UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Fourier Analysis of Iterative Methods for Elliptic Problems

Tony F. Chan
Howard C. Elman

January 1987

Revised February 1988

CAM Report 87-04

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

Abstract. We present a Fourier method for analyzing stationary iterative methods and preconditioners for discretized elliptic boundary value problems. Similar to the von Neumann stability analysis of hyperbolic and parabolic problems, the approach is easier to apply and reveals more details about convergence properties than standard techniques, and can be applied in a systematic way to a wide class of numerical methods. Although the analysis is applicable only to periodic problems, the results essentially reproduce those of classical convergence and condition number analysis for problems with other boundary conditions, such as the Dirichlet problem. In addition, they give suggestive new evidence of the strengths and weaknesses of methods such as incomplete factorization preconditioners in the Dirichlet case.

1. Introduction

Iterative methods constitute an indispensable tool for solving large sparse linear systems of equations, such as those arising from the discretization of elliptic partial differential equations. Among the more common examples of such techniques are stationary methods such as the Jacobi, Gauss-Seidel and SOR methods [34,37], and preconditioned conjugate-gradient or semi-iterative methods [7,9,21], which use an approximate factorization of the coefficient matrix to improve the conditioning of the problem. For the "model problem," the discrete Poisson equation

$$-\Delta u = f \tag{1.1}$$

posed on the unit square $\Omega = \{0 \le x, y \le 1\}$ with Dirichlet boundary conditions and discretized by finite differences, there are rigorous theoretical results giving bounds on convergence rates for all these methods.

In particular, let

$$Au = b \tag{1.2}$$

denote a linear system where A is symmetric positive definite, let Q be some nonsingular splitting operator, and let the splitting be represented as

$$A = Q - R. \tag{1.3}$$

We consider stationary methods of the form

$$u^{(m+1)} = Q^{-1}Ru^{(m)} + Q^{-1}b,$$

and preconditioned conjugate gradient methods, where for symmetric positive definite $Q = LL^T$, the conjugate gradient method is used to solve the preconditioned system $[L^{-1}AL^{-T}][L^Tu] = L^{-1}b$. Let $e^{(m)} = u - u^{(m)}$ denote the error for the m'th iterate computed by any such method. The number of iterations needed to make the relative error

$$\|e^{(m)}\|/\|e^{(0)}\| \tag{1.4}$$

less than a specified tolerance is approximately inversely proportional to the asymptotic rate of convergence, R_∞. For stationary methods, $R_\infty = -\ln(\rho)$, where $\rho = \rho(Q^{-1}R)$,

the spectral radius of the iteration matrix, and for PCG, $R_\infty = -\ln((\sqrt{\kappa}-1)/(\sqrt{\kappa}+1))$ where κ is the condition number of $L^{-1}AL^{-T}$. Here the norms for (1.4) are the Euclidean norm $\|v\|_2 = (v^T v)^{1/2}$ for the stationary methods and the A-norm $\|v\|_A = (v^T Av)^{1/2}$ for the preconditioned conjugate gradient methods.[1]

Consider the case where (1.2) comes from the Dirichlet problem discretized by second order finite differences on a uniform $n \times n$ grid in Ω. Let $h = \frac{1}{n+1}$ and $N = n^2$. The asymptotic convergence rates for a representative set of iterative methods, as functions of h, are given in Table 1. For stationary methods, these results are classical; their derivation and history can be found in Varga [34], pp. 201ff, Young [37], pp. 127ff and 464ff, and Axelsson [3], pp. 37-39. The relaxation parameters ω_b and ω_1 are the optimal and "good" choices for SOR and SSOR, respectively, as presented in [34, 37]. The asymptotic analysis for SSORCG is the same as that of the SSOR semi-iterative method, as discussed in [37], p. 472. The use of the ILU factorization as a preconditioner for CG is presented in Meijerink and van der Vorst [29], although no condition number analysis is given there. Asymptotic results showing that ILU preconditioned finite difference operators have condition number $O(h^{-2})$ (the same as A) appear in Gustafsson [19]; the specific coefficient $\sqrt{17}\pi$ given in the table is a lower bound, derived in Chandra [7], p. 247. Analysis showing that the MILU preconditioned systems have $O(h^{-1})$ condition number (for general elliptic operators) appears in Dupont, Kendall and Rachford [11]; further analysis is given in the papers by Axelsson [2], Dupont [10] and Gustafsson [19]. The coefficient $2\sqrt{\pi}$ in the convergence rate corresponds to a near-optimal choice of the MILU iteration parameter, which follows from the analysis in [2]. For a summary of some of the early developments of preconditioners, see Golub and O'Leary's annotated bibliography [17].

	Method	Convergence Rate
Stationary	Jacobi	$\frac{\pi^2}{2}h^2$
	Gauss-Seidel	$\pi^2 h^2$
	SOR(ω_b)	$2\pi h$
	SSOR(ω_1)	πh
PCG	ILUCG	$\sqrt{17}\pi h$
	MILUCG	$2\sqrt{\pi}\sqrt{h}$
	SSORCG(ω_1)	$2\sqrt{\pi}\sqrt{h}$

Table 1: Asymptotic convergence rates for the Dirichlet problem.

One unsatisfying aspect of the derivations of these results is the degree to which they depend on "hard analysis," i.e. the establishing of complicated sets of inequalities leading

[1] For both classes of methods, R_∞ is the limiting value as $m \to \infty$ of $-\frac{1}{m}\ln \beta_m$, where β_m is an upper bound for (1.4). This notation is standard for stationary methods, where $\beta_m = \|(Q^{-1}R)^m\|$ [34,37]. For PCG, $\beta_m = 2[(\sqrt{\kappa}-1)/(\sqrt{\kappa}+1)]^m$, but using the limit is actually a slight abuse of notation since PCG converges in a finite number of steps. In the context of Table 1, however, this number of steps is $O(h^{-2})$, typically much larger than R_∞^{-1} for PCG, so that R_∞ is still a useful measure.

to bounds on spectral radii or condition numbers. Moreover, the analyses for individual methods tend to be specialized for those methods, so that it is somewhat difficult to gain insight into one method from the analysis of another. For example, although from an intuitive point of view the SOR and SSOR iterative methods appear to be closely related, the analyses of the two methods are different. Determination of the optimal SOR iteration parameter comes from a specific relationship between the eigenvalues of the SOR and Jacobi iteration matrices [34,37], whereas (for the natural ordering) there is no such result for SSOR. Similarly, the analyses of relaxation methods say little about the incomplete factorizations, and conversely. (Cf. however the variable-ω generalized SSOR technique, which is equivalent to the MILU factorization [2].) Moreover, the performance of CG depends on both the extreme eigenvalues and the *distribution* of eigenvalues [3]; in general, the analyses of preconditioners provides virtually no information concerning the latter issue. Because of this, subtleties of behavior of preconditioned CG are not understood.

A heuristic explanation for why these analyses must be difficult is that the coefficient matrix A and splitting operator Q typically do not share a common set of eigenvectors, which makes it difficult to analyze the spectrum of $Q^{-1}A$ or of $Q^{-1}R$. This phenomenon can be seen by considering the classical stationary methods. An orthogonal set of eigenvectors for the discrete Dirichlet operator A on an $n \times n$ grid consists of the n^2 vectors $\{v^{(s,t)} \mid 1 \leq s,t, \leq n\}$, whose $((k-1)n+j)$'th component is

$$v_{jk}^{(s,t)} = \sin \frac{s\pi j}{n+1} \sin \frac{t\pi k}{n+1}. \tag{1.5}$$

Thus, the spectral decomposition of A is a discrete finite Fourier sine series. As shown by Frankel [15], the Gauss-Seidel and SOR iteration matrices have eigenvectors $w^{(s,t)}$, where

$$w_{jk}^{(s,t)} = \lambda_{st}^{\frac{j+k}{2}} v_{jk}^{(s,t)}, \tag{1.6}$$

and λ_{st} is the corresponding eigenvalue of the iteration matrix.[2] Thus, the eigenvectors can still be expressed in terms of trigonometric (sine) functions (so that they bear some resemblance to a Fourier series), but they differ from the eigenvectors of A by a componentwise multiplicative factor. (The eigenvectors for the Jacobi method are the same as those of A.) As for preconditioners, their analyses avoid the consideration of eigenvectors entirely, and instead consist of case by case studies of the extreme eigenvalues of $Q^{-1}A$, for different Q. In this paper, we introduce a Fourier analysis for iterative methods and preconditioners applied to discretized elliptic partial differential equations, which has the property that for model problems all the operators under consideration share a common set of orthonormal eigenvectors. As a result, the methodology can be used in a uniform manner to study convergence properties of a broad collection of methods, and all eigenvalues can be determined essentially by inspection.

[2] This relationship between the eigenvectors of the Jacobi iteration matrix and those of Gauss-Seidel and SOR was also shown by Young in his thesis [36], in much more general form. In particular, Young's results (which never appeared in print) are not limited to constant coefficient operators or simple boundary conditions.

Fourier methods are a standard tool for the analysis of both differential equations and discrete solution methods for time dependent problems. A classic example is the von Neumann stability analysis. Consider for example the Cauchy problem for the heat equation (see Richtmyer and Morton [30], pp. 9ff):

$$\frac{\partial u}{\partial t}(x,t) = \frac{\partial^2 u}{\partial x^2}(x,t) \qquad -\infty \le x \le +\infty, \qquad u(x,0) \text{ given.}$$

Let $\{u_j^n\}$ be defined by some finite difference scheme. For the discrete solution to be of any use, it should be bounded (in time), since the continuous solution is damped out as t increases. To reflect this requirement, the (strict) von Neumann stability analysis procedure imposes the condition that each discrete Fourier component of the solution of the form

$$u_j^n = e^{im(j\Delta x)}\xi(m)^n \tag{2.1}$$

satisfies the condition $|\xi(m)| \le 1$. For example, if Euler's method

$$u_j^{n+1} = \frac{\Delta t}{(\Delta x)^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) + u_j^n \tag{2.2}$$

is used for the difference scheme, then substitution of (2.1) into (2.2) shows that (2.1) is a solution when the amplification factor $\xi(m)$ satisfies

$$\xi(m) = \frac{\Delta t}{(\Delta x)^2}(e^{im\Delta x} - 2 + e^{-im\Delta x}) + 1. \tag{2.3}$$

That is, $\xi(m) = 1 - 4\frac{\Delta t}{(\Delta x)^2}\sin^2(\frac{m\Delta x}{2})$, and the von Neumann requirement is satisfied for all m provided $\frac{\Delta t}{(\Delta x)^2} \le \frac{1}{2}$. Von Neumann analysis has been applied extensively and successfully as a general guideline in many applications; see e.g. the comments by Roache [31], pp. 50ff. The Fourier approach has also been used to construct discrete schemes for solving PDEs with spectral methods [18], which compute solutions in the form of a finite linear combination of discrete Fourier components of the form (2.1).

Fourier methodology is also a standard analytic tool for elliptic problems, see e.g. Weinberger [35] ch. IV for an elementary treatment. For example, the eigenfunctions for the Laplace operator on the unit square with homogeneous Dirichlet boundary conditions are

$$v^{(s,t)} = \sin s\pi x \ \sin t\pi y, \qquad s, t = 1, 2, \ldots,$$

of which the vectors (1.5) are the discrete analogues. More generally, the Laplace equation $\Delta u = 0$ on a rectangle with Dirichlet boundary conditions $u = g$ has a uniformly convergent Fourier series solution when g has a uniformly convergent Fourier series on each side of the rectangle. In a recent analysis, Bube and Strikwerda [5] use Fourier transforms of difference operators to derive regularity estimates of their solutions, which in turn can be used to analyze the convergence of the discrete solution. Fourier analysis has also been used in the development of numerical methods for elliptic problems. For finite difference methods, discrete Fourier series such as (1.5) are exploited by one type of so-called "fast

direct" elliptic equation solvers, whose efficiency derives from the fast Fourier transform, see e.g. Swarztrauber [33]. These techniques can also be used as preconditioners for iterative methods for solving nonseparable problems, resulting in convergence rates that are asymptotically bounded independent of mesh, e.g. Concus and Golub [8], Elman and Schultz [12].

There has been some use of Fourier methods for the analysis of numerical methods for discrete elliptic problems. The most notable example is Brandt's "local mode analysis" for use with multigrid methods [4], where a heuristic analysis is used to demonstrate that the Gauss-Seidel iteration reduces high frequency errors rapidly. Kettler [24] uses a similar approach to study PCG as a multigrid smoother, and Jameson [22] analyzes a modified Runge-Kutta marching scheme used as a smoother in a multigrid algorithm for transonic flow problems. Other examples include an analysis recently used for "local" (i.e. variable ω) relaxation schemes by Kuo et. al. [25,26]; and an analysis of the stability of (complex) factorizations of the discrete Laplacian by Liniger [28]. All of these techniques ignore the effects of boundary conditions, but the behavior predicted by them agrees with numerical experiments.

In general, though, Fourier analysis has not been a popular tool for studying numerical methods for elliptic problems. We suspect that it never caught on for stationary methods because it is not rigorously applicable except for constant coefficient operators with periodic boundary conditions, whereas the classical analysis does a thorough job of explaining their performance for general problems. (Cf. Section 5 for some exceptions to the restriction on boundary conditions. See also [20] for a perturbation analysis relating the one-dimensional periodic and Dirichlet problems, and [1,27] for generalizations of the classical SOR analysis.) They have rarely been applied to preconditioners because, in addition to the restrictions on boundary conditions, the preconditioning matrices do not look like constant coefficient (i.e. constant diagonal) operators even when the continuous problem has constant coefficients. The only exception seems to be [24]. However, there only the smoothing rate is needed, which is governed by the convergence rate of the middle frequency of the error, arguably less sensitive to the effect of boundary conditions. Our results show that the Fourier approach works even for predicting the behaviour of the low frequencies.

