such that the problem

$$u \in V_{\mathscr{E}}: \qquad c(u, v) = g(v), \quad \forall v \in V_{\mathscr{E}} \tag{3.6}$$

for a given right-hand side g can be easily solved, and

$$m_1 c(u, u) \leqslant a(u, u) \leqslant m_2 c(u, u), \quad \forall u \in V_{\mathscr{E}} \tag{3.7}$$

holds with $m_1 > 0$, $m_2 < +\infty$. We will assume that m_1 and m_2 are the largest and the smallest possible, respectively, so that (3.7) holds, and denote the condition number

$$\kappa = m_2/m_1.$$

Then the preconditioned conjugate gradient method requires at most $O(\sqrt{\kappa})$ iterations to reduce the error by a fixed amount, see, e.g. [11]. Each iteration requires the solution of the problem (3.6).

The form $c(u, v)$ will be called a preconditioner.

3.2. General construction of preconditioner

We now describe a general principle for constructing the preconditioner. For each element K, define the local space

$$V_K = \{u|_K : u \in V_{\mathscr{E}}\}.$$

By integration in (3.2) over the elements K, we can write

$$a(u, v) = \sum_K a_K(u, v), \quad a_K(u, v) = \int\int_K (\nabla u)^{\mathrm{T}} A \, \nabla v.$$

Denote the energy seminorm on K by

$$|u|_K = (a_K(u, u))^{1/2}.$$

Now let V_K be split in a direct upper sum of subspaces

$$V_K = V_0 \oplus \cdots \oplus V_{n_K},$$

so every $v_K \in V_K$ can be written uniquely as the sum of $n_K + 1$ components,

$$v_K = v_{K,0} + \cdots + v_{K,n_K}, \quad v_{K,i} \in V_i. \tag{3.8}$$

We then define

$$c(u, v) = \sum_K c_K(u, v), \quad c_K(u, v) = a_K(u_{K,0}, v_{K,0}) + \cdots + a_K(u_{K,n_K}, v_{K,n_K}).$$

Obviously, if for all $K \in \mathscr{E}$,

$$m_{K,1} c_K(u, u) \leqslant a_K(u, u) \leqslant m_{K,2} c_K(u, u), \quad \forall u \in V_K, \tag{3.9}$$

then (3.7) holds with some

$$m_1 \geqslant \min_K m_{K,1}, \qquad m_2 \leqslant \max_K m_{K,2}.$$

Thus the condition number $K = m_2/m_1$ can be bounded element by element and independently of the number of elements. Let $m_{K,1}$ and $m_{K,2}$ be the largest and smallest, respectively, such that (3.9) holds, and denote the local condition number

$$\kappa_K = m_{K,2}/m_{K,1}.$$

The following theorem, given in [15, Theorem 4.1], shows that to bound the condition number, it is sufficient and necessary to bound the decomposition (3.8) in energy. For other versions and related results, see [2,13,16,19].

Theorem 3.1. *Let b_K be the least number such that*

$$\sum_{i=0}^{n_K} |u_{K,i}|_K^2 \leqslant b_K |u|_K^2, \quad \forall u \in V_K, \tag{3.10}$$

Then

$$b_K \leqslant \kappa_K \leqslant (n_K + 1) b_K.$$

In the following sections, we give specific definitions of the decomposition (3.1), and use Theorem 3.1 to estimate the condition number κ_K.

4. Preconditioning by bilinear elements

Let $K = G_K(\hat{K}_{\varepsilon_K}) \in \mathscr{E}$ be an element as in Fig. 2. Define the decomposition of the local space V_K by specifying the associated projections: Let

$$u_K = \sum_{i=0}^{5} u_{K,i}, \quad u_K \in V_K, \tag{4.1}$$

where

- $u_{K,0}$ is the mapped bilinear interpolation of u; that is, $u_{K,0}(v_i) = u_K(v_i)$, $i = 1, 2, 3, 4$, and $u_{K,0} \circ G_K \in P_1(\hat{K}_{\varepsilon_K})$;
- for $i = 1, 2, 3, 4$,

$$u_{K,i} = \begin{cases} u - u_{K,0} & \text{on the side } s_i, \\ 0 & \text{on } \partial K \setminus s_i, \end{cases}$$

  $u_{K,i} \in V_K$, and $|U_{K,i}|_K$ is minimal for all such functions;
- $u_{K,5} = u - \sum_{i=0}^{4} u_{K,i}$.

