BIT manuscript No.
(will be inserted by the editor)

Analysis of a new dimension-wise splitting iteration with selective relaxation for saddle point problems

Martin J. Gander · Qiang Niu · Yingxiang Xu

Received: date / Accepted: date

Abstract We propose a new Dimension-wise Splitting with Selective Relaxation (DSSR) method for saddle point systems arising from the discretization of the incompressible Navier-Stokes equations. Using Fourier analysis, we determine the optimal choice of the relaxation parameter that leads to the best performance of the iterative method for the Stokes and the steady Oseen equations. We also explore numerically the influence of boundary conditions on the optimal choice of the parameter, the use of inner and outer iterations, and the performance for a lid driven cavity flow.

Keywords splitting iterations · optimized relaxation parameter · Stokes · Oseen

Mathematics Subject Classification (2000) 65F10 · 65N22

1 Introduction

We consider the steady incompressible Navier-Stokes equations

$$\begin{aligned} -\nu\Delta\mathbf{u}+\mathbf{u}\cdot\nabla\mathbf{u}+\nabla p &= \mathbf{f}, \quad \text{in } \Omega,\\ \nabla\cdot\mathbf{u} &= 0, \quad \text{in } \Omega,\\ \mathbf{u} &= \mathbf{g}, \quad \text{on } \partial\Omega, \end{aligned} \tag{1.1}$$

M.J. Gander
Section de Mathématiques, Université de Genève, 2-4 rue du Lièvre, CP 64, CH-1211, Geneva, Switzerland. E-mail: martin.gander@unige.ch

Q. Niu
Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China. E-mail: qiang.niu@xjtlu.edu.cn, partially supported by NSFC-11301420

Y. Xu
Corresponding author. School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China. E-mail: yxxu@nenu.edu.cn, partially supported by NSFC-11201061, 11471047, 11271065, CPSF-2012M520657 and the Science and Technology Development Planning of Jilin Province 20140520058JH

where $\mathbf{u}$ is the velocity field, p is the pressure, ν is the given kinematic viscosity determined by the fluid (ν is inversely proportional to the Reynolds number), and the domain Ω is an open bounded domain in $\mathbb{R}^2$ (or $\mathbb{R}^3$).

Equation (1.1) is nonlinear because of the transport term $\mathbf{u}\cdot\nabla\mathbf{u}$. A popular strategy for dealing with this nonlinearity is to use a relaxation for the nonlinear term: find $(\mathbf{u}^k, p^k)$ with an initial guess on the velocity field $\mathbf{u}^0$ satisfying $\nabla\mathbf{u}^0 = 0$ by solving

$$\begin{aligned} -\nu\Delta\mathbf{u}^k + \mathbf{u}^{k-1}\cdot\nabla\mathbf{u}^k + \nabla p^k &= \mathbf{f}, \quad \text{in } \Omega, \\ \nabla\cdot\mathbf{u}^k &= 0, \quad \text{in } \Omega, \\ \mathbf{u}^k &= \mathbf{g}, \quad \text{on } \partial\Omega. \end{aligned} \tag{1.2}$$

Discretizing equation (1.2) using a finite element or finite difference method leads to a so-called saddle point problem

$$H\mathbf{z} = \mathbf{b}, \tag{1.3}$$

where in 2D

$$H = \begin{pmatrix} A_1 & 0 & B_1^T \\ 0 & A_2 & B_2^T \\ -B_1 & -B_2 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{z} = \begin{pmatrix} \mathbf{u} \\ p \end{pmatrix},$$

and H is a sparse matrix in $\mathbb{R}^{n\times n}$, $\mathbf{b}\in\mathbb{R}^n$ is the right hand side that depends on the given data $\mathbf{f}$ and $\mathbf{g}$, and we are solving for the unknown $\mathbf{z}\in\mathbb{R}^n$.

Problem (1.3) is a typical saddle point problem, which arises also in many other applications throughout computational science and engineering, and the iterative solution of such problems has been the focus of intense research efforts, see for example the review [10]. In addition to the block preconditioners based on approximate commutators [15, 18] and least squares [16], one of the successful iterative methods for solving saddle point problems is the Hermitian and skew-Hermitian splitting (HSS) method invented by Bai, Golub and Ng [5], where the system matrix H is split into its Hermitian and non-Hermitian parts, and then an ADI like iterative scheme is used, solving only Hermitian and skew-Hermitian problems, and a relaxation parameter α is introduced to obtain rapid convergence. Like all stationary iterative methods, HSS also defines naturally a preconditioner that can be used to accelerate a Krylov method.

