BIT
2003, Vol. 43, No. 4, pp. 001–019

0006-3835/00/4004-0001 $15.00

OPTIMIZATION OF THE HERMITIAN AND SKEW-HERMITIAN SPLITTING ITERATION FOR SADDLE-POINT PROBLEMS*

MICHELE BENZI[1]†, MARTIN J. GANDER[2], and GENE H. GOLUB[3]‡

[1] *Department of Mathematics and Computer Science, Emory University Atlanta, GA 30322, USA. email: benzi@mathcs.emory.edu*

[2] *Department of Mathematics and Statistics, McGill University Montreal, QC, Canada H3A 2K6. email: mgander@math.mcgill.ca*

[3] *Scientific Computing and Computational Mathematics Program, Stanford University Stanford, CA 94305-9025, USA. email: golub@sccm.stanford.edu*

DEDICATED TO DAVID M. YOUNG ON HIS 80TH BIRTHDAY

Abstract.

We study the asymptotic rate of convergence of the alternating Hermitian/skew-Hermitian iteration for solving saddle-point problems arising in the discretization of elliptic partial differential equations. By a careful analysis of the iterative scheme at the continuous level we determine optimal convergence parameters for the model problem of the Poisson equation written in div-grad form. We show that the optimized convergence rate for small mesh parameter h is asymptotically $1 - O(h^{1/2})$. Furthermore we show that when the splitting is used as a preconditioner for a Krylov method, a different optimization leading to two clusters in the spectrum gives an optimal, h-independent, convergence rate. The theoretical analysis is supported by numerical experiments.

AMS subject classification: 65F10, 65N22.

Key words: HSS iteration, saddle-point problems, Fourier analysis, rates of convergence

1 Introduction.

We consider the solution of large sparse linear systems of the form

$$\begin{bmatrix} A & B^* \\ B & O \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix} \tag{1.1}$$

where $A \in \mathbb{C}^{n\times n}$ is Hermitian and positive semidefinite, $B \in \mathbb{C}^{m\times n}$ has full rank $m \le n$, $f \in \mathbb{C}^n$, and $g \in \mathbb{C}^m$. Here B^* denotes the conjugate transpose of B. We further assume that A and B have no nontrivial null vectors in common, which guarantees the existence and uniqueness of the solution of (1.1). Linear systems

*Received October 2002. Revised July 2003. Communicated by Daniel B. Szyld.
†This work was supported in part by NSF grant DMS-0207599.
‡Research supported, in part, by DOE-FC02-01ER4177.

of this type arise in a number of applications, including the numerical solution of elliptic PDEs and Stokes problems by mixed finite element methods; see, e.g., [5, 8, 9]. In practice, the matrices and vectors in (1.1) are usually real. Here we allow complex entries because we analyze the problem later in the frequency (Fourier) domain.

In this paper we study the rate of convergence of the Hermitian/skew-Hermitian splitting (HSS) iteration [1] applied to (1.1). The use of HSS as a stationary iteration for solving (1.1) has been first proposed in [3], where it was shown that the iteration converges for a large class of problems (of which (1.1) is a special case). In the same paper, it was also shown that the HSS iteration can provide an effective preconditioner for Krylov subspace methods applied to (1.1).

Here we focus on the simple model problem of the Poisson equation in order to gain insight on how to choose an optimal value of the convergence parameter when the method is used either as a fixed-point iteration, or as a preconditioner for a Krylov method (GMRES). We base our approach on an analysis of the algorithm at the continuous level, in the spirit of [11], [12] and [10]; see also the basic reference [6]. The continuous (Fourier-based) analysis of the iteration operator provides insight into the choice of convergence parameters, and yields estimates for the asymptotic rate of convergence in terms of the discretization parameter (mesh size) h, as $h \to 0$. Furthermore, the spectral analysis can be used to show that taking a small value for the optimization parameter yields an optimal preconditioner for GMRES, with convergence in 2-3 iterations independent of h.

Other approaches for choosing the HSS iteration parameter can be found in the recent references [2] and [4].

2 The Hermitian/skew-Hermitian (HSS) iteration.

We begin by writing the saddle-point problem (1.1) in *non-Hermitian form* $\mathcal{A}\,\mathbf{x} = \mathbf{b}$, where

$$\mathcal{A} = \begin{bmatrix} A & B^* \\ -B & O \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} u \\ p \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} f \\ -g \end{bmatrix}.$$

Now observe that $\mathcal{A} = \mathcal{H} + \mathcal{S}$, where $\mathcal{H} = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*)$ and $\mathcal{S} = \frac{1}{2}(\mathcal{A} - \mathcal{A}^*)$ are the Hermitian and skew-Hermitian parts of $\mathcal{A}$, respectively. In our case, we have

$$\mathcal{H} = \begin{bmatrix} A & O \\ O & O \end{bmatrix} \quad \text{and} \quad \mathcal{S} = \begin{bmatrix} O & B^* \\ -B & O \end{bmatrix}.$$

Let $\alpha > 0$ be a parameter and consider the following two splittings of $\mathcal{A}$,

$$\mathcal{A} = (\mathcal{H} + \alpha\mathcal{I}) - (\alpha\mathcal{I} - \mathcal{S}), \qquad \mathcal{A} = (\mathcal{S} + \alpha\mathcal{I}) - (\alpha\mathcal{I} - \mathcal{H}),$$

where $\mathcal{I}$ denotes the $(n+m)$-by-$(n+m)$ identity matrix. Note that

$$\mathcal{H} + \alpha\mathcal{I} = \begin{bmatrix} A + \alpha I_n & O \\ O & \alpha I_m \end{bmatrix} \quad \text{and} \quad \mathcal{S} + \alpha\mathcal{I} = \begin{bmatrix} \alpha I_n & B^* \\ -B & \alpha I_m \end{bmatrix}$$

are both nonsingular matrices. The HSS algorithm is obtained by alternating between these two splittings. Given an initial guess $\mathbf{x}^0 = (u^0, p^0)$, the Hermitian/skew-Hermitian iteration computes a sequence $\{\mathbf{x}^\ell\}$ as follows:

$$\begin{cases} (\mathcal{H} + \alpha\mathcal{I})\mathbf{x}^{\ell+\frac{1}{2}} = (\alpha\mathcal{I} - \mathcal{S})\mathbf{x}^\ell + \mathbf{b}, \\ (\mathcal{S} + \alpha\mathcal{I})\mathbf{x}^{\ell+1} = (\alpha\mathcal{I} - \mathcal{H})\mathbf{x}^{\ell+\frac{1}{2}} + \mathbf{b}. \end{cases} \tag{2.1}$$

The first half-step of algorithm (2.1) requires the solution of linear systems of the form

$$(A + \alpha I_n)u^{\ell+\frac{1}{2}} = \alpha u^\ell + f - B^* p^\ell. \tag{2.2}$$

Once the solution $u^{\ell+\frac{1}{2}}$ of (2.2) has been obtained, we compute

$$p^{\ell+\frac{1}{2}} = p^\ell + \frac{1}{\alpha}(Bu^\ell - g).$$

Note that the coefficient matrix in (2.2) is Hermitian positive definite (HPD); furthermore, for PDE problems, the matrix $A + \alpha I$ is typically well-conditioned independent of the mesh size h. Indeed, if we normalize A so that its largest eigenvalue is $\lambda_{\max}(A) = 1$, then for the spectral condition number of $A + \alpha I$ we have the upper bound

$$\kappa(A + \alpha I) = \frac{1 + \alpha}{\lambda_{\min}(A) + \alpha} \leq 1 + \frac{1}{\alpha},$$

independent of the size of the problem. System (2.2) can be solved by any method for HPD systems, like a sparse Cholesky factorization or the conjugate gradient (CG) algorithm.