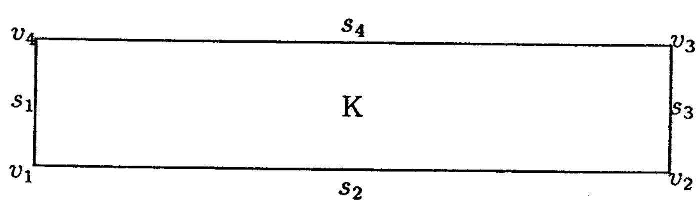

Fig. 2. Thin element.

By summation over all elements, we can see that the preconditioner $c(u, v)$ can be written as

$$c(u, v) = c(u_0, v_0) + \sum_e c(u_e, v_e) + \sum_K c(u_{K,5}, v_{K,5}),$$

where $u = u_0 + \Sigma_e u_e + \Sigma_K u_{K,5}$, u_0 is the piecewise mapped bilinear interpolation of u, u_e is a function determined by its values on one edge e of the finite element mesh, nonzero only on the two elements adjacent to the mesh, and $u_{K,5}$ is a function nonzero only on element K. The solution of the system (3.6) thus decouples into solution of a system, the matrix of which is the stiffeness matrix of the discretization of (3.1) by bilinear elements, and one subproblem for each side e and one for the interior of each element K. For more details and variants of the method, see [2,14]. Here we only note that the components $u_{K,5}$ can be eliminated in advance to obtain a mathematically equivalent preconditioner for a problem with fewer variables. This preconditioner was proposed and studied in [2], where it was proved that $\kappa_K \leqslant C(1 + \log^2 p)$ in the case $\varepsilon = 1$. Here we show how κ_K deteriorates for $\varepsilon \to 0$.

Theorem 4.1. *For the preconditioner defined by the decomposition* (4.1), *it holds that*

$$C_1 \varepsilon^{-2} \leqslant \kappa_K \leqslant C_2 \varepsilon^{-2}(1 + \log^2 p), \tag{4.2}$$

where $C_1 > 0$ and C_2 do not depend on ε or p.

Proof. In view of the assumptions on the mapping G_K, we may assume without loss of generality that

$$K = \hat{K}_\varepsilon.$$

Because of (3.3) and (3.4), we may also assume without loss of generality that

$$a_K(u, u) = |u|_{1,K}^2.$$

In view of Theorem 3.1, we need only to show that for the minimal b_K in (3.10),

$$b_K \leqslant C\varepsilon^{-2}(1 + \log^2 p), \tag{4.3}$$

$$b_K \geqslant C_1 \varepsilon^{-2}, \quad C_1 > 0. \tag{4.4}$$

In [2], it was proved that (4.3) holds for $\varepsilon = 1$. We give the proof here for the sake of completeness. Let $u \in V_K$. Because adding a constant to u will change only the component $u_{K,0}$ by the same constant, we may assume without loss of generality that $\iint_K u = 0$. It follows from Theorem 2.3 and Lemma 2.2 that

$$|u_{K,0}|_{0,\infty,K} \leqslant C(1 + \log^{1/2} p)|u|_{1,K},$$

$$|u_{K,0}|_{1,K} \leqslant C(1 + \log^{1/2} p)|u|_{1,K}.$$

Now $u(v_i) = u_{K,0}(v_i)$, $i = 1, 2, 3, 4$, and from the trace theorem and Theorem 2.5, there exist functions $w_i \in P_p(K)$, $i = 1, 2, 3, 4$, so that

$$w_i = \begin{cases} u - u_{K,0} & \text{on } s_i, \\ 0 & \text{on } \partial K \setminus s_i, \end{cases}$$

and

$$|w_i|_{1,K} \leqslant C(1 + \log p)|u|_{1,K}.$$

By the definition of $u_{K,i}$, we have $|u_{K,i}|_{1,K} \leqslant |w_i|_{1,K}$. The estimate of $u_{K,5}$ follows from the triangle inequality, and we may conclude that (3.10) holds with $b_K \leqslant C(1+\log^2 p)$ for $\varepsilon = 1$. The general case of (4.3) now follows using (2.5).