In the present work, we examine the model problem (1.1) with periodic boundary conditions, and define a discrete approximation and splittings (1.3) by analogy with operators for other boundary conditions. These matrices all share the same set of orthogonal eigenvectors, and it is easy to examine spectral radii and condition numbers. Although the analysis is only exact for periodic boundary conditions, there is a strong correspondence with results for other boundary conditions. In particular, the orders of magnitude of asymptotic convergence rates for the Dirichlet problem are reproduced exactly by the periodic analysis. Moreover, the Fourier methodology provides insights into subtleties of behavior of methods, especially preconditioning techniques, not available from existing analysis. Thus, our analysis can be used like the von Neumann analysis as a practical tool to help determine whether or not a method is effective.

In Section 2, we present the periodic model problem and outline the methodology that will be used throughout the paper. In Section 3, we show how this methodology can be

applied to the Jacobi, Gauss-Seidel, SOR and SSOR stationary iterative methods, and in Section 4, we consider the ILU, MILU, SSOR and alternating direction DKR approximate factorization preconditioners [6]. In particular, we show that the standard results for both stationary methods [37] and incomplete factorizations [6,7,11,19] are reproduced essentially verbatim by the Fourier analysis. Finally, in Section 5, we present a heuristic analysis and experimental evidence demonstrating that the Fourier results can provide information and insights into methods for the model problem with other boundary conditions.

2. Framework of Analysis

Consider the Poisson equation (1.1) on the unit square, with periodic boundary conditions

$$u(x,0) = u(x,1), \qquad u(0,y) = u(1,y). \tag{2.4}$$

The eigenfunctions of the Laplacian with these boundary conditions are

$$u(x,y) = e^{ix2\pi s} e^{iy2\pi t}, \tag{2.5}$$

where s and t are integers, and the corresponding eigenvalues are

$$(2\pi s)^2 + (2\pi t)^2. \tag{2.6}$$

Discretizing this problem by centered finite differences on a uniform $(n+1) \times (n+1)$ grid gives rise to a system of linear equations

$$Au = b \tag{2.7}$$

of order $N = (n+1)^2$. It is convenient to represent vectors u of order N as a doubly indexed array $\{u_{jk}\}$, $0 \le j,k \le n$. Alternatively, u is a function defined on the mesh points $\{(jh,kh) | 0 \le j,k \le n\}$, with $u_{jk} = u(jh,kh)$, where $h = 1/(n+1)$. If the difference operators are scaled by h^2, then the equation of (2.7) corresponding to the (j,k) grid point is

$$4u_{jk} - u_{j+1,k} - u_{j-1,k} - u_{j,k+1} - u_{j,k-1} = b_{jk}, \tag{2.8}$$

where $b_{jk} = h^2 f(jh,kh)$. Because of the periodic boundary conditions, the indexing of (2.8) is performed in mod $n+1$ arithmetic so that $u_{n+1,k} = u_{0k}$ and $u_{j0} = u_{j,n+1}$. The coefficient matrix A has the form

$$\begin{pmatrix} P & B & & & B \\ B & P & B & & \\ & & \ddots & & \\ & & & & B \\ B & & & B & P \end{pmatrix},$$

where

$$P = \begin{pmatrix} 4 & -1 & & & -1 \\ -1 & 4 & -1 & & \\ & & \ddots & & -1 \\ -1 & & -1 & & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & & & \\ & -1 & & \\ & & \ddots & \\ & & & -1 \end{pmatrix}.$$

Consider the vector $u^{(s,t)}$ defined by

$$u_{jk}^{(s,t)} = e^{i\frac{j}{n+1}2\pi s} e^{i\frac{k}{n+1}2\pi t} = e^{ij\theta_s} e^{ik\phi_t}, \tag{2.9}$$

where

$$\theta_s = \frac{2\pi s}{n+1}, \quad \phi_t = \frac{2\pi t}{n+1} \qquad 0 \le s,t \le n. \tag{2.10}$$

This is an eigenvector of A analogous to the eigenfunction (2.5), and for integers $0 \le s,t \le n$, $\{u^{(s,t)}\}$ comprises a set of orthogonal eigenvectors for A that span $\mathbf{C}^N$. After substitution of (2.9) into the recurrence on the left side of (2.8), a straightforward computation shows that

$$Au^{(s,t)} = \lambda u^{(s,t)},$$

where

$$\lambda = \lambda_{st} \;=\; 4 - 2\cos\theta_s - 2\cos\phi_t \;=\; 4\Big(\sin^2\frac{\theta_s}{2} + \sin^2\frac{\phi_t}{2}\Big), \tag{2.13}$$

i.e. λ_{st} is the eigenvalue corresponding to eigenvector $u^{(s,t)}$. (To see more clearly the relationship between the eigenvalues (2.6) and (2.13), note that the eigenvalues of $\frac{1}{h^2}A$ are $\frac{4}{h^2}(\sin^2(\theta_s/2) + \sin^2(\phi_s/2))$, which for small s and t are approximately equal to the continuous eigenvalues (2.6).) Note that eigenvalues for the eight indices

$$\begin{array}{llll} (s,t), & (t,s), & (s,n+1-t), & (n+1-t,s), \\ (n+1-s,t), & (n+1-t,s), & (n+1-s,n+1-t), & (n+1-t,n+1-s), \end{array}$$

are all equal, so that most eigenvalues of A are of multiplicity eight.

We will define splittings of the form (1.3) for the periodic problem (2.7) by analogy with versions for the Dirichlet problem. All the splitting operators Q can be described in terms of computational molecules on the underlying grid. For example, for the Gauss-Seidel and SOR iterative methods, Q is given by a matrix L in which the row corresponding to the (j,k) grid point has nonzero entries in the columns corresponding to the $(j,k),(j-1,k)$ and $(j,k-1)$ points. Similarly, for the ILU and MILU incomplete factorizations, Q has the form LU in which L has the nonzero structure just described and the row of U for the (j,k) grid point contains nonzeros in the columns corresponding to (j,k), $(j+1,k)$ and $(j,k+1)$. The computational molecules for A, L and U are shown in Figure 1.

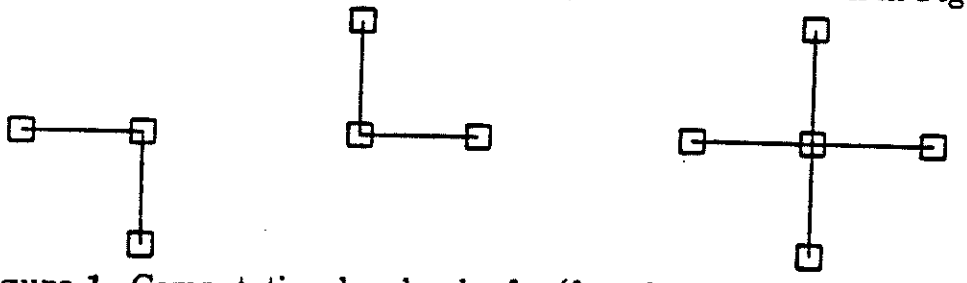

Figure 1. Computational molecules for (from left to right) A, L, and U.

As with A, indexing is performed in mod $n+1$ arithmetic. As a result, L (i.e. each version of L) is not a lower triangular matrix, but instead is a block matrix of (block)

order $n+1$, with nonzero structure

$$\begin{pmatrix} X & & & Y \\ Y & X & & \\ & \ddots & & \\ & & Y & X \end{pmatrix}. \tag{2.14}$$

X and Y are of order $n+1$, X has nonzero structure

$$\begin{pmatrix} \times & & & \times \\ \times & \times & & \\ & \ddots & & \\ & & \times & \times \end{pmatrix}, \tag{2.10}$$

and Y is a diagonal matrix. (Here "$\times$" denotes a nonzero entry.) That is, L is block lower triangular except for a nonzero diagonal block in the upper right corner, and the diagonal blocks of L are lower triangular except for a nonzero in the upper right corner. U has the same nonzero structure as L^T. In the following, we use the word "method" loosely in conjunction with these splittings, although in reality the splittings are designed only as analytic tools, and not as the basis for numerical methods for solving (1.1) and (2.1). Indeed, since L and U are not triangular, applying the actions of their inverses is more complicated than for other boundary conditions.

We remark that there is a close relationship between the spectrum of the discrete periodic problem (2.4) and that of discretized problems with other boundary conditions, such as Dirichlet and mixed Dirichlet-Neumann conditions. Consider the Dirichlet case, in which the boundary conditions (2.1) are replaced by

$$u(x,y) = g(x,y), \quad (x,y) \in \partial\Omega. \tag{2.11}$$

Discretizing on a uniform grid with n interior points in each direction results in a linear system of order n^2 whose eigenvalues are [11]

$$\tilde{\lambda}_{st} = 4\Big(\sin^2 \frac{\tilde{\theta}_s}{2} + \sin^2 \frac{\tilde{\phi}_s}{2}\Big), \tag{2.12}$$

where

$$\tilde{\theta}_s = \frac{\pi s}{n+1}, \quad \tilde{\phi}_t = \frac{\pi t}{n+1}, \qquad 1 \le s,t \le n.$$

The corresponding eigenvectors are $\tilde{u}_{jk}^{(s,t)} = \sin(j\tilde{\theta}_s)\sin(k\tilde{\phi}_t)$. Most of the eigenvalues are of multiplicity two, and comparison with (2.8) shows that roughly one fourth of the eigenvalues of the Dirichlet problem are also eigenvalues of the periodic problem. Equivalently, on a given mesh of width h the Dirichlet problem admits eigenvalues with roughly twice as many Fourier modes in each component as does the periodic problem. The smallest modes for the Dirichlet and periodic problems are πh and $2\pi h$, respectively (see [19]).

These observations will be used as the basis of a heuristic analysis connecting results for the periodic problem to problems with Dirichlet boundary conditions (see Section 5).[3]

Finally, observe that for the continuous periodic problem (1.1)/(2.1) the eigenpair (1.1) in the case $s = t = 0$ is $\lambda = 0$, $u = 1$. Hence, the problem (1.1)/(2.1) is not well posed: if v is a solution, then $v + c$ is also a solution for any constant c. Similarly, the discrete eigenpair (2.8) for $s = t = 0$ is $\lambda = 0$, $u \equiv 1$, so that A is singular. (All other eigenvalues are nonzero, so that A has rank $N-1$.) In addition, in some cases the splitting matrix Q is also singular (see Section 4). Consequently, it is not meaningful to talk about Q^{-1}, $Q^{-1}R$, or the condition number of $Q^{-1}A$. In the Fourier analysis below, we will restrict our attention to the nonzero modes in each component of A and Q, i.e. to the cases $1 \le s, t \le n$. These modes are analogues of the lowest modes for the Dirichlet problem. Thus, the smallest nonzero eigenvalue of A that we will consider is $\lambda_{min} = 8\sin^2(\pi h) \approx 8\pi^2 h^2$, and the largest eigenvalue (for $\theta_s = \phi_t \approx \pi$) is $\lambda_{max} \approx 8$. We will define analogues of spectral radii and condition numbers in terms of these restricted sets of eigenvalues and eigenvectors. Note that the extreme periodic eigenvalues (and therefore the conditioning with respect to the restricted eigenvalue set) for a mesh of width $\frac{h}{2}$ are the same as those of the Dirichlet problem for mesh width h. Below, we will use this correspondence to apply our results to iterative methods for the Dirichlet problem.

3. Stationary Iterative Methods

In this section, we define the splittings for the periodic problem that correspond to the Jacobi, Gauss-Seidel, SOR and SSOR stationary methods for (2.7) and we perform a Fourier convergence analysis of each of them. Let $\{\psi_{st}\}_{s,t=0}^{n}$ denote the eigenvalues of the splitting operator Q and let $\{\mu_{st}\}_{s,t=0}^{n}$ denote the eigenvalues of R. As we will show below, each of the splitting operators is nonsingular and Q, R, and A all share the same set of orthonormal eigenvectors. Hence, the eigenvalues of $Q^{-1}R$ are $\{\psi_{st}/\mu_{st}\}_{s,t=0}^{n}$, and the spectral radius of $Q^{-1}R$ (with respect to the restricted set of modes) is

$$\rho = \max_{1 \le s,t \le n} \mu_{st}/\psi_{st}. \tag{3.1}$$

We must determine the value of (3.1) for each of the splittings.

Let

$$A = D - (L + L^T), \tag{3.2}$$

where $D = diag(A)$ and $D - L$ has the nonzero structure (2.14) - (2.15). Then the Jacobi splitting is given by $Q = D$, $R = L + L^T$. Applying these operators to the eigenvector $u^{(s,t)}$ of (2.9) gives

$$Du^{(s,t)} = 4u^{(s,t)}, \quad Lu^{(s,t)} = (e^{-i\theta_s} + e^{-i\phi_t})u^{(s,t)}, \quad L^T u^{(s,t)} = (e^{i\theta_s} + e^{i\phi_t})u^{(s,t)}.$$

[3] Other boundary conditions give rise to similar properties. For example, if (2.11) is replaced by the pure Neumann condition $u_x = 0$ on the vertical boundaries $x = 0$, $x = 1$, and first order differencing is used for u_x, then the eigenvalues are as in (2.12) except $s = 0$ and $s = n + 1$ are also included, and the eigenvectors are $\cos(j\tilde{\theta}_s)\sin(k\tilde{\phi}_t)$. See [35] for further discussion along these lines.