A substantial development followed this seminal work, with an important focus on how one should choose the relaxation parameter α to get fast convergence for a given class of problems. For example, Bai, Golub and Li [3] optimized the spectral radius of the iteration matrix and obtained an eigenvalue-based optimized relaxation parameter α^*. In [2], the authors proposed a new variant of HSS called accelerated Hermitian and skew-Hermitian splitting (AHSS), which improved the HSS method by introducing an extra relaxation parameter in order to shift the non-Hermitian part differently. These two parameters were then optimized for performance, and the optimal choice was shown to be related to the singular values of a certain matrix arising in the iteration. A more concrete estimate was given by Bai [1] with an analysis at the continuous level, by optimizing an approximation of the spectral radius of the iteration matrix, which leads to quasi-optimal parameters. The main difficulty in these approaches based on the matrix H itself is to obtain the estimates on the quantities on which the optimized relaxation parameter depends on.

A different idea is to relate the iteration back to the underlying physical operator, from which the matrix H was obtained by discretization. This was proposed by Benzi, Gander and Golub [9], where the HSS iteration was related back to the underlying Laplace operator written in mixed form. This can be achieved using Fourier analysis, and can lead to accurate predictions for the optimized relaxation parameter. This analysis also revealed a new feature of HSS, when used as a preconditioner for a Krylov method: instead of minimizing the spectral radius, a different choice of the relaxation parameter leads to two very small and tight clusters in the spectrum of the HSS preconditioned operator, which then leads to convergence in two iterations when solving the preconditioned system using a Krylov method. Fourier analysis can thus be very helpful for understanding the performance of iterative methods, see also [19, 21,20,7,22,23,8].

More recently, new stationary iterative methods were proposed for saddle point problems. In particular, Benzi and Guo [11] proposed the dimensional splitting preconditioner for saddle point problems arising from the discretization of the Navier-Stokes equation, where there is also a relaxation parameter, but the optimization problem was not addressed. Benzi et al. [12] proposed a further variant called Relaxed Dimensional Factorization (RDF) preconditioner, directly to be used with a Krylov method. They did consider the optimization of the relaxation parameter, but without studying the preconditioner as a stationary iterative method.

In this paper, we present and analyze a new method based on a Dimension-wise Splitting with Selective Relaxation (DSSR) that can be used both as an iterative solver and a preconditioner for a Krylov method, see also [14]. We prove first that the DSSR iteration converges unconditionally for any relaxation parameter $\alpha > 0$ when the method is applied to the 2D Stokes problem. We also show how Fourier analysis can be used to optimize the relaxation parameter of the method, both to minimize the spectral radius and to form tight clusters in the spectrum of the preconditioned operator. We find that the optimal choice for the relaxation parameter α is inversely proportional to the viscosity ν, and with this choice the convergence rates are robust in ν. We then extend our results to the 3D Stokes and Oseen equations. Using numerical experiments, we illustrate that the theoretically optimized parameters are still asymptotically optimal and robust also in the presence of Dirichlet boundary conditions, where the Fourier analysis is not valid any more. We also show numerically that DSSR is competitive with RDF, which was recently shown to outperform many other preconditioners, and test the use of inexact inner solves, which leads to inner and outer iterations.

2 Dimension-wise splitting with selective relaxation (DSSR)

The dimension-wise splitting with selective relaxation is based on the decomposition of the system matrix in (1.3) into $H = H_1 + H_2$ [11,12], where

$$H_1 = \begin{pmatrix} A_1 & 0 & B_1^T \\ 0 & 0 & 0 \\ -B_1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad H_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & A_2 & B_2^T \\ 0 & -B_2 & 0 \end{pmatrix}. \tag{2.1}$$

To add a relaxation parameter θ, we define the diagonal matrices

$$E_1 := \mathtt{diag}(0, I, \theta I) \quad \text{and} \quad E_2 := \mathtt{diag}(I, 0, (1-\theta)I)$$

with $0 < \theta < 1$, where the block sizes of the zero matrices 0 and the identity matrices I correspond to the block sizes of the submatrices in (2.1).

We can thus define two splittings of the system matrix H with further relaxation parameter $\alpha > 0$,

$$H = (\alpha E_1 + H_1) - (\alpha E_1 - H_2) \quad \text{and} \quad H = (\alpha E_2 + H_2) - (\alpha E_2 - H_1).$$

