UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

The Interface Probing Technique in Domain Decomposition

Tony F.C. Chan
Tarek Mathew

February 1991
(Revised May, 1991)
CAM Report 91-02

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

UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

The Interface Probing Technique in Domain Decomposition

Tony F.C. Chan
Tarek Mathew

February 1991
(Revised May, 1991)
CAM Report 91-02

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

THE INTERFACE PROBING TECHNIQUE IN DOMAIN DECOMPOSITION *

TONY F. C. CHAN AND TAREK P. MATHEW †

Abstract. The interface probing technique, which was developed and used by Chan and Resasco, and Keyes and Gropp, is an algebraic technique for constructing interface preconditioners in domain decomposition algorithms. The basic technique is to approximate interface matrices by matrices having a specified sparsity pattern. The construction involves only matrix vector products, and thus the interface matrix need not be known explicitly. A special feature is that the approximations adapts to the variations in the coefficients of the equations and the aspect ratios of the subdomains. This preconditioner can then be used in conjunction with many standard iterative methods, such as conjugate gradient methods.

In this paper, we summarize some old results and also present new ones, both algebraic and analytic, about the interface probing technique and its applications to interface operators. Comparisons are made with some optimal preconditioners such as the Golub-Mayer preconditioner.

Key Words. Interface probe, domain decomposition, elliptic equations, preconditioners.

AMS subject classifications: 65N20, 65F10.

1. Introduction. Many domain decomposition methods can be viewed as techniques for constructing preconditioners for solving linear systems arising from the discretization of elliptic partial differential equations. These preconditioners are based on the solution of smaller problems on subregions of the domain, together with a preconditioner for the reduced problem on the interface separating the subregions. Many algorithms have been developed having optimal or almost optimal rates of convergence with respect to mesh parameters, such as those of Bramble, Pasciak and Schatz [4], [6], [7], [8], Dryja and Widlund [20], [21], and Smith [32].

An important component in these methods is the preconditioner for the reduced problem on the interface of a two subdomain problem. For such problems, various interface preconditioners have been developed having rates of convergence independent of the mesh size h, see Bjørstad and Widlund [3], Bramble, Pasciak and Schatz [5], Dryja [19], and Golub and Mayer [24]. However, most of these interface preconditioners do not account for the coefficients of the problem or the geometry of the subdomains and hence, their performance can be sensitive to variations in these other parameters.

Versions of the interface probe were proposed by Chan and Resasco [15] and Eisenstat [22], and further extended and used by Keyes and Gropp [28], [29], as a technique for constructing preconditioners for interface problems. It differs from the other preconditioners mentioned, however, in that it is an algebraic technique, and as it turns out, one of its advantages is that the technique results in preconditioners which adapt to large variations in the coefficients and to most variations in the aspect ratios of the subdomains, though it doesn't adapt optimally to changes in mesh size. The basic idea behind this technique is to approximate the interface matrix by a matrix having a specified sparsity pattern chosen to capture the strongest coupling

* Original version: December 31, 1990, Revised: May 3, 1991

† Department of Mathematics, University of California at Los Angeles, Los Angeles, CA. 90024. This work was supported in part by the Department of Energy under contract DE-FG03-87ER25037, by the Army Research Office under contract DAAL03-88-K-0085, by the National Science Foundation under grant FDP NSF ASC 9003002 and by the Air Force Office for Scientific Research under grant AFOSR-90-0271 (subcontract to UCLA UKRF-4-24384-90-87).

of the interface operator, using a few matrix vector products of the interface matrix (which is usually not known explicitly) with carefully chosen probe vectors.

Since the probing method is an algebraic method, it can easily be applied to any operator having decay properties, and is not necessarily restricted to 2nd order problems. It has been used successfully to construct preconditioners to 4th order problems, to the Navier-Stokes equations, see Chan [10], Tsui [33], to convection-diffusion problems, see Chan and Keyes [13], and to problems where inexact solvers are used in the subdomain solves, see Chan and Goovaerts [11]. Preconditioners having other sparsity patterns (nonbanded) have been constructed in applications to domain decomposition algorithms involving many subdomains, see Chan and Mathew [14].

Our purpose in this paper is to survey various algebraic and analytic properties of the interface probing technique (including many new results). Most of our results deal with the specific case of tridiagonal approximations. In Section 2, we describe properties of the reduced interface matrix of an elliptic problem for a two subdomain decomposition. A brief survey of standard interface preconditioners is presented in Section 3, and a version of the interface probing technique is described in Section 4. In Section 5, we discuss various purely algebraic properties of the probe approximation together with conditions under which the approximations preserve symmetry and non-singularity. In addition, we discuss other versions of the probing technique. Section 6 concerns applications of the probing technique to the interface operator in domain decomposition. There we show for a model elliptic problem that the rate of convergence of a tridiagonal probe preconditioner is $O(h^{-1/2})$. The results thus indicate that the tridiagonal probe preconditioner performs as well asymptotically as the optimal *tridiagonal* preconditioner for the interface matrix, which has been conjectured by Greenbaum and Rodrigue [25] to have a rate of convergence bounded below by $O(h^{-1/2})$. We also present results concerning dependence of the probe preconditioned system on the aspect ratios of the subdomains, and on the scaling of the coefficients. Some numerical and theoretical comparisons are made with the Golub-Mayer preconditioner. Finally, in Section 7 we summarize the main properties of the probing technique.

2. A model elliptic problem and properties of the interface system. We consider the following 2nd order self adjoint elliptic operator L on a polygonal domain Ω in R^2, with Dirichlet boundary conditions:

$$Lu = -\frac{\partial}{\partial x}\left(a(x,y)\frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial y}\left(b(x,y)\frac{\partial u}{\partial y}\right) = f(x,y) \quad \text{in } \Omega, \quad \text{with } u = 0 \quad \text{on } \partial\Omega, \tag{1}$$

where $a(x,y)$ and $b(x,y)$ are assumed to be uniformly positive functions on the domain Ω.

We discretize equation (1) using the standard 5-point difference approximation on a uniform mesh of width h, with nodes (x_i, y_j) lying in Ω, where $x_{i+1} = x_i + h$ and $y_{j+1} = y_j + h$, see Varga [34]:

$$(L_h u^h)_{ij} = \begin{array}{l}(a_{i+\frac{1}{2},j} + a_{i-\frac{1}{2},j} + b_{i,j+\frac{1}{2}} + b_{i,j-\frac{1}{2}})u^h_{ij} - a_{i+\frac{1}{2},j}u^h_{i+1,j} \\ -a_{i-\frac{1}{2},j}u^h_{i-1,j} - b_{i,j+\frac{1}{2}}u^h_{i,j+1} - b_{i,j-\frac{1}{2}}u^h_{i,j-1}\end{array} = h^2 f_{ij}, \tag{2}$$

where $u^h_{ij} \equiv u^h(x_i, y_j) \approx u(x_i, y_j)$ is the discrete solution and $a_{i\pm\frac{1}{2},j} \equiv a(x_i \pm \frac{h}{2}, y_j)$, etc. This results in a symmetric positive definite linear system $L_h u^h = f$.

Let the domain Ω be partitioned into two non-overlapping subregions Ω_1 and Ω_2 with interface Γ denoting the intersection of their boundaries:

$$\Gamma \equiv \partial\Omega_1 \cap \partial\Omega_2.$$

If we group the unknowns in the interior of Ω_1 in u_1, and those in the interior of Ω_2 in u_2, and finally those on the interface Γ in u_3, then in this new ordering of unknowns, L_h has the following block structure:

$$\begin{pmatrix} L_{11} & 0 & L_{13} \\ 0 & L_{22} & L_{23} \\ L_{13}^T & L_{23}^T & L_{33} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}, \tag{3}$$

where, L_{11} and L_{22} denote discretizations corresponding to local problems on Ω_1 and Ω_2, respectively, etc. If we eliminate the interior unknowns, we obtain the following reduced system for u_3:

$$\begin{aligned} &Su_3 = f_3 - L_{13}^T L_{11}^{-1} f_1 - L_{23}^T L_{22}^{-1} f_2, \\ &\text{where} \quad S \equiv L_{33} - L_{13}^T L_{11}^{-1} L_{13} - L_{23}^T L_{22}^{-1} L_{23}. \end{aligned} \tag{4}$$

The matrix S is the reduced interface operator, and is also referred to as a Schur complement or a capacitance matrix. It is expensive to compute, requiring n solves on each subdomain, where n is the number of unknowns on the interface Γ. Therefore most domain decomposition methods are based on the solution of the reduced interface system (4) by means of a preconditioned conjugate gradient method. This requires only matrix vector products with S, which can be computed at a cost of one solve on each subdomain without computing S. Then, it is possible to solve problem (4) in less than n iterations, if the new system is well conditioned, and much research has gone into the construction of efficient preconditioners for S. Once u_3 is determined by solving system (4), the solution in the interior of the subdomains, u_1 and u_2 can be obtained at the cost of one solve in each subdomain, using the first two block rows of equation (3). Alternatively, once a preconditioner for S is available, it is easy to construct a preconditioner for the global matrix L_h, involving approximate solves on the subdomains, rather than exact solves on the subdomains, see for instance, [5].

Much is known about the properties of the interface matrix S, which is easily seen to be symmetric and positive definite. S is a discrete approximation to a Steklov-Poincaré operator coupling the subdomain problems through a *transmission* boundary condition, see [1], [31]. This Steklov-Poincaré operator can be shown to be spectrally equivalent to the pseudo-differential operator given by the square root of the Laplacian on the interface. Hence, it can be shown that S is spectrally equivalent to the square root of the discrete Laplacian on the interface Γ, and its condition number can be shown to grow at a rate $O(h^{-1})$, as $h \to 0$, see [3]. Some details of this will be given in Section 3 on preconditioners. S is a dense matrix, however, its entries can be shown to decay away from the diagonal at a rate $|S_{ij}| = O(|i-j|^{-2})$, see [24]. Other properties of S will be described following the definition of *discrete harmonic functions* below, and in Theorem 2.2.

Definition. A grid function w^h satisfying $(L^h w^h)_{ij} = 0$, for nodes ij lying in the interior of Ω_1 and Ω_2, will be referred to as a *piecewise discrete harmonic function.*

Piecewise discrete harmonic functions w^h for discretization (2), can easily be shown to satisfy the following discrete strong maximum principle:

LEMMA 2.1. *If w^h is a non-constant, piecewise discrete harmonic function, then its nodal values on the interior of the subdomains is strictly bounded above and below by the maximum and minimum, respectively, of the nodal values of w^h on the boundary of the subdomains. i.e., if $W_{min} \leq w_{ij}^h \leq W_{max}$, $\forall ij \in \partial\Omega_1 \cup \partial\Omega_2$, and if w^h is piecewise discrete harmonic, then, either $w_{ij}^h \equiv W_{max} = W_{min}$, or $W_{min} < w_{ij}^h < W_{max}$, $\forall ij \in$ interior of Ω.*

Proof. See [26]. □

Remark. The maximum principle is valid only on the *computational* domain, i.e, on the nodes that are connected to other nodes by the stencil. For instance, if the 5-point discretization is used on a rectangular domain, the 4 corner nodes are not coupled to the other nodes, and they are not considered to be part of the computational domain.

The Schur complement is related to *piecewise discrete harmonic functions* as follows. Given a grid function w defined on the interface Γ, if we let Ew denote the *piecewise discrete harmonic extension* of w into the two subdomains, i.e.,

$$Ew \equiv (-L_{11}^{-1}L_{13}w, -L_{22}^{-1}L_{23}w, w)^T,$$

then Sw is obtained from (3) by applying the stencil L_h to Ew on the nodes lying on Γ:

$$L_h Ew = (0, 0, (L_{33} - L_{13}^T L_{11}^{-1} L_{13} - L_{23}^T L_{22}^{-1} L_{23})w)^T = (0, 0, Sw). \tag{5}$$

The above properties can be used to prove strict diagonal dominance, and other properties of S, stated in Theorem 2.2. We recall that a matrix C is said to be diagonally dominant, if

$$|C_{ii}| \;\geq\; \textstyle\sum_{j \neq i} |C_{ij}|, \qquad \text{for } i = 1, \ldots, n. \tag{6}$$

C is said to be strictly diagonally dominant, if a strict inequality holds in (6) for all rows i. We also recall that a matrix C is said to be an M-matrix, if $C_{ij} \leq 0$, for $i \neq j$, and if $C_{ij}^{-1} \geq 0$, for all i, j.