The second half-step of algorithm (2.1) requires the solution of a linear system of the form

$$\begin{cases} \alpha u^{\ell+1} + B^* p^{\ell+1} = (\alpha I_n - A)u^{\ell+\frac{1}{2}} + f \equiv f^\ell, \\ -Bu^{\ell+1} + \alpha p^{\ell+1} = \alpha p^{\ell+\frac{1}{2}} - g \equiv g^\ell. \end{cases} \tag{2.3}$$

This linear system can be solved in various ways, including the CG-like method for shifted skew-Hermitian systems described in [7], or using a Schur complement reduction to eliminate $u^{\ell+1}$ from (2.3), leading to a Hermitian positive definite linear system in m unknowns of the form

$$(BB^* + \alpha^2 I_m)p^{\ell+1} = Bf^\ell + \alpha g^\ell, \tag{2.4}$$

after which we compute $u^{\ell+1} = \frac{1}{\alpha}(f^\ell - B^* p^{\ell+1})$. Note that if α is sufficiently large, the coefficient matrix in (2.4) becomes diagonally dominant, with condition number bounded independent of h, and an iterative method like CG applied to (2.4) can be expected to converge rapidly. As an alternative to CG, system (2.4) can be efficiently and accurately solved by the LSQR algorithm [13].

Also note that when the HSS iteration is used as a preconditioner, there is no need to solve the systems (2.2) and (2.3) exactly. Generally speaking, using inexact solves instead of exact ones makes this approach more competitive [1]. In this paper, however, we limit ourselves to the case where the linear systems (2.2) and (2.3) are solved to high accuracy, as this greatly simplifies the analysis.

To analyze the convergence of (2.1), we eliminate the intermediate vector $\mathbf{x}^{\ell+\frac{1}{2}}$ and write the iteration in fixed point form as

$$\mathbf{x}^{\ell+1} = \mathcal{T}_\alpha \mathbf{x}^\ell + \mathbf{c} \tag{2.5}$$

where

$$\mathcal{T}_\alpha := (\mathcal{S} + \alpha\mathcal{I})^{-1}(\alpha\mathcal{I} - \mathcal{H})(\mathcal{H} + \alpha\mathcal{I})^{-1}(\alpha\mathcal{I} - \mathcal{S}) \tag{2.6}$$

is the iteration matrix of the method, and

$$\mathbf{c} := (\mathcal{S} + \alpha\mathcal{I})^{-1}[\mathcal{I} + (\alpha\mathcal{I} - \mathcal{H})(\mathcal{H} + \alpha\mathcal{I})^{-1}]\mathbf{b}.$$

It is easy to verify that if we set

$$\mathcal{M}_\alpha := \frac{1}{2\alpha}(\mathcal{H} + \alpha\mathcal{I})(\mathcal{S} + \alpha\mathcal{I}),$$

then we can rewrite the iteration (2.1) in *correction form*,

$$\mathbf{x}^{\ell+1} = \mathbf{x}^\ell + \mathcal{M}_\alpha^{-1}\mathbf{r}^\ell, \quad \mathbf{r}^\ell = \mathbf{b} - \mathcal{A}\mathbf{x}^\ell.$$

Note that $\mathcal{T}_\alpha = \mathcal{I} - \mathcal{M}_\alpha^{-1}\mathcal{A}$. Furthermore, the use of a Krylov subspace method to accelerate the convergence of the HSS iteration is equivalent to the application of the Krylov method to the preconditioned system $\mathcal{M}_\alpha^{-1}\mathcal{A}\mathbf{x} = \mathcal{M}_\alpha^{-1}\mathbf{b}$ in the case of left preconditioning, or to $\mathcal{A}\mathcal{M}_\alpha^{-1}\mathbf{y} = \mathbf{b}$, $\mathbf{y} = \mathcal{M}_\alpha\mathbf{x}$ for right preconditioning; see [3].

The fixed point iteration (2.5) converges for arbitrary initial guesses $\mathbf{x}^0$ and right-hand sides $\mathbf{b}$ to the solution $\mathbf{x} = \mathcal{A}^{-1}\mathbf{b}$ if and only if $\rho(\mathcal{T}_\alpha) < 1$. It is shown in [3] that if the sub-matrix A in (1.1) is HPD, then $\rho(\mathcal{T}_\alpha) < 1$ for all $\alpha > 0$: the iteration is unconditionally convergent.

The parameter α should be chosen so as to maximize the rate of convergence of the iterates. Since the asymptotic rate of convergence is governed by the spectral radius of $\mathcal{T}_\alpha$, it makes sense to try to choose α so as to make $\rho(\mathcal{T}_\alpha)$ as small as possible. In general, this is a difficult problem. In the next section we analyze the iteration at the continuous level for a model problem, in an attempt to gain insight into the choice of α. We also estimate what kind of asymptotic rate of convergence we can expect.

3 Analysis at the continuous level

We focus our attention on the Poisson equation

$$\Delta p \equiv \text{div grad } p = g. \tag{3.1}$$

The solution p and the right-hand side g in (3.1) are real-valued functions defined on an open subset $\Omega \subseteq \mathbb{R}^d$ (here $d = 1, 2$). The differential equation is complemented by appropriate boundary conditions.

In order to cast the Poisson equation (3.1) in saddle-point form we rewrite it as a first-order system, as is customary in mixed finite elements. Introducing the vector-valued function $\mathbf{u} = \text{grad } p$, equation (3.1) can be written as the system

$$\mathbf{u} - \text{grad } p = \mathbf{0}, \quad \text{div } \mathbf{u} = g$$

which in matrix form becomes a saddle point problem of type (1.1), namely

$$\begin{bmatrix} 1 & -\mathrm{grad} \\ \mathrm{div} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ g \end{bmatrix}. \tag{3.2}$$

We can now formally consider the application of the HSS iteration to the saddle point problem (3.2), when it is written in non-self-adjoint form

$$\begin{bmatrix} 1 & -\mathrm{grad} \\ -\mathrm{div} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -g \end{bmatrix}. \tag{3.3}$$

We start with the Poisson equation written in the form (3.3) in one spatial dimension,

$$\begin{bmatrix} 1 & -\partial_x \\ -\partial_x & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ -g \end{bmatrix} \tag{3.4}$$

on the infinite domain $x \in \mathbb{R}$. Note that although this system looks deceptively symmetric, it is not self-adjoint, since the adjoint of ∂_x is $-\partial_x$. Taking a Fourier transform in the x-direction with the Fourier parameter k, we obtain the transformed system

$$\begin{bmatrix} 1 & -ik \\ -ik & 0 \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{g} \end{bmatrix} \tag{3.5}$$

for which the (continuous) HSS iteration matrix according to (2.6) is

$$\hat{\mathcal{T}}_\alpha = \frac{1}{(k^2+\alpha^2)(1+\alpha)} \begin{bmatrix} -(1+\alpha)k^2 - \alpha^2 + \alpha^3 & 2k\alpha^2 i \\ 2k\alpha^2 i & (1-\alpha)k^2 + \alpha^2 + \alpha^3 \end{bmatrix}.$$

This iteration matrix has for each frequency parameter value k the eigenvalues

$$\lambda_{1,2}(\alpha, k) = \frac{\alpha^3 - \alpha k^2 \pm \sqrt{k^4 + 2\alpha^2(1-2\alpha^2)k^2 + \alpha^4}}{(k^2+\alpha^2)(1+\alpha)} \tag{3.6}$$

(where λ_1 corresponds to taking the '+' sign in (3.6)) and thus the spectral radius $\rho(\alpha, k) := \max(|\lambda_1(\alpha, k)|, |\lambda_2(\alpha, k)|)$.