Note here that both

$$\alpha E_1 + H_1 = \begin{pmatrix} A_1 & 0 & B_1^T \\ 0 & \alpha I & 0 \\ -B_1 & 0 & \alpha\theta I \end{pmatrix} \quad \text{and} \quad \alpha E_2 + H_2 = \begin{pmatrix} \alpha I & 0 & 0 \\ 0 & A_2 & B_2^T \\ 0 & -B_2 & \alpha(1-\theta) I \end{pmatrix}$$

are nonsingular matrices, provided that the boundary conditions imposed on the velocity components lead to invertible discrete Laplace operators. The DSSR algorithm consists of the alternating solution between these two splittings: given an initial guess $\mathbf{z}^0 = (\mathbf{u}^0, p^0)$, the iteration computes a sequence of approximations $\mathbf{z}^k$ defined by

$$\begin{aligned} (\alpha E_1 + H_1)\mathbf{z}^{k+\frac{1}{2}} &= (\alpha E_1 - H_2)\mathbf{z}^k + \mathbf{b}, \\ (\alpha E_2 + H_2)\mathbf{z}^{k+1} &= (\alpha E_2 - H_1)\mathbf{z}^{k+\frac{1}{2}} + \mathbf{b}. \end{aligned} \tag{2.2}$$

The DSSR algorithm can be rewritten in fixed point form by eliminating the intermediate vector $\mathbf{z}^{k+\frac{1}{2}}$, and we obtain

$$\mathbf{z}^{k+1} = M_\alpha \mathbf{z}^k + N_\alpha \mathbf{b}, \tag{2.3}$$

where

$$M_\alpha = (\alpha E_2 + H_2)^{-1}(\alpha E_2 - H_1)(\alpha E_1 + H_1)^{-1}(\alpha E_1 - H_2)$$

and

$$N_\alpha = (\alpha E_2 + H_2)^{-1}[(\alpha E_2 - H_1)(\alpha E_1 + H_1)^{-1} + I].$$

Theorem 2.1 *The fixed point $\mathbf{z}^*$ of (2.3) is solution to (1.3).*

Proof It is easy to verify by a direct calculation that

$$\begin{aligned} \alpha H &= \alpha(H_1 + H_2) \\ &= (\alpha E_1 + H_1)(\alpha E_2 + H_2) - (\alpha E_2 - H_1)(\alpha E_1 - H_2) \\ &= (\alpha E_1 + H_1)(\alpha E_2 + H_2) - (\alpha E_2 - H_1)(\alpha E_1 + H_1)(\alpha E_1 + H_1)^{-1}(\alpha E_1 - H_2) \\ &= (\alpha E_1 + H_1)(\alpha E_2 + H_2) - (\alpha E_1 + H_1)(\alpha E_2 - H_1)(\alpha E_1 + H_1)^{-1}(\alpha E_1 - H_2). \end{aligned}$$

Multiplying this equation by $(\alpha E_2 + H_2)^{-1}(\alpha E_1 + H_1)^{-1}$ from the left on both sides yields

$$\alpha(\alpha E_2 + H_2)^{-1}(\alpha E_1 + H_1)^{-1} H = I - M_\alpha.$$

We can now multiply both sides from the right by the fixed point vector $\mathbf{z}^*$ of the iteration (2.3), and obtain

$$\alpha(\alpha E_2 + H_2)^{-1}(\alpha E_1 + H_1)^{-1} H \mathbf{z}^* = (\alpha E_2 + H_2)^{-1}((\alpha E_2 - H_1)(\alpha E_1 + H_1)^{-1} + I)\mathbf{b}.$$

Now since the right hand side of this equation satisfies

$$\begin{aligned}
&(\alpha E_2+H_2)^{-1}((\alpha E_2-H_1)(\alpha E_1+H_1)^{-1}+I)\mathbf{b}\\
&=(\alpha E_2+H_2)^{-1}((\alpha E_2-H_1)(\alpha E_1+H_1)^{-1}+(\alpha E_1+H_1)(\alpha E_1+H_1)^{-1})\mathbf{b}\\
&=(\alpha E_2+H_2)^{-1}(\alpha E_1+\alpha E_2)(\alpha E_1+H_1)^{-1}\mathbf{b}\\
&=\alpha(\alpha E_2+H_2)^{-1}(\alpha E_1+H_1)^{-1}\mathbf{b},
\end{aligned}$$

we obtain the desired result $H\mathbf{z}^*=\mathbf{b}$, which concludes the proof. □

The DSSR iteration (2.2) can also be rewritten in *correction form*, namely

$$\mathbf{z}^{k+1}=\mathbf{z}^k+P_\alpha^{-1}\mathbf{r}^k,\quad \mathbf{r}^k=\mathbf{b}-H\mathbf{z}^k,$$

where $P_\alpha:=\frac{1}{\alpha}(\alpha E_1+H_1)(\alpha E_2+H_2)$. Note that $M_\alpha=I-P_\alpha^{-1}H$. In addition, we know that using a Krylov subspace method to accelerate the convergence of the DSSR iteration is equivalent to the application of the Krylov subspace method to the preconditioned system $P_\alpha^{-1}H\mathbf{z}=P_\alpha^{-1}\mathbf{b}$ in the case of left preconditioning, or to $HP_\alpha^{-1}\mathbf{y}=\mathbf{b}$, $\mathbf{y}=P_\alpha\mathbf{z}$ for right preconditioning.