THEOREM 2.2. *The Schur complement S, defined in equation (4) for discretization (2), is an M-matrix and satisfies:*

1. *$S_{km} < 0$ for $k \neq m$,*
2. *$S_{kk} > 0$ for all k,*
3. *$S_{kk} - \sum_{m \neq k} |S_{km}| > 0$ for $k = 1, \ldots, n$, (i.e., S is strictly diagonally dominant).*

Proof. All of the above statements are easily proved using the strong maximum principle and the relation between the Schur complement and discrete harmonic functions. Let (i_k, j_k) be the index of the k'th node on Γ and let δ_k denote the Kronecker Delta function on Γ, which is 1 on the k'th node on Γ, and zero on all other nodes of Γ. Then, applying the stencil of L_h to $E\delta_k$ at the m-th node on Γ, we obtain:

$$\begin{aligned} S_{mk} \;=\; (L_h E\delta_k)_{i_m,j_m} \;=\; & (a_{i_k+\frac{1}{2},j_k} + a_{i_k-\frac{1}{2},j_k} + b_{i_k,j_k+\frac{1}{2}} + b_{i_k,j_k-\frac{1}{2}})(E\delta_k)_{i_m,j_m} \\ & -a_{i_k+\frac{1}{2},j_k}(E\delta_k)_{i_m+1,j_m} - a_{i-\frac{1}{2},j}(E\delta_k)_{i_m-1,j_m} \\ & -b_{i,j+\frac{1}{2}}(E\delta_k)_{i_m,j_m+1} - b_{i,j-\frac{1}{2}}(E\delta_k)_{i_m,j_m-1}. \end{aligned} \tag{7}$$

Since $E\delta_k = \delta_k$ on Γ we have that $(E\delta_k)_{i_m,j_m} = 0$, for nodes $m \neq k$ on Γ. And so

equation (7) becomes: for $m \neq k$

$$\begin{aligned} S_{mk} &= (a_{i_k+\frac{1}{2},j_k} + a_{i_k-\frac{1}{2},j_k} + b_{i_k,j_k+\frac{1}{2}} + b_{i_k,j_k-\frac{1}{2}})0 \\ &\quad -a_{i_k+\frac{1}{2},j_k}(E\delta_k)_{i_m+1,j_m} - a_{i-\frac{1}{2},j}(E\delta_k)_{i_m-1,j_m} \\ &\quad -b_{i,j+\frac{1}{2}}(E\delta_k)_{i_m,j_m+1} - b_{i,j-\frac{1}{2}}(E\delta_k)_{i_m,j_m-1} \end{aligned}$$

which is negative since the coefficients $a_{i_k+\frac{1}{2},j_k}, a_{i_k-\frac{1}{2},j_k}, b_{i_k,j_k+\frac{1}{2}}, b_{i_k,j_k-\frac{1}{2}}$ are positive, and $0 \leq (E\delta_k)_{ij} \leq 1$ by the maximum principle, with strict inequality holding for the nodes (i,j) lying in the interior of the subdomains (because $E\delta_k$ is non-constant). Thus, we have proved (1).

To prove (2), we apply the stencil to $E\delta_k$ at the k'th node on Γ:

$$\begin{aligned} S_{kk} &= (a_{i_k+\frac{1}{2},j_k} + a_{i_k-\frac{1}{2},j_k} + b_{i_k,j_k+\frac{1}{2}} + b_{i_k,j_k-\frac{1}{2}})1 \\ &\quad -a_{i_k+\frac{1}{2},j_k}(E\delta_k)_{i_k+1,j_k} - a_{i-\frac{1}{2},j}(E\delta_k)_{i_k-1,j_k} \\ &\quad -b_{i,j+\frac{1}{2}}(E\delta_k)_{i_k,j_k+1} - b_{i,j-\frac{1}{2}}(E\delta_k)_{i_k,j_k-1} \\ &> 0 \end{aligned},$$

since by the strong version of the maximum principle $0 \leq (E\delta_k)_{ij} \leq 1$ at all nodes, with strict inequality at the nodes (i,j) lying in the interior of the subdomains.

We now show that S is strictly diagonally dominant. Let $\mathbf{1} \equiv (1,1,\ldots,1)^T$ on Γ. Then, since $S_{km} \leq 0$ for $k \neq m$, by (5):

$$S_{kk} - \sum_{m \neq k} |S_{km}| = S_{kk} + \sum_{m \neq k} S_{km} = (S\mathbf{1})_k = (L_h E\mathbf{1})_{i_k j_k}.$$

By the discrete maximum principle, $0 < (E\mathbf{1})_{ij} < 1$, for all nodes ij lying in the interior of the subdomains, so we obtain:

$$\begin{aligned} (L_h E\mathbf{1})_{i_k,j_k} &= (a_{i_k+\frac{1}{2},j_k} + a_{i_k-\frac{1}{2},j_k} + b_{i_k,j_k+\frac{1}{2}} + b_{i_k,j_k-\frac{1}{2}})1 \\ &\quad -a_{i_k+\frac{1}{2},j_k}(E\mathbf{1})_{i_k+1,j_k} - a_{i-\frac{1}{2},j}(E\mathbf{1})_{i_k-1,j_k} \\ &\quad -b_{i,j+\frac{1}{2}}(E\mathbf{1})_{i_k,j_k+1} - b_{i,j-\frac{1}{2}}(E\mathbf{1})_{i_k,j_k-1} \\ &> 0 \end{aligned}$$

This proves (3).

Since S is symmetric and positive definite, and since $S_{ij} \leq 0$ for $i \neq j$, it follows that S is an M-matrix, see Varga [34]. □

3. Some well known preconditioners for S. The rate of convergence and the efficiency of most domain decomposition algorithms depend on the choice of preconditioners M for S, both in the case of two subdomains and in the case of many subdomains. (Though our studies in this paper are restricted to the case of two subdomains, we mention here that in the case of many subdomains, a preconditioner for S can be built in terms of preconditioners for the reduced interface operator of two subdomain problems, see [4], [20].) In this section, we mention some of the preconditioners that have been proposed for S in the case of two subdomains. They include preconditioners by Axelsson and Polman [2], Bjørstad and Widlund [3], Bramble, Pasciak and Schatz [5], Chan [9], Chan and Hou [12], Chan and Keyes [13], Dryja [19], Golub and Mayer [24], Keyes and Gropp [28], [29], and Funaro, Quarteroni and Zanolli [17].

The rate of convergence of these preconditioned systems are determined by the quotient of the maximum and minimum eigenvalues of $M^{-1}S$. We use the term condition number and use $\kappa(M^{-1}S)$ to denote this ratio. Many of the above mentioned

preconditioners, M, are known to have *optimal* convergence rates with respect to mesh size variation, i.e.,

$$\kappa(M^{-1}S) = O(1) \quad \text{independent of } h.$$

Of these, several are based on the property that the Steklov-Poincaré operator coupling the subproblems is spectrally equivalent to the pseudo-differential operator given by the square root of the Laplacian on the interface, i.e., the operator obtained when in the eigenfunction expansion of the Laplace operator, the eigenvalues are replaced by its square roots. In the discrete case, this can be implemented efficiently using discrete sine transforms, as they form the eigenfunctions of the discrete one dimensional Laplacian:

$$-\Delta_h = tridiag(-1, 2, -1) = W\Lambda_1 W,$$

where $W = W^{-1} = W^T$ is the $n \times n$ sine transform matrix with entries

$$W_{ij} = \sqrt{\frac{2}{n+1}} sin(\frac{ij\pi}{n+1})$$

where $\Lambda_1 = diag(4\sin^2(\frac{i\pi}{2(n+1)}))$. Such preconditioners for S have the form,

$$M = W\Lambda W^{-1},$$

where W is the same as above, but where Λ is a diagonal matrix which approximates $\Lambda_1^{1/2}$. Such preconditioners can be inverted in $O(n\log(n))$ operations, if the fast sine transform is used. We list two such commonly used preconditioners, which differ in their choice of eigenvalues:

1. The Dryja preconditioner [19], is $M_D \equiv W\Lambda_1^{1/2}W^{-1}$. Equivalently, $M_D = (-\Delta_h)^{1/2}$.

2. The Golub-Mayer preconditioner [24], $M_{GM} \equiv W(\Lambda_1 + \frac{\Lambda_1^2}{4})^{1/2}W^{-1}$, for the same Λ_1 used in M_D. We note that there is a close connection between the Golub-Mayer preconditioner and "Dirichlet-Neumann" preconditioners, see [3], [9].

A similar idea using properties of the boundary trace operator to construct effective preconditioners has been used by Glowinski and Pironneau [23] to solve the biharmonic problem.

Both M_D and M_{GM}, have been shown to be *optimal* with respect to mesh refinement, see [3]. However, their performance could be sensitive to variations in the aspect ratios of the subdomains Ω_1 and Ω_2, and variations in the coefficients of L. Numerical results are presented in Tables 1, 2 and 3 of Section 6 which illustrates the dependence on the mesh size h, aspect ratios of the subdomains, and coefficients, respectively. Some theory for model problems is also presented in Section 6.

4. The Interface Probing Preconditioner. The probing technique was introduced in Chan and Resasco [15], Keyes and Gropp [28], and Eisenstat [22], as an algebraic technique for constructing sparse approximation to the interface operator S in the two subdomain case. One of its advantages is to account for the deterioration in the performance of some of the previously mentioned preconditioners, with respect to coefficients and aspect ratios. The main idea is to approximate S by a matrix having a specified sparsity pattern using matrix vector products of S with a few carefully chosen probe vectors. The sparsity pattern is chosen to capture the strongest coupling

of the interface operator S, and is usually banded. The motivation for using a sparse approximation to the interface operator S is the observation that it has weak global coupling, i.e., the entries of S decay rapidly away from the diagonal. For instance, for model problems and geometries, it has been shown that

$$|S_{ij}| = O(\frac{1}{|i-j|^2}),$$

for i, j away from the diagonal, see Golub and Mayers [24]. Thus, if M is a banded approximation to S, we would hope or expect that M be an effective preconditioner. However, since S is seldom known explicitly, it is not possible to choose the exact band of S, and so we construct a banded approximation M to S using matrix vector products of S (which is computable even if S is not explicitly known) with a few carefully chosen *probe vectors*. It turns out that the banded approximation to S obtained by probing, often leads to a better preconditioner than using the exact band of S, even if these were known, because the row sums of the probed approximation tend to approximate the row sums of S, unlike the row sums of the bands of S. Note that the name interface probing originates from the fact that the approximation to S is constructed using matrix vector products of S with a few test vectors defined on the interface Γ, to *probe* the "large" entries of S. Each such matrix vector product of S involves the inversion of L_h on each Ω_i, with a few *probing* boundary conditions on Γ, given by the *probe vectors*. Once the matrix-vector products are found, the construction of the preconditioner M from the vector outputs is done at very little expense, with the number of operations being linearly proportional to the number of non-zero elements in M.

Consider then the problem of constructing a banded approximation M_d of upper and lower bandwidth d, to an arbitrary square matrix $C \in R^{n \times n}$ (C may be nonsymmetric in general, and may not be known explicitly). We introduce the notation

$$M_d = \text{PROBE}(C, d),$$

to denote that M_d is constructed from C using the PROBE procedure. We make the following two observations. First, as noted by Curtiss, Powell and Reid [16], if C were a banded matrix having upper and lower bandwidth d, and C were not explicitly known, then the entries of C can be reconstructed from its action on $2d+1$ carefully chosen test vectors (instead of the standard choice of the n unit vectors e_i, forming the columns of the identity matrix). Second, if the matrix C is close to banded, i.e., its entries are small away from a band, we can compute its action on the same test vectors mentioned above and use these just as in the case where C is exactly banded, to construct a banded *approximation* to C (which we hope is good).