To prove (4.4), consider a polynomial $f(x)$ such that $f(0)=f(1)=0$ and let $u(x, y)=f(x)$ on K. Then the component $u_{K,0}=0$, and

$$|u|_{1,K}^2 = \varepsilon\int_0^1 |f'|^2 = \varepsilon |f|_{1,(0,1)}^2. \tag{4.5}$$

Let $v \in H^1(K)$ be determined by the boundary conditions

$$v = \begin{cases} f & \text{on } s_1, \\ 0 & \text{on } \partial K \setminus s_1, \end{cases} \tag{4.6}$$

and by the requirement that $|v|_{1,K}$ is minimal. Then the component $u_{K,1}$ satisfies

$$|u_{K,1}|_{1,K} \geqslant |v|_{1,K}, \tag{4.7}$$

and v is the solution of the boundary value problem $\Delta v = 0$ in K with the boundary conditions (4.6). Expanding f in a sine series,

$$f(x) = \sum_{n=1}^{\infty} a_n \sin(n\pi x),$$

we can write v as

$$v(x, y) = \sum_{n=1}^{\infty} a_n \frac{\sinh(n\pi(\varepsilon - y))}{\sinh(n\pi\varepsilon)} \sin(n\pi x).$$

Now using Green's theorem,

$$|v|_{1,K}^2 = \int\int_K (\nabla v)^{\mathrm{T}} \nabla v = \int_{\partial K} v\frac{\partial v}{\partial n} = -\int_0^1 v\frac{\partial v}{\partial y}\bigg|_{y=0} \mathrm{d}x.$$

By orthogonality of sine functions, it follows that

$$|v|_{1,K}^2 = \tfrac{1}{2}\sum_{n=1}^{\infty} n\pi a_n^2 \frac{\cosh(n\pi\varepsilon)}{\sinh(n\pi\varepsilon)} \geqslant \frac{1}{\varepsilon}\sum_{n=1}^{\infty} \tfrac{1}{2}a_n^2 = \frac{1}{\varepsilon}\| f \|_{0,(0,1)}^2$$

using the inequality

$$\frac{\cosh t}{\sinh t} > \frac{1}{t}, \quad t > 0,$$

and Parseval's equality. Thus using (4.5), we get

$$\frac{|v|_{1,\hat{K}_\varepsilon}^2}{|u|_{1,\hat{K}_\varepsilon}^2} \geqslant \frac{1}{\varepsilon^2}\frac{\| f \|_{0,(0,1)}^2}{|f|_{1,(0,1)}^2}.$$

Choosing $f(x) = x(1-x)$ and using (4.7), we get (4.4). □

We have calculated the condition numbers for $p = 2$ to 16 and $\varepsilon = 1$ to $\varepsilon = 2^{-13}$. The results are in Fig. 3. We can see that the condition numbers indeed grow as ε^{-2}. The influence of p on

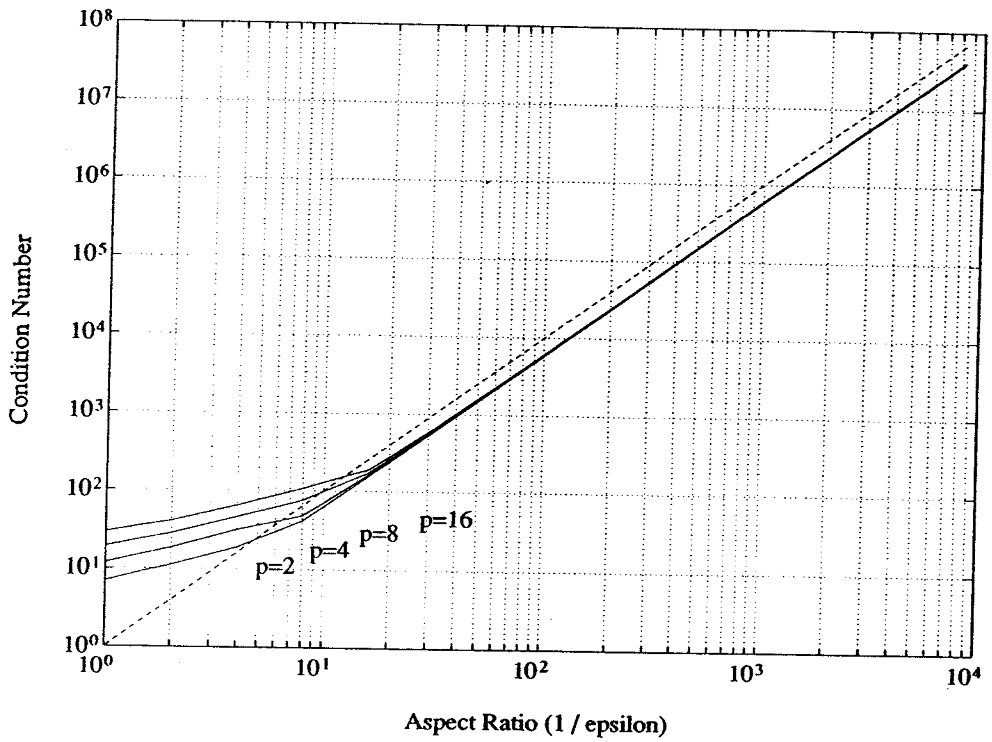

Fig. 3. Condition numbers for linear preconditioning.

the condition number is seen only for relatively small aspect ratios; for very small ε, the influence of ε prevails. The numerical results are in complete correspondence with Theorem 4.1.