3 Convergence analysis of DSSR at the continuous level

It is well known that the iterative scheme (2.2) converges for any initial guess $\mathbf{z}^0$, if the spectral radius of M_α is strictly less than one, $\rho(M_\alpha)<1$. In order to obtain fast convergence, the parameters θ and α should be chosen to make the spectral radius small. This is in general a difficult problem at the linear algebra level. One can however optimize the parameters in the method using the fact that the matrices are discretizations of underlying differential operators, see e.g. [9].

3.1 Analysis of DSSR for the Stokes equation in 2D

We start by considering the Stokes equation in the unbounded domain $\mathbb{R}^2$,

$$\begin{aligned}
-\nu\Delta u+\partial_x p&=f_1,\\
-\nu\Delta v+\partial_y p&=f_2,\\
\partial_x u+\partial_y v&=0.
\end{aligned}\tag{3.1}$$

From the analysis in Appendix A we see that for the Stokes problem, i.e. without the advection term, the best parameter choice for θ is $\theta=\frac{1}{2}$, and we thus only have one parameter left to optimize. Writing the Stokes equation (3.1) in matrix form yields

$$\begin{pmatrix}-\nu\Delta & 0 & \partial_x\\ 0 & -\nu\Delta & \partial_y\\ \partial_x & \partial_y & 0\end{pmatrix}\begin{pmatrix}u\\ v\\ p\end{pmatrix}=\begin{pmatrix}f_1\\ f_2\\ 0\end{pmatrix}.$$

Following the approach in [9], we take a Fourier transform in both the x and y directions and obtain for each Fourier mode (k_1,k_2) the linear system

$$\begin{pmatrix}\nu(k_1^2+k_2^2) & 0 & ik_1\\ 0 & \nu(k_1^2+k_2^2) & ik_2\\ ik_1 & ik_2 & 0\end{pmatrix}\begin{pmatrix}\hat{u}\\ \hat{v}\\ \hat{p}\end{pmatrix}=\begin{pmatrix}\hat{f}_1\\ \hat{f}_2\\ 0\end{pmatrix}.$$

This has the great advantage that we can now study the performance of the DSSR algorithm for each Fourier mode separately. Note however that this neglects boundary conditions in the analysis, which we will see play a certain role as well for the iteration.

The Fourier transformed linear operator $\hat{H}$ corresponding to the Stokes equation (3.1) is thus given by

$$\hat{H} = \begin{pmatrix} \nu(k_1^2+k_2^2) & 0 & ik_1 \\ 0 & \nu(k_1^2+k_2^2) & ik_2 \\ ik_1 & ik_2 & 0 \end{pmatrix},$$

and we can also write the dimension-wise splitting operators H_1 and H_2 after the Fourier transform,

$$\hat{H}_1 = \begin{pmatrix} \nu(k_1^2+k_2^2) & 0 & ik_1 \\ 0 & 0 & 0 \\ ik_1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \hat{H}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \nu(k_1^2+k_2^2) & ik_2 \\ 0 & ik_2 & 0 \end{pmatrix}.$$

Thus, the Fourier transformed iteration matrix of the DSSR algorithm is given by

$$\hat{M}_\alpha = (\alpha\hat{E}_2+\hat{H}_2)^{-1}(\alpha\hat{E}_2-\hat{H}_1)(\alpha\hat{E}_1+\hat{H}_1)^{-1}(\alpha\hat{E}_1-\hat{H}_2),$$

with $\hat{E}_1 = \texttt{diag}(0,1,1/2)$ and $\hat{E}_2 = \texttt{diag}(1,0,1/2)$. The eigenvalues of the Fourier transformed iteration matrix $\hat{M}_\alpha$ can now simply be calculated, and we get

$$\lambda_{1,2}(k_1,k_2) = 0, \quad \lambda_3(k_1,k_2) = \frac{\alpha\nu(k_1^2+k_2^2)-2k_1^2}{\alpha\nu(k_1^2+k_2^2)+2k_1^2} \cdot \frac{\alpha\nu(k_1^2+k_2^2)-2k_2^2}{\alpha\nu(k_1^2+k_2^2)+2k_2^2}.$$

Thus the spectral radius of $\hat{M}_\alpha$ is given by

$$\rho(k_1,k_2,\nu,\alpha) = |\lambda_3(k_1,k_2)|.$$

Note that ρ depends only on k_1^2 and k_2^2, and thus it is sufficient to analyze the algorithm for positive Fourier frequencies $k_1,k_2 > 0$. A direct calculation shows that

$$\liminf_{k_1,k_2\to 0} \lambda_3(k_1,k_2) = \frac{\alpha\nu-2}{\alpha\nu+2} = \liminf_{k_1,k_2\to\infty} \lambda_3(k_1,k_2).$$

Together with the fact that $\rho(k_1,k_2,\nu,\alpha) < 1$ for any fixed k_1,k_2, we thus arrive at the following convergence theorem.