Hence, the eigenvalues of the Jacobi iteration matrix are

$$\eta_{st} = \frac{e^{i\theta_s} + e^{-i\theta_s} + e^{i\phi_t} + e^{-i\phi_t}}{4} = \frac{\cos(\theta_s) + \cos(\phi_t)}{2}. \tag{3.3}$$

The largest values occur when $s = t = 1$, so that the spectral radius is

$$\rho_J = \cos(2\pi/(n+1)) = \cos(2\pi h).$$

Note that the eigenvalues of D, L and L^T can be be identified by simply examining the computational molecules and the corresponding (constant) matrix entries for each of these operators. For example, at any (j,k) mesh point, L uses the neighboring $(j-1,k)$ and $(j,k-1)$ points, with corresponding matrix values equal to one, so the resulting eigenvalue is $(1\times)(e^{-i\theta} + e^{-i\phi})$. This technique of determining eigenvalues by inspection applies to all of the operators of this paper (including A). It is analogous to the determination of the amplification factors (or symbols) for the von Neumann stability analysis from the difference scheme, e.g. compare the coefficients and the subscripts, respectively, of (2.2) with the coefficients and exponent signs of (2.3).

For the Gauss-Seidel splitting, $Q = D - L$ and $R = L^T$. Consequently, the eigenvalues of the Gauss-Seidel iteration matrix are

$$\gamma_{st} = \frac{e^{i\theta_s} + e^{i\phi_t}}{4 - (e^{-i\theta_s} + e^{-i\phi_t})}.$$

A straightforward algebraic manipulation gives

$$|\gamma_{st}|^2 = \frac{1 + \cos(\theta_s - \phi_t)}{8 - 4(\cos(\theta_s) + \cos(\phi_t)) + 1 + \cos(\theta_s - \phi_t)}.$$

The maximum value occurs at $s = t = 1$, which results in a spectral radius of

$$\rho_{GS} = \frac{1}{\sqrt{1 + 8\sin^2(\pi h)}}.$$

The SOR splitting is defined by

$$Q = \frac{1}{\omega}D - L, \qquad R = \frac{(1-\omega)}{\omega}D + L^T,$$

where $\omega > 0$ is the relaxation parameter. Therefore, the eigenvalues of the SOR iteration matrix are

$$\sigma_{st} = \frac{4(1-\omega) + \omega(e^{i\theta_s} + e^{i\phi_t})}{4 - \omega(e^{-i\theta_s} + e^{-i\phi_t})}.$$

Then

$$|\sigma_{st}|^2 = \frac{(1-\omega) + \frac{\omega^2}{8}(1 + \cos(\theta_s - \phi_t)) - \omega(1-\omega)(\sin^2(\theta_s/2) + \sin^2(\phi_t/2))}{(1-\omega) + \frac{\omega^2}{8}(1 + \cos(\theta_s - \phi_t)) + \omega(\sin^2(\theta_s/2) + \sin^2(\phi_t/2))}. \tag{3.4}$$

Note that (3.4) has the form

$$|\sigma_{st}|^2 = \frac{g(\theta_s, \phi_t) + \omega(\omega - 1) f(\theta_s, \phi_t)}{g(\theta_s, \phi_t) + \omega f(\theta_s, \phi_t)},$$

where $f > 0$ and both the numerator and denominator are positive. If $0 < \omega < 2$, then $|\omega(1-\omega)| < \omega$, which implies that $\omega f > |\omega(\omega-1)|f$. Therefore, the SOR spectral radius ρ_{SOR} is less than 1. This is essentially Kahan's result [23], and we now restrict our attention to ω in this range. It is easy to verify that the spectral radius occurs at $s = t = 1$. Therefore,

$$\rho_{\mathrm{SOR}}(\omega)^2 = \frac{(\omega-1)^2 - 8\omega(1-\omega)\sin^2(\pi h)}{(\omega-1)^2 + 8\omega\sin^2(\pi h)}.$$

Differentiating this expression with respect to ω, we find that the value of ω that minimizes $\rho_{\mathrm{SOR}}(\omega)$ is

$$\omega_* = \frac{2}{1 + 2\sin(\pi h)}.$$

Substituting ω_* into $\rho_{\mathrm{SOR}}(\omega)$ gives

$$\rho_{\mathrm{SOR}}(\omega^*) = \sqrt{\frac{1 - \sin(\pi h)}{1 + \sin(\pi h)}}.$$

Finally, for the SSOR stationary splitting,

$$Q = \frac{1}{\omega(2-\omega)}(D - \omega L)D^{-1}(D - \omega L^T),$$

and

$$R = \frac{1}{\omega(2-\omega)}((\omega-1)D - \omega L)D^{-1}((\omega-1)D - \omega L^T).$$

Therefore, the SSOR eigenvalues are

$$\tau_{st} = \left(\frac{4(1-\omega) + \omega(e^{i\theta_s} + e^{i\phi_t})}{4 - \omega(e^{-i\theta_s} + e^{-i\phi_t})}\right)\left(\frac{4(1-\omega) + \omega(e^{-i\theta_s} + e^{-i\phi_t})}{4 - \omega(e^{i\theta_s} + e^{i\phi_t})}\right). \tag{3.5}$$

The two factors on the right hand side of (3.5) are complex conjugates of one another, and the first factor is exactly σ_{st} of (3.4). Therefore $|\tau_{st}| = |\sigma_{st}|^2$, and the same arguments as for SOR give

1. If $0 < \omega < 2$, then $\rho_{\mathrm{SSOR}}(\omega) < 1$.
2. The optimal value of ω for SSOR is $\omega^* = \frac{2}{1+2\sin(\pi h)}$.
3. The minimum spectral radius for SSOR is $\rho_{\mathrm{SSOR}}(\omega^*) = \frac{1-\sin(\pi h)}{1+\sin(\pi h)}$.

	Fourier			Classical		
Method	ω^*	$\rho(\omega^*)$	$R(\omega^*)$	ω^*	$\rho(\omega^*)$	$R(\omega^*)$
Jacobi	—	$\cos(\pi h)$	$\frac{\pi^2}{2}h^2$	—	$\cos(\pi h)$	$\frac{\pi^2}{2}h^2$
Gauss-Seidel	—	$\frac{1}{\sqrt{1+8\sin^2(\frac{\pi h}{2})}}$	$\pi^2 h^2$	—	$\cos^2(\pi h)$	$\pi^2 h^2$
SOR	$\frac{2}{1+2\sin(\frac{\pi h}{2})}$	$\sqrt{\frac{1-\sin(\frac{\pi h}{2})}{1+\sin(\frac{\pi h}{2})}}$	$\frac{\pi}{2}h$	$\frac{2}{1+\sin(\pi h)}$	$\frac{1-\sin(\pi h)}{1+\sin(\pi h)}$	$2\pi h$
SSOR	$\frac{2}{1+2\sin(\frac{\pi h}{2})}$	$\frac{1-\sin(\frac{\pi h}{2})}{1+\sin(\frac{\pi h}{2})}$	πh	$\frac{2}{1+2\sin(\frac{\pi h}{2})}$	$\frac{1-\sin(\frac{\pi h}{2})}{1+\sin(\frac{\pi h}{2})}$	πh

Table 2: Comparison of Fourier results and classical results for stationary methods.

We collect the results from the Fourier analysis of stationary methods in Table 2. We also include the known results for the Dirichlet problem. From our observations at the end of Section 2, there is a correspondence between the spectrum of the discrete periodic problem with mesh size $\frac{h}{2}$ and that of the discrete Dirichlet problem with mesh size h. Thus, in the table we show the Fourier results for a mesh size $\frac{h}{2}$. Comparison of the Fourier and classical results reveals an extraordinary agreement. The Fourier and classical results for the Jacobi and SSOR methods agree exactly, as do the asymptotic convergence rates for the Gauss-Seidel method. For the SOR splitting, the values of ω^* agree asymptotically, and so do the exponents of h in the asymptotic rate of convergence. The only disagreement is a factor of four in the coefficients of h in the asymptotic convergence rates.

4. Preconditioners

In this section, we define the ILU, MILU, SSOR and ADDKR incomplete factorizations for (2.7) and perform a spectral analysis of the preconditioned systems. The standard analysis of preconditioners examines the condition number of the preconditioned operator $Q^{-1/2}AQ^{-1/2}$, i.e. the ratio of the maximum and minimum eigenvalues of $Q^{-1}A$. The condition number can be derived from upper and lower bounds on the Rayleigh quotient

$$\frac{(v, Q^{-1/2}AQ^{-1/2}v)}{(v,v)} \tag{4.1}$$

(see [7,11,19]). The eigenvectors of A with unit norm are

$$v^{(s,t)} = h^2 u^{(s,t)},$$

where $u^{(s,t)}$ is given by (2.9). Let V denote the orthonormal matrix whose columns are $v^{(s,t)}$, and let Λ denote the diagonal matrix containing the corresponding eigenvalues (2.13). Again, as we show below, the eigenvectors of A are also eigenvectors of each preconditioner Q that we are studying. If Ψ denotes the diagonal matrix of eigenvalues of Q, then

$$A = V\Lambda V^*, \qquad Q = V\Psi V^*. \tag{4.2}$$

For nonsingular Q, it is easily verified from (4.2) that the extreme values of (4.1) are given by the minimum and maximum values of $\mu_{st} = \lambda_{st}/\psi_{st}$. We refer to the quantities μ_{st} as the "preconditioned eigenvalues," and we define the "restricted" condition number $\kappa^{(Q)}$ for the preconditioned problem to be the ratio μ_{max}/μ_{min}, where μ_{max} and μ_{min} are the extreme eigenvalues for the set of restricted modes,

$$\mu_{max} = \max_{1 \le s,t \le n} \lambda_{st}/\psi_{st}, \qquad \mu_{min} = \min_{1 \le s,t \le n} \lambda_{st}/\psi_{st}.$$

These expressions and $\kappa^{(Q)}$ are well-defined for both singular and nonsingular Q.

4.1. The ILU Factorization

The ILU factorization is defined to be the product $Q = LU$, where L has the nonzero structure (2.14) - (2.15) and U has the structure of L^T, such that the entries of Q have the same values as those of A wherever A is nonzero. For this and all other incomplete factorizations, we use the convention that U has unit diagonal. By formally multiplying the factors and matching the entries of Q and A, we find that the defining condition is imposed by choosing the nonzero off-diagonal entries of L to be equal to the corresponding entries of A and the nonzero entries of U to be equal to the corresponding entries of A premultiplied by the inverse of the diagonal of L. For the discrete Laplacian, the off-diagonals of L are identically -1 and those of U in the (j,k) row are $-1/\alpha_{jk}$. The diagonal entries must then satisfy (in mod $n+1$) arithmetic

$$\alpha_{jk} = 4 - 1/\alpha_{j-1,k} - 1/\alpha_{j,k-1}, \qquad 0 \le j,k \le n.$$

These equations are satisfied by $\alpha_{jk} \equiv \alpha = 2 + \sqrt{2}$. Hence, the ILU factors are constant coefficient matrices. They are also strictly diagonally dominant, so that (in contrast with A), L, U and Q are nonsingular.

The preconditioning matrix Q is equal to $A + R$, where

$$R = \begin{pmatrix} Z & E^T & & & E \\ E & Z & E^T & & \\ & & \ddots & & \\ & & & E^T & \\ E^T & & & E & Z \end{pmatrix},$$

Z is a matrix of all zeros,

$$E = \begin{pmatrix} 0 & \eta & & & \\ & 0 & \eta & & \\ & & \ddots & & \\ & & & \eta & \\ \eta & & & & 0 \end{pmatrix} \tag{4.3}$$

and $\eta = 1/\alpha = 1/(2+\sqrt{2})$. For any eigenvector $u = u^{(s,t)}$ of A given by (2.9),

$$\begin{aligned} [Ru]_{jk} &= \frac{1}{2+\sqrt{2}}(u_{j-1,k+1} + u_{j+1,k-1}) = \frac{1}{2+\sqrt{2}} e^{ij\theta} e^{ik\phi} (e^{i(\theta-\phi)} + e^{-i(\theta-\phi)}) \\ &= \frac{2}{2+\sqrt{2}} \cos(\theta - \phi)\, u_{jk}. \end{aligned}$$

As a result,

$$Qu^{(s,t)} = \psi_{st} u^{(s,t)},$$

where

$$\psi_{st} = 4\Big(\sin^2 \frac{\theta_s}{2} + \sin^2 \frac{\phi_t}{2}\Big) + \frac{2}{2+\sqrt{2}} \cos(\theta_s - \phi_t). \tag{4.4}$$

Thus, A, R and Q all share the same set of orthonormal eigenvectors. (As observed in Section 3, these eigenvalues could also be determined by directly applying L and U as difference operators to $u^{(s,t)}$, or by simply ascertaining the symbols of L and U in terms of their effect on the computational molecule. This approach shows that L and U also have the same eigenvectors.)

The condition number $\kappa^{(I)}$ for the ILU preconditioning is the ratio of the maximum and minimum nonzero values of

$$\mu^{(I)} = \frac{\lambda_{st}}{\psi_{st}} = \frac{4\,(\sin^2(\theta_s/2) + \sin^2(\phi_t/2))}{4\,(\sin^2(\theta_s/2) + \sin^2(\phi_t/2)) + \frac{2}{2+\sqrt{2}} \cos(\theta_s - \phi_t)} \tag{4.5}$$

where θ_s and ϕ_t are as in (2.10) with $s, t \geq 1$. The following result gives an asymptotic bound for $\kappa^{(I)}$.

Theorem 4.1. For the ILU preconditioned operator, $\kappa^{(I)} = O(h^{-2})$.

We defer a proof to the Appendix. This result coincides with the analogous asymptotic bound for the condition number of the ILU preconditioned Dirichlet operator [7,19].