5. Improved preconditioner

The lower bound on the condition number κ_K in the proof of Theorem 4.1 was obtained by splitting a function constant in the short direction in two components which correspond to the two long sides s_2 and s_4. This cannot happen in the preconditioner defined by the following decomposition.

For $K = G_K(\hat{K}_\varepsilon)$, we now define the decomposition of V_K by

$$u_K = \sum_{i=0}^{3} u_{K,i}, \tag{5.1}$$

where

- $u_{K,0}$ is mapped bilinear interpolation of u: $u_{K,0}(v_i) = u_K(v_i)$, $i = 1, 2, 3, 4$, and $u_{K,0} \circ G_K \in P_1(\hat{K}_\varepsilon)$;
- $u_{K,1}$, and $u_{K,3}$ are defined by the boundary conditions

$$u_{K,i} = \begin{cases} u - u_{K,0} & \text{on } s_i, \\ 0 & \text{on } \partial K \setminus s_i, \end{cases}$$

$i = 1, 3$, and by the requirement that $u_{K,i} \in V_K$ and $|u_{K,i}|_K$ is the minimal possible;

- $u_{K,2} = u_{K,0} - u_{K,1} - u_{K,3}$.

The preconditioner problem (3.6) now decomposes as in Section 4, except that there will be also subproblems corresponding to strips of the long edges and the interiors of adjacent thin elements. Because the matrix of such a subproblem is essentially the matrix of a one-dimensional problem, with low bandwidth or envelope size, direct solution of the auxiliary problem (3.6) is still quite cheap.

We can show that now the condition number κ_K deteriorates only as ε^{-1} rather than as ε^{-2}.

Theorem 5.1. *For the preconditioner defined by the decomposition* (4.1), *it holds that*

$$C_1 \min\{p, \varepsilon^{-1}\} \leqslant \kappa_K \leqslant C_2 \varepsilon^{-1}(1 + \log^2 p),$$

where $C_1 > 0$ *and* C_2 *do not depend on* ε *or* p.

Proof. As in the proof of Theorem 4.1, we may assume without loss of generality that

$$K = \hat{K}_\varepsilon, \qquad a_K(u, u) = |u|_{1,K}^2.$$

Because of Theorem 3.1, we only need to show that the minimal b_K in (3.10) satisfies

$$b_K \leqslant C\varepsilon^{-1}(1 + \log^2 p), \tag{5.2}$$

$$b_k \geqslant C_1 \min\{p, \varepsilon^{-1}\}, \quad C_1 > 0. \tag{5.3}$$

Let $u \in V_K = P_p(K) = P_p(\hat{K}_\varepsilon)$. Because adding a constant to u will change only the component $u_{K,0}$ by the same constant, we may assume without loss of generality that $\iint_K u = 0$. Then we have from Theorem 2.3,

$$|u(v_i)| \leqslant C\varepsilon^{-1/2}(1 + \log^{1/2} p)|u|_{1,K}, \quad i = 1, 2, 3, 4, \tag{5.4}$$

and from Theorem 2.4,

$$|u(v_1) - u(v_4)| \leqslant C(1 + \log^{1/2} p)|u|_{1,K}, \tag{5.5}$$

$$|u(v_2) - u(v_3)| \leqslant C(1 + \log^{1/2} p)|u|_{1,K}. \tag{5.6}$$

It follows from (5.4)–(5.6) and Lemma 2.2 that

$$|u_{K,0}|_{1,K} \leqslant C\varepsilon^{-1/2}(1 + \log^{1/2} p)|u|_{1,K}. \tag{5.7}$$

Further, (5.5) and (5.6) give

$$|u_{K,0}|_{1/2,s_i} \leqslant C(1 + \log^{1/2} p)|u|_{1,K}, \quad i = 1, 3, \tag{5.8}$$

because $u_{K,0}$ is linear on all sides, and because of the scaling invariance property (2.4). From (5.8) and Theorem 2.1, we have

$$|u - u_{K,0}|_{1/2,s_i} \leqslant C(1 + \log^{1/2} p)|u|_{1,K}, \quad i = 1, 3. \tag{5.9}$$