To optimize the asymptotic performance of the stationary HSS iteration, we need to determine the parameter α which minimizes the spectral radius over all relevant frequency parameters k of the numerical computation. We first note that ρ depends on k^2 only and thus it is sufficient to analyze the frequencies $k \geq 0$. In addition on a numerical grid, k cannot vary arbitrarily, since it is bounded from above by a maximum frequency $k_{\max}$ which can be estimated by $k_{\max} = \pi/h$ where h is the mesh parameter. It can also be bounded from below by $k_{\min} > 0$ ($k_{\min} = 0$ would make the problem singular); for example if the domain is bounded and has homogeneous Dirichlet conditions, a sine expansion would have the lowest mode $\sin(\pi x/L)$ where L is the length of the domain, and thus $k_{\min} = \pi/L$. Other boundary conditions lead to similar estimates, for example for homogeneous Dirichlet on one side and homogeneous Neumann on

the other, one could use $k_{\min} = \pi/(2L)$. Hence to optimize the performance of the HSS algorithm, we have to solve the min-max problem

$$\min_{\alpha>0} \left(\max_{k_{\min} \le k \le k_{\max}} \rho(\alpha, k) \right). \tag{3.7}$$

To solve this min-max problem, we need several Lemmas. The first two are about monotonicity properties of the functions $\lambda_j(\alpha, k)$, $j = 1, 2$.

LEMMA 3.1. *For a given $k \in (0, \frac{1}{2}]$ and $\alpha \ge \sqrt{k}$, $\lambda_1(\alpha, k)$ is a real and strictly increasing function of α.*

PROOF. Factoring the discriminant d in (3.6), we obtain

$$d = (k^2 + \alpha^2(1+2k))(k^2 + \alpha^2(1-2k)) > 0, \quad \text{for } 0 < k \le \tfrac{1}{2} \tag{3.8}$$

and hence λ_1 is a real function for all $\alpha > 0$. Taking a partial derivative with respect to α, we obtain

$$\frac{\partial \lambda_1}{\partial \alpha} = \frac{(4\alpha^3k^2+\alpha^4-k^4+4\alpha^2k^2)\sqrt{d} - (3\alpha^4k^2+\alpha^6+8\alpha^3k^4+4\alpha^4k^4-4\alpha^6k^2+3k^4\alpha^2+k^6)}{(1+\alpha)^2(\alpha^2+k^2)^2\sqrt{d}}.$$

To show that this derivative is positive, it suffices to show that the numerator is positive. Since $\alpha \ge \sqrt{k}$ and $k \le \frac{1}{2}$, the factor in front of $\sqrt{d}$ in the numerator is positive, because $\alpha^4 - k^4 \ge k^2(1-k^2) > 0$. Taking the square of the first term and subtracting the square of the second term in the numerator leads after simplifying to the expression

$$4\alpha^2k^2(1+\alpha)^2(\alpha^2+k^2)^2((1-4k^2)\alpha^4 + (2\alpha^2-3k^2)k^2) > 0,$$

which is positive, because $1 - 4k^2 \ge 0$ for $k \le \frac{1}{2}$ and with $\alpha \ge \sqrt{k}$ we have $2\alpha^2 - 3k^2 \ge k(2-3k) > 0$ for $k \le \frac{1}{2}$. Hence the numerator in the derivative is positive and thus λ_1 is a strictly increasing function of α. ∎

LEMMA 3.2. *For a given $k \ge 1$ and $0 < \alpha < \frac{k}{\sqrt{2k-1}}$, $\lambda_2(\alpha, k)$ is a real and strictly increasing function of α.*

PROOF. From the second factor of the discriminant d given in (3.8) we see that $d > 0$ with the assumption on α in the lemma, and hence λ_2 is real. Taking a derivative with respect to α of λ_2 in this range of α leads to

$$\frac{\partial \lambda_2}{\partial \alpha} = \frac{(4\alpha^3k^2+\alpha^4-k^4+4\alpha^2k^2)\sqrt{d} + (3\alpha^4k^2+3k^4\alpha^2+8\alpha^3k^4+\alpha^6-4\alpha^6k^2+k^6+4\alpha^4k^4)}{(\alpha^2+\alpha^3+k^2\alpha+k^2)^2\sqrt{d}}.$$

Since the denominator is positive, it suffices to show that the numerator is positive. Now the second term in the numerator is positive, since

$$\alpha^4k^4 - \alpha^6k^2 = \alpha^4k^2(k-\alpha)(k+\alpha) > 0$$

is positive with $k \ge 1$ and $\alpha < \frac{k}{\sqrt{2k-1}} \le k$. Subtracting the first term squared in the numerator from the second term squared leads to

$$4\alpha^2k^2(1+\alpha)^2(\alpha^2+k^2)^2(-\alpha^4 + 4\alpha^4k^2 - 2\alpha^2k^2 + 3k^4) > 0,$$

which is positive, since $\alpha < k$, which shows that the numerator is positive and therefore λ_2 is a strictly increasing function of α. ☐

The next Lemma gives the precise expression of the spectral radius $\rho(\alpha, k)$ for all possible parameters α and k.

LEMMA 3.3. *If $0 < \alpha \leq 1$, then the spectral radius is given by*

$$\rho(\alpha, k) = \begin{cases} \lambda_1(\alpha, k) & \text{if } k \leq \alpha, \\ -\lambda_2(\alpha, k) & \text{if } k \geq \alpha. \end{cases}$$

If $\alpha > 1$, then the spectral radius is given by

$$\rho(\alpha, k) = \begin{cases} \lambda_1(\alpha, k) & \text{if } k \leq \alpha(\alpha - \sqrt{\alpha^2 - 1}), \\ \sqrt{\frac{\alpha-1}{\alpha+1}} & \text{if } \alpha(\alpha - \sqrt{\alpha^2 - 1}) \leq k \leq \alpha(\alpha + \sqrt{\alpha^2 - 1}), \\ -\lambda_2(\alpha, k) & \text{if } k \geq \alpha(\alpha + \sqrt{\alpha^2 - 1}). \end{cases}$$

PROOF. For $0 < \alpha \leq 1$, the discriminant d in (3.6) satisfies

$$d = k^4 + 2\alpha^2(1 - 2\alpha^2)k^2 + \alpha^4 \geq 0$$

and hence both eigenvalues are real for all k. If $k \leq \alpha$, then the common term in the eigenvalues is non-negative, and hence $\lambda_1 > 0$ is the larger eigenvalue in modulus, which determines the spectral radius, whereas for $k \geq \alpha$ the common term in the eigenvalues is non-positive, and hence $\lambda_2 < 0$ is the larger eigenvalue in modulus, and $-\lambda_2$ gives the spectral radius.

If $\alpha > 1$, then the discriminant d is non-positive for $\alpha(\alpha - \sqrt{\alpha^2 - 1}) \leq k \leq \alpha(\alpha + \sqrt{\alpha^2 - 1})$. In that case the spectral radius is given by the modulus $|\lambda_1| = |\lambda_2| = \sqrt{\frac{\alpha-1}{\alpha+1}} < 1$ independent of k. For $k \leq \alpha(\alpha - \sqrt{\alpha^2 - 1})$ the discriminant is non-negative and since

$$\frac{1}{2} < \alpha(\alpha - \sqrt{\alpha^2 - 1}) < 1 \quad \text{for } \alpha > 1 \tag{3.9}$$

we have by a similar argument as above that the spectral radius is given by λ_1, whereas for $k \geq \alpha + \sqrt{\alpha^2 - 1}$ the discriminant is also non-negative and the spectral radius is given by $-\lambda_2$. ☐

The following Lemma is essential for the solution of the min-max problem, because it shows that the maximum cannot be attained in the interior of the frequency domain $[k_{\min}, k_{\max}]$.