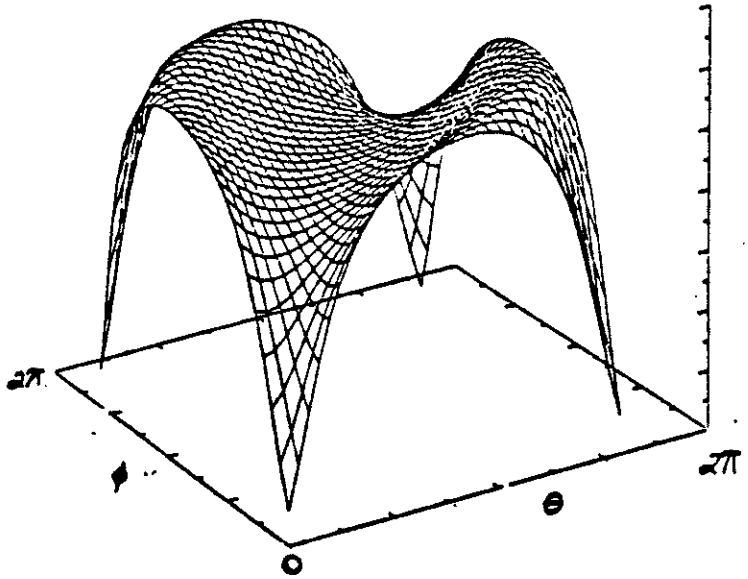

Figure 2. Surface plot of ILU preconditioned eigenvalues, $n = 30$.

In addition to providing condition numbers, the formula (4.5) gives a clear picture of the distribution of eigenvalues and the effect of the ILU preconditioner. The denominators ψ_{st} of (4.5), i.e. the eigenvalues of the ILU preconditioning matrix Q, are all $O(1)$. Hence, the extreme ILU preconditioned eigenvalues correspond precisely to the extreme modes of the original matrix. These are the smallest ones, occurring wherever $\sin^2(\theta_s/2)$ and $\sin^2(\phi_t/2)$ is small, i.e. at the four corners of the box $\{0 < \theta_s, \phi_t < 2\pi\}$. Away from these

corners, all the eigenvalues of Q and A are of order one, so the preconditioned operator is well-behaved. A surface plot of (4.5) (for $n = 30$) that confirms this is shown in Figure 2.

4.2. The MILU Factorization

The MILU factorization is defined so that the entries of Q have the same values as those of A for all off-diagonal indices at which A is nonzero, and the sum of the entries of each row of the error matrix $R = Q - A$ equals ch^2, where c is a nonnegative constant that is independent of h. These conditions are imposed by choosing the off-diagonal entries of L and U to be the same as in the ILU factorization, and the diagonal values of L to satisfy

$$\alpha_{jk} = 4 + ch^2 - 1/\alpha_{j-1,k} - 1/\alpha_{j,k-1} - 1/\alpha_{j-1,k+1} - 1/\alpha_{j+1,k-1}. \tag{4.6}$$

This expression is satisfied by

$$\alpha_{jk} \equiv \alpha = 2 + \frac{ch^2}{2} + \frac{1}{2}\sqrt{8ch^2 + (ch^2)^2} \tag{4.7}$$

If $c = 0$, then $\alpha = 2$. The MILU factors also have constant coefficients, they are diagonally dominant and hence nonsingular for $c > 0$, and for $c = 0$ they have a zero eigenvalue with eigenvector equal to the constant vector.

The error matrix has the form

$$R = \begin{pmatrix} D & E^T & & & E \\ E & D & E^T & & \\ & & \ddots & & \\ & & & & E^T \\ E^T & & & E & D \end{pmatrix},$$

where $D = diag(-2/\alpha + ch^2)$, E has the form (4.3), $\eta = 1/\alpha$ and α is defined by (4.7). The eigenvalues of Q corresponding to the eigenvectors $u^{(s,t)}$ are

$$\psi_{st} = 4\left(\sin^2\frac{\theta_s}{2} + \sin^2\frac{\phi_t}{2}\right) + \frac{2}{\alpha}\cos(\theta_s - \phi_t) - \frac{2}{\alpha} + ch^2.$$

If $c = 0$, then $\alpha = 2$. In this case, for $s = t = 0$, $\psi_{00} = 0$ and Q is singular. The restricted condition number $\kappa^{(M)}$ for the MILU preconditioning is the ratio of maximum and minimum values of

$$\mu^{(M)} = \frac{\lambda_{st}}{\psi_{st}} = \frac{4(\sin^2(\theta_s/2) + \sin^2(\phi_t/2))}{4(\sin^2(\theta_s/2) + \sin^2(\phi_t/2)) + (2/\alpha)(\cos(\theta_s - \phi_t) - 1) + ch^2}, \tag{4.8}$$

for $1 \le s, t \le n$.

The following result gives a bound for $\kappa^{(M)}$.

Theorem 4.2. For the MILU preconditioned operator, if $c > 0$ then $\kappa^{(M)} = O(h^{-1})$, and if $c = 0$, then $\kappa^{(M)} = O(h^{-2})$.

See the Appendix for a proof. The analysis of the MILU preconditioned Dirichlet operator gives the same result for $c > 0$. We know of no theoretical result in the Dirichlet case for $c = 0$, although we have observed empirically that the condition number behaves like $O(h^{-1})$ there also. We will comment on this difference between the periodic and (observed) Dirichlet results in Section 5.

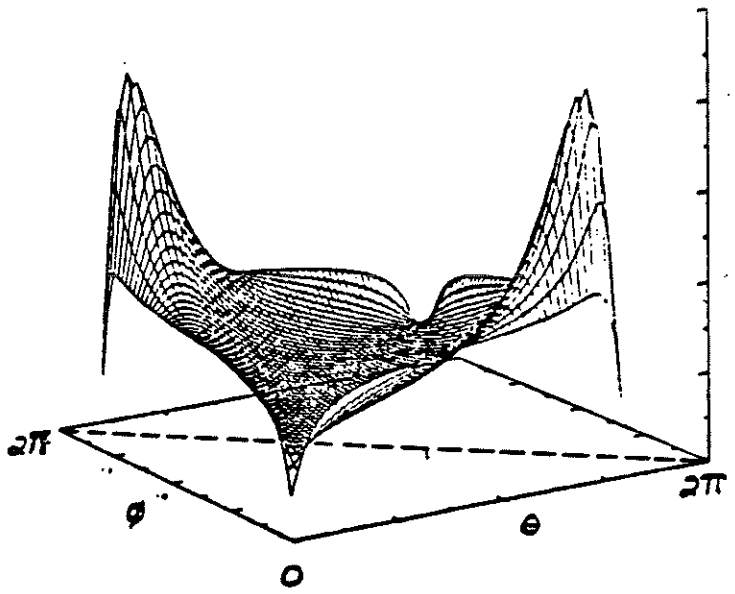

Figure 3. Surface plot of MILU preconditioned eigenvalues, $n = 50$, $c = 80$.

As in the case of the ILU preconditioner, the MILU formula (4.8) shows clearly the eigenvalue distributions and the effect of the MILU preconditioner. A surface plot of (4.8) for $n = 50$ and $c = 80$ is shown in Figure 3. Moreover, examination of this expression shows subtleties of behavior for the periodic problem not revealed by analyses of the Dirichlet problem. Let $\mathcal{A}$ denote the box $\{(\theta, \phi) | 0 \leq \theta, \phi \leq 2\pi\}$ in two-dimensional (θ, ϕ)-space, and let $\mathcal{S} = \{(\theta, \phi) | \phi = 2\pi - \theta\}$, the transverse diagonal across $\mathcal{A}$. (In Figure 3, $\mathcal{A}$ is the square depicted in the horizontal plane and $\mathcal{S}$ is the dotted line. See also Figure A.1 in the Appendix.) Then the following observations hold; see the Appendix for elucidation.

1. The smallest eigenvalues of the MILU preconditioned operator are of order one, and the asymptotically extreme eigenvalues are the large ones, of order h^{-1} for $c > 0$ and of order h^{-2} for $c = 0$. Examples of extreme large eigenvalues occur near the endpoints of $\mathcal{S}$; for $c > 0$, they occur when $\theta_s \approx 2\sqrt{h}$ (i.e. $s \approx \sqrt{n+1}/\pi$), and for $c = 0$, when $\theta_s = O(h)$.
2. Indeed, the only extreme eigenvalues occur near the endpoints of $\mathcal{S}$. That is, if $\mathcal{D}$ is any domain containing the corners of $\mathcal{S}$, then for all θ_s and ϕ_t outside of $\mathcal{D}$, the eigenvalues corresponding to θ_s and ϕ_t are of asymptotic order one as $h \to 0$.
3. The effect on conditioning of the ch^2 term used in the definition of the MILU factorization can be clearly seen. When $c > 0$, the condition number is $O(h^{-1})$ and when $c = 0$ it is $O(h^{-2})$. The latter result differs from empirical observations for the Dirichlet problem (see Section 5), although it turns out to entail a delicate cancellation.

4.3. The SSOR Factorization

Let $A = D - (L + L^T)$ as in (3.2). Then the SSOR factorization is given by

$$Q = (D - \omega L)D^{-1}(D - \omega L^T),$$

where $\omega \in [0, 2]$ is a scalar. It is easily verified that the SSOR eigenvalues are

$$\psi_{st} = \left(4 - \omega + \frac{\omega^2}{2}\right) + 4\omega\left(\sin^2\frac{\theta_s}{2} + \sin^2\frac{\phi_t}{2}\right) + \frac{\omega^2}{2}\cos(\theta_s - \phi_t). \tag{4.9}$$

In principle one would like to choose ω to minimize the condition number $\kappa^{(S)}(\omega)$. Unfortunately, this turns out to involve rather complicated calculations. Instead we proceed in two stages. First, we determine the optimal value of ω that minimizes $\kappa^{(S)}(\omega)$ on the line $\mathcal{S}$ only. This one-dimensional problem is much more tractable and we find:

Lemma 4.1. Let $\mu_{st} = \lambda_{st}/\psi_{st}$ denote the SSOR preconditioned eigenvalues. Then

$$\min_{\omega\in[0,2]} \frac{\max_{(\theta_s,\phi_t)\in\mathcal{S}} \mu_{st}}{\min_{(\theta_s,\phi_t)\in\mathcal{S}} \mu_{st}} = O(h^{-1}),$$

with the optimal value of ω given by $\omega^* = 2/(1 + 2\sin(\pi h))$.

See the Appendix for a proof. For this choice of ω, we then have the following bound on the SSOR condition number.

Theorem 4.3. For the SSOR preconditioned operator, if $\omega = \omega^*$ then $\kappa^{(S)} = O(h^{-1})$.

Proof. The proof of this result is essentially the same as the proof of Theorem 4.2 for the MILU preconditioning (see the Appendix): for $\omega = \omega^*$, it can be shown that the SSOR preconditioned eigenvalues $\mu^{(S)}$ satisfy

$$\frac{1}{5} \le \mu^{(S)} \le \frac{(1+2s)^2}{16s + 8s^2} = O(h^{-1}),$$

where $s = \sin(\pi h/2) = O(h)$. We omit the details. Q.E.D.

This asymptotic bound on the condition number again coincides with the results for the Dirichlet problem. It also shows that for any ω, the SSOR preconditioned operator has condition number at least $O(h^{-1})$, since any other value gives a condition number at least that large on $\mathcal{S}$. In empirical observations of the SSOR-preconditioned periodic eigenvalues $\mu^{(S)} = \lambda_{st}/\psi_{st}$, we also find that the maximum and minimum values occur on $\mathcal{S}$, suggesting that ω^* is indeed the *true* optimal value. Finally, ω^* coincides with the near-optimal value $2/(1+2\sin(\pi h/2))$ for the Dirichlet problem [37], scaled to account for a factor of two difference in modes analogous to the difference between the Dirichlet and periodic Fourier modes.

4.4. The ADDKR Factorization

The previous results show that the Fourier analysis not only confirms the classical results for Dirichlet problems, but also reveals more details about the iterative methods,

such as the eigenvalue distribution. But perhaps the most powerful use of the Fourier approach is to use these new insights to design better methods. We will show such an example in this section.

The ADDKR incomplete factorization [6] combines a standard incomplete factorization with an incomplete factorization for a permuted version of A in which the order of the grid points in one direction (without loss of generality, the x-direction) is reversed. For motivation, consider the stationary method

$$u \leftarrow u + \tau Q_1^{-1}(b - Au).$$

where $Q_1 = L_1 U_1$ is some incomplete factorization, and τ is a scalar parameter. By convention, the unknowns of (2.7) are ordered in the natural order with horizontal lines ordered from left to right. Since the factorization has a preferred direction on the grid, this sweep does not annihilate errors uniformly on the grid. This phenomenon can be seen from Figure 3, where the eigenvalue μ is large ($O(h^{-1})$) near $(\theta, \phi) = (\theta_1, \phi_n)$ and small ($O(1)$) near $(\theta, \phi) = (\theta_1, \phi_1)$. It therefore seems natural to combine this method with one that complements this behaviour, i.e. has large eigenvalues near (θ_1, ϕ_1) and small ones near (θ_1, ϕ_n). It turns out this can be achieved by changing the direction of the ordering of the grid points in defining the preconditioner. Let P denote the permutation that reorders the horizontal lines from right to left, and let $\tilde{L}_2 \tilde{U}_2$ denote the incomplete factorization for PAP^T; equivalently, $Q_2 = L_2 U_2 = [P^T \tilde{L}_2 P][P^T \tilde{U}_2 P]$ is another approximate factorization for A. Then a second sweep

$$u \leftarrow u + \tau Q_2^{-1}(b - Au)$$

will tend to annihilate errors left over from the first one.