Theorem 3.1 *The DSSR iteration (2.2) applied to a consistently discretized version of the Stokes problem (3.1) on an unbounded domain converges for any $\alpha > 0$, provided the mesh size is small enough.*

To optimize the performance of the DSSR iteration, we need to choose the relaxation parameter α to minimize the maximum of the spectral radius $\rho(k_1,k_2,\nu,\alpha)$ over all Fourier frequencies (k_1,k_2), which means that we have to solve the min-max problem

$$\min_{\alpha>0} \max_{k_1,k_2>0} \rho(k_1,k_2,\nu,\alpha). \tag{3.2}$$

Theorem 3.2 *The parameter value $\alpha^* = \frac{\sqrt{3}}{\nu}$ solves the min-max problem (3.2), and the corresponding optimized spectral radius of $\hat{M}_\alpha$ satisfies*

$$\max_{k_1,k_2>0} \rho(k_1,k_2,\nu,\alpha^*) = \frac{2-\sqrt{3}}{2+\sqrt{3}}.$$

Proof A direct calculation shows that

$$\frac{\partial \lambda_3}{\partial k_1} = -\frac{32k_1k_2^2\alpha\nu(k_1^4-k_2^4)}{(\alpha\nu(k_1^2+k_2^2)+2k_2^2)^2(\alpha\nu(k_1^2+k_2^2)+2k_1^2)^2}, \tag{3.3}$$

which is positive for $0<k_1<k_2$, and negative for $k_1>k_2>0$. Similarly we obtain that

$$\frac{\partial \lambda_3}{\partial k_2} = \frac{32k_2k_1^2\alpha\nu(k_1^4-k_2^4)}{(\alpha\nu(k_1^2+k_2^2)+2k_2^2)^2(\alpha\nu(k_1^2+k_2^2)+2k_1^2)^2}, \tag{3.4}$$

which is positive for $k_1>k_2>0$, and negative for $0<k_1<k_2$. We therefore obtain that at the zero of the two functions (3.3) and (3.4), $k_1=k_2$, the function λ_3 attains its maximum. Inserting $k_1=k_2$ shows that the maximum value of λ_3 is given by $\lambda_3(k_1,k_1) = \left(\frac{\alpha\nu-1}{\alpha\nu+1}\right)^2$, independently of k_1.

Since $\frac{\partial\lambda_3}{\partial k_1} > 0$ for $k_1<k_2$, λ_3 is monotonically increasing in k_1 for $k_1\in(0,k_2)$, and since $\frac{\partial\lambda_3}{\partial k_2} > 0$ for $k_2<k_1$, λ_3 is monotonically increasing in k_2 for $k_2\in(0,k_1)$. Thus, $\lim_{k_1\to 0}\lambda_3 = \frac{\alpha\nu-2}{\alpha\nu+2} = \lim_{k_2\to 0}\lambda_3$ is a potential minimum of λ_3 for $k_1,k_2>0$.

Next, since $\frac{\partial\lambda_3}{\partial k_1} < 0$ for $k_1>k_2$, λ_3 is monotonically decreasing in k_1 for $k_1>k_2$, and since $\frac{\partial\lambda_3}{\partial k_2} < 0$ for $k_2>k_1$, λ_3 is monotonically decreasing in k_2 for $k_2>k_1$. Thus $\lim_{k_1\to\infty}\lambda_3 = \frac{\alpha\nu-2}{\alpha\nu+2} = \lim_{k_2\to\infty}\lambda_3$ is another potential minimum of λ_3 for $k_1,k_2>0$.

We therefore see that $\rho(k_1,k_2,\nu,\alpha)$ has two potential extrema: $\left|\frac{\alpha\nu-2}{\alpha\nu+2}\right|$ when k_1 or k_2 tend to 0 or ∞, and $\left(\frac{\alpha\nu-1}{\alpha\nu+1}\right)^2$ for $k_1=k_2$.