LEMMA 3.4. *For any given $\alpha > 0$, the spectral radius $\rho(\alpha, k)$ attains its maximum in k on the boundary either at $k_{\min}$ or $k_{\max}$.*

PROOF. When λ_1 is real, a partial derivative of $\lambda_1(\alpha, k)$ with respect to k gives

$$\frac{\partial \lambda_1}{\partial k} = -\frac{4\alpha^3 k(\alpha^3 - k^2\alpha + \sqrt{d})}{(\alpha^2 + k^2)^2(1 + \alpha)\sqrt{d}}.$$

This expression is negative in both cases of Lemma 3.3 where the spectral radius $\rho = \lambda_1$, and thus ρ is decreasing when k is increasing in these cases. Similarly a

partial derivative of $\lambda_2(\alpha,k)$ with respect to k gives

$$\frac{\partial \lambda_2}{\partial k} = \frac{4\alpha^3 k(\alpha^3 - k^2\alpha - \sqrt{d})}{(\alpha^2+k^2)^2(1+\alpha)\sqrt{d}}.$$

This expression is negative in both cases of Lemma 3.3 where the spectral radius $\rho = -\lambda_2$, and thus ρ is increasing when k is increasing in these cases. Hence if $0 < \alpha \leq 1$, $\rho(\alpha,k)$ is first a decreasing function of k until $k = \alpha$, and then an increasing function of k. It can therefore only attain its maximum on the boundary, at $k_{\min}$ or $k_{\max}$. If $\alpha > 1$, then $\rho(\alpha,k)$ is first a decreasing function of k, until the discriminant becomes negative, then ρ stays constant, independent of k, until the discriminant becomes positive again and then ρ is an increasing function of k. Therefore also for $\alpha > 1$ the spectral radius ρ can only attain its maximum on the boundary, at $k_{\min}$ or $k_{\max}$. □

To solve the min-max problem (3.7), it therefore suffices to analyze the spectral radius $\rho(\alpha,k)$ on the boundaries at $k_{\min}$ and $k_{\max}$. We denote these quantities in the sequel by $r_1(\alpha) := \rho(\alpha,k_{\min})$ and $r_2(\alpha) := \rho(\alpha,k_{\max})$.

LEMMA 3.5. *For $k_{\max} \geq 1$ and $0 < \alpha \leq \frac{k_{\max}}{\sqrt{2k_{\max}-1}}$ the function $r_2(\alpha)$ is a strictly decreasing function of α.*

PROOF. First we show that under the conditions of the lemma, $r_2(\alpha) = -\lambda_2(\alpha,k_{\max})$. For $0 < \alpha \leq 1$, we have by Lemma 3.3 that $r_2(\alpha) = -\lambda_2(\alpha,k_{\max})$, since $k_{max} \geq 1 \geq \alpha$. For $1 < \alpha \leq \frac{k_{\max}}{\sqrt{2k_{\max}-1}}$, the condition $\alpha \leq \frac{k_{\max}}{\sqrt{2k_{\max}-1}}$ together with $k_{\max} \geq 1$ implies that $k_{\max} \geq \alpha(\alpha + \sqrt{\alpha^2-1})$ and thus again by Lemma 3.3 we have $r_2(\alpha) = -\lambda_2(\alpha,k_{\max})$. Now by Lemma 3.2 we have that λ_2 is a real, strictly increasing function of α, and thus $r_2(\alpha) = -\lambda_2(\alpha,k_{\max})$ is a real and strictly decreasing function of α. □

LEMMA 3.6. *For $k_{\max} \geq 1$ and $\alpha > \frac{k_{\max}}{\sqrt{2k_{\max}-1}}$ the function $r_2(\alpha)$ is a strictly increasing function of α.*

PROOF. Using $\alpha > \frac{k_{\max}}{\sqrt{2k_{\max}-1}} \geq 1$ and Lemma 3.3, we have $r_2(\alpha) = \sqrt{\frac{\alpha-1}{\alpha+1}}$ and since

$$\frac{d}{d\alpha}\sqrt{\frac{\alpha-1}{\alpha+1}} = \frac{1}{(1+\alpha)\sqrt{\alpha^2-1}} > 0,$$

r_2 is a strictly increasing function of α. □

LEMMA 3.7. *If $k_{\max} \geq 1$ and $\frac{k_{\max}}{2k_{\max}-1} \leq k_{\min} < k_{\max}$ then for $\alpha^* = \frac{k_{\max}}{\sqrt{2k_{\max}-1}}$ we have $r_1(\alpha^*) = r_2(\alpha^*)$.*

PROOF. By Lemma 3.3 we have that $r_2(\alpha^*) = \sqrt{\frac{\alpha^*-1}{\alpha^*+1}}$. It suffices therefore to show that under the conditions of the lemma $r_1(\alpha^*)$ also equals this value. Since $\alpha^* \geq 1$ with $k_{\max} \geq 1$, the second case of Lemma 3.3 applies to r_1. Now $k_{\min} < k_{\max} = \alpha^*(\alpha^* + \sqrt{(\alpha^*)^2-1})$ and also

$$k_{\min} \geq \frac{k_{\max}}{2k_{\max}-1} = \alpha^*(\alpha^* - \sqrt{(\alpha^*)^2-1}),$$

where the last equality follows by a short calculation, which implies by Lemma 3.3 that also $r_1(\alpha^*) = \sqrt{\frac{\alpha^*-1}{\alpha^*+1}}$, concluding the proof. □

THEOREM 3.8. *If $k_{\max} \geq 1$ and $\frac{k_{\max}}{2k_{\max}-1} \leq k_{\min} < k_{\max}$, then the optimal parameter of the HSS iterative method for the one-dimensional Poisson equation (3.4) and the optimized convergence rate are given by*

$$\alpha^* = \frac{k_{\max}}{\sqrt{2k_{\max}-1}}, \quad \rho^* = \frac{k_{\max}-1}{k_{\max}+\sqrt{2k_{\max}-1}}. \tag{3.10}$$

PROOF. The optimal parameter is defined as the solution of the min-max problem (3.7). By Lemma 3.4 the maximum is attained on the boundary. Under the conditions given in the theorem, Lemma 3.5 and Lemma 3.6 imply that the minimum of $r_2(\alpha)$ is attained at α^*. Hence at the solution of the min-max problem, the optimal spectral radius can only be bigger than or equal to $r_2(\alpha^*)$. But at α^* we have under the conditions of the theorem that $r_1(\alpha^*) = r_2(\alpha^*)$, and hence α^* is the unique solution of the min-max problem. ∎

This first theorem gives the optimal solution for values of $k_{\min}$ which are bigger than $\frac{k_{\max}}{2k_{\max}-1}$, which for large $k_{\max}$ goes to $\frac{1}{2}$. To find the solution of the min-max problem for smaller $k_{\min}$, we need three more technical lemmas.

LEMMA 3.9. *For $0 < k_{\min} \leq \frac{1}{2}$ we have*

$$r_1(\alpha) = \begin{cases} \lambda_1(\alpha, k_{\min}), & \alpha \geq k_{\min}, \\ -\lambda_2(\alpha, k_{\min}), & 0 < \alpha \leq k_{\min} \end{cases}$$

PROOF. For $0 < \alpha \leq 1$ the result holds by Lemma 3.3. For $\alpha > 1$ we have $k_{\min} \leq \frac{1}{2} < \alpha(\alpha - \sqrt{\alpha^2 - 1})$ using (3.9) and therefore by Lemma 3.3 again $r_1(\alpha) = \lambda_1(\alpha, k_{\min})$, which concludes the proof. ∎

LEMMA 3.10. *For $0 < k_{\min} \leq \frac{1}{2}$, $k_{\max} \geq 1$ and $0 < \alpha \leq \sqrt{k_{\min}k_{\max}}$ the function $r_2(\alpha)$ is strictly decreasing when α is increasing.*