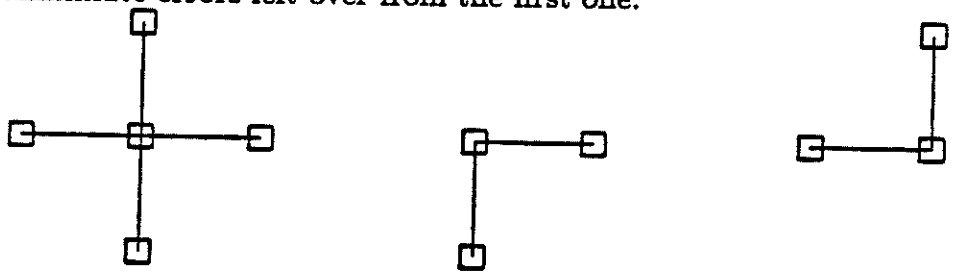

Figure 4. Computational molecules for (from left to right) A, L_2 and U_2.

L_1 and U_1 are defined exactly as in the MILU factoriation, except for the diagonal entries, where h^2 is replaced by h^p for some $p \in [0, 2]$. Hence, the diagonal values are given by

$$\alpha = 2 + \frac{ch^p}{2} + \frac{1}{2}\sqrt{8ch^p + (ch^p)^2}. \tag{4.10}$$

L_2 and U_2 are defined in an identical manner, except that their computational molecules are as in Figure 4. That is, L_2 has the form (2.14) but in which X has the form of the transpose of (2.15). Let $R_1 = Q_1 - A$ and $R_2 = Q_2 - A$. As shown in [6], the splitting operator

$$Q = Q_1(A + R_1 + R_2)^{-1} Q_2$$

corresponds to performing a sweep based on Q_1 followed by a sweep based on Q_2. The ADDKR preconditioning uses Q as the preconditioner for (2.7). We are interested in the

ratio of maximum and minimum eigenvalues of $Q^{-1}A$, although in this case, Q is not symmetric and this ratio is not the condition number.

Exactly as for the preconditioners discussed above, the eigenvalues of Q_1 are

$$\psi_{st}^{(1)} = 4\left(\sin^2(\theta_s/2) + \sin^2(\theta_t/2)\right) + (2/\alpha)(\cos(\theta_s - \phi_t) - 1) + ch^p,$$

and those of Q_2 are

$$\psi_{st}^{(2)} = 4\left(\sin^2(\theta_s/2) + \sin^2(\theta_t/2)\right) + (2/\alpha)(\cos(\theta_s + \phi_t) - 1) + ch^p,$$

where α is as in (4.10). Note that the only difference between these two expressions is in the sign of ϕ_t. The eigenvalues of $A + R_1 + R_2$ are

$$\psi_{st} = 4\left(\sin^2(\theta_s/2) + \sin^2(\theta_t/2)\right) + (2/\alpha)(\cos(\theta_s - \phi_t) + \cos(\theta_s + \phi_t) - 2) + 2ch^p.$$

Hence, the ADDKR preconditioned eigenvalues are

$$\mu_{st}^{(A)} = \frac{4\left(\sin^2(\theta_s/2) + \sin^2(\theta_t/2)\right)\psi_{st}}{\psi_{st}^{(1)}\psi_{st}^{(2)}}. \tag{4.11}$$

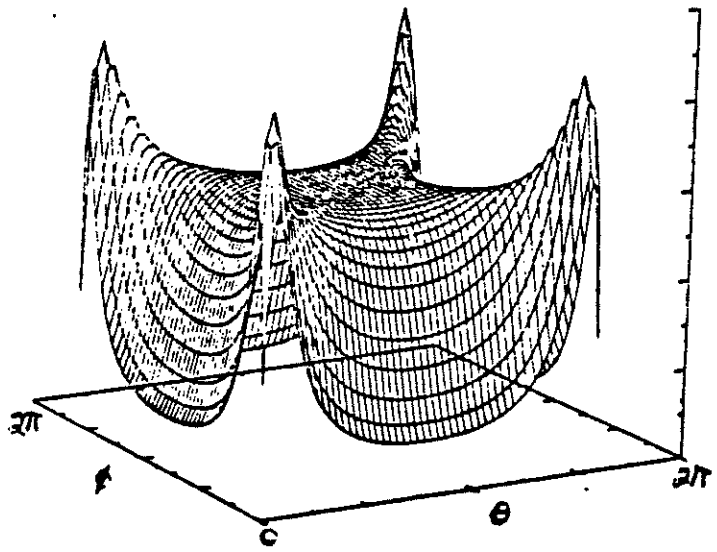

Figure 5. Surface plot of ADDKR preconditioned eigenvalues, $p = \frac{4}{3}$, $n = 50$, $c = 1$.

Using techniques analogous to those for the MILU preconditioning, we can show that the maximum eigenvalue of (4.11) is bounded above by $2 = O(1)$, for any p. Although we have not been able to derive a rigorous lower bound for μ, we have been able to get a good idea of its value by examining (4.11) empirically. Let $\mu(\theta, \phi)$ be used to denote $\mu_{st}^{(A)}$. It is straightforward to show that μ is symmetric with respect to reflection over each of the lines $\theta = \pi$, $\phi = \pi$, $\theta = \phi$, $\theta = 2\pi - \phi$, i.e.

$$\mu(\theta, \phi) = \mu(2\pi - \theta, \phi) = \mu(\phi, \theta) = \mu(2\pi - \phi, 2\pi - \theta).$$

Therefore, we can restrict our attention to the triangular region of the first quadrant of $\mathcal{A}$ bounded by $\theta = \phi$ and $\theta = \pi$ (see Figure A.1). We have observed empirically that the minimum value occurs on the horizontal line $\phi = \phi_1 = 2\pi h$, and that on this line, $\mu(\theta, 2\pi h)$ takes on its minimum value (as a function of θ) at one of the endpoints $\theta = 2\pi h$, $\theta \approx \pi$. It is easily shown that asymptotically, $\mu(2\pi h, 2\pi h) = \frac{8\pi^2}{c} h^{2-p}$, and $\mu(\pi, 2\pi h) = 2\sqrt{2c}\, h^{p/2}$. The asymptotically optimal value of $p = \frac{4}{3}$ can then be determined by equating the two exponents of h. We therefore have the following

Conjecture. For the value $p = \frac{4}{3}$, the ratio of maximum and minimum eigenvalues of the ADDKR preconditioned periodic operator has the asymptotic value $O(h^{-2/3})$, and this is the smallest asymptotic value for all p in the interval $[0, 2]$.

This optimal choice of p agrees with the empirically determined optimal value for the Dirichlet problem [6]. A surface plot of (4.11) is given in Figure 5.

5. Relation to other Boundary Conditions

All the results presented so far apply only to the periodic problem. However, as we have observed throughout our presentation, these results are very similar to analogous results for the Dirichlet problem. In this section, we present an analysis and further numerical evidence relating the periodic analysis to the Dirichlet problem.

The results of the previous sections concern difference operators M_p defined for the periodic problem. That is, if A_p, Q_p and R_p denote the coefficient matrix and splitting operators for the periodic problem, then $M_p = Q_p^{-1} R_p$ for stationary methods and $M_p = Q_p^{-1} A_p$ for preconditioning methods, and all the operators of both classes of methods share the eigenvectors $\{u^{(s,t)}\}$ of (2.13). Let $\lambda_p^{(s,t)}$ denote the eigenvalues of M_p corresponding to $u^{(s,t)}$, and depending on context, let K_p represent either A_p or R_p. (Here we are ignoring the possible singularity of Q_p. To avoid reference to Q_p^{-1}, we could also say that $\{(\lambda_p^{(s,t)}, u^{(s,t)})\}$ are the solutions to the generalized eigenvalue problem $K_p u = \lambda Q_p u$.)

Consider the vectors

$$v^{(s,t)} \equiv (u^{(s,n+1-t)} + u^{(n+1-s,t)}) - (u^{(s,t)} + u^{(n+1-s,n+1-t)}),$$
$$w^{(s,t)} \equiv (u^{(s,n+1-t)} + u^{(n+1-s,t)}) + (u^{(s,t)} + u^{(n+1-s,n+1-t)}),$$

which satisfy

$$v_{jk}^{(s,t)} = 4\sin(j\theta_s)\sin(k\phi_t), \qquad w_{jk}^{(s,t)} = 4\cos(j\theta_s)\cos(k\phi_t).$$

Fron the periodicity of (all) the preconditioned operators, we have $\lambda_p^{(s,t)} = \lambda_p^{(n+1-s,n+1-t)}$, and therefore

$$\begin{aligned} M_p v^{(s,t)} &= \lambda_p^{(s,n+1-t)}(u^{(s,n+1-t)} + u^{(n+1-s,t)}) - \lambda_p^{(s,t)}(u^{(s,t)} + u^{(n+1-s,n+1-t)}) \\ &= \left(\frac{\lambda_p^{(s,t)} + \lambda_p^{(s,n+1-t)}}{2}\right) v^{(s,t)} + \left(\frac{\lambda_p^{(s,n+1-t)} - \lambda_p^{(s,t)}}{2}\right) w^{(s,t)}. \end{aligned}$$

Equivalently,

$$K_p v^{(s,t)} = \lambda_+^{(s,t)} Q_p v^{(s,t)} + \lambda_-^{(s,t)} Q_p w^{(s,t)}, \tag{5.1}$$

where

$$\lambda_+^{(s,t)} = \frac{\lambda_p^{(s,t)} + \lambda_p^{(s,n+1-t)}}{2}, \qquad \lambda_-^{(s,t)} = \frac{\lambda_p^{(s,n+1-t)} - \lambda_p^{(s,t)}}{2}.$$

Now assume that n is odd, $n = 2m + 1$ for some positive integer m. Consider a partitioning of the unit square into four equal quadrants (ordered counterclockwise starting from the lower left) and an ordering of any grid function u as

$$u = (u_1, u_2, u_3, u_4, u_5)^T,$$

where u_i, $1 \le i \le 4$, corresponds to the grid points interior to the i'th quadrant and u_5 corresponds to the grid points on the interface separating the quadrants. Let $v^{(s,t)}$ and $w^{(s,t)}$ be partitioned in this manner.

Consider the Dirichlet problem with homogeneous boundary conditions on the first quadrant, i.e. the square $\Omega_1 = [0, \frac{1}{2}] \times [0, \frac{1}{2}]$. The coefficient matrix A_d is the discrete Laplacian defined on an $m \times m$ grid. We define $M_d = Q_d^{-1} K_d$ to be the Dirichlet operator analogous to M_p on the square Ω_1. That is, Q_d (or its factors) and K_d are defined on the stencils of Figures 1 and 6 on the interior of Ω_1, and the coefficients of Q_d and K_d are precisely those of their periodic analogues at the corresponding stencil points.[4] Since $v^{(s,t)}$ vanishes on the boundary of Ω_1, we have

$$(K_p v^{(s,t)})_1 = K_d v_1^{(s,t)} \quad \text{and} \quad (Q_p v^{(s,t)})_1 = Q_d v_1^{(s,t)},$$

so that by (5.1),

$$K_d v_1^{(s,t)} = \lambda_+^{(s,t)} Q_d v_1^{(s,t)} + \lambda_-^{(s,t)} (Q_p w^{(s,t)})_1.$$

If

$$\lambda_p^{(s,t)} = \lambda_p^{(s,n+1-t)}, \tag{5.2}$$

then $\lambda_-^{(s,t)} = 0$ and $K_d v_1^{(s,t)} = \lambda_p^{(s,t)} Q_d v_1^{(s,t)}$. Relation (5.2) holds for the eigenvalues of both the Jacobi iteration matrix (see (3.3)) and the ADDKR preconditioned matrix (see (4.11)). Hence we have the following result.

Theorem 5.1. If $\lambda_p^{(s,t)} = \lambda_p^{(s,n+1-t)}$, then $\lambda_p^{(s,t)}$ is an eigenvalue of M_d. In particular, for both the Jacobi method and the ADDKR preconditioner, the eigenvalues $\lambda_p^{(s,t)}$ of M_p, $1 \le s, t \le m$, are precisely the m^2 eigenvalues of the corresponding Dirichlet matrix M_d for the quadrant Ω_1.

[4] Note that the resulting triangular matrices for the ILU, MILU and ADDKR incomplete factorizations have constant values on each of their bands, so that these are actually slightly different factorizations than the standard ones for the Dirichlet problem. It has been observed empirically that the standard factors are very close to these constant coefficient factors, see [13].

For the other methods under consideration, $\lambda_p^{(s,t)} \neq \lambda_p^{(s,n+1-t)}$ and $(Q_p w^{(s,t)})_1$ is not necessarily small on Ω_1, so that $\lambda_+^{(s,t)}$ will not be a good approximation to an eigenvalue of M_d. However, the strong correlation between the results for the periodic problem and the classical results for the Dirichlet problem suggests that the periodic analysis gives a good indication of the performance of the corresponding method for the Dirichlet problem. We now give two examples of numerical evidence supporting this conjecture.

First, we show that the Fourier analysis predicts the *distribution* of the eigenvalues for the ILU preconditioned Dirichlet operator. Figure 6 plots the Dirichlet eigenvalues for $h = \frac{1}{16}$ (computed in double precision using EISPACK [16]) and the periodic eigenvalues for $h = \frac{1}{32}$. For the figure, both sets of eigenvalues have been sorted in increasing order, and the horizontal axis represents the index of the Dirichlet eigenvalues or the index of the periodic eigenvalues scaled by $\frac{225}{961}$. As the figure shows, the two distributions are almost identical.