We show below that solving the equi-oscillation equation gives the best choice,

$$\left(\frac{\alpha\nu-1}{\alpha\nu+1}\right)^2 = \left|\frac{\alpha\nu-2}{\alpha\nu+2}\right| \quad\Longrightarrow\quad \alpha^* = \frac{\sqrt{3}}{\nu}, \tag{3.5}$$

with the corresponding spectral radius $\max_{k_1,k_2>0}\rho(k_1,k_2,\nu,\alpha) = \frac{2-\sqrt{3}}{2+\sqrt{3}} \approx 0.0718$. To prove that α^* given in (3.5) indeed solves the min-max problem (3.2), note that $\left(\frac{\alpha\nu-1}{\alpha\nu+1}\right)^2$ is monotonically increasing in α for $\alpha\in(\frac{1}{\nu},\infty)$, and thus $\alpha>\alpha^*$ would lead to a larger spectral radius for $\hat{M}_\alpha$. Similarly, $\frac{\alpha\nu-2}{\alpha\nu+2}$ is monotonically decreasing in α for $2/\nu>\alpha>0$, and thus $\alpha<\alpha^*$ would lead to a larger spectral radius for $\hat{M}_\alpha$, which concludes the proof. □

If DSSR is used as a preconditioner for a Krylov subspace method, one can use the same relaxation parameter α^* from (3.5) that was optimized for the stationary iteration, since minimizing the spectral radius corresponds to clustering the eigenvalues of the preconditioned operator around 1. However for DSSR there exists also a different option: one can form two tight separate clusters with an appropriate choice of the optimization parameter α, as we show next.

Lemma 3.1 *The eigenvalues $\lambda_3(k_1,k_2)$ for $k_1,k_2>0$ are real and contained in the interval $T=\left(\frac{\alpha\nu-2}{\alpha\nu+2},(\frac{\alpha\nu-1}{\alpha\nu+1})^2\right]$.*

Proof This result follows directly from the proof of Theorem 3.2. □

Note that the length of the interval T, which is given by $\frac{4}{(\alpha\nu+1)^2(\alpha\nu+2)}>0$ tends to zero as $\alpha\nu$ tends to infinity.

Lemma 3.2 *There exists a polynomial $p_2(x)$ of order 2 which satisfies $p_2(0)=1$, $p_2(1)=0$ and on the shifted interval $T_s=\left(1-\frac{\alpha\nu-2}{\alpha\nu+2},1-(\frac{\alpha\nu-1}{\alpha\nu+1})^2\right]$*

$$\max_{x\in T_s}|p_2(x)|\le q_2(\alpha,\nu),$$

where

$$\begin{aligned} q_2(\alpha,\nu) &= \frac{(\alpha\nu)^3-4(\alpha\nu)^2+5(\alpha\nu)-2}{2(\alpha\nu)^5+4(\alpha\nu)^4-5(\alpha\nu)^3-12(\alpha\nu)^2-3(\alpha\nu)-2} \\ &= \frac{1}{2(\alpha\nu)^2}+O(\frac{1}{(\alpha\nu)^3}). \end{aligned} \tag{3.6}$$

Proof We assume that the polynomial $p_2(x)$ has the form $p_2(x)=1+bx+cx^2$, since it satisfies $p_2(0)=1$. Using the condition $p_2(1)=0$ we get

$$1+b+c=0. \tag{3.7}$$

Since this polynomial is convex, it will attain on the interval T_s its maximum on the left end point $1-(\frac{\alpha\nu-1}{\alpha\nu+1})^2$ and its minimum on the right end point $1-\frac{\alpha\nu-2}{\alpha\nu+2}$. Thus setting the absolute values of $p_2(x)$ at $1-(\frac{\alpha\nu-1}{\alpha\nu+1})^2$ and $1-\frac{\alpha\nu-2}{\alpha\nu+2}$ equal, together with (3.7), we can derive the values of b and c, and thus the polynomial $p_2(x)$. We therefore obtain the maximum value of $p_2(x)$ on the interval to be

$$p_2\left(1-\left(\frac{\alpha\nu-1}{\alpha\nu+1}\right)^2\right)=\frac{(\alpha\nu)^3-4(\alpha\nu)^2+5(\alpha\nu)-2}{2(\alpha\nu)^5+4(\alpha\nu)^4-5(\alpha\nu)^3-12(\alpha\nu)^2-3(\alpha\nu)-2}.$$

Expanding in $\alpha\nu$ for $\alpha\nu$ large we find (3.6), which concludes the proof. □

Since the DSSR preconditioned Stokes system is diagonalizable with an orthogonal basis, we get a convergence theorem for GMRES with $p_2(x)$ as residual polynomial:

Theorem 3.3 *GMRES applied to the preconditioned fixed point system $(I-M_\alpha)\mathbf{z}=N_\alpha\mathbf{b}$ from (2.3) converges in the case of a fine enough discretized Stokes system (3.1) on an unbounded domain in at most two iteration steps to a given small tolerance τ if the optimized parameter α^* satisfies $q_2(\alpha^*)=\tau$, which implies asymptotically that $\alpha^*\approx\frac{1}{\nu}\sqrt{\frac{1}{2\tau}}$.*

One can also estimate the decay of a higher order residual polynomial, for example of degree three:

Lemma 3.3 *There exists a polynomial $p_3(x)$ of order 3 which satisfies $p_3(0) = 1$, $p_3(1) = 0$ and on the shifted interval T_s*

$$\max_{x \in T_s} |p_3(x)| \leq q_3(\alpha, \nu),$$

where $q_3(\alpha,\nu)$ is explicitly known and has the asymptotic expansion $q_3(\alpha,\nu) = \frac{1}{8(\alpha\nu)^4} + O\left(\frac{1}{(\alpha\nu)^5}\right)$.