PROOF. With $k_{\min} \leq \frac{1}{2}$ we have

$$\sqrt{k_{\min}k_{\max}} \leq \sqrt{\frac{k_{\max}}{2}} = \frac{k_{\max}}{\sqrt{2k_{\max}}} \leq \frac{k_{\max}}{\sqrt{2k_{\max}-1}}$$

and therefore by Lemma 3.5 the result follows. ∎

LEMMA 3.11. *For $0 < k_{\min} \leq \frac{1}{2}$, $k_{\max} \geq 1$ and $\alpha > \sqrt{k_{\min}k_{\max}}$ the function $r_1(\alpha)$ is strictly increasing when α is increasing.*

PROOF. With $k_{\max} \geq 1$ we have $\alpha > \sqrt{k_{\min}k_{\max}} \geq \sqrt{k_{\min}}$. For $\alpha > 1$, we have with (3.9) that $k_{\min} \leq \frac{1}{2} \leq \alpha(\alpha - \sqrt{\alpha^2 - 1})$. Hence by Lemma 3.3 we obtain $r_1(\alpha) = \lambda_1(\alpha, k_{\min})$ for any $\alpha > \sqrt{k_{\min}k_{\max}}$. Applying Lemma 3.1 the result follows. ∎

THEOREM 3.12. *If $k_{\max} \geq 1$ and $0 < k_{\min} \leq \frac{1}{2}$, then the optimal parameter of the HSS preconditioner for the one-dimensional Poisson equation (3.4) is $\alpha^* = \sqrt{k_{\min}k_{\max}}$ and the optimized convergence rate is*

$$\rho^* = \frac{(k_{\max}-k_{\min})\sqrt{k_{\max}k_{\min}} + \sqrt{(k_{\max}+k_{\min})^2 - 4k_{\max}^2k_{\min}^2}}{(k_{\max}+k_{\min})(1+\sqrt{k_{\max}k_{\min}})}. \tag{3.11}$$

PROOF. The optimal parameter is defined as the solution of the min-max problem (3.7). We first note that at α^* we have $r_1(\alpha^*) = r_2(\alpha^*)$. Since for $\alpha > \alpha^*$ Lemma 3.11 shows that $r_1(\alpha) > r_1(\alpha^*)$ and for $\alpha < \alpha^*$ Lemma 3.10 shows that $r_2(\alpha) > r_2(\alpha^*)$ the proof is complete. ∎

COROLLARY 3.13. *The optimal parameter α^* and the optimized convergence rate of the HSS stationary iteration for the one-dimensional Poisson problem behave in the case of Theorem 3.8 asymptotically for small mesh parameter h like*

$$\alpha^* = \sqrt{\frac{\pi}{2}}\frac{1}{\sqrt{h}} + O(\sqrt{h}), \quad \rho^* = 1 - \sqrt{\frac{2}{\pi}}\sqrt{h} + O(h). \tag{3.12}$$

In the case of Theorem 3.12, we have

$$\alpha^* = \sqrt{\pi k_{\min}}\frac{1}{\sqrt{h}}, \quad \rho^* = 1 - \frac{1-\sqrt{1-4k_{\min}^2}}{\sqrt{\pi k_{\min}}}\sqrt{h} + O(h). \tag{3.13}$$

PROOF. The results are obtained by using the estimate $k_{\max} = \pi/h$ and expanding for small h. ∎

If HSS is used as a preconditioner for a Krylov method, one could use the same optimized parameter derived for the stationary method, since minimizing the spectral radius corresponds to clustering the eigenvalues of the preconditioned operator around 1. But for HSS there is a better option: one can form two very tight separate clusters with an appropriate choice of the optimization parameter α.

LEMMA 3.14. *For $0 < \alpha < 1$ the eigenvalues $\lambda_j(\alpha, k)$, $j = 1, 2$ and $k \in [k_{\min}, k_{\max}]$, are real and contained in the two disjoint intervals $[a_1, b_1]$ and $[a_2, b_2]$, where, independently of $k_{\max}$, we have*

$$a_1 = \frac{1-\alpha}{1+\alpha},\ b_1 = \lambda_1(\alpha, k_{\min}), \quad a_2 = -1,\ b_2 = \lambda_2(\alpha, k_{\min}).$$

PROOF. The eigenvalues are real because the discriminant $d \geq 0$ for $0 < \alpha < 1$ and the intervals are, due to the monotonicity result shown in Lemma 3.4, $[\lambda_j(\alpha, k_{\max}), \lambda_j(\alpha, k_{\min})]$, $j = 1, 2$. In the lower bound, we can take the limit as $k_{\max}$ goes to infinity to get intervals independent of $k_{\max}$, because these limits are finite, which leads to the a_j given in the lemma. Finally the intervals are disjoint, because for $0 < \alpha < 1$ the λ_j have opposite signs, since their product equals the constant term in the quadratic determining the eigenvalues, which is $(\alpha - 1)(\alpha + 1)(\alpha^2 + k^2)^2 < 0$. ∎

LEMMA 3.15. *For $0 < \alpha < 1$ there exists a polynomial $p(x)$ of degree 2 which satisfies $p(0) = 1$ and on the shifted intervals given in Lemma 3.14*

$$\max_{x \in (1-[a_1,b_1]) \cup (1-[a_2,b_2])} |p(x)| \leq q(\alpha, k_{\min}),$$

where $q(\alpha, k_{\min})$ is explicitly known and satisfies $q(\alpha, k_{\min}) = \frac{1}{2k_{\min}^2}\alpha^2 + O(\alpha^3)$.

PROOF. Let $p(x)$ be the quadratic interpolation polynomial at the three points $(0,1)$, $(1-\frac{a_1+b_1}{2},0)$ and $(1-\frac{a_2+b_2}{2},0)$. Since this polynomial is convex, it will attain its maximum on the two shifted intervals either at $1-b_1$ or $1-a_2$, where it has the values

$$\begin{aligned} p_1 &:= p(1-b_1) = \frac{(2b_1-a_2-b_2)(b_1-a_1)}{(a_1+b_1-2)(a_2+b_2-2)} \geq 0, \\ p_2 &:= p(1-a_2) = \frac{(2a_2-a_1-b_1)(a_2-b_2)}{(a_1+b_1-2)(a_2+b_2-2)} \geq 0. \end{aligned}$$

Using now the definitions of a_j and b_j, $j=1,2$, from Lemma 3.14, we find that p_1 and p_2 depend on α and $k_{\min}$, and thus the maximum of the polynomial is bounded by $q(\alpha,k_{\min}) := \max(p_1,p_2)$. Expanding in α, we find

$$p_1 = \frac{1}{2k_{\min}^2}\alpha^2 - \frac{1}{k_{\min}^2}\alpha^3 + O(\alpha^4), \quad p_2 = \frac{1}{2k_{\min}^2}\alpha^2 - \frac{1}{4k_{\min}^4}\alpha^4 + O(\alpha^5),$$

which concludes the proof. □

THEOREM 3.16. *GMRES applied to the one-dimensional Poisson equation written in div-grad form preconditioned with HSS converges in at most two steps to a given small tolerance τ if the optimization parameter satisfies $0 < \alpha < \min(\bar{\alpha},1)$, where $\bar{\alpha}$ satisfies $q(\bar{\alpha},k_{\min}) = \tau$, $\bar{\alpha} \approx \sqrt{2\tau}k_{\min}$.*

PROOF. Since GMRES finds at each step the smallest polynomial in modulus on the spectrum of the operator, it will find in its second step a polynomial which is at least as small in modulus as $p(x)$ from Lemma 3.15. Hence the GMRES polynomial is for a given α smaller than $q(\alpha,k_{\min})$ from Lemma 3.15. Now if $\alpha \leq \bar{\alpha}$, where $\bar{\alpha}$ is defined by $q(\bar{\alpha},k_{\min}) = \tau$, then by Lemma 3.15 the polynomial $p(x)$ is smaller than τ in modulus, which completes the proof. □

This result implies that α can be chosen so as to have h-independent convergence. An illustration of the dependence of the spectrum of the preconditioned operator on the optimization parameter can be found in Figure 3.1, where both the continuous Fourier spectrum and the discrete spectrum from the discretization in the numerical section are shown. First, one can see that the Fourier analysis predicts extremely well the spectrum of the discretized preconditioned operator, and thus we can expect the results of the analysis to be decisive for the discretized problem. Second, the sequence of images shows clearly that the optimal value of α for HSS as an iterative method is very different from the optimal value of α for HSS used as a preconditioner for GMRES. For the iterative application, one needs to minimize the spectral radius, and the optimal α puts the eigenvalues of the preconditioner symmetrically onto a circle, whereas for GMRES, it is best to produce two tight clusters for performance.