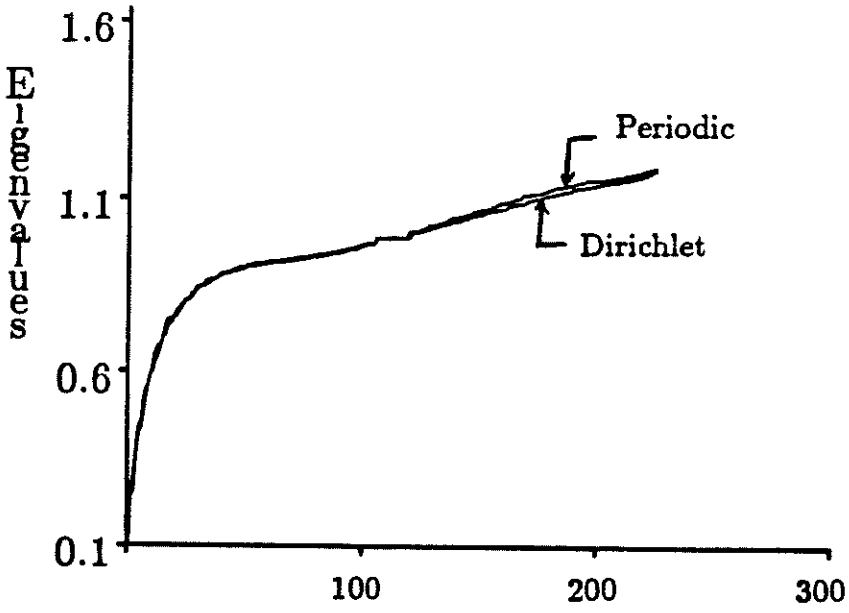

Figure 6. Distribution of ILU preconditioned eigenvalues for Dirichlet (h=1/16) and periodic (h=1/32) operators.

Second, we show that the Fourier analysis predicts the dependence of the MILU-condition number on the parameter c. It can be shown from (A.8) by elementary calculus that for the periodic problem, the maximum eigenvalue of the MILU preconditioned matrix is

$$\mu_{max} \approx \frac{1}{\sqrt{2c}\,h}, \tag{5.3}$$

occurring when $\phi_t = 2\pi - \theta_s$ and $\sin^2(\theta_s/2) \approx \sqrt{c/8}\,h$. Similarly, it can be shown that the minimum eigenvalue is

$$\mu_{min} \approx \frac{1}{1 + c/(8\pi^2)}, \tag{5.4}$$

occurring for $\theta_s = \theta_1 = \phi_t = \phi_1 = 2\pi/(n+1)$. Hence, as a function of c, the restricted condition number is

$$\kappa^{(M)}(c) \approx \frac{1 + c/(8\pi^2)}{\sqrt{2c}\,h}, \tag{5.5}$$

Using elementary calculus, it is straightforward to show that $\kappa^{(M)}$ has a minimum value of $1/(2\pi h)$ at $c \approx 8\pi^2$.

We wish to apply these observations to the MILU-preconditioned discrete Dirichlet problem on the unit square. First, note that the recurrence (4.6) for the diagonal values $\{\alpha_{jk}\}$ is not satisfied exactly for the MILU factorization of the Dirichlet operator. For indices $j = 1$ or n and $k = 1$ or n, the entries $\alpha_{j-1,k}$ or $\alpha_{j+1,k}$ and $\alpha_{j,k-1}$ or $\alpha_{j,k+1}$ do not appear. As noted above and shown in [13], α_{jk} is close to the constant value α of (4.7) for most j, k. Thus, we could consider the alternative incomplete factorization for the Dirichlet operator in which the constant α is used in place of the varying $\{\alpha_{jk}\}$. We denote this constant coefficient incomplete factorization by MILU*. We will compare the results (5.3) – (5.5) with the corresponding values for the preconditioned Dirichlet operator, using both the *true* MILU preconditioning and the MILU* preconditioning.

Let the Dirichlet problem be defined on a mesh of width h_d, so that for the Dirichlet MILU preconditioner Q, the row sum of $Q - A$ is $c_d h_d^2$. We seek a correspondence between this problem and the periodic problem for mesh size $h_p = \frac{h_d}{2}$. For the correspondence between the MILU-preconditioned periodic problem and the (constant coefficient) MILU*-preconditioned Dirichlet problem, we adopt the convention that both factorizations use the same value of α. That is, we use c_d and h_d in place of c and h in (4.7) for the Dirichlet problem, and c_p and h_p for the periodic problem, and then we equate the two values of α obtained. The result is $c_p = 4c_d$. We will also compare the performance of the true Dirichlet MILU preconditioner to the periodic version with these pairs of values c_d and c_p.

In Figure 7, we plot the minimum eigenvalues for the Dirichlet and periodic problems for $h_d = \frac{1}{26}$, and in Figure 8, we plot the maximum eigenvalues for both $h_d = \frac{1}{26}$ and $h_d = \frac{1}{52}$. (The minimum values for $h_d = \frac{1}{52}$ are nearly identical to those plotted in Figure 7.) In Figure 9, we plot the condition numbers for the three preconditioned operators, for both values of h_d. (Again, the data for the periodic problem comes from using $h_p = \frac{h_d}{2}$ and $c_p = 4c_d$ in the MILU symbol (4.8).) The eigenvalues for the Dirichlet problems were computed as the eigenvalues of the preconditioned matrix $B = AQ^{-1}$ using Arnoldi's method with Chebyshev acceleration [32]. The stopping criterion for the eigenvalue iteration was $\|Bv - \lambda v\|_2 \leq .5 \times 10^{-3}$, where $\|v\|_2 = 1$. The Dirichlet eigenvalues were computed on a VAX-8600 in double precision Fortran; the periodic eigenvalues (4.8) were computed in single precision.

Figures 7 - 9 show the results for the three examples to be qualitatively very similar. In particular, we can say the following:

1. The maximum MILU*-eigenvalues are nearly indistinguishable from the periodic eigenvalues, and the true Dirichlet MILU-eigenvalues tend to the other sets as c_d grows. The minimum eigenvalues (which do not vary asymptotically with h_d) for the three problems are qualitatively similar and tend to one another with increasing c_d, although (somewhat surprisingly) the true MILU-eigenvalues are closer to the periodic ones than the MILU*-values.

2. The minimum periodic eigenvalues are smaller than both sets of Dirichlet values, and the maximum periodic eigenvalues are larger than the Dirichlet ones, so that the periodic condition number is an upper bound for the Dirichlet condition numbers.

3. The optimal value $c_p \approx 8\pi^2$ determined above gives an optimal $c_d \approx 2\pi^2$ for the Dirichlet problem, which is the same value derived from bounds on the condition number by the analyses of [2,19].[5] The actual minimum values for the Dirichlet curves of Figure 9 are slightly smaller, but the dependence of conditioning on c clearly follows the same general pattern for the two types of boundary conditions.

We end with one further observation from these results that reveals the usefulness of the Fourier analysis. The folklore for the MILU-preconditioned Dirichlet problem holds that the condition number for $c_d = 0$ is also $O(h_d^{-1})$, although this has never been proved. The computed condition numbers when $c_d = 0$ for the three methods considered here are as follows:

h_d	Dirichlet MILU	Dirichlet MILU*	Periodic
1/26	7.5	39.8	27.4
1/52	15.7	84.8	110.0

Thus, the values for the MILU-preconditioned Dirichlet operator agree with the folklore,[6] and the MILU*-conditioning also appears to grow like h_d^{-1}. Our analysis of the periodic problem gives the condition number for $c_p = 0$ as $O(h_p^{-2})$, but this result is strongly dependent on certain exact cancellations (see the Appendix). We suspect that these cancellations do not occur in the Dirichlet case, and that the better performance is a consequence of this.

[5] We elaborate on this point as follows. First, in both [2] and [19], the preconditioning parameter is scaled by $diag(A)$, i.e. for the model problem, $4ch^2$ is added to the diagonal instead of ch^2. The optimal value for the scaled modification in [2] is therefore $\frac{\pi^2}{2}$. Moreover, in scrutinizing these results, we discovered that the optimal choice of $\frac{\pi^2}{8}$ reported by Gustafsson [19] is actually in error; the correct value for the analysis of [19] is also $\frac{\pi^2}{2}$.

[6] To the best of our knowledge, no experiments demonstrating this have been reported previously.

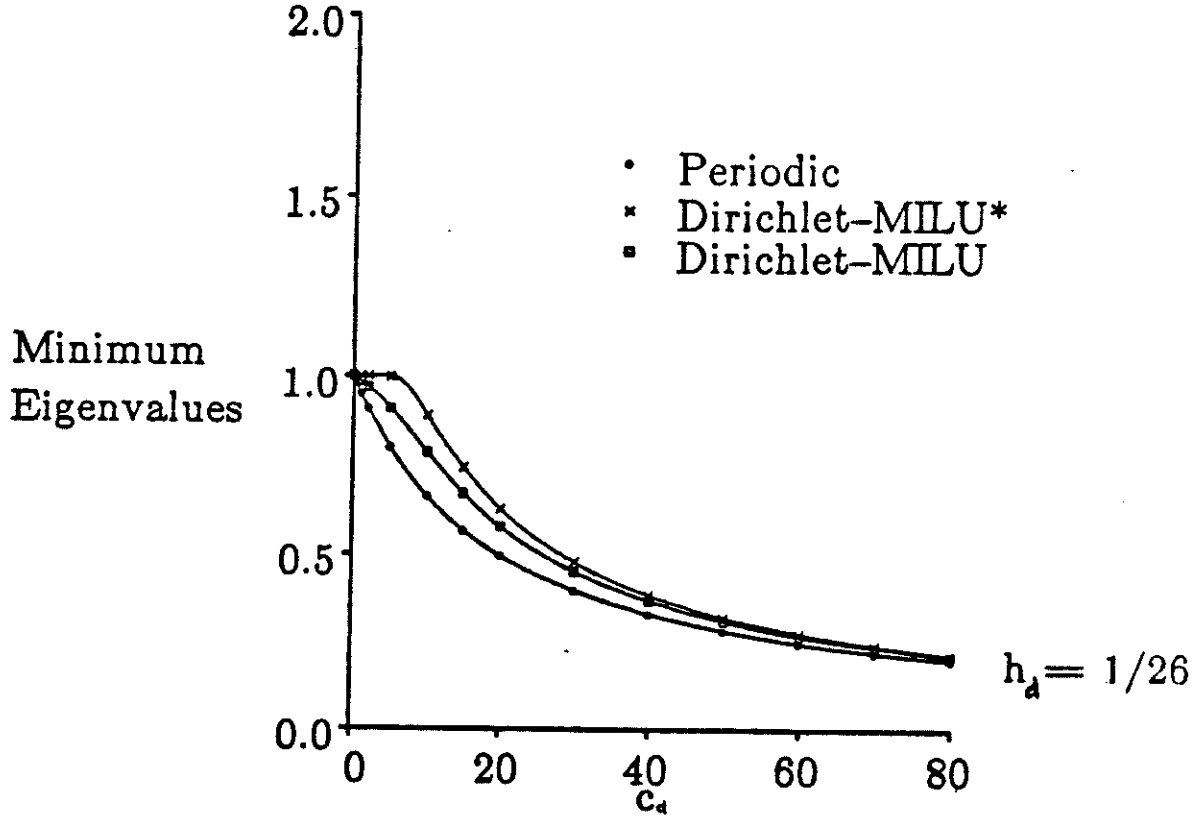

Figure 7. Minimum eigenvalues of the MILU preconditioned operators, $h_d = 1/26$.

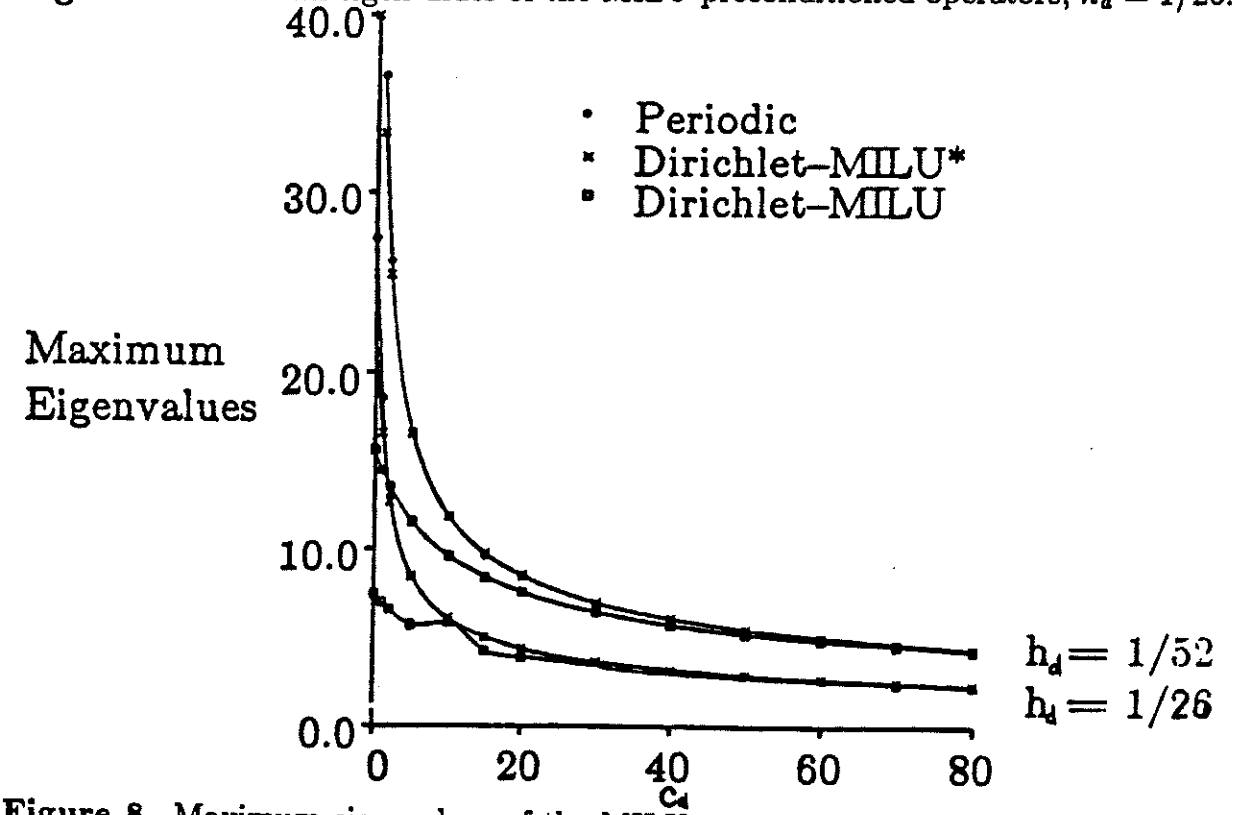

Figure 8. Maximum eigenvalues of the MILU preconditioned operators, $h_d = 1/26$ and $h_d = 1/52$.