Proof Similar to the proof of Lemma 3.2 we assume that the polynomial $p_3(x)$ has the form $p_3(x) = 1 + bx + cx^2 + dx^3$. Since we require $p_3(1) = 0$ we get

$$1 + b + c + d = 0. \tag{3.8}$$

Denote by $x_1 = 1 - (\frac{\alpha\nu-1}{\alpha\nu+1})^2$ and $x_2 = 1 - \frac{\alpha\nu-2}{\alpha\nu+2}$. Let $x_m = (x_1 + x_2)/2$. Note that one can appropriately choose the polynomial $p_3(x)$ such that $p_3(x_1)$ and $p_3(x_2)$ are positive and $p_3(x_m)$ is negative. We then set the absolute values of $p_3(x)$ at these three points equal, i.e. $p_3(x_1) = -p_3(x_m) = p_3(x_2)$, which, together with (3.8) gives the expressions of b, c and d, and thus the polynomial $p_3(x)$.

A direct calculation shows that the minimum of the polynomial $p_3(x)$ is given by $x^* = \frac{c-\sqrt{c^2-3bd}}{-3d}$ and $p_3(x^*) = -\frac{1}{8(\alpha\nu)^4} + O\left(\frac{1}{(\alpha\nu)^5}\right)$. Thus we arrive at our result by noting that $p_3(x_1) = p_3(x_2) = \frac{1}{8(\alpha\nu)^4} + O\left(\frac{1}{(\alpha\nu)^5}\right)$. □

Theorem 3.4 *For the same configuration as in Theorem 3.3, GMRES converges in at most three steps to a given small tolerance τ if the optimized parameter $\bar{\alpha}$ satisfies $q_3(\bar{\alpha}) = \tau$, which means asymptotically $\bar{\alpha} \approx \frac{1}{\nu}\left(\frac{1}{8\tau}\right)^{\frac{1}{4}}$.*

Remark 3.1 This last theorem shows that we do not need to choose the relaxation parameter very big for good performance of GMRES.

3.2 Analysis of DSSR for the Stokes equation in 3D

We now use the techniques from Subsection 3.1 to analyze the performance of DSSR for the 3D Stokes equation, i.e. the saddle point system (1.3) with

$$H = \begin{pmatrix} A_1 & 0 & 0 & B_1^T \\ 0 & A_2 & 0 & B_2^T \\ 0 & 0 & A_3 & B_3^T \\ -B_1 & -B_2 & -B_3 & 0 \end{pmatrix}. \tag{3.9}$$

Using the splitting

$$H_1 = \begin{pmatrix} A_1 & 0\,0 & B_1^T \\ 0 & 0\,0 & 0 \\ 0 & 0\,0 & 0 \\ -B_1 & 0\,0 & 0 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_2 & 0 & B_2^T \\ 0 & 0 & 0 & 0 \\ 0 & -B_2 & 0 & 0 \end{pmatrix} \quad \text{and} \quad H_3 = \begin{pmatrix} 0\,0 & 0 & 0 \\ 0\,0 & 0 & 0 \\ 0\,0 & A_3 & B_3^T \\ 0\,0 & -B_3 & 0 \end{pmatrix},$$

the DSSR iteration is now a three level iteration,

$$\begin{aligned}(\alpha E_1 + H_1)\mathbf{z}^{k+\frac{1}{3}} &= (\alpha E_1 - H_2 - H_3)\mathbf{z}^k + \mathbf{b},\\ (\alpha E_2 + H_2)\mathbf{z}^{k+\frac{2}{3}} &= (\alpha E_2 - H_1 - H_3)\mathbf{z}^{k+\frac{1}{3}} + \mathbf{b},\\ (\alpha E_3 + H_3)\mathbf{z}^{k+1} &= (\alpha E_3 - H_1 - H_2)\mathbf{z}^{k+\frac{2}{3}} + \mathbf{b},\end{aligned} \tag{3.10}$$

where we choose now $E_1 = \texttt{diag}(0,1,1,1)$, $E_2 = \texttt{diag}(1,0,1,1)$ and $E_3 = \texttt{diag}(1,1,0,1)$. In fixed point form, the DSSR iteration reads

$$\mathbf{z}^{k+1} = M_\alpha \mathbf{z}^k + N_\alpha \mathbf{b},$$

where $M_\alpha = (\alpha E_3 + H_3)^{-1}(\alpha E_3 - H_1 - H_2)(\alpha E_2 + H_2)^{-1}(\alpha E_2 - H_1 - H_3)(\alpha E_1 + H_1)^{-1}(\alpha E_1 - H_2 - H_3)$ and $N_\alpha = (\alpha E_3 + H_3)^{-1}\big(I + (\alpha E_3 - H1 - H2)(\alpha E_2 + H_2)^{-1}(I + (\alpha E_2 - H_1 - H_3)(\alpha E_1 + H_1)^{-1})\big)$.