For the Poisson equation in two dimensions, we start again with the equation written in non-self-adjoint, div-grad form,

$$\begin{bmatrix} 1 & 0 & -\partial_x \\ 0 & 1 & -\partial_y \\ -\partial_x & -\partial_y & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix} \tag{3.14}$$

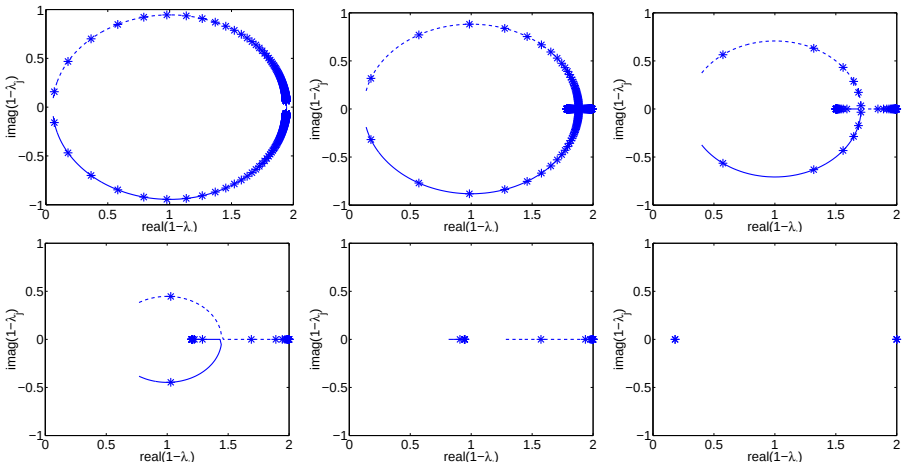

Figure 3.1: The spectrum of the HSS preconditioned 1d-Poisson problem. The solid line is $1-\lambda_1(\alpha,k)$, the dashed line $1-\lambda_2(\alpha,k)$ and the stars are the eigenvalues of the discretized preconditioned operator with mesh parameter $h=1/200$ described in the numerical section. From top left to bottom right, we have $\alpha = 17.73, 8, 3, 1.5, 0.9, 0.1$, where the first value is optimal for the iterative HSS, and the last one is good for HSS as a preconditioner for GMRES, since there are two tight clusters.

on the infinite domain $(x,y)\in\mathbb{R}^2$. Taking a Fourier transform in the x and y-direction with the Fourier parameters k and l, we obtain the transformed system

$$\begin{bmatrix} 1 & 0 & -ik \\ 0 & 1 & -il \\ -ik & -il & 0 \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{v} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\hat{g} \end{bmatrix} \tag{3.15}$$

for which the HSS iteration matrix according to (2.6) is

$$\hat{\mathcal{T}}_\alpha = \frac{\begin{bmatrix} (\alpha-1)(\alpha^2+l^2)-(\alpha+1)k^2 & -2kl\alpha & 2ik\alpha^2 \\ -2kl\alpha & (\alpha-1)(\alpha^2+k^2)-(\alpha+1)l^2 & 2il\alpha^2 \\ 2ik\alpha^2 & 2il\alpha^2 & (1-\alpha)(k^2+l^2)+\alpha^2+\alpha^3 \end{bmatrix}}{(1+\alpha)(\alpha^2+k^2+l^2)}.$$

The iteration matrix $\hat{\mathcal{T}}_\alpha$ has for each frequency parameter value k the eigenvalues

$$\lambda_{1,2} = \frac{\alpha^3-\alpha(k^2+l^2)\pm\sqrt{(k^2+l^2)^2+2\alpha^2(1-2\alpha^2)(k^2+l^2)+\alpha^4}}{(1+\alpha)(\alpha^2+k^2+l^2)},\ \lambda_3 = \frac{\alpha-1}{\alpha+1}$$

and the spectral radius $\rho(\alpha,k,l) := \max(|\lambda_1(\alpha,k,l)|, |\lambda_2(\alpha,k,l)|, |\lambda_3(\alpha,k,l)|)$.

THEOREM 3.17. *Let* $K_{\max} := \sqrt{k_{\max}^2+l_{\max}^2} \geq 1$. *If* $\frac{K_{\max}}{\sqrt{2K_{\max}-1}} \leq K_{\min} := \sqrt{k_{\min}^2+l_{\min}^2} < K_{\max}$, *then the optimal parameter of the HSS preconditioner for*

the two-dimensional Poisson equation (3.14) and the optimized convergence rate are

$$\alpha^* = \frac{K_{\max}}{\sqrt{2K_{\max}-1}}, \quad \rho^* = \frac{K_{\max}-1}{K_{\max}+\sqrt{2K_{\max}-1}}. \tag{3.16}$$

If $0 < K_{\min} \leq \frac{1}{2}$*, then we have* $\alpha^* = \sqrt{K_{\min}K_{\max}}$ *and*

$$\rho^* = \frac{(K_{\max}-K_{\min})\sqrt{K_{\max}K_{\min}}+\sqrt{(K_{\max}+K_{\min})^2-4K_{\max}^2K_{\min}^2}}{(K_{\max}+K_{\min})(1+\sqrt{K_{\max}K_{\min}})}. \tag{3.17}$$

PROOF. If we can show that the modulus of the additional eigenvalue λ_3 is always bounded by the modulus of the other two, then the proof in one dimension is still valid, since the eigenvalues λ_1 and λ_2 coincide with the one-dimensional ones after the change of variable $K = \sqrt{k^2+l^2}$. To see that the modulus of λ_3 is bounded by the modulus of λ_1 for all K and α, we have to distinguish three cases. First for $0 < \alpha \leq 1$ both λ_1 and λ_3 are negative and the difference in modulus is

$$-\lambda_1+\lambda_3 = \frac{-\alpha^2+2\alpha K^2-K^2+\sqrt{-4\alpha^4K^2+K^4+\alpha^4+2\alpha^2K^2}}{(1+\alpha)(\alpha^2+K^2)}$$

which is positive, since

$$-4\alpha^4K^2+K^4+\alpha^4+2\alpha^2K^2-(-\alpha^2+2\alpha K^2-K^2)^2 = -4\alpha K^2(-1+\alpha)(\alpha^2+K^2) \geq 0$$

for $0 < \alpha \leq 1$. The next case is $1 < \alpha \leq \frac{K}{\sqrt{2K-1}}$, since for these values of α the eigenvalue λ_1 is still real. Now $\lambda_3 > 0$ and $\lambda_1 < 0$, and their difference

$$-\lambda_1-\lambda_3 = \frac{-2\alpha^3+\alpha^2+K^2+\sqrt{-4\alpha^4K^2+K^4+\alpha^4+2\alpha^2K^2}}{(1+\alpha)(\alpha^2+K^2)}$$

is also positive, since the term in front of the square root, $e(\alpha) := -2\alpha^3+\alpha^2+K^2$ is positive for $1 < \alpha \leq \frac{K}{\sqrt{2K-1}}$, as one can see by first checking the two endpoints, $e(1) = K^2-1 > 0$ since we must have $K > 1$ for the second case to exist, and $e(\frac{K}{\sqrt{2K-1}}) = 2\frac{(-1+\sqrt{2K-1})K^3}{(2K-1)^{(3/2)}} > 0$, and then showing that the function in between is monotone, since $e'(\alpha) = -\alpha(2\alpha-1)$ does not change sign in the interval for α under consideration. Finally for the third case, $\alpha > \frac{K}{\sqrt{2K-1}}$, taking the difference of the moduli squared, we find

$$|\lambda_1|^2-|\lambda_3|^2 = 2\frac{\alpha-1}{(1+\alpha)^2} > 0 \quad \text{for } \alpha > 1.$$