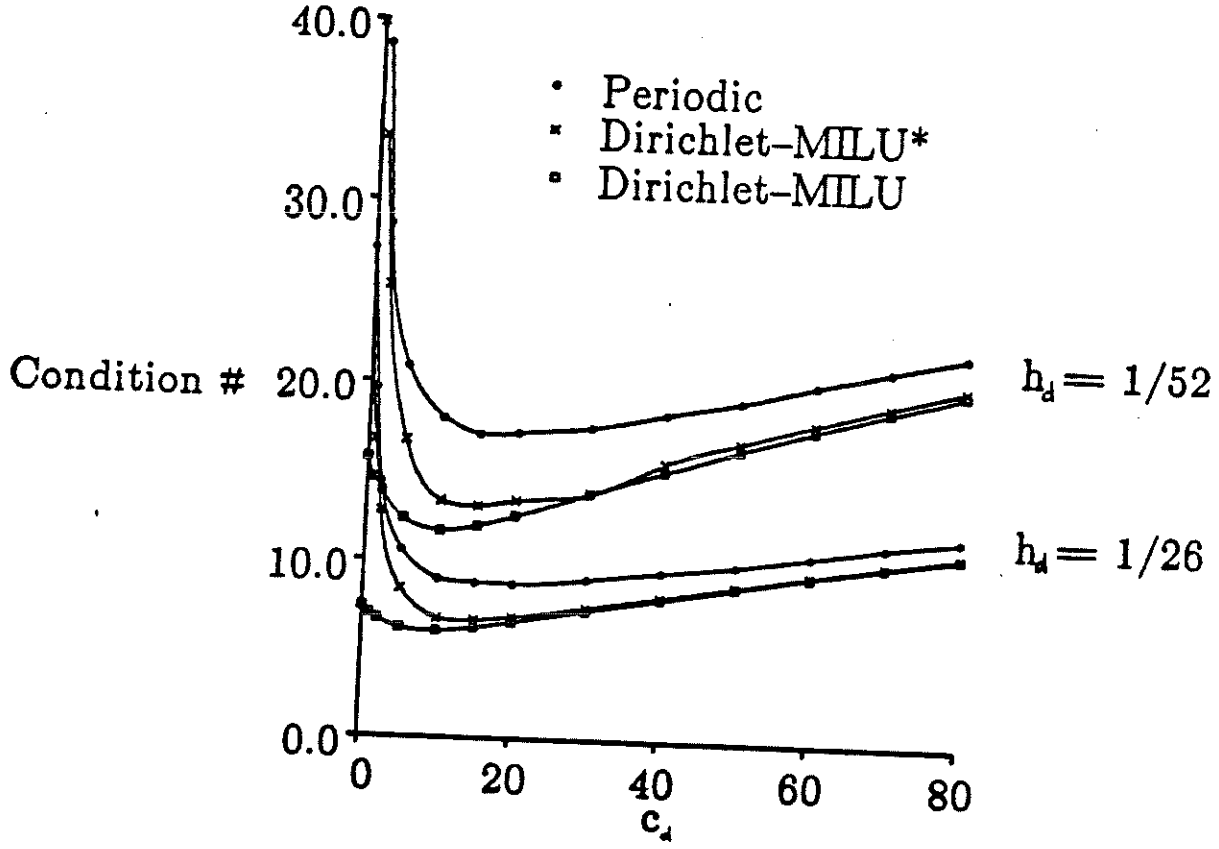

Figure 9. Condition numbers of the MILU preconditioned operators, $h_d = 1/26$ and $h_d = 1/52$.

Appendix

In this section, we fill in the technical details omitted from the main text: we prove Theorems 4.1 and 4.2, provide evidence for the observations made at the end of Section 4.2, and prove Lemma 4.1.

Proof of Theorem 4.1. We bound $\mu^{(I)}$ of (4.5) by deriving explicit bounds for λ_{st} of (2.13) and ψ_{st} of (4.4). It holds immediately that

$$\lambda_{st} \leq 8, \qquad \psi_{st} \leq 8 + \frac{2}{2+\sqrt{2}}. \tag{A.1}$$

For the inequalities in the other direction, first note that since $1 \leq s,t \leq n$,

$$\frac{\pi}{n+1} \leq \frac{\theta_s}{2} \leq \frac{n\pi}{n+1},$$

so that

$$\sin\frac{\theta_s}{2} \geq \sin\frac{\pi}{n+1} = \sin(\pi h).$$

Hence

$$\lambda_{st} \geq 8\sin^2(\pi h) = O(h^2).$$

Combining this with the second inequality of (A.1) gives $\mu^{(I)} \geq O(h^2)$. To bound ψ_{st} below, we use the identity

$$\cos(\theta_s - \phi_t) = \left(1 - 2\sin^2\frac{\theta_s}{2}\right)\left(1 - 2\sin^2\frac{\phi_t}{2}\right) + 4\left(\sin\frac{\theta_s}{2}\sin\frac{\phi_t}{2}\right)\sqrt{\left(1-\sin^2\frac{\theta_s}{2}\right)\left(1-\sin^2\frac{\phi_t}{2}\right)}. \tag{A.2}$$

Substituting $x = \sin(\theta_s/2)$, $y = \sin(\phi_t/2)$ gives

$$\cos(\theta_s - \phi_t) = 1 - 2(x^2+y^2) + 4x^2y^2 + 4xy\sqrt{(1-x^2)(1-y^2)}. \tag{A.3}$$

It then follows from (4.4) that

$$\begin{aligned}\psi_{st} &= \frac{2}{2+\sqrt{2}}\left(2(1+\sqrt{2})(x^2+y^2) + 1 + 4x^2y^2 + 4xy\sqrt{(1-x^2)(1-y^2)}\right)\\ &> \frac{2}{2+\sqrt{2}}\left(2(x^2+y^2) + 1 - 4|x||y|\right) = \frac{2}{2+\sqrt{2}}\left(2(|x|-|y|)^2 + 1\right)\\ &\geq \frac{2}{2+\sqrt{2}},\end{aligned}$$

where we have used the fact that $xy\sqrt{(1-x^2)(1-y^2)} \geq -|x||y|$. Combining this with the first inequality of (A.1) gives $\mu^{(I)} < 4(2+\sqrt{2}) = O(1)$, whence $\kappa^{(I)} \leq O(h^{-2})$.

To see that this bound is tight, consider the case of $\theta_s = \phi_t$. Then (4.5) simplifies to

$$\mu^{(I)} = \frac{8\sin^2(\theta_s/2)}{8\sin^2(\theta_s/2) + 2/(2+\sqrt{2})}.$$

If $\theta_s = \theta_1 = 2\pi h$, then $\mu^{(I)} \approx \frac{8\pi^2 h^2}{8\pi^2 h^2 + 2/(2+\sqrt{2})} = O(h^2)$; and if $\theta_s \approx \pi$ ($s \approx (n+1)/2$), then $\mu^{(I)} \approx \frac{8}{8+2/(2+\sqrt{2})} = O(1)$. Q.E.D.

Proof of Theorem 4.2. Letting $x = \sin(\theta_s/2)$, $y = \sin(\phi_t/2)$ and using (A.3), we rewrite (4.8) in terms of x and y as

$$\mu^{(M)} = \frac{4(x^2+y^2)}{(4-4/\alpha)(x^2+y^2) + (8/\alpha)(x^2y^2 + xy\sqrt{(1-x^2)(1-y^2)}\,) + ch^2}. \qquad (A.4)$$

Note that both x and y are nonzero and bounded in absolute value by one, and $\min|x| = \min|y| = \sin(\pi h) \le \tilde{c}h$ for all small h, where $\tilde{c} \approx \pi$ is independent of h.

For the lower bound on $\mu^{(M)}$, taking absolute values and applying the triangle equality to the representation of ψ_{st} in the denominator of (A.4) gives

$$\begin{aligned} 0 \le \psi_{st} = |\psi_{st}| &\le (4+4/\alpha)(x^2+y^2) + (8/\alpha)(x^2y^2 + |x||y|) + ch^2 \\ &\le 10(x^2+y^2) + ch^2, \end{aligned}$$

where we have used the facts that $1/\alpha \le 1/2$, $x^2y^2 \le |x||y|$ and $|x||y| \le (x^2+y^2)/2$. Hence,

$$\mu^{(M)} \ge \frac{2}{5}\,\frac{1}{1 + ch^2/10(x^2+y^2)} \ge \frac{1}{1 + c/(20(\tilde{c}h)^2)} = O(1).$$

For the upper bound on $\mu^{(M)}$, from the representation of ψ_{st} in the denominator of (A.4), we have

$$\psi_{st} \ge (4-4/\alpha)(x^2+y^2) - (8/\alpha)|x||y|(1-|x||y|) + ch^2. \qquad (A.5)$$

By assumption, x and y are both nonzero, so that

$$\mu^{(M)} \le \frac{1}{(1-\frac{1}{\alpha}) - \frac{2}{\alpha}\frac{|x||y|(1-|x||y|)}{x^2+y^2} + \frac{ch^2}{x^2+y^2}}. \qquad (A.6)$$

But

$$0 \le 1 - |x||y| \le 1 - \tilde{c}^2h^2.$$

Substituting this inequality into (A.6) and using the fact that $|x||y|/(x^2+y^2) \le 1/2$ gives

$$\mu^{(M)} \le \frac{1}{1-(1/\alpha)-(1/\alpha)(1-\tilde{c}^2h^2)+ch^2} = \frac{1}{1-2/\alpha+O(h^2)}.$$

Since we are concerned only with the asymptotic behavior as $h \to 0$, it is sufficient to consider the high order parts of $2/\alpha$. In particular, $\sqrt{8ch^2 + (ch^2)^2} \approx 2h\sqrt{2c}$ for small h, so that $\alpha = 2 + h\sqrt{2c} + O(h^2)$ and $2/\alpha = 2/(2 + h\sqrt{2c} + O(h^2)) = 1 - h\sqrt{2c}/2 + O(h^2)$. Consequently,

$$\mu^{(M)} \leq \frac{1}{h\sqrt{2c}/2 + O(h^2)}.$$

Thus, for $c > 0$, $\mu^{(M)} \leq O(h^{-1})$ and $\kappa^{(M)} \leq O(h^{-1})$; but for $c = 0$ the $O(h)$ term disappears from the denominator so that $\mu^{(M)} \leq O(h^{-2})$ and $\kappa^{(M)} \leq O(h^{-2})$.

These asymptotic bounds on $\mu^{(M)}$ are also achieved. For example, if $\theta_s = \phi_t \approx \pi$, then $\mu^{(M)} = 8/(8 + ch^2) = O(1)$ for any c, so that the lower bound is reached. For the upper bound, consider the case of $\phi_t = 2\pi - \theta_s$, so that

$$\cos(\theta_s - \phi_t) = \cos(2\theta_s) = 1 - 8\sin^2(\theta_s/2) + 8\sin^4(\theta_s/2). \tag{A.7}$$

Substitution into (4.8) gives

$$\mu^{(M)} = \frac{8\sin^2(\theta_s/2)}{4h\sqrt{2c}\sin^2(\theta_s/2) + 8\sin^4(\theta_s/2) - 4h\sqrt{2c}\sin^4(\theta_s/2) + ch^2}. \tag{A.8}$$

If $c = 0$, then this expression simplifies to $\sin^2(\theta_s/2)/\sin^4(\theta_s/2)$, and the particular choice of $\theta_s = 2\pi h$ leads to $\mu^{(M)} = O(h^{-2})$. If $c > 0$, then $\theta_s = 2\pi h$ results in $\mu^{(M)} = O(1)$, but $\theta_s \approx 2\sqrt{h}$ (i.e. $s \approx \sqrt{n+1}/\pi$) gives $\mu^{(M)} = O(h^{-1})$. Hence when $c = 0$, $\kappa^{(M)} = O(h^{-2})$, and when $c > 0$, $\kappa^{(M)} = O(h^{-1})$. Q.E.D.

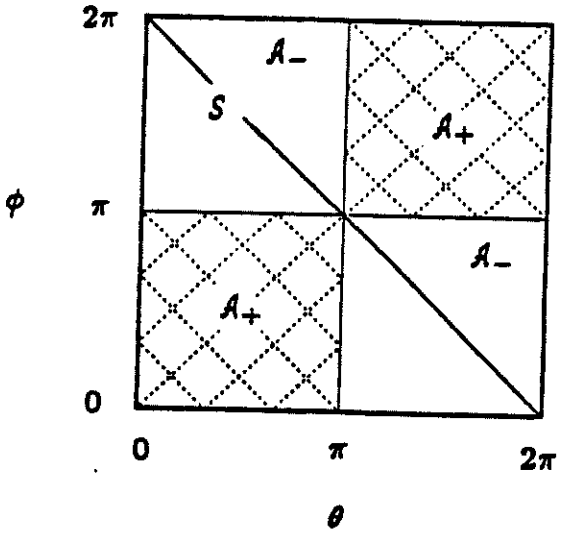

Figure A.1. Division of $\mathcal{A} = \{(\theta, \phi) | 0 \leq \theta, \phi \leq 2\pi\}$.