We illustrate the PROBE procedure for the case $d = 1$, in which case M_1 is a tridiagonal matrix, and the following three probe vectors are commonly used: $v_1 = (1,0,0,1,0,0,\ldots)^T$, $v_2 = (0,1,0,0,1,0,\ldots)^T$ and $v_3 = (0,0,1,0,0,1,\ldots)^T$. Since M_1 is tridiagonal, it can easily be checked that all its nonzero entries appear in the vectors

$M_1 v_i$, $i = 1, 2, 3$, as illustrated below:

$$(8)\quad \begin{pmatrix} m_{11} & m_{12} & & & & \\ m_{21} & m_{22} & m_{23} & & & \\ & m_{32} & m_{33} & m_{34} & & \\ & & m_{43} & m_{44} & m_{45} & \\ & & & m_{54} & m_{55} & \ddots \\ & & & & \ddots & \ddots \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & 0 \\ m_{21} & m_{22} & m_{23} \\ m_{34} & m_{32} & m_{33} \\ m_{44} & m_{45} & m_{43} \\ m_{54} & m_{55} & m_{56} \\ \vdots & \vdots & \vdots \end{pmatrix}$$

The probe algorithm reconstructs the nonzero entries of M_1 by equating the right hand side of (8) to the corresponding entries in the vectors $[Cv_1, Cv_2, Cv_3]$.

By construction of M_1, we would obtain $M_1 = C$, if C is indeed tridiagonal. Note that the product Cv_1 is the sum of columns 1, 4, 7, etc. of C, while Cv_2 is the sum of columns 2, 5, 8, etc., of C, and Cv_3 is the sum of columns 3, 6, 9, etc., of C. Therefore, if C were not tridiagonal but close to being tridiagonal, then the sum of every third column of C would contain contributions from other nonzero entries in other columns. If these off band entries of C are small, then their effects may be ignored if we only wish to construct an *approximation* to the tridiagonal band of C.

Also note that the procedure for constructing M_1 =PROBE$(C, 1)$ can be viewed as an attempt to determine a tridiagonal matrix M_1 which has the same action on the set of vectors v_i as C, i.e.,

$$(9)\qquad M_1 v_i = C v_i, \quad i = 1, 2, 3.$$

Though equation (9) is satisfied if C is tridiagonal, it may not be satisfied for arbitrary C, since a simple count gives $3n$ equations (one equation from each component of the three test vectors) for the $3n - 2$ unknowns on the tridiagonal band of $M_1 \in R^{n \times n}$. However, as illustrated in equation (8), a unique tridiagonal matrix M_1 can always be determined by choosing $3n - 2$ appropriate equations in (9), which pick out the non-zero entries of M_1. The remaining two equations in (9) may not be satisfied for arbitrary C. In addition, the tridiagonal matrix M_1 may not be symmetric, even if C is symmetric. We present an example of a 4×4 symmetric matrix C for which equations (9) is not satisfied, and for which M_1 is non-symmetric.

$$\begin{pmatrix} m_{11} & m_{12} & 0 & 0 \\ m_{21} & m_{22} & m_{23} & 0 \\ 0 & m_{32} & m_{33} & m_{34} \\ 0 & 0 & m_{43} & m_{44} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 & 5 \\ 0 & 10 & 0 & 2 \\ 0 & 0 & 10 & 0 \\ 5 & 2 & 0 & 10 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$(10)\qquad \Rightarrow \begin{pmatrix} m_{11} & m_{12} & 0 \\ m_{21} & m_{22} & m_{23} \\ m_{34} & m_{32} & m_{33} \\ m_{44} & 0 & m_{43} \end{pmatrix} = \begin{pmatrix} 15 & 0 & 0 \\ 2 & 10 & 0 \\ 0 & 0 & 10 \\ 15 & 2 & 0 \end{pmatrix}$$

$$\text{Therefore } C = \begin{pmatrix} 10 & 0 & 0 & 5 \\ 0 & 10 & 0 & 2 \\ 0 & 0 & 10 & 0 \\ 5 & 2 & 0 & 10 \end{pmatrix} \Longrightarrow \text{PROBE}(C, 1) = \begin{pmatrix} 15 & 0 & 0 & 0 \\ 2 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 15 \end{pmatrix}$$

From equations (10) we obtain that $m_{12} = 0$ and that $m_{21} = 2$, and thus M_1 is not symmetric. Moreover, equating the entries in row 4 column 2, we obtain the inconsistent equation that $0 = 2$, etc. However, if the matrix C is close to being tridiagonal, then M_1 and C will have almost the same action on v_i.

The procedure in equation (8) for constructing the tridiagonal PROBE$(C, 1)$ can easily be generalized to the case of banded matrices of upper and lower band width d. We denote the general banded approximation by PROBE(C, d) (which can be nonsymmetric in general, see Section 5.4). As mentioned earlier, this requires $2d + 1$ probe vectors. Rather than describe this procedure, which is a straightforward generalization, we provide a Matlab code for constructing PROBE(A, d). Given a matrix A, the following procedure returns the probed preconditioner M =PROBE(A, d) of upper and lower band width d.

MATLAB CODE FOR PROBE(A, d)

```
function M=probe(A,d)
n=length(A); k=min([2*d+1,n]);
if k == 1
  M=diag(A*ones(n,1));
else
  v= rem([1:n]'*ones(1,k),k) == ones(n,1)*[1:(k-1), 0];
  av=A*v;
  M=zeros(A);
  for c=1:k,
     for i=c:k:n,
        M(max([i-d 1]):min([i+d n]),i) = av(max([i-d 1]):min([i+d n]),c);
     end
  end
end
```

It is also possible to extend the probing technique to construct approximations having a specified structure, which is not necessarily sparse. For instance, probe approximations which are Toeplitz or Circulant matrices have been constructed see Keyes [27], Li [30]. Probing has also been used to compute the eigenvalues of a matrix M_S approximating the interface operator S, see Chan and Keyes [13]. For instance, in [13], M_S is assumed to have the following form: $M_S = WDW^{-1}$, where W is the discrete sine transform matrix, and D is a diagonal matrix consisting of the eigenvalues of M_S. There are several ways to choose D. In [13], they let

$$D = \text{PROBE}(W^{-1}SW, 0). \tag{11}$$

If D is chosen suitably, we obtain preconditioners M_S which are spectrally equivalent to S, [13], [14]. This is called the *spectral probe* method.

5. Algebraic properties of banded probes. Although there have been many experimental studies and successful applications of the probing technique, there has not been much focus on the algebraic and analytical properties of these methods. In this Section, we summarize some new as well as old results on the algebraic properties of the tridiagonal ($d = 1$) probe preconditioners.

5.1. Linearity. We note that by construction, the probe preconditioner is linearly dependent on the matrix C. i.e.,

$$\text{PROBE}(\alpha C_1 + C_2, d) = \alpha\text{PROBE}(C_1, d) + \text{PROBE}(C_2, d).$$

This property implies that probing S in equation (4) or the two terms $L_{13}^T L_{11}^{-1} L_{13}$ and $L_{23}^T L_{22}^{-1} L_{23}$ separately produce the same results (since L_{33} can be obtained from L).

5.2. Nonsingularity. The PROBE approximation can sometimes be singular even if the original matrix is nonsingular. However, under certain conditions, nonsingularity of the preconditioner can be proved. The following result concerns the preservation of diagonal dominance.

THEOREM 5.1. *If the i^{th} row of C is (strictly) diagonally dominant, then the i^{th} row of the PROBE approximation $M = PROBE(C,d)$ is also (strictly) diagonally dominant. From this it follows that $PROBE(C,d)$ will be nonsingular if either C is strictly diagonally dominant or if C is irreducibly diagonally dominant and $PROBE(C,d)$ is irreducible.*

Proof. We consider only the case $d = 1$. (The proof for general d is similar.) Recall that the entries M_{ij}, for $|i-j| \leq 1$, are defined by:

$$M_{ij} = \sum_{k:(k-j) mod\ 3=0} C_{ik}.$$

Using the definition of M_{ij} we obtain that

$$|M_{ii}| - |M_{ii-1}| - |M_{ii+1}|$$

$$= |C_{ii} - \sum_{\substack{k \neq i: \\ (k-i) mod\ 3 = 0}} C_{ik}| - |\sum_{\substack{k: \\ (k-i) mod\ 3 = 1}} C_{ik}| - |\sum_{\substack{k: \\ (k-i) mod\ 3 = 2}} C_{ik}|$$

Applying the triangle inequality to all three terms, we obtain that

$$|M_{ii}| - |M_{ii-1}| - |M_{ii+1}| \geq |C_{ii}| - \sum_{k \neq i} |C_{ik}|$$

which is nonnegative due to the diagonal dominance of C. Since M is tridiagonal, this is the same as

$$|M_{ii}| \geq \sum_{j \neq i} |M_{ij}|, \tag{12}$$

which proves that M is diagonally dominant. Note that the inequalities can be replaced by strict inequalities, if S is strictly diagonally dominant. □

Unfortunately, PROBE(C,d) can result in singular approximations if the given matrix is not *strictly* diagonally dominant, as the following example illustrates.

$$C = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Longrightarrow \text{probe}(C,1) = \begin{pmatrix} 0 & 0 & & \\ 0 & 1 & 0 & \\ & 0 & 1 & 0 \\ & & 0 & 1 \end{pmatrix}.$$

The next example illustrates that even if C is symmetric positive definite, PROBE$(C,0)$ need not be.

$$C = \begin{pmatrix} 1 & -2 \\ -2 & 10 \end{pmatrix} \Longrightarrow \text{PROBE}(C,0) = \begin{pmatrix} -1 & 0 \\ 0 & 8 \end{pmatrix}.$$

However, in our applications to the interface matrices S in elliptic problems, we are able to prove the following result.

THEOREM 5.2. *If S is the interface operator (Schur complement) corresponding to the discrete elliptic operator L_h defined in equation (2), then*

1. *PROBE$(S, 0)$ is a diagonal matrix with positive diagonal entries.*
2. *PROBE(S, d) is nonsingular and strictly diagonally dominant. Hence, it will be symmetric positive definite if PROBE(S, d) is symmetric.*

Proof. The proof follows trivially from Theorem 2.2 and Theorem 5.1. □

5.3. Symmetry. Recall from example (10) that PROBE$(C, 1)$ can be non-symmetric even if C is symmetric. In some preconditioned conjugate gradient methods, it is desirable to have preconditioners which preserve the symmetry of the coefficient matrix. One possible remedy which preserves the bandwidth of PROBE, is to take the symmetric part of the resulting PROBE(.,.) matrix, i.e., define:

$$\text{symmetrised-PROBE}(C, 1) \equiv \left(\text{PROBE}(C, 1) + \text{PROBE}(C, 1)^T\right)/2.$$

Unfortunately, symmetrised-PROBE(.,.) does not preserve diagonal dominance of even strictly diagonally dominant matrices, as the following example illustrates:

$$(13)\ C = \begin{pmatrix} 100 & 0 & 0 & 0 & 50 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 50 & 0 & 0 & 0 & 100 \end{pmatrix} \Rightarrow \text{PROBE}(C, 1) = \begin{pmatrix} 100 & 50 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 50 & 100 \end{pmatrix}$$

$$\Rightarrow \text{symmetrised-PROBE}(C, 1) = \begin{pmatrix} 100 & 25 & 0 & 0 & 0 \\ 25 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 25 \\ 0 & 0 & 0 & 25 & 100 \end{pmatrix}.$$

However, such problems occur very rarely in applications to interface matrices, where there is a decay in the entries of the matrix we probe. As an alternative to the symmetrised-PROBE, the following minmodsym-PROBE can also be used to obtain a symmetric banded approximation, preserving diagonal dominance. The off diagonal entries (i, j) and (j, i) of minmodsym-PROBE(C, d) is chosen to be the (i, j) or (j, i) entry of PROBE(C, d) having smaller modulus. i.e., if we let M =PROBE(C, d), then:

$$\text{minmodsym-PROBE}(C, d)_{ij} \equiv \begin{cases} M_{ij} & \text{if } |M_{ij}| = \min\{|M_{ij}|, |M_{ji}|\} \\ M_{ji} & \text{if } |M_{ji}| = \min\{|M_{ij}|, |M_{ji}|\} \end{cases}$$

However, the minmodsym procedure is no longer linear. As the next theorem indicates, this procedure preserves symmetry and strict diagonal dominance.

THEOREM 5.3. *If C is symmetric and strictly diagonally dominant, and $C_{ii} > 0$ then*

$$M = \textit{minmodsym-PROBE}(C, d)$$

is symmetric positive definite and strictly diagonally dominant.