Hence the new third eigenvalue does never enter the argument in the convergence rate and the proof in one dimension is valid after the variable substitution $K = \sqrt{k^2+l^2}$. □

COROLLARY 3.18. *The optimal parameter* α^* *and the optimized convergence rate of the HSS stationary iteration for the two-dimensional Poisson problem*

	Iterative	GMRES		
h	HSS	No Prec.	HSS	HSS Krylov
1/25	46	48	13	2
1/50	63	98	17	2
1/100	91	198	22	2
1/200	127	398	28	2
1/400	179	798	37	2
1/800	252	1598	49	2

Table 4.1: One-dimensional case. Comparison of optimized stationary HSS iteration, GMRES without preconditioner, GMRES with the optimized HSS iteration as a preconditioner and HSS optimized for GMRES.

behave in the first case of Theorem 3.17 asymptotically for small mesh parameter h like

$$\alpha^* = \frac{\sqrt{\pi}}{2^{1/4}}\frac{1}{\sqrt{h}} + O(\sqrt{h}), \quad \rho^* = 1 - \frac{2^{1/4}}{\sqrt{\pi}}\sqrt{h} + O(h). \tag{3.18}$$

In the second case of Theorem 3.17, we have

$$\alpha^* = 2^{1/4}\sqrt{\pi\sqrt{k_{\min}^2+l_{\min}^2}}\frac{1}{\sqrt{h}} + O(\sqrt{h}), \ \rho^* = 1 - \frac{1-\sqrt{1-4(k_{\min}^2+l_{\min}^2)}}{2^{1/4}\sqrt{\pi\sqrt{k_{\min}^2+l_{\min}^2}}}\sqrt{h} + O(h). \tag{3.19}$$

There is also a clustering result in 2 dimensions when HSS is used as a preconditioner for a Krylov method.

THEOREM 3.19. *GMRES applied to the two-dimensional Poisson equation written in div-grad form and preconditioned with HSS converges in at most two steps to a given small tolerance τ, if the optimization parameter satisfies $0 < \alpha < \min(\bar{\alpha}, 1)$, where $\bar{\alpha} \approx \sqrt{2\tau(k_{\min}^2 + l_{\min}^2)}$.*

PROOF. Since the additional eigenvalue λ_3 is contained in the first cluster of Lemma 3.14, Lemma 3.15 holds also in the two-dimensional case and the result follows like in one dimension. ∎

4 Numerical experiments.

The first numerical experiment corresponds to the one-dimensional Poisson equation. We choose as the domain the interval $[0,1]$ with a homogeneous Neumann condition on the left and a Dirichlet condition on the right and a forcing function $g(x) = \sin \pi x$. We discretize the problem using finite differences with a forward difference for the grad part and a backward difference for the div part. We take a random vector as the initial guess and we stop the iteration when the initial residual has been reduced by at least three orders of magnitude. Table 4.1 shows the number of iterations the HSS method takes depending on the mesh parameter h. As one can see, the stationary HSS method converges significantly faster than (full) GMRES without preconditioner, already for very moderate values of the mesh parameter h. (Note that GMRES takes exactly as

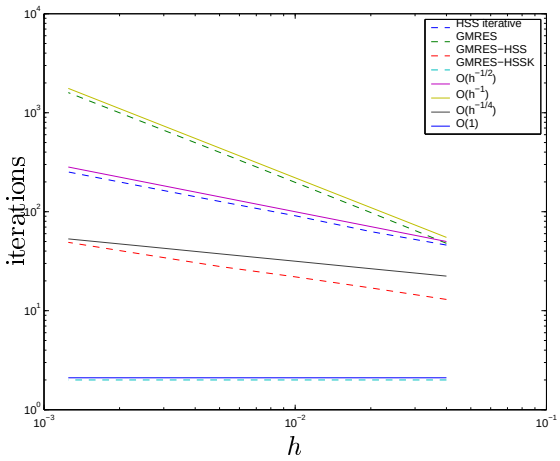

Figure 4.1: Asymptotic performance of HSS as a stationary iteration and as a preconditioner when either the optimization for the iterative version or for the Krylov method is used, compared to unpreconditioned GMRES.

many iterations as the number of unknowns in the saddle-point problem.) This improves further when HSS is used as a preconditioner for GMRES. Convergence is fastest however when HSS is optimized for GMRES. Here we have used $\alpha = 10^{-2}$ to form two tight, separate clusters. In that case two iterations are enough uniformly in the mesh parameter h to converge to the tolerance 10^{-3}, as predicted in Theorem 3.16. In Figure 4.1 we show the iteration numbers in a loglog plot which shows that the continuous analysis predicts the discrete asymptotic convergence rates very well. Only for the GMRES-HSS accelerated method the additional square root is not quite obtained from the Krylov method.

We next show that the continuous optimization leads to good estimates for the optimal parameter α for the discretized problem: in Figure 4.2 on the left we have run the stationary HSS iteration for many parameters α and show the required number of iterations to converge to the tolerance 10^{-3}. One can see that the analysis predicts the optimal parameter α^* rather well. To illustrate Theorem 3.16 further, we computed for a fixed mesh size $h = 1/100$ and a variable tolerance τ the numerical value of $\bar{\alpha}$, such that for $\alpha < \bar{\alpha}$ GMRES preconditioned with HSS converges in two iterations. Figure 4.2 shows on the right that Theorem 3.16 predicts this bound well for tolerances $\tau \geq 10^{-10}$. For smaller values of the tolerance roundoff is probably affecting the numerical result. For all practical purposes, the combination of GMRES with the optimized HSS preconditioner behaves like a direct solver.

The next set of numerical experiments is for the two-dimensional Poisson equation on the unit square with forcing term $g(x, y) = \sin \pi x \sin \pi y$. We impose Neumann boundary conditions for $x = 0$, $x = 1$ and homogeneous Dirichlet boundary conditions for $y = 0$, $y = 1$. Here we used a zero initial guess and

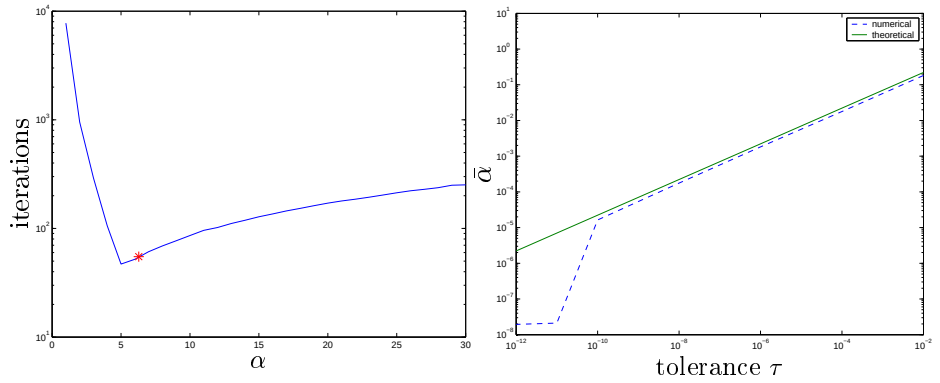

Figure 4.2: Comparison of the analytically determined optimal α with the numerical performance in an actual code on the left. On the right a comparison of the theoretical and numerical limiting value $\bar{\alpha}$, such that for $\alpha < \bar{\alpha}$ GMRES preconditioned with HSS converges in two steps.