Now Theorem 2.4 and the definition of $u_{K,0}$ imply that

$$\|u - u_{K,0}\|_{0,\infty,s_i} \leqslant C(1 + \log^{1/2} p)|u|_{1,K}, \quad i = 1, 3. \tag{5.10}$$

Theorem 2.5 and (5.9), (5.10) now give the existence of functions $w_i \in P_p(K)$ such that for $i = 1$, 3

$$w_i = \begin{cases} u - u_{K,0} & \text{on } s_i, \\ 0 & \text{on } \partial K \setminus s_i, \end{cases}$$

and

$$|w_i|_{1,K} \leqslant C\varepsilon^{-1/2}(1 + \log p)|u|_{1,K}.$$

By the definition of $u_{K,i}$, we have $|u_{K,i}|_{1,K} \leqslant |w_i|_{1,K}$, which together with (5.7) and the triangle inequality concludes the proof of (5.2).

To prove (5.3), consider the function

$$u(x, y) = x^p y/\varepsilon$$

on $K = \hat{K}_\varepsilon$. Then the bilinear interpolation of u is $u_{K,0}(x, y) = xy/\varepsilon$. By a simple computation,

$$|u|_{1,K}^2 = \frac{1}{\varepsilon}\frac{1}{2p+1} + \frac{\varepsilon p^2}{3(2p-1)}.$$

Thus

$$|u_{K,0}|_{1,K}^2 > \frac{1}{3\varepsilon}$$

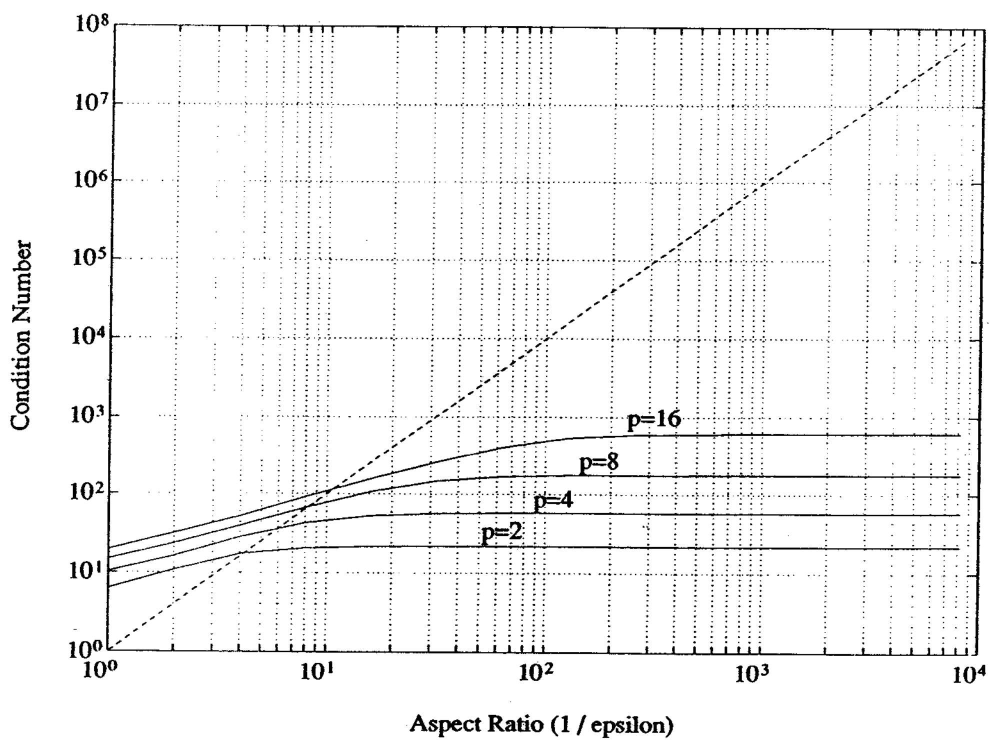

Fig. 4. Condition numbers for improved preconditioning.

and if $p=[1/\varepsilon]$,

$$|u|_{1,K}^2 \leqslant C,$$

which proves (5.3). □

Figure 4 summarizes numerically evaluated numbers κ_K for $p=2$ to 16 and $\varepsilon=1$ to 2^{-13}. The results are in a complete agreement with Theorem 5.1.