Proof. Symmetry follows by construction. Diagonal dominance is preserved since PROBE(C,d) preserves diagonal dominance, and since the off diagonal entries of minmodsym-PROBE(C,d) are chosen to decrease the modulus of the off diagonal terms of PROBE(C,d). □

There is an alternative procedure to compute symmetric approximations, due to Keyes and Gropp [28], [29], in which a banded, symmetric approximation having upper and lower bandwidth d is constructed based on using $d+1$ probe vectors. We will refer to this as the symmetric-PROBE and denote it by symmetric-PROBE(.,.). The procedure is linear, and it is computationally less expensive than PROBE(.,d). We illustrate the procedure by an example for the case $d=1$, i.e., to construct a symmetric tridiagonal approximation. In this case the probe vectors are $v_1=(1,0,1,0,\ldots)^T$ and $v_2=(0,1,0,1,\ldots)^T$, and we have:

$$(14)\quad \begin{pmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & b_3 & a_3 & b_4 & \\ & & b_4 & a_4 & \ddots \\ & & & \ddots & \ddots \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \end{pmatrix} = \begin{pmatrix} a_1 & b_2 \\ b_2+b_3 & a_2 \\ a_3 & b_3+b_4 \\ b_4+b_5 & a_4 \\ \vdots & \vdots \end{pmatrix} = [Cv_1, Cv_2].$$

The symmetric tridiagonal approximation symmetric-PROBE$(C,1) = \mathit{tridiag}(b_i, a_i, b_{i+1})$ is obtained from the probed output vectors Cv_1, Cv_2, as indicated in the following algorithm:

SYMMETRIC-PROBE ALGORITHM.

For $i=1,\ldots,n$

$$a_i = \begin{cases} (Cv_1)_i & \text{if } i \text{ is odd} \\ (Cv_2)_i & \text{if } i \text{ is even} \end{cases}$$

$b_2 = (Cv_2)_1$

For $i=3,\ldots,n$

$$b_i = \begin{cases} (Cv_1)_{i-1} - b_{i-1} & \text{if } i \text{ is odd} \\ (Cv_2)_{i-1} - b_{i-1} & \text{if } i \text{ is even} \end{cases}$$

System (14) consists of $2n$ equations for $2n-1$ unknowns, and is therefore an overdetermined system. However, if the matrix C we probe is symmetric, it can be shown that the resulting system is consistent, see [28], [29], and from this it then follows that $Cv_i =$ symmetric-PROBE$(C,1)v_i$ for $i=1,2$. The general banded symmetric-PROBE follows easily by using $d+1$ probe vectors, with 1's every $(d+1)$th column. Note that symmetric-PROBE(.,d) requires d less probe vectors than PROBE(.,d) and hence less subdomain solves.

Unfortunately, the symmetric-PROBE(.,.) does not preserve diagonal dominance or positive definiteness in general, as the following example illustrates:

$$\text{If } C = \begin{pmatrix} 3 & -1 & 0 & -2 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -2 & 0 & -1 & 4 \end{pmatrix} \text{ then PROBE}(C,1) = \begin{pmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & 1 & -2 & 1 \\ & & -1 & 2 \end{pmatrix},$$

which preserves diagonal dominance, but

$$\text{symmetric-PROBE}(C,1) = \begin{pmatrix} 3 & -3 & & \\ -3 & 2 & 1 & \\ & 1 & 2 & -2 \\ & & -2 & 4 \end{pmatrix},$$

which is not diagonally dominant. However, such difficulties are encountered only rarely in applications for matrices having decay properties.

An idea similar to probing has been used by Axelsson and Polman [2] to construct a tridiagonal symmetric approximation M_{AP} to a symmetric matrix C, based on the test vectors $v_1 = (1, \ldots, 1)^T$ and $v_2 = (1, 2, 3, \ldots, n)^T$. They have shown that their resulting approximation M_{AP} satisfies $M_{AP}v_i = Cv_i$, for $i = 1, 2,$. The Axelsson-Polman PROBE preserves diagonal dominance. In addition, under certain assumptions, they have been able to obtain lower bounds for the spectrum of M_{AP} in terms of C.

6. Probe preconditioners in domain decomposition. The preceeding section dealt with purely algebraic properties of the probing technique, valid for arbitrary matrices. In this section we study convergence properties of the tridiagonal probing technique applied to the interface matrix S in equation (4) for a model elliptic problem known to have a condition number growing at a rate h^{-1}, as $h \to 0$. Our studies focus on how the mesh size, aspect ratios of the subdomains, and the variations in the coefficients affect the rate of convergence of the probe preconditioned system for the interface matrix S.

In particular, we show that an application of a version of the tridiagonal probe preconditioner results in a condition number which grows at a rate $h^{-1/2}$ as $h \to 0$. We also show that this condition number is generally insensitive to variations in the aspect ratios of the subdomains. Finally, we consider how this condition number depends on the coefficients of the elliptic problem. There we present theoretical bounds for the condition number of the preconditioned system when the coefficients are scaled by positive scalar constants on each subdomain. In all these cases, we present both theoretical and numerical comparisons of the convergence rates of the Golub-Mayer preconditioner with the probed preconditioner.

6.1. Eigen-decomposition of the Schur complement S and the probe preconditioner for a model elliptic problem. The model problem we consider is the Laplacian on the rectangle $\Omega = [0, l_1 + l_2] \times [0, 1]$, as illustrated in Figure 1, with Dirichlet boundary conditions on the vertical boundaries and *periodic* boundary conditions on the horizontal boundaries:

$$(15) \qquad \begin{cases} -\Delta u &= f & \text{in } \Omega \\ u(x,y) &= 0 & \text{for } y \in [0,1] \text{ and } x = 0 \\ u(x,y) &= 0 & \text{for } y \in [0,1] \text{ and } x = l_1 + l_2 \\ u(x,0) &= u(x,1) & \text{for } x \in (0, l_1 + l_2). \end{cases}$$

Ω is partitioned into two subdomains $\Omega_1 = [0, l_1] \times [0, 1]$ and $\Omega_2 = [l_1, l_1 + l_2] \times [0, 1]$, with the interface $\Gamma = \{l_1\} \times [0, 1]$. Problem (15) is discretized by the 5 point Laplacian on an $(m_1 + m_2 + 3) \times (n + 2)$ grid, which includes the boundary nodes, with mesh size $h = 1/(n + 1)$, $l_1 = (m_1 + 1)h$ and $l_2 = (m_2 + 1)h$. Note that there are $n + 1$ distinct unknowns on each vertical line, due to periodicity, and there are $m_1 + m_2 + 1$ unknowns in the interior of each horizontal line.

For this model problem, the eigen-decomposition of the Schur complement S, as well as the eigen-decomposition of a suitable tridiagonal PROBE of the Schur complement, can be computed exactly. The eigen-decomposition of S can be found using the discrete eigenfunctions of the 5-point Laplacian, see Chan [9] and Donato [18], and is given below:
$S = F diag(\lambda_0, \ldots, \lambda_n) F^{-1}$, where

$\lambda_0 = \left(\frac{1}{m_1+1} + \frac{1}{m_2+1}\right)$ and

$\lambda_j = \left(\frac{1+\gamma_j^{m_1+1}}{1-\gamma_j^{m_1+1}} + \frac{1+\gamma_j^{m_2+1}}{1-\gamma_j^{m_2+1}}\right)\sqrt{\sigma_j + \frac{\sigma_j^2}{4}}$ for $j = 1, \ldots, n$ with σ_j and γ_j defined by

$\sigma_j = 4\sin^2(j\pi h)$,

$\gamma_j \equiv \frac{1+\frac{\sigma_j}{2} - \sqrt{\sigma_j + \frac{\sigma_j^2}{4}}}{1+\frac{\sigma_j}{2} + \sqrt{\sigma_j + \frac{\sigma_j^2}{4}}}$,

$F = [f_0, \ldots, f_n]$ is a unitary matrix ($F^{-1} = F^H$), with $f_j = \sqrt{h}(1, e^{2\pi j h}, \ldots, e^{2\pi n j h})^T$.

Note that if l_1, $l_2 \to \infty$, then $\lambda_0 \to 0$, and the interface matrix becomes singular. Therefore, we restrict to the case where either l_1 or l_2 is $O(1)$ independent of h.

For this model problem, $S \in R^{(n+1)\times(n+1)}$ is symmetric and circulant (due to periodicity), and so rather than defining M =symmetrised-PROBE$(S,1)$ discussed in Section 4 (which would result in a tridiagonal but not circulant approximation), we construct a tridiagonal, circulant approximation $M_{CP} \in R^{(n+1)\times(n+1)}$ by using a variant of the tridiagonal PROBE which we denote M_{CP} =circulant-PROBE$(S,1)$. We describe it for the case $n+1$ being even, in which case we need just one probe vector $\equiv (1,0,1,0,\ldots,1,0)^T$:

$$M_{CP} = \begin{pmatrix} \alpha & -\beta & & -\beta \\ \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots \\ -\beta & & -\beta & \alpha \end{pmatrix} \quad \text{with} \quad M_{CP}\begin{pmatrix} 1 \\ 0 \\ 1 \\ \vdots \end{pmatrix} = \begin{pmatrix} \alpha \\ -2\beta \\ \alpha \\ \vdots \end{pmatrix} := S\begin{pmatrix} 1 \\ 0 \\ 1 \\ \vdots \end{pmatrix}. \tag{16}$$

The values α and β are easily found. The following lemma contains estimates for $\alpha - 2\beta$ and β.

LEMMA 6.1. *The row sum $\alpha - 2\beta$ of M_{CP} =circulant-PROBE$(S,1)$ satisfies:*

$$\alpha - 2\beta = \sum_{i=0}^{n} S_{ji} = \lambda_0 = \left(\frac{h}{l_1} + \frac{h}{l_2}\right) \qquad \text{for } j = 0, \ldots, n,$$

and $1 \le \beta \le 2$.

Proof. Our proof will be based on the fact that S and M_{CP} have the same row sums. We use two probe vectors $v_1 = (1,0,1,0,\ldots,1,0)^T$ and $v_2 = (0,1,0,1,\ldots,0,1)^T$. First, it can be easily verified that if $Sv_1 = M_{CP}v_1$, (which holds by construction of M_{CP}), then $Sv_2 = M_{CP}v_2$. Since $v_1 + v_2 = (1,1,\ldots,1)^T$, we obtain: $M_{CP}(1,\ldots,1)^T = S(1,\ldots,1)^T$, i.e., the row sums must be equal. Using the fact that the row sum of M_{CP} is $\alpha - 2\beta$, we obtain that:

$$\alpha - 2\beta = \sum_{i=0}^{n} S_{ji} = (S\mathbf{1})_j = \lambda_0 = \left(\frac{h}{l_1} + \frac{h}{l_2}\right) > 0, \qquad \text{for } j = 0, \ldots, n,$$

since $(1,\ldots,1)^T$ is an eigenvector of S corresponding to eigenvalue λ_0.

To prove that $1 \le \beta \le 2$, we use the expression for -2β given in equation (16), and use the alternative expression for the entries of S_{ij} in terms of *discrete harmonic extensions*, as described in equation (5) of Section 2, to obtain that:

$$-2\beta = (L_h E v_1)_{ij},$$

where (i,j) is any node on Γ with 0 nodal value for the probe vector v_1, and E denotes the discrete harmonic extension and L_h denotes the discretization of the elliptic operator. Applying the 5-point Laplacian at node (i,j) on Γ results in:

$$-2\beta = -2 - (Ev_1)_{i,j+1} - (Ev_1)_{i,j-1}.$$

By the maximum principle the entries of Ev_1 lie between 0 and 1 at all other nodes, thus it follows that $-2 \geq -2\beta \geq -4$ and the result follows. □

The eigen-decomposition of M_{CP} can be explicitly found for the model problem.