	Iterative	GMRES		
h	HSS	No Prec.	HSS	HSS Krylov
1/10	66	54	14	2
1/25	103	140	19	2
1/50	146	286	25	2
1/100	207	574	34	2

Table 4.2: Two-dimensional case. Comparison of optimized stationary HSS iteration, GMRES without preconditioner, GMRES with the optimized HSS iteration as a preconditioner, and HSS optimized for GMRES.

stopping tolerance 10^{-6}. The results, shown in Table 4.2, are in agreement with our theoretical analysis. In particular, note that convergence is attained in two iterations (independent of h) when HSS is optimized for GMRES (here we used $\alpha = 10^{-3}$ as suggested by Theorem 3.19; using $\alpha = 10^{-2}$ results in three GMRES iterations). Clearly, there is a trade-off between rapid convergence of the preconditioned GMRES iteration and the amount of work that is necessary to solve the linear systems (2.1). A larger value of $\alpha < 1$, while still resulting in h-independent convergence, makes the linear systems (2.1) more diagonally dominant and thus easier to solve. When α is increased from 10^{-3} to 0.9 the number of iterations slowly grows from two to six, independently of h. The *a priori* determination of an optimal value of α that minimizes the total work (rather than the number of iterations) appears to be difficult, especially in view of the fact that the actual cost depends on the method used to solve the linear systems in (2.1). For the model problems studied here, the first system is diagonal, but the second one is not as trivial to solve. Here we consider the case where a Schur complement reduction is used, leading to (2.4). We have performed numerical experiments where we used the conjugate gradient method preconditioned with

	Inner PCG Iterations		
h	$\alpha = \alpha^*$	$\alpha = 10^{-1}$	$\alpha = 10^{-3}$
1/10	56	12	8
1/25	95	18	12
1/50	150	27	18
1/100	238	51	34

Table 4.3: Two-dimensional case. Total number of inner PCG iterations for different choices of α.

an incomplete Cholesky factorization with drop tolerance $\tau = 10^{-4}$ for the solution of (2.4). We iterate on (2.4) until the relative residual has been reduced under 10^{-12}, although in practice inexact solves with variable residual tolerance can be used for the inner iterations; see the examples and analysis in [1].

As expected, when using $\alpha = \alpha^*$ as given in Theorem 3.17 we found that the preconditioned CG method applied to (2.4) converges in a number of iterations which is bounded independent of h. By taking h sufficiently small (up to $h = 1/700$, corresponding to $m = 490,000$ unknowns in (2.4)), we determined this bound to be 13. On the other hand, when using $\alpha = 10^{-3}$, which results in a total of two (outer) preconditioned GMRES iterations, we found experimentally that the number of (inner) PCG iterations necessary to solve (2.4) grew as $h \to 0$, at least for the range of values of h that we tried. We also tried several values of α up to $\alpha = 1$, but we found that the total number of PCG iterations was always higher than with $\alpha = 10^{-3}$.

In Table 4.3 we report the total number of inner PCG iterations for three choices of α. It can be seen that using $\alpha = 10^{-3}$ is the most efficient of the three options. This is true even if one takes into account the fact that each inner PCG iteration is cheaper for $\alpha = \alpha^*$, due to the fact that the coefficient matrix $BB^* + \alpha^2 I_m$ in (2.4) is more diagonally dominant and therefore the incomplete Cholesky factors are sparser than for $\alpha = 10^{-3}$. However, since the number of inner PCG iterations is not h-independent, the overall method does not display optimal (linear) arithmetic complexity. This can be remedied by using an optimal solver for the solution of the systems (2.4), such as multigrid.

In addition to the simple test problems shown here, we experimented with anisotropic problems with constant coefficients. After applying a simple diagonal scaling (see [3]), we obtained results similar to those reported in Table 4.2. Again, we found that choosing $\alpha \in (0,1)$ results in an h-independent preconditioner for GMRES. In particular, $\alpha = 10^{-3}$ led to convergence in just two iterations. The results of numerical experiments on more complicated problems, including diffusion-type problems with discontinuous coefficients and problems arising from Stokes and Navier–Stokes equations, can be found in [3]; see also [2].

5 Conclusions.

The purpose of this paper was to shed some light on the problem of choosing the convergence parameter α in the HSS iteration applied to saddle-point problems arising from elliptic PDEs written in first order form. To this end, we focused on the simple model problem of the Poisson equation, for which we can use Fourier transforms in order to completely analyze the iteration operator at the continuous level.

The same technique can be used to study more complicated problems with constant coefficients, such as anisotropic problems and the Stokes problem.

A very important issue, not addressed in this paper, is the impact of inexact (iterative) inner solves. For some problems, it has been shown in [1] that the asymptotic rate of convergence of the (stationary) HSS iteration can be maintained by carefully tuning the inner stopping tolerance. It would be interesting to see to what extent this remains true when the HSS iteration is used as a preconditioner for a Krylov method, particularly in the context of saddle-point problems.

Acknowledgment.

We would like to thank Zhong-Zhi Bai and Michael Ng for their crucial role in developing the HSS algorithm and for calling our attention to it. Part of this work was carried out during the Milovy Conference on Computational Linear Algebra with Applications. We thank the local organizers for fostering a stimulating and friendly atmosphere. Finally, we would like to thank the referees for their careful reading of the manuscript and suggestions.

REFERENCES

1. Z. Z. Bai, G. H. Golub, and M. K. Ng, *Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 603–626.
2. Z. Z. Bai, G. H. Golub, and J. Y. Pan, *Preconditioned Hermitian and skew-Hermitian splitting methods for non-Hermitian positive semidefinite linear systems*, Technical Report SCCM-02-12, Scientific Computing and Computational Mathematics Program, Department of Computer Science, Stanford University, 2002.
3. M. Benzi and G. H. Golub, *An iterative method for generalized saddle point problems*, Technical Report SCCM-02-14, Scientific Computing and Computational Mathematics Program, Department of Computer Science, Stanford University, 2002.
4. D. Bertaccini, G. H. Golub, S. Serra Capizzano, and C. Tablino Possio, *Preconditioned HSS method for the solution of non-Hermitian positive definite linear systems*, Technical Report SCCM-02-11, Scientific Computing and Computational Mathematics Program, Department of Computer Science, Stanford University, 2002.
5. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, New York etc., 1991.

6. T. F. Chan and H. C. Elman, *Fourier analysis of iterative methods for elliptic problems*, SIAM Rev., 31 (1989), pp. 20–49.
7. P. Concus and G. H. Golub, *A generalized conjugate gradient method for nonsymmetric systems of linear equations*, in Computing Methods in Applied Sciences and Engineering, Lecture Notes in Econom. and Math. Systems 134, R. Glowinski and J. L. Lions, eds., Springer-Verlag, Berlin, 1976, pp. 56–65.
8. H. C. Elman, *Preconditioners for saddle point problems arising in computational fluid dynamics*, Appl. Numer. Math., 43 (2002), pp. 75–89.
9. M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Application to the Solution of Boundary-Value Problems*, Stud. Math. Appl., Vol. 15, North-Holland, Amsterdam, 1983.
10. M. J. Gander, F. Magoules, and F. Nataf, *Optimized Schwarz methods without overlap for the Helmholtz equation*, SIAM J. Sci. Comput., 24 (2002), pp. 38–60.
11. M. J. Gander and F. Nataf, *AILU: a preconditioner based on the analytic factorization of the elliptic operator*, Numer. Linear Algebra Appl., 7 (2000), pp. 505–526.
12. M. J. Gander and F. Nataf, *AILU for Helmholtz problems: a new preconditioner based on the analytic parabolic factorization*, J. Comput. Acoust., 9 (2001), pp. 1499-1506.
13. C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Soft., 8 (1982), pp. 43–71.