Acknowledgement

We are indebted to Professor Alan W. Craig from the University of Durham, United Kingdom, for very helpful discussions in an early stage of this work. We would also like to thank Professor Olof Widlund from New York University and the anonymous referees for their very helpful comments.

References

[1] I. Babuška, The p and $h-p$ versions of the finite element method: The state of the art, in: D.L. Dwoyer, M.Y. Hussaini and R.G. Voigt, eds., *Finite Elements, Theory and Applications* (Springer, New York, 1988).

[2] I. Babuška, A.W. Craig, J. Mandel and J. Pitkäranta, Efficient preconditioning for the p-version finite element method in two dimensions, *SIAM J. Numer. Anal.* 28 (1991) 624–662.

[3] I. Babuška and H.C. Elman, Some aspects of parallel implementation of the finite element method on message passing architecture, *J. Comput. Appl. Math.* 27 (1989) 157–187.

[4] I. Babuška, M. Griebel and J. Pitkäranta, The problem of selecting the shape functions for a p-type finite element, *Internat. J. Numer. Methods Engrg.* 28 (1989) 1891–1908.

[5] I. Babuška, B.A. Szabó and I.N. Katz, The p-version of the finite element method, *SIAM J. Numer. Anal.* 18 (1981) 515–545.

[6] J.H. Bramble, J.E. Pasciak and A.H. Schatz, The construction of preconditioners for elliptic problems by substructuring I, Math. Comp. 47 (1986) 103–134.

[7] T.F. Chan, Analysis of preconditioners for domain decomposition, *SIAM J. Numer. Anal.* 24 (1987) 382–390.

[8] T.F. Chan, T.Y. Hou and P.L. Lions, Geometry-independent convergence results for domain decomposition algorithms, Tech. Report CAM 89-11, UCLA, Los Angeles, CA (1989).

[9] T.F. Chan and D.C. Resasco, Analysis of domain decomposition preconditioners on irregular regions, Tech. Report CAM 87-05, UCLA, Los Angeles, CA (1987); also in: *Proceedings Sixth IMACS International Symposium on Computer Methods for Partial Differential Equations*, Lehigh University, Betlehem, PA (1987).

[10] P. Concus, G.H. Golub and D.P. O'Leary, A generalized conjugate gradient method for the numerical solution of elliptic differential equations, in: J.R. Bunch and D.J. Rose, eds., *Sparse Matrix Computations* (Academic Press, New York, 1976) 309–332.

[11] G.H. Golub and C.F. Van Loan, *Matrix Computations* (John Hopkins University Press, Baltimore, MD, 2nd ed., 1989).

[12] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards* 49 (1952) 409–436.

[13] P.L. Lions, On the Schwarz alternating method, I, in: R. Glowinski, G.H. Golub, G.A. Meurant and J. Périaux, eds., *Domain Decomposition Methods for Partial Differential Equations* (SIAM, Philadelphia, PA, 1988) 1–42.

[14] J. Mandel, Hierarchical preconditioning and partial orthogonalization for the p-version finite element method in: *T.F. Chan, R. Glowinski, J. Periaux and O.B. Widlund, eds., Third International Symposium on Domain Decomposition Methods for Partial Differential Equations* (SIAM, Philadelphia, PA, 1990).

[15] J. Mandel, Iterative solvers by substructuring for the p-version finite element method, in: *Proceedings International Conference on Spectral and High Order Methods for Partial Differential Equations*, Como, Italy (1989); also: *Comput. Methods Appl. Mech. Engrg.* 80 (1990) 117–128.

[16] J. Mandel, Two-level domain decomposition preconditioning for the p-version finite element method in three dimensions, in: *Proceedings Fourth Copper Mountain Conference on Multigrid Methods*, Copper Mountain, CO (1989); also: *Internat. J. Numer. Methods Engrg.* 29 (1990) 1095–1108.

[17] J. Mandel, A domain decomposition method for p-version finite elements in three dimensions, in: *Proceedings 7th International Conference on Finite Element Methods in Flow Problems*, Huntsville, AL (1989).

[18] J. Nečas, *Les Méthodes Directes en Théorie des Équations Elliptiques* (Academia, Prague, 1967).

[19] O.B. Widlund, Optimal iterative refinement methods, in: T.F. Chan, R. Glowinski, G.A. Meurant, J. Périaux and O. Widlund, eds., *Domain Decomposition Methods for Partial Differential Equations* II, (SIAM, Philadelphia, PA, 1989).

[20] O.C. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, London, 3rd ed., 1977).