LEMMA 6.2. *The circulant-PROBE matrix M_{CP} is diagonalized by the Discrete Fourier Transform F and has eigen-decomposition:*

$$M_{CP} = F \ \ diag\,(\lambda_0 + \beta\sigma_j) \ \ F^{-1}. \tag{17}$$

Proof. Since M_{CP} is circulant, it is diagonalized by the Discrete Fourier Transform F. Its eigenvalues can be determined by applying M_{CP} to each column of F, for $j = 0, \ldots, n$:

$$(M_{CP} f_j)_i = -\beta e^{(i-1)j2\pi h} + \alpha e^{ij2\pi h} - \beta e^{(i+1)j2\pi h}$$

$$= \left(\alpha - 2\beta + \beta 4\sin^2(j\pi h)\right) e^{ij2\pi h}$$

$$= (\lambda_0 + \beta\sigma_j)\, e^{ij2\pi h} = (\lambda_0 + \beta\sigma_j)\,(f_j)_i \text{ for } j = 0, \ldots, n,$$

where $\sigma_j = 4\sin^2(j\pi h)$. Thus, $M_{CP} = F \ \ diag(\lambda_0 + \beta\sigma_j) \ \ F^{-1}$. □

Since M_{CP} and S are diagonalized by F, we obtain that:

$$M_{CP}^{-1} S = F \ \ diag\left(\frac{\lambda_j}{\lambda_0 + \beta\sigma_j}\right) \ \ F^{-1}.$$

For convenience, we define $\Phi_j \equiv \lambda_j/(\lambda_0 + \beta\sigma_j)$ for $j = 0, \ldots, n$. Then the condition number of $M_{CP}^{-1}S$ is determined by the quotient of the maximum and minimum of Φ_j for $j = 0, \ldots, n$. This quotient is 1 when $j = 0$, since $\sigma_0 = 0$. To determine bounds for the extrema when $j = 1, \ldots, n$, we replace the discrete optimization problem by the optimization problem for its natural continuous extension, $\Phi(\sigma)$, where

$$\Phi(\sigma) = \left(\frac{1 + \gamma(\sigma)^{l_1/h}}{1 - \gamma(\sigma)^{l_1/h}} + \frac{1 + \gamma(\sigma)^{l_2/h}}{1 - \gamma(\sigma)^{l_2/h}}\right) \frac{\sqrt{\sigma + \frac{\sigma^2}{4}}}{\left(\frac{h}{l_1} + \frac{h}{l_2}\right) + \beta\sigma},$$

with $\gamma(\sigma)$ defined by

$$\gamma(\sigma) \equiv \frac{1 + \frac{\sigma}{2} - \sqrt{\sigma + \frac{\sigma^2}{4}}}{1 + \frac{\sigma}{2} + \sqrt{\sigma + \frac{\sigma^2}{4}}}, \tag{18}$$

and where the discrete values of σ_j for j varying from 1 to n was replaced by the continuous variable $\sigma \in [4\sin^2(\pi h), 4]$, and the eigenvalues $\Phi_j = \lambda_j/(\lambda_0 + \beta\sigma_j)$ was replaced by its continuous counterparts $\Phi(\sigma) = \lambda(\sigma)/(\lambda_0 + \beta\sigma)$.

An upper bound for the condition number of the preconditioned system can thus be expressed in terms of $\Phi(\sigma)$ as follows:

$$\kappa(M_{CP}^{-1}S) \leq \frac{\max\{\max_\sigma \Phi(\sigma), 1\}}{\min\{\min_\sigma \Phi(\sigma), 1\}}, \qquad \text{for } \sigma \in [4\sin^2(\pi h), 4], \tag{19}$$

and so the bounds for $\Phi(\sigma)$ determine the rate of convergence of the preconditioned system.

6.2. Dependence on mesh size h. We now consider bounds for the eigenvalues $\Phi(\sigma)$, for $\sigma \in [4\sin^2(\pi h), 4]$. Because there should be at least one line of unknowns in the interior of each subdomain, $2h \leq l_1$ and $2h \leq l_2$. The eigenvalues $\Phi(\sigma)$ can be rewritten as: $\Phi(\sigma) = \mu(\sigma)H(\sigma)$, where

$$\mu(\sigma) \equiv \left(\frac{1+\gamma(\sigma)^{l_1/h}}{1-\gamma(\sigma)^{l_1/h}} + \frac{1+\gamma(\sigma)^{l_2/h}}{1-\gamma(\sigma)^{l_2/h}}\right), \quad \text{and} \quad H(\sigma) \equiv \frac{\sqrt{\sigma + \frac{\sigma^2}{4}}}{\left(\frac{h}{l_1}+\frac{h}{l_2}\right) + \beta\sigma}. \tag{20}$$

In the following Lemma, we list some properties of $\Phi(\sigma)$, $\mu(\sigma)$, $H(\sigma)$ and $\gamma(\sigma)^{\frac{l_i}{h}}$.

LEMMA 6.3. *The following hold:*

1. *There exists positive constants c_1 and c_2 independent of h, l_1 and l_2 such that $\gamma(\sigma)^{\frac{l_i}{h}}$ satisfies:*

$$0 \leq e^{-\frac{l_i}{h}c_1\sqrt{\sigma}} \leq \gamma(\sigma)^{\frac{l_i}{h}} \leq e^{-\frac{l_i}{h}c_2\sqrt{\sigma}} \leq 1, \qquad \text{for } \sigma \in [4\sin^2(\pi h), 4].$$

2. *For the constants c_1 and c_2 given in (1) above, the function $\mu(\sigma)$ satisfies:*

$$2 + \sum_{i=1}^{2} \frac{2e^{-\frac{l_i}{h}c_1\sqrt{\sigma}}}{1-e^{-\frac{l_i}{h}c_1\sqrt{\sigma}}} \leq \mu(\sigma) \leq 2 + \sum_{i=1}^{2} \frac{2e^{-\frac{l_i}{h}c_2\sqrt{\sigma}}}{1-e^{-\frac{l_i}{h}c_2\sqrt{\sigma}}}.$$

3. *$H(\sigma)$ has a unique critical point at $\sigma = \sigma^\star \equiv \frac{\frac{h}{l_1}+\frac{h}{l_2}}{\beta - \left(\frac{h}{2l_1}+\frac{h}{2l_2}\right)}$, where the maximum of the function is attained. $H(\sigma)$ is a monotonically decreasing function for $\sigma > \sigma^\star$ and is monotonically increasing for $\sigma < \sigma^\star$.*
4. *$\mu'(\sigma) \leq 0$ for $\sigma \in [4\sin^2(\pi h), 4]$, and is thus a decreasing function.*
5. *$\Phi'(\sigma) \leq 0$ for $\sigma \geq \sigma^\star$.*

Proof. To prove (1), we first write

$$\gamma(\sigma)^{\frac{l_i}{h}} = e^{\frac{l_i}{h}\log(\gamma(\sigma))},$$

and expand $\log(\gamma(\sigma))$ in a Taylor series in $\sqrt{\sigma}$, with remainder term and evaluate the remainder term to obtain uniform error bounds. We outline the steps. Substituting the expression for $\gamma(\sigma)$ into $\log(\gamma(\sigma))$, we obtain:

$$\log(\gamma(\sigma)) = \log(1 - z(\sigma)), \qquad \text{where } z(\sigma) \equiv \frac{2\sqrt{\sigma + \frac{\sigma^2}{4}}}{1 + \frac{\sigma}{2} + 2\sqrt{\sigma + \frac{\sigma^2}{4}}},$$

where, for $\sigma \in [0, 4]$, the function $z(\sigma) \in [0, \frac{4\sqrt{2}}{3+4\sqrt{2}}]$. Expanding $\log(1-z)$ in a Taylor series with remainder, about $z = 0$, it can be easily shown that:

$$-z\left(\frac{59+24\sqrt{2}}{18}\right) \leq \log(1-z) \leq -z, \qquad \text{for } z \in [0, \frac{4\sqrt{2}}{3+4\sqrt{2}}].$$

Expanding $z(\sigma)$ in a series in $\sqrt{\sigma}$, we obtain:

$$\frac{2\sqrt{\sigma}}{3+4\sqrt{2}} \leq z(\sigma) \leq 2\sqrt{2}\sqrt{\sigma}, \quad \text{for } \sigma \in [0,4].$$

Combining the preceeding results, we obtain that $\log(\gamma(\sigma))$ is uniformly equivalent to $\sqrt{\sigma}$ for $\sigma \in I_1$. Substituting this as argument for the exponential function, which is monotone increasing, we obtain the uniform upper and lower bounds given in (1), with specific values for the constants c_1 and c_2. We omit the details.

The proof of (2) follows easily from (1) by using the definition of $\mu(\sigma)$.

(3). The derivitive of $H(\sigma)$ is easily verified to be:

$$H'(\sigma) = \frac{\left(\beta - \frac{h}{2l_1} - \frac{h}{2l_2}\right)\left(-\sigma + \frac{\frac{h}{l_1}+\frac{h}{l_2}}{\beta - \frac{h}{2l_1} - \frac{h}{2l_2}}\right)}{2\sqrt{\sigma + \frac{\sigma^2}{4}}\left(\frac{h}{l_1} + \frac{h}{l_2} + \beta\sigma\right)^2}.$$

From this, we see that the only critical point of $H(\sigma)$ occurs at $\sigma^\star = \frac{\frac{h}{l_1}+\frac{h}{l_2}}{\beta - \left(\frac{h}{2l_1}+\frac{h}{2l_2}\right)}$. That this critical point corresponds to a maximum is easily shown by observing that $H'(\sigma)$ is positive to the left of the critical point and that it is negative to the right of the critical point.

(4) is easily proved using the expression for the derivitive of of $\mu(\sigma)$:

$$\mu'(\sigma) = \sum_{i=1}^{2} \frac{2\left(\frac{l_i}{h}\right)\gamma'(\sigma)\gamma(\sigma)^{\frac{l_i}{h}-1}}{\left(1-\gamma(\sigma)^{\frac{l_i}{h}}\right)^2},$$

which is non-positive since

$$\gamma'(\sigma) = \frac{-1}{\sqrt{\sigma + \frac{\sigma^2}{4}}\left(1 + \frac{\sigma}{2} + \sqrt{\sigma + \frac{\sigma^2}{4}}\right)^2} \leq 0.$$

(5). To prove (5), we consider the expression for $\Phi'(\sigma) = \mu'(\sigma)H(\sigma) + \mu(\sigma)H'(\sigma)$. We note that $\mu(\sigma)$ and $H(\sigma)$ are non-negative, and that $\mu'(\sigma) \leq 0$ from part (4) of the proof. Using the expression for $H'(\sigma)$ given in part (3), we see that $H'(\sigma) \leq 0$ for $\sigma \geq \sigma^\star$. Combining these results, we obtain that $\Phi'(\sigma) \leq 0$ for $\sigma \geq \sigma^\star$. □

The following theorem contains the main result of this section.

THEOREM 6.4. *If* $\min\{l_1, l_2\}$ *is* $O(1)$ *independent of* h, *then for small enough* h, *the eigenvalues* $\Phi(\sigma)$ *of* $M_{CP}^{-1}S$ *satisfy:*

$$C_1 \leq \Phi(\sigma) \leq C_2\sqrt{\frac{l_1 l_2}{(l_1+l_2)}}h^{-1/2}.$$

Proof. Essentially, the bounds for $\Phi(\sigma)$ will be shown to be determined by bounds for the maximum and minimum of $H(\sigma)$, though there are some difficulties due to the presence of the $\mu(\sigma)$ term which can become large for small values of l_1 and l_2. $H(\sigma)$ is the quotient of a square root function representing the eigenvalues of S and a linear function representing the eigenvalues of the probe approximation, and its maximum and minimum can be computed explicitly. The details are now described.

We consider two cases separately. In case (1), we assume that the aspect ratios l_1 and l_2 are both strictly greater than 1. In this case, the function $\mu(\sigma)$ will be shown to be uniformly bounded and the bounds for $\Phi(\sigma)$ are obtained by finding bounds for the maximum and minimum of $H(\sigma)$. This can be done using results in Lemma 6.3. In case (2), we assume that at least one of the aspect ratios l_1 or l_2 is smaller than or equal to 1. In this case, the function $\mu(\sigma)$ is not uniformly bounded, and the proof used in case (1) has to be modified. We prove the bounds for $\Phi(\sigma)$ by considering two subintervals separately. The details are outlined below. For convenience, throughout the proof we let C_1 and C_2 denote some generic positive constants independent of h, l_1 and l_2.