We now provide justification for the remarks made at the end of Section 4.2 concerning the MILU preconditioning. Extreme eigenvalues on the transverse diagonal $\mathcal{S}$ were exhibited in the proof of Theorem 4.2. To see that the only extreme eigenvalues are near the corners of $\mathcal{S}$, note that in (A.2), the square root is actually $\pm\cos(\theta_s/2)\cos(\phi_t/2)$, where the positive square root applies in the region $\mathcal{A}_+ = \{0 \leq \theta, \phi \leq \pi\} \cup \{\pi \leq \theta, \phi \leq 2\pi\}$,

and the negative square root applies in the complement $\mathcal{A}_- = \mathcal{A} - \mathcal{A}_+$. These regions are depicted in Figure A.1. Similarly, the sign of the square root in (A.4) depends on the location in $\mathcal{A}$ of the modes determining x and y. It follows immediately from (A.4) that $\mu^{(M)} \leq 2$ on $\mathcal{A}_+$. Moreover, (A.5) implies that

$$\psi_{st} \geq 2(|x| - |y|)^2 + \frac{8}{\alpha}x^2y^2.$$

Away from the endpoints of $\mathcal{S}$ in $\mathcal{A}_-$, $|x|$ and $|y|$ are both bounded below by a constant ξ independent of h. Hence, $\mu^{(M)} = O(1)$ away from these corners.

The effect on the preconditioned operator of the ch^2 term used in the definition of the MILU factorization is clearly evident from (A.8). When $c = 0$, the coefficient of the $\sin^2(\theta_s/2)$ term in the denominator is zero, and the result is a preconditioned eigenvalue of order h^{-2} when θ_s is $O(h)$. In contrast, for nonzero c, the value $\theta_s = O(h)$ produces an eigenvalue of order one. This result for $c = 0$ differs from empirical results for the Dirichlet problem, but closer examination of (4.8) and (A.4) reveals this phenomenon to be a very delicate matter. If (θ_s, ϕ_t) is in $\mathcal{S}$, then $x = \sin(\theta_s/2) = \sin(\phi_t/2) = y$. Let $y = ax$ be determined by a pair of values of θ_s and ϕ_t near $\mathcal{S}$ in the corners of $\mathcal{A}$. The numerator of (4.8) is at least $O(h^2)$, so it is possible to have an eigenvalue of $O(h^{-2})$ only if the denominator is not larger than $O(h^4)$. Note that the negative square root is used everywhere near the ends of $\mathcal{S}$. When $c = 0$ (so that $\alpha = 2$), the denominator of (A.4) is

$$2(1 + a^2)x^2 + 4a^2x^4 - 4ax^2\sqrt{(1 - x^2)(1 - a^2x^2)}. \tag{A.9}$$

But

$$\sqrt{(1 - x^2)(1 - a^2x^2)} = (1 - x^2)\sqrt{1 + \frac{1 - a^2}{1 - x^2}x^2} \approx 1 - x^2 + \frac{(1 - a^2)x^2}{2}.$$

Consequently, (A.9) is approximately

$$2(a - 1)^2x^2 + O(x^4),$$

which is at least $O(h^2)$ unless $a = 1$. Thus, any perturbation away from $\mathcal{S}$ pushes the preconditioned eigenvalues from the extreme $O(h^{-2})$ values to $O(1)$. Thus, although the restricted condition number for the MILU preconditioned system with $c = 0$ is $O(h^{-2})$, the extreme eigenvalues appear in a very constrained set on $\mathcal{S}$. In contrast, there are many examples of (small) extreme eigenvalues for the ILU preconditioned operator. A similar argument also shows that all the extreme eigenvalues for $c > 0$ (of order h^{-1}) occur on $\mathcal{S}$.

Finally, the optimal ω and SSOR condition number on $\mathcal{S}$ are determined as follows:

Proof of Lemma 4.1. With $x = \sin(\theta_s/2)$, $y = \sin(\phi_t/2)$, (4.9) can be rewritten as

$$\psi_{st} = (4 - \omega + \omega^2) + (4\omega - \omega^2)(x^2 + y^2) + 2\omega^2(x^2y^2 + xy\sqrt{(1 - x^2)(1 - y^2)}).$$

Relation (A.7) implies that on the line $\mathcal{S}$, the SSOR eigenvalues are

$$(2 - \omega)^2 + 4\omega(2 - \omega)x^2 + 4\omega^2x^4,$$

and the eigenvalues of the SSOR preconditioned operator on $\mathcal{S}$ are

$$\mu = \mu(x,\omega) = \frac{8}{(2-\omega)^2/x^2 + 4\omega(2-\omega) + 4\omega^2 x^2}, \qquad (A.10)$$

where $\sin^2(\pi h) \leq x^2 \leq 1$. We can use (A.10) to derive the value of ω that minimizes the condition number on $\mathcal{S}$. First, it can be shown by elementary calculus that for any ω, the maximum value of μ is $\mu_{max}(\omega) = 1/(\omega(2-\omega))$, occurring when $x^2 = (2-\omega)/\omega$. Moreover, for fixed ω, μ is a convex function of x^2, where $\sin^2(\pi h) \leq s^2 \leq 1$. Hence, its minimum must occur at one of the endpoints $\sin^2(\pi h)$ or 1. Thus, the condition number on $\mathcal{S}$ is the larger of

$$\kappa_1(\omega) = \frac{\mu_{max}(\omega)}{\mu_1(\omega)}, \qquad \kappa_2(\omega) = \frac{\mu_{max}(\omega)}{\mu_2(\omega)},$$

where

$$\mu_1(\omega) = \frac{8}{(2-\omega)^2 + 4\omega(2-\omega) + 4\omega^2},$$

$$\mu_2(\omega) = \frac{8}{(2-\omega)^2/\sin^2(\pi h) + 4\omega(2-\omega) + 4\omega^2\sin^2(\pi h)}$$

are the values of μ at the endpoints. As functions of ω, κ_1 and κ_2 are both convex functions that tend to ∞ as ω approaches both 0 and 2. Moreover, by setting $\kappa_1 = \kappa_2$, it can be verified that they intersect at just one point, given by $\omega^* = 2/(1 + 2\sin(\pi h))$. Hence, on $\mathcal{S}$, $\kappa_1(\omega)$ and $\kappa_2(\omega)$ have the form given in Figure A.2 and the minimum value of $\max\{\kappa_1(\omega), \kappa_2(\omega)\}$ occurs at ω^*. Moreover, $\mu_{max}(\omega^*) = (1 + 2\sin(\pi h))^2/(8\sin(\pi h)) = O(h^{-1})$ and $\mu_{min}(\omega^*) = (1 + 2\sin(\pi h))^2/(2(1 + \sin(\pi h))^2) = O(1)$, so that the condition number is $O(h^{-1})$. Q.E.D.

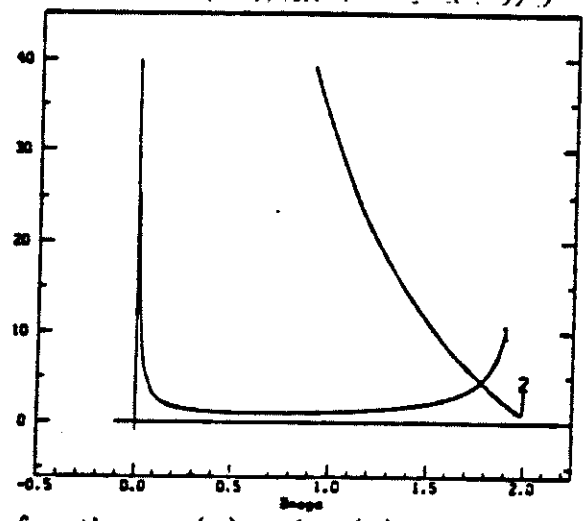

Figure A.2. Convex functions $\kappa_1(\omega)$ and $\kappa_2(\omega)$ determining SSOR condition number on $\mathcal{S}$.

References

[1] L. M. Adams, R. J. LeVeque and D. M. Young, Analysis of the SOR iteration for the 9-point Laplacian, ICASE Report, 1986. To appear in *SIAM J. Numer. Anal.*

[2] O. Axelsson, A generalized SSOR method, *BIT* 13:443-467, 1972.

[3] O. Axelsson, Solution of linear systems of equations: iterative methods, in V. A. Barker, Ed., *Sparse Matrix Techniques*, Springer-Verlag, New York, 1976, pp. 1-51.

[4] A. Brandt, Multi-level adaptive solutions to boundary-value problems, *Math. Comp.* 31: 333-390, 1977.

[5] K. P. Bube and J. C. Strikwerda, Interior regularity estimates for elliptic systems of difference equations, *SIAM J. Numer. Anal.* 20: 653-670, 1983.

[6] T. F. Chan, K. R. Jackson and B. Zhu, Alternating-direction incomplete factorizations, *SIAM J. Numer. Anal.* 20: 239-257, 1983.

[7] R. Chandra, *Conjugate Gradient Methods for Partial Differential Equations*, Ph. D. Thesis, Dept. of Computer Science, Yale Univ., 1978.

[8] P. Concus and G. H. Golub, Use of fast direct methods for the efficient numerical solution of nonseparable elliptic equations, *SIAM J. Numer. Anal.* 10: 1103-1120, 1973.

[9] P. Concus, G. H. Golub and D. P. O'Leary, A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, in J. R. Bunch and D. J. Rose, Eds, *Proceedings of the Symposium on Sparse Matrix Computations*, Academic Press, New York, 1975, pp. 309-332.

[10] T. Dupont, A factorization procedure for the solution of elliptic difference equations, *SIAM J. Numer. Anal.* 5:753-782, 1968.

[11] T. Dupont, R. P. Kendall and H. H. Rachford Jr., An approximate factorization procedure for solving self-adjoint elliptic difference equations, *SIAM J. Numer. Anal.* 5:559-573, 1968.

[12] H. C. Elman, A stability analysis of incomplete LU factorizations, *Math. Comp.* 47:191-217, 1986.

[13] H. C. Elman and M. H. Schultz, Preconditioning by fast direct methods for nonself-adjoint nonseparable elliptic equations, *SIAM J. Numer. Anal.* 23: 44-57, 1986.

[14] G. E. Forsythe and W. R. Wasow, *Finite Difference Methods for Partial Differential Equations*, John Wiley and Sons, New York, 1960.

[15] S. P. Frankel, Convergence rates of iterative treatments of partial differential equations, *Math. Comp.* 4:65-75, 1950.

[16] B. S. Garbow, J. M. Boyle, J. J. Dongarra and C. B. Moler, *Matrix Eigensystem Routines: EISPACK Guide Extension*, Springer-Verlag, New York, 1972.

[17] G. H. Golub and D. P. O'Leary, *Some History of the Conjugate Gradient and Lanczos Algorithms: 1948-1976*, Computer Science Series Technical Report 1859, Univ. of Maryland, 1987.

[18] D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods*, NSF-CBMS Monograph No. 26, SIAM, Philadelphia, 1977.

[19] I. Gustafsson, A class of first order factorizations, *BIT* 18:142-156, 1978.

[20] W. Hackbusch, Multigrid convergence for a singular perturbation problem, *Lin. Alg. Appl.* 58:125-145, 1984.

[21] L. A. Hageman and D. M. Young, *Applied Iterative Methods*, Academic Press, New York, 1981.

[22] T. Jameson, Multigrid algorithms for compressible flow calculations, in W. Hackbusch and U. Trottenberg, Eds., *Multigrid Methods II*, Lecture Notes in Mathematics 1228, Springer-Verlag, New York, 1985, pp. 166-201.

[23] W. Kahan, *Gauss-Seidel Methods of Solving Large Systems of Linear Equations*, Ph. D. Thesis, Univ. of Toronto, 1958.

[24] R. Kettler, Analysis and comparison of relaxation schemes in robust multigrid and preconditioned conjugate gradient methods, in W. Hackbusch and U. Trottenberg, Eds., *Multigrid Methods*, Lecture Notes in Mathematics 960, Springer-Verlag, New York, 1982, pp. 502-534.

[25] C.-C. Kuo, *Parallel Algorithms and Architectures for Solving Elliptic Partial Differential Equations*, Report LIDS-TH-1432, Laboratory for Information and Decision Systems, MIT, 1985.

[26] C.-C. J. Kuo, B. C. Levy and R. R. Musicus, *A Local Relaxation Method for Solving Elliptic PDEs on Mech-Connected Arrays*, Report LIDS-TH-1508, Laboratory for Information and Decision Systems, MIT, 1986. To appear in *SIAM J. Sci. Stat. Comput.*

[27] R. J. LeVeque and L. N. Trefethen, *Fourier Analysis of the SOR Iteration*, Numerical Analysis Report 86-6, Dept. of Mathematics, MIT, 1986.

[28] W. Liniger, On factored discretizations of the Laplacian for the fast solution of Poisson's equations on general regions," *BIT* 24:592-608, 1984.

[29] J. A. Meijerink and H. A. van der Vorst, An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix, *Math. Comp.* 31:148-162, 1977.

[30] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, Interscience, New York, 1967.

[31] P. J. Roache, *Computational Fluid Dynamics*, Hermosa Publishers, Albuquerque, 1982.

[32] Y. Saad, Chebyshev acceleration techniques for solving nonsymmetric eigenvalue problems, *Math. Comp.* 42:567-588, 1984.

[33] P. N. Swarztrauber, The methods of cyclic reductin, fourier analysis and the FACR algorithm for the discrete solution of Poisson's equation on a rectangle, *SIAM Review* 19: 490-501, 1977.

[34] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, New York, 1962.

[35] H. F. Weinberger, *A First Course in Partial Differential Equations with Complex Variables and Transform Methods*, Blaisdell, New York, 1965.

[36] D. M. Young, *Iterative Methods for Solving Partial Difference Equations of Elliptic Type*, Ph. D. Thesis, Harvard Univ., 1950.

[37] D. M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.