Case (1). In this case, l_1 and l_2 are both assumed greater than 1. Then, since $\frac{\sqrt{\sigma}}{h} \geq \frac{2\sin(\pi h)}{h} \geq C_1$, it follows that $\frac{l_i\sqrt{\sigma}}{h} \geq C_1$. Substituting this into the expression for the bounds for $\mu(\sigma)$, given in part 2 of Lemma 6.3, we obtain uniform upper and lower bounds:

$$C_1 \leq \mu(\sigma) \leq C_2, \qquad \text{for } \sigma \in [4\sin^2(\pi h), 4].$$

Since $\Phi(\sigma) = \mu(\sigma)H(\sigma)$, we can obtain bounds for $\Phi(\sigma)$ by considering bounds for $H(\sigma)$. Since by assumption $\min\{l_1, l_2\}$ is $O(1)$ independent of h, it follows that for small enough h, we have $\sigma^\star = O(h)$, and therefore $\sigma^\star > 4\sin^2(\pi h)$. Thus the maximum of $H(\sigma)$ occurs in the interior of the interval $[4\sin^2(\pi h), 4]$, by part 3 of Lemma 6.3. In this case, $H(\sigma)$ is monotone increasing to the left of $\sigma^\star$, and monotone decreasing to the right of $\sigma^\star$. Thus, we obtain that:

$$\min\{H(4\sin^2(\pi h)), H(4)\} \leq H(\sigma) \leq H(\sigma^\star), \qquad \text{for } \sigma \in [4\sin^2 \pi h), 4].$$

Substituting the expression for $\sigma^\star$ into $H(\sigma)$, it can be easily shown that:

$$H(\sigma^\star) \leq C_2\sqrt{\frac{l_1 l_2}{l_1 + l_2}}h^{1/2}.$$

At $\sigma = 4\sin^2(\pi h)$, it can be shown that:

$$H(4\sin^2(\pi h)) \geq \frac{1}{\frac{l_1+l_2}{l_1 l_2} + \beta h},$$

which becomes large if both l_1 and l_2 becomes large. At $\sigma = 4$, it can easily be shown that:

$$\frac{\sqrt{8}}{9} \leq H(4).$$

Thus the minimum of $H(\sigma)$ is always $O(1)$. Combining these bounds with the uniform bounds for $\mu(\sigma)$, we obtain that:

$$C_1 \leq \Phi(\sigma) \leq C_2\sqrt{\frac{l_1 l_2}{l_1 + l_2}}h^{-1/2}.$$

Case (2). Here we assume that either l_1 or l_2 is smaller than or equal to 1. For definiteness, let us suppose that $l_1 \leq 1$ and $l_1 \leq l_2$. In this case, $\mu(\sigma)$ may no longer be uniformly bounded. Indeed, $\mu(4\sin^2(\pi h))$ can be of size $O(\frac{1}{l_1})$. Consequently, we do not consider bounds for $H(\sigma)$ and $\mu(\sigma)$ separately, as they lead to bounds which

are larger than is the case. Instead, we find uniform bounds for $\Phi(\sigma)$ on the two sub-intervals $I_1 \equiv [4\sin^2(\pi h), \sigma^\star]$ and $I_2 \equiv [\sigma^\star, 4]$, separately (as mentioned before, when either l_1 or l_2 is $O(1)$, for small enough h we have $\sigma^\star > 4\sin^2(\pi h)$). In case $(2a)$, we obtain bounds for $\Phi(\sigma)$ on the interval I_1 using uniformly valid expansions for $\Phi(\sigma)$. In case $(2b)$, we obtain bounds for $\Phi(\sigma)$ on the interval I_2, using the fact that $\Phi(\sigma)$ is monotone decreasing on I_2, by Lemma 6.3. The details of both subcases are given below.

Case $(2a)$. On interval I_1, we will first show that $\Phi(\sigma)$ satisfies:

$$(21) \quad C_1 \frac{l_1\sqrt{\sigma}}{h}\left(1+\frac{e^{-\frac{c_1 l_1}{h}\sqrt{\sigma}}}{1-e^{-\frac{c_1 l_1}{h}\sqrt{\sigma}}}\right) \le \Phi(\sigma) \le C_2 \frac{\sqrt{\sigma} l_1}{h}\left(1+\frac{e^{-\frac{c_2 l_1}{h}\sqrt{\sigma}}}{1-e^{-\frac{c_2 l_1}{h}\sqrt{\sigma}}}\right).$$

To show this, we note that:

$$\frac{h}{2l_1} \le \sigma^\star = \frac{h/l_1 + h/l_2}{\beta - \left(\frac{h}{2l_1}+\frac{h}{2l_2}\right)} \le \frac{4h}{l_1},$$

since $2h \le l_1 \le l_2$, and $1 \le \beta \le 2$. From this it follows that:

$$\beta\sigma \le \beta\sigma^\star \le 4\beta\frac{h}{l_1} \le 8\frac{h}{l_1}, \qquad \text{for } \sigma \in I_1.$$

This result can be used to obtain bounds for the denominator of $H(\sigma)$:

$$(22) \qquad \frac{h}{l_1} \le \frac{h}{l_1} + \frac{h}{l_2} + \beta\sigma \le 10\frac{h}{l_1}, \qquad \text{for } \sigma \in I_1,$$

which in turn gives bounds for $H(\sigma)$:

$$(23) \qquad \frac{C_1 l_1\sqrt{\sigma}}{h} \le H(\sigma) \le \frac{C_2 l_1\sqrt{\sigma}}{h}, \qquad \text{for } \sigma \in I_1.$$

Next, we obtain bounds for $\mu(\sigma)$ by modifying part (2) of Lemma 6.3, in which the sums are replaced with bounds for each term. We easily obtain:

$$(24) \quad 1+\frac{e^{-\frac{c_1 l_1}{h}\sqrt{\sigma}}}{1-e^{-\frac{c_1 l_1}{h}\sqrt{\sigma}}} \le \mu(\sigma) \le 4+\frac{4e^{-\frac{c_2 l_1}{h}\sqrt{\sigma}}}{1-e^{-\frac{c_2 l_1}{h}\sqrt{\sigma}}}, \qquad \text{for } \sigma \in [4\sin^2(\pi h), 4].$$

Combining equations (23) and (24), we obtain the bounds for $\Phi(\sigma) = \mu(\sigma)H(\sigma)$ given in equation (21).

Note that the upper and lower bounds for $\Phi(\sigma)$ in equation (21) can be expressed in terms of a single function $T(z)$:

$$\frac{C_1}{c_1}T\left(\frac{c_1 l_1\sqrt{\sigma}}{h}\right) \le \Phi(\sigma) \le \frac{C_2}{c_2}T\left(\frac{c_2 l_1\sqrt{\sigma}}{h}\right) \quad \text{, where } T(z) \equiv z\left(1+\frac{e^{-z}}{1-e^{-z}}\right),$$

for $z \equiv \frac{c_i l_1\sqrt{\sigma}}{h}$ varying in the interval $[\frac{2l_1 c_i}{h}\sin(\pi h), \frac{l_1 c_i}{h}\sqrt{\sigma^\star}]$ which is a subset of the interval $[0, \frac{l_1 c_i}{h}\sqrt{\sigma^\star}]$. It is easily verified that

$$T'(z) = \frac{e^z\left(e^z - 1 - z\right)}{\left(e^z-1\right)^2} > 0, \qquad \text{for } z > 0.$$

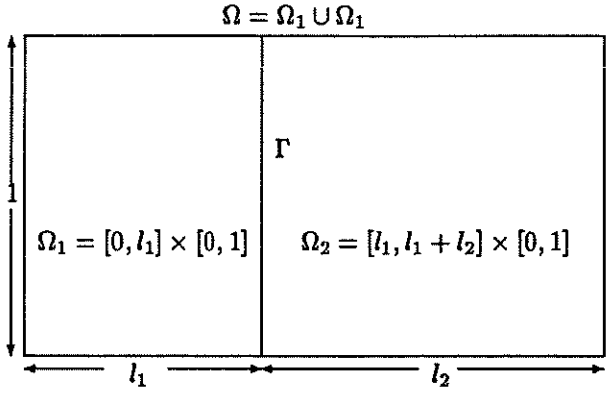

FIG. 1. *A model domain.*

Thus, $T(z)$ is a monotone increasing function satisfying:

$$T(0) \leq T(z) \leq T(\frac{c_2 l_1 \sqrt{\sigma^*}}{h}), \quad \text{for } z \in [0, \frac{c_2 l_1 \sqrt{\sigma^*}}{h}].$$

We note that the lower bound is $T(0) = 1$. It is easily shown for sufficiently large z, say $z \geq 1$, that $T(z) \leq C_2 z$, for some positive constant C_2. Since, $\frac{h}{2l_1} \leq \sigma^* \leq 4\frac{h}{l_1}$, we obtain that $c_2\sqrt{\sigma^*}\frac{l_1}{h} \geq c_2\sqrt{\frac{l_1}{2h}} \geq 1$ for sufficiently small h, and that $T(\frac{c_2 l_1 \sqrt{\sigma^*}}{h}) \leq C_1 \frac{c_2 l_1 \sqrt{\sigma^*}}{h} \leq C_2 \sqrt{l_1} h^{-1/2}$. Substituting this in the expression for $\Phi(\sigma)$, we obtain that

$$C_1 \leq \Phi(\sigma) \leq C_2 \sqrt{l_1} h^{-1/2}, \quad \text{for } \sigma \in I_1.$$

Since $\frac{l_1}{2} \leq \frac{l_1 l_2}{l_1 + l_2}$, it follows that this is our desired result.

Case (2b). Finally, we consider bounds for $\Phi(\sigma)$, when $\sigma \in I_2$. Since $\sigma \geq \sigma^*$, we obtain by part 5 of Lemma 6.3 that $\Phi'(\sigma) \leq 0$ and thus

$$\Phi(\sigma^*) \geq \Phi(\sigma) \geq \Phi(4), \quad \text{for } \sigma \in I_2.$$

Since $\Phi(\sigma^*) \leq C_2 \sqrt{l_1} h^{-1/2}$, and $\Phi(4) \geq C_2$, we obtain the same bounds for $\Phi(\sigma)$ on the interval I_2 as on the interval I_1. Since $\frac{l_1}{2} \leq \frac{l_1 l_2}{l_1 + l_2}$, it follows that this is our desired result. □

Next we present some sample numerical results which compares the symmetrised tridiagonal probe preconditioner M_1 = symmetrised-PROBE$(S, 1)$ with the Golub-Mayer preconditioner M_{GM} for the following elliptic problem on the domain Ω of Figure 1:

$$(25) \quad Lu = -\frac{\partial}{\partial x}\left(e^{\theta_1 xy}\frac{\partial u}{\partial x}\right) - \frac{\partial}{\partial y}\left(e^{\theta_2 xy}\frac{\partial u}{\partial y}\right) = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega.$$

The 5 point centered scheme (2) was used on a $n \times n$ grid, with subdomains of size $m_1 \times n$ and $m_2 \times n$, where $m_1 + m_2 = n$. The condition number and the iterations required to reduce the residual by a factor of 10^{-7} (in the Euclidean norm) are listed for varying choices of h, l_1, l_2, θ_1 and θ_2.

Table 1 lists the condition numbers and the number of iterations required for both preconditioners as the mesh width h is varied, with the values of l_1, l_2, θ_1

TABLE 1
h dependence for $\theta_1 = 2$, $\theta_2 = -2$ and $l_1 = l_2 = 1/2$.

n	$\kappa(M_1^{-1}S)$	Iters		$\kappa(M_{GM}^{-1}S)$	Iters		$\kappa(M_{SGM}^{-1}S)$	Iters
10	1.22	6		1.80	7		-	-
20	1.62	8		1.85	7		2.29	9
30	1.97	9		1.87	7		2.34	10
40	2.28	10		1.88	7		2.38	9

and θ_2 fixed as indicated. Since the Golub-Mayer preconditioner is independent of possibly highly varying coefficients, we also used a scaled Golub-Mayer preconditioner $M_{SGM} \equiv D^{1/2} M_{GM} D^{1/2}$, where D is the diagonal of the matrix L_{33}. As expected, $\kappa(M_{GM}^{-1}S)$ is uniformly bounded for varying h, whereas $\kappa(M_1^{-1}S)$ depends mildly on h (about $O(h^{-1/2})$, consistent with Theorem 6.4), even though the boundary conditions are different. The cross-over is about $n = 20$ or 30. For this case, M_{SGM} performs slightly worse than M_{GM}.

Based on studies of optimal preconditioners of a given sparsity pattern, Greenbaum and Rodrigue [25] had conjectured that the optimal symmetric positive definite tridiagonal preconditioner for the interface matrix S has condition number bounded below by $O(h^{-1/2})$. Our numerical results indicate that the tridiagonal probe preconditioner performs as well asymptotically as the optimal tridiagonal preconditioner (which of course cannot be computed easily in general) for S.

6.3. Dependence on aspect ratios. We shall next discuss the dependence of $\kappa(M_{CP}^{-1}S)$ on the aspect ratios l_1 and l_2. The result from Theorem 6.4 for the model problem indicates that the condition number of the probe preconditioned system does depend on the aspect ratios of the two subdomains:

$$\kappa(M_{CP}^{-1}S) \leq C\sqrt{\frac{l_1 l_2}{l_1 + l_2}} h^{-1/2} \leq C\sqrt{\min\{l_1, l_2\}} h^{-1/2}. \tag{26}$$

For instance, if $l_1 \leq l_2$, then $\kappa \leq C\sqrt{l_1} h^{-1/2}$, and this can grow if l_1 becomes large, and thus the performance of the probe preconditioner can deteriorate if both aspect ratios become large. However, we note that in this case S itself becomes close to singular. On the other hand, if $l_1 = O(1)$, and l_2 is allowed to vary independent of l_1, then $\kappa(M_{CP}^{-1}S)$ can be bounded independent of l_2.

In comparison, for the Dirichlet case, the Golub-Mayer preconditioned system has been shown (see Bjørstad and Widlund [3] and Chan [9]) to have a condition number

$$\kappa(M_{GM}^{-1}S) \leq C\left(1 + \frac{1}{l_1} + \frac{1}{l_2}\right).$$

This bound can become large if either of the aspect ratios l_1 or l_2 becomes small. However, the performance is good when both the aspect ratios are $O(1)$ or larger.

Table 2 illustrates the varying performance of M_{GM}, M_{SGM} and M_1 with respect to aspect ratios of the subdomains, for test problem (25) on a unit square with a $(10+m_2)\times 40$ grid, partitioned into two subdomains of size 10×40 and $m_2 \times 40$, with $\theta_1 = 2$ and $\theta_2 = -2$ and using Dirichlet boundary conditions. Note that $\kappa(M_1^{-1}S)$ appears to

TABLE 2
Aspect ratio dependence for $h = 1/40$, $m_1 = 10$, m_2 given, $\theta_1 = 2$ and $\theta_2 = -2$.

m_2	$\kappa(M_1^{-1}S)$	Iters		$\kappa(M_{GM}^{-1}S)$	Iters		$\kappa(M_{SGM}^{-1}S)$	Iters
8	1.87	9		1.79	8		2.91	12
6	1.76	9		1.97	9		3.57	14
4	1.60	8		2.37	10		4.62	16
2	1.37	7		3.81	12		6.47	18

decrease like $O(\sqrt{m_2})$ as $m_2 \to 0$, as predicted by (26), whereas $\kappa(M_{GM}^{-1}S)$ deteriorates mildly. Again, M_{SGM} performed slightly worse than M_{GM} for this problem.

Unlike the theoretical bounds presented in Theorem 6.4 for the model periodic case, which showed that the condition number of the probe preconditioned system grows as $\sqrt{\min\{l_1, l_2\}}$ for fixed h, the numerical results for the *Dirichlet* case indicate that the condition number of $M_1^{-1}S$ are bounded independent of l_1 or l_2, and for small l behaves like $O(\sqrt{l})$. This will be illustrated in Figure 2, of Section 6.4.

6.4. Dependence on scalings of the coefficients. In this subsection, we focus on the performance of the preconditioners for various scalings of the coefficients. Here, some of the results are not restricted to the model problem. First, we consider the operator L of problem (1) with coefficients $a(x, y)$ and $b(x, y)$. As before, Ω is partitioned into two subdomains Ω_1 and Ω_2. We note that the Schur complement S can be written as

$$S = S^{(1)} + S^{(2)} = (L_{33}^{(1)} - L_{13}^T L_{11}^{-1} L_{13}) + (L_{33}^{(2)} - L_{23}^T L_{22}^{-1} L_{23}),$$

where $L_{33}^{(i)}$ denotes the contribution to L_{33} from subdomain Ω_i, and $L_{33} = L_{33}^{(1)} + L_{33}^{(2)}$.

The coefficients of the original operator L are modified to obtain a scaled operator $L(\rho)$ as follows: $a(x, y)$ and $b(x, y)$ are multiplied by a positive constant scaling ρ in subdomain Ω_1, thereby making the coefficients possibly discontinuous across the interface Γ. The Schur complement for the scaled operator $L(\rho)$ will be denoted by $S(\rho)$, and it is easily seen that: $S(\rho) = \rho S^{(1)} + S^{(2)}$. If M denotes any preconditioner for $S = S(1)$, then the following theorem gives an upper bound for the condition number of the preconditioned system $M^{-1}S(\rho)$:

THEOREM 6.5. *The condition number of the preconditioned system $M^{-1}S(\rho)$ is bounded by* $\max\{\kappa(M^{-1}S^{(1)}), \kappa(M^{-1}S^{(2)})\}$.

Proof. Let λ_i^{min} and λ_i^{max} denote the lower and upper bounds for the following Rayleigh quotients:

$$\lambda_i^{min} \leq \frac{x^T S^{(i)} x}{x^T M x} \leq \lambda_i^{max}.$$

From this it follows that

$$\rho\lambda_1^{min} + \lambda_2^{min} \leq \frac{x^T S(\rho) x}{x^T M x} \leq \rho\lambda_1^{max} + \lambda_2^{max}.$$

Thus the condition number of the preconditioned system is bounded by

$$(\rho\lambda_1^{max} + \lambda_2^{max})/(\rho\lambda_1^{min} + \lambda_2^{min})$$

which is easily shown to be a monotone function of ρ with the asymptotes given by $\lambda_1^{max}/\lambda_1^{min}$ and $\lambda_2^{max}/\lambda_2^{min}$. □

This theorem indicates that the scaling ρ could possibly affect the conditioning of $M^{-1}S(\rho)$ adversely only if $\kappa(M^{-1}S^{(1)})$ and $\kappa(M^{-1}S^{(2)})$ are significantly different. The upper bounds on the worst possible conditioning, however, depends only on M, $S^{(1)}$ and $S^{(2)}$ and these are independent of ρ. Though in the case of two subdomains, a simple scaling of the interface preconditioner will not affect the preconditioning, in the case of more than two subdomains, however, proper scaling of the preconditioner on each of the edges constituting the interface is required for efficient preconditioning, see for instance [4]. This can also be done by suitable use of probing, see [14].

Applying Theorem 6.5 to the case of the scaled version of a model Dirichlet problem preconditioned by the Golub-Mayer preconditioner, we obtain that:

COROLLARY 6.6. *If $M = M_{GM}$ is used to precondition $S(\rho)$, then $\kappa(M_{GM}^{-1}S(\rho))$ can vary between $O(1+\frac{1}{l_1})$ and $O(1+\frac{1}{l_2})$, depending on ρ.*

The preceeding case corresponded to preconditioners which did not adapt to the scale of each term in the Schur complement $S(\rho)$. In case the preconditioner M adapts to the scaling ρ, as in the case of probe preconditioners which are linearly dependent on the matrices they approximate, then we obtain different upper bounds for the preconditioned system, as given in the following:

THEOREM 6.7. *If the preconditioner for $S(\rho) \equiv \rho S^{(1)} + S^{(2)}$ is of the form: $M(\rho) = \rho M^{(1)} + M^{(2)}$, and if*

$$\lambda_{min}^i \leq \frac{x^T S^{(i)} x}{x^T M^{(i)} x} \leq \lambda_{max}^i, \qquad \text{for } i = 1, 2,$$

then,

$$\kappa(M(\rho)^{-1}S(\rho)) \leq \frac{\max\{\lambda_{max}^1, \lambda_{max}^2\}}{\min\{\lambda_{min}^1, \lambda_{min}^2\}}.$$

Proof. The proof follows trivially from the assumptions. □

Thus, if the bounds for the subdomain problems are independent of the aspect ratios, then the scaled version will also be independent of the aspect ratios. For the scaled version of the model operator L of equation (15), with $S(\rho) = \rho S^{(1)} + S^{(2)}$, we easily obtain that $M_{CP}(\rho) = \rho M^{(1)} + M^{(2)}$, where $S^{(i)} = F\ diag(\lambda_j^{(i)})\ F^{-1}$ and

$$\lambda_j^{(i)} = \left(\frac{1+\gamma(\sigma_j)^{l_i/h}}{1-\gamma(\sigma_j)^{l_i/h}}\right)\sqrt{\sigma_j + \frac{\sigma_j^2}{4}}, \quad \text{and} \quad \gamma(\sigma_j) \equiv \frac{1+\frac{\sigma_j}{2}-\sqrt{\sigma_j+\frac{\sigma_j^2}{4}}}{1+\frac{\sigma_j}{2}+\sqrt{\sigma_j+\frac{\sigma_j^2}{4}}},$$

with $\lambda_0^i = h/l_i$, and where

$$M^{(i)} = F diag\left(\frac{h}{l_i} + \beta_i \sigma_j\right) F^{-1}, \qquad \text{for } j = 0, \ldots, n,$$

where $1 \geq \beta_i \geq 1/2$. In this case, we obtain the following bounds for the condition number of the preconditioned system $\kappa(M_{CP}(\rho)^{-1}S(\rho))$:

COROLLARY 6.8. *If $M_{CP}(\rho) =$ circulant-PROBE$(S(\rho), 1)$ is used to precondition $S(\rho)$ for the scaled version of the model problem, then*

$$\kappa(M_{CP}^{-1}(\rho)S(\rho)) \leq C\sqrt{\max\{l_1, l_2\}}h^{-1/2},$$

for a constant C independent of ρ, h, and the aspect ratios l_1 and l_2.

Proof. By using linearity of the probing procedure, it is easy to verify that $(M_{CP}^{(i)})^{-1}S^{(i)}$ is the same as the preconditioned system $M_{CP}^{-1}S$ when both subdomains have the same aspect ratio l_i. Thus, it follows that:

$$C_1 \leq \lambda((M^{(i)})^{-1}S^{(i)}) \leq C\sqrt{l_i}h^{-1/2},$$

for $i = 1, 2$, where $\lambda(.)$ denotes the eigenvalues of the matrix argument. Applying Theorem 6.7, the desired result follows. □

This theoretical result indicates that the probe preconditioner in the model problem can be sensitive to scalings of the coefficients only if the aspect ratios of the two subdomains are significantly different, in which case the condition number can vary from $O(\sqrt{\min\{l_1, l_2\}}h^{-1/2})$ to $O(\sqrt{\max\{l_1, l_2\}}h^{-1/2})$. However, in the case that $\max\{l_1, l_2\}$ is large, then $\kappa(S(\rho)) = O(\max\{l_1, l_2\})$ which is also large. If both l_1 and l_2 are $O(1)$, then the scalings do not affect the convergence of the preconditioner.

In the Dirichlet case, however, numerical results seem to indicate that the probe preconditioned system performs better than as suggested in Corollary 6.8. In Figure 2, we illustrate the results for the model scaled Poisson problem with Dirichlet boundary conditions with varying l_1, l_2 and ρ. The results indicate that the condition number of the probe preconditioned system can be bounded independently of l_1, l_2 and ρ. Based on Figure 2, we conjecture that the condition number of the probe preconditioned system satisfies:

$$\kappa(M_1^{-1}S) \leq C\sqrt{\min\{l_1, l_2, 1\}}h^{-1/2},$$

for the Dirichlet case. This is similar to the bound obtained for the model periodic case in Theorem 6.4, except that $\min\{l_1, l_2\}$ is replaced by $\min\{l_1, l_2, 1\}$ which is uniformly bounded for large l_1 and l_2.

We now present numerical results on the rate of convergence of the probe preconditioned system and both the regular and scaled version of the Golub-Mayer preconditioned system, in the case of continuous, but highly varying coefficients. The tests were carried out on problem (25) with *Dirichlet* boundary conditions, with the parameters as shown. The results are presented in Table 3. The probing preconditioners seem to adapt well to such variations in the coefficients of L, while the performance of other preconditioners which are independent of the coefficients, like M_{GM}, deteriorate. However, unlike the results in Table 1 and 2, the scaled version M_{SGM} improves over the performance of M_{GM} significantly. This may be due to the isotropy of the coefficients in Table 3. More tests seem to be needed to study the effect of scaling on optimal preconditioners such as M_{GM}.

7. Summary. Unlike various interface preconditioners in domain decomposition, the probing preconditioners are constructed as algebraic approximations to the interface operator. They have the disadvantage of being non-spectrally equivalent with respect to mesh size variation. However, since the techniques are algebraic in nature, they can and have been applied to construct preconditioners to more general differential operators for which optimal preconditioners are not known.

We have shown that under certain conditions which are often valid in applications, the probing technique leads to nonsingular approximations. In addition, the preconditioners are linearly dependent on the matrices they approximate and preserve diagonal dominance. However, not all the probing techniques preserve symmetry of the matrices they approximate, and symmetric positive definiteness is generally not preserved.

TABLE 3
Dependence on coefficients for $\theta_1 = \theta_2$, $n = 20$, and $l_1 = l_2 = 1/2$.

θ_1	$\kappa(M_1^{-1}S)$	Iters	$\kappa(M_{GM}^{-1}S)$	Iters	$\kappa(M_{SGM}^{-1}S)$	Iters
0	1.68	7	1.09	3	1.09	3
2	1.67	8	2.48	12	1.11	4
4	1.66	8	6.17	17	1.18	4
6	1.63	8	15.37	21	1.28	5

For a model elliptic problem we have shown that the probing technique has some desireable properties: it reduces the condition number of the interface operator from $O(h^{-1})$ to $O(h^{-1/2})$. Moreover, the probing technique is also fairly robust with respect to aspect ratios and coefficient variations, though there could be some mild dependence for large aspect ratios. However, for the Dirichlet problem, our numerical results indicate that the rates are bounded independent of the aspect ratios l_1, l_2 and the scaling ρ but retains the $O(h^{-1/2})$ dependence.

In summary, if h is not very small, and the aspect ratios and coefficients are highly varying, then probing can provide a competitive alternative to other available interface preconditioners.

Acknowledgement. The authors would like to thank Professor Olof Widlund for his helpful discussions on various topics in this paper. They would also like to thank the referees for their helpful suggestions.

REFERENCES

[1] V. I. Agoshkov. Poincaré-Steklov operators and domain decomposition methods in finite dimensional spaces. In Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 73–112, Philadelphia, 1988. SIAM.

[2] O. Axelsson and B. Polman. Block preconditioning and domain decomposition methods, I. Technical Report 8735, Catholic University, Nijmegen, December 1987.

[3] Petter E. Bjørstad and Olof B. Widlund. Iterative methods for the solution of elliptic problems on regions partitioned into substructures. *SIAM J. Numer. Anal.*, 23(6):1093 – 1120, 1986.

[4] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, I. *Math. Comp.*, 47(175):103– 134, 1986.

[5] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. An iterative method for elliptic problems on regions partitioned into substructures. *Math. Comp.*, 46(173):361–369, 1986.

[6] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, II. *Math. Comp.*, 49:1–16, 1987.

[7] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, III. *Math. Comp.*, 51:415 – 430, 1988.

[8] James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, IV. *Math. Comp.*, 53:1 – 24, 1989.

[9] Tony F. Chan. Analysis of preconditioners for domain decomposition. *SIAM J. Numer. Anal.*, 24(2):382–390, 1987.

[10] Tony F. Chan. Boundary probe preconditioners for fourth order elliptic problems. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Domain Decomposition Methods*, pages 160–167, Philadelphia, 1989. SIAM.

[11] Tony F. Chan and Danny Goovaerts. A note on the efficiency of domain decomposed incomplete factorizations. *SIAM J. Sci. Stat. Comput.*, 11(4):794 – 803, July 1990.

[12] Tony F. Chan and Thomas Y. Hou. Domain decomposition interface preconditioners for general 2nd order elliptic problems. Technical Report CAM 88-16, Department of Mathematics,

UCLA, 1990.

[13] Tony F. Chan and David F. Keyes. Interface preconditioning for domain-decomposed convection-diffusion operators. Technical report, CAM 89-28, Department of Mathematics, University of California Los Angeles, 1989. Pages 245 – 262, Proceedings of the Third International Symposium on Domain Decomposition Methods, Houston, Texas, April, 1989.

[14] Tony F. Chan and Tarek P. Mathew. An application of the probing technique to the vertex space method in domain decomposition. Technical Report CAM 90-22, Department of Mathematics, UCLA, 1990. To appear in Proceedings of 5th Domain Decomposition Conference, Moscow, April 1990.

[15] Tony F. Chan and Diana C. Resasco. A survey of preconditioners for domain decomposition. Technical Report /DCS/RR-414, Yale University, 1985.

[16] A. R. Curtis, M. J. Powell, and J. K. Reid. On the estimation of sparse Jacobian matrices. *J. Inst. Maths. Applics.*, 13:117–120, 1974.

[17] A. Quarteroni D. Funaro and P. Zanolli. An iterative procedure with interface relaxation for domain decomposition methods. *SIAM J. Numer. Anal.*, 25:1213 –1236, 1988.

[18] June Donato. Eigendecompositions of the capacitance matrix for Poisson's equation on a strip. 1988. Department of Mathematics, UCLA, Course M285J Project.

[19] Maksymilian Dryja. A capacitance matrix method for Dirichlet problem on polygon region. *Numer. Math.*, 39:51 – 64, 1982.

[20] Maksymilian Dryja and Olof B. Widlund. An additive variant of the Schwarz alternating method for the case of many subregions. Technical Report 339, also Ultracomputer Note 131, Department of Computer Science, Courant Institute, 1987.

[21] Maksymilian Dryja and Olof B. Widlund. Some domain decomposition algorithms for elliptic problems. In *Proceedings of the Conference on Iterative Methods for Large Linear Systems held in Austin, Texas, October 1988, to celebrate the Sixty-fifth Birthday of David M. Young, Jr., Academic Press, Orlando, Florida, 1989.*, 1989.

[22] S. C. Eisenstat. Personal Communication, 1985.

[23] R. Glowinski and O. Pironneau. Numerical methods for the first biharmonic equation and for the two-dimensional stokes problem. *Siam Review*, 21(2):167 – 212, 1979.

[24] Gene Golub and D. Mayers. The use of preconditioning over irregular regions. In R. Glowinski and J. L. Lions, editors, *Computing Methods in Applied Sciences and Engineering, VI*, pages 3–14, Amsterdam, New York, Oxford, 1984. North-Holland. Proceedings of a conference held in Versailles, France, December 12-16,1983.

[25] Anne Greenbaum and Garry Rodrigue. Optimal preconditioners of a given sparsity pattern. *BIT*, 29:610 – 634, 1989.

[26] Eugene Isaacson and Herbert B. Keller. *Analysis of Numerical Methods.* John Wiley and Sons, 1966.

[27] David E. Keyes. Domain decomposition methods for the parallel computation of reacting flows. *Comput. Phys. Comm.*, 53:s181 – s200, 1989.

[28] David E. Keyes and William D. Gropp. A comparison of domain decomposition techniques for elliptic partial differential equations and their parallel implementation. *SIAM J. Sci. Stat. Comput.*, 8(2):s166 – s202, 1987.

[29] David E. Keyes and William D. Gropp. Domain decomposition techniques for the parallel solution of nonsymmetric systems of elliptic bvps. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Domain Decomposition Methods*, Philadelphia, 1989. SIAM.

[30] Hongjun Li. Toeplitz preconditioner in domain decomposition methods. 1988. UCLA, M285J Project Report.

[31] Alfio Quarteroni and Alberto Valli. Theory and applications of steklov-poincaré operators for boundary-value problems: the heterogeneous operator case. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Proceedings of 4th International Conference on Domain Decomposition Methods, Moscow*, Philadelphia, 1990. SIAM.

[32] Barry F. Smith. An optimal domain decomposition preconditioner for the finite element solution of linear elasticity problems. Technical Report 482, Department of Computer Science, Courant Institute, 1989. To Appear in Proceedings of Copper Mountain Conference on Iterative Methods, SIAM Journal of Scientific and Statistical Computing.

[33] Waikin Tsui. *Domain Decompostion of Biharmonic and Navier-Stokes Equations.* PhD thesis, University of California at Los Angeles, 1991.

[34] Richard S. Varga. *Matrix Iterative Analysis.* Prentice-Hall, 1962.

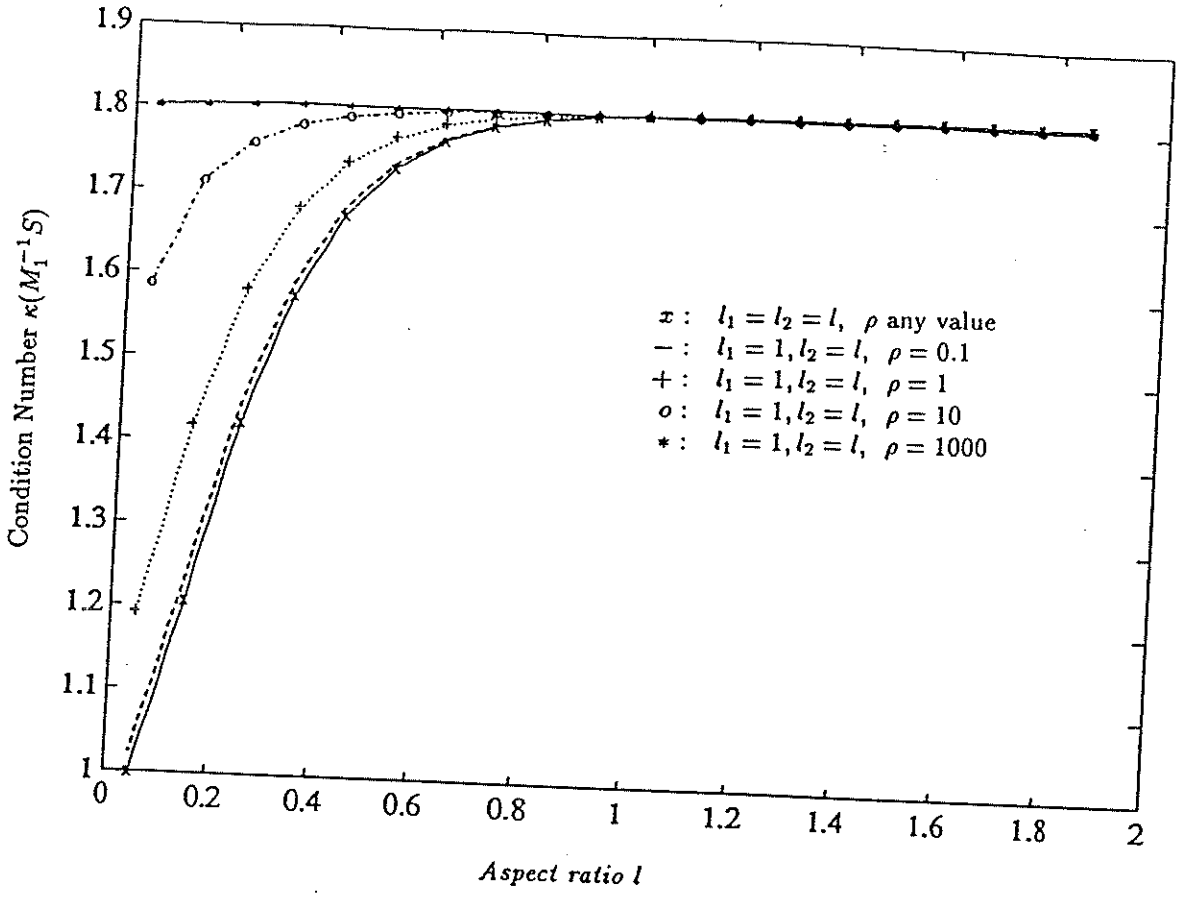

FIG. 2. *Probing for model Dirichlet problem with varying l_1, l_2 and ρ, and $n = 20$.*