Similar to Theorem 2.1, we can show that the fixed point of the DSSR iteration (3.10) solves the saddle point system (1.3) with the coefficient matrix H given by (3.9). The DSSR iteration (3.10) can also be rewritten in correction form,

$$\mathbf{z}^{k+1} = \mathbf{z}^k + P_\alpha^{-1}\mathbf{r}^k, \quad \mathbf{r}^k = b - H\mathbf{z}^k,$$

where $P_\alpha = (\alpha E_1 + H_1)\big((\alpha E_1 + H_1) + (\alpha E_3 - H_1 - H_2)(\alpha E_2 + H_2)^{-1}(\alpha(E_1 + E_2) - H_3)\big)^{-1}(\alpha E_3 + H_3)$. Note again that $M_\alpha = I - P_\alpha^{-1}H$.

Similar to the analysis performed in Subsection 3.1, we find that the optimized relaxation parameter α_{opt} is given by[1]

$$\alpha_{opt} \approx \frac{2.0439}{\nu},$$

which leads to the maximum of the convergence factor $\max \rho \approx 0.0525$.

3.3 Analysis of DSSR for the steady Oseen problem

To study the performance of DSSR for the steady Oseen problem (1.2), Fourier analysis cannot be applied directly any more because of the variable coefficients from the relaxed advection term. To gain some insight, we assume that $\mathbf{u}^{k-1}$ is a constant, $\mathbf{u}^{k-1} := (u_0, v_0)$. In the implementation, we can then for example use averages, $u_0 = \frac{1}{|\Omega|}\int_\Omega u^{k-1} dxdy$ and $v_0 = \frac{1}{|\Omega|}\int_\Omega v^{k-1} dxdy$ in the formula of the optimized relaxation parameter α_{opt} we will obtain. We thus consider the steady Oseen problem with $\mathbf{u}^{k-1} = (u_0, v_0)$ in the unbounded domain $\mathbb{R}^2$,

$$\begin{pmatrix} -\nu\Delta + u_0\partial_x + v_0\partial_y & 0 & \partial_x \\ 0 & -\nu\Delta + u_0\partial_x + v_0\partial_y & \partial_y \\ \partial_x & \partial_y & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ p \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ 0 \end{pmatrix}.$$

By a Fourier transform in the x and y directions we obtain as in Subsection 3.1 the transformed coefficient matrix $\hat{H}$, the corresponding iteration matrix $\hat{M}_\alpha$ with $\hat{E}_1 =$

[1] 2.0439 is a root of the polynomial $12x^4 - 36x^3 + 11x^2 + 23x + 5$.

$\mathtt{diag}(0,1,\theta)$ and $\hat{E}_2 = \mathtt{diag}(1,0,1-\theta)$, and thus the eigenvalues of $\hat{M}_\alpha$, which are $\lambda_{1,2}(k_1,k_2) = 0$ and

$$\lambda_3(k_1,k_2) = \frac{\alpha\nu\big((\theta-1)(k_1^2+k_2^2)+i(\theta-1)(k_1u_0+k_2v_0)\big)+k_1^2}{\alpha\nu\big((\theta-1)(k_1^2+k_2^2)+i(\theta-1)(k_1u_0+k_2v_0)\big)-k_2^2}\cdot \frac{\alpha\nu(\theta(k_1^2+k_2^2)+i\theta(k_1u_0+k_2v_0))-k_2^2}{\alpha\nu\big(\theta(k_1^2+k_2^2)+i\theta(k_1u_0+k_2v_0)\big)+k_1^2}.$$

The spectral radius of $\hat{M}_\alpha$ is thus given by

$$\rho(k_1,k_2,\nu,\alpha,\theta) = |\lambda_3(k_1,k_2)|,$$

and to optimize the parameters α and θ, we then need to solve the min-max problem

$$\min_{\alpha>0,0<\theta<1}\max_{(k_1,k_2)\in\mathbb{K}}\rho(k_1,k_2,\nu,\alpha,\theta), \tag{3.11}$$

where $\mathbb{K} = [k_{1,\min},k_{1,\max}]\times[k_{2,\min},k_{2,\max}]$ is the frequency domain involved in the calculation. When a uniform grid is used with mesh size h, $k_{1,\min}$ and $k_{2,\min}$ can be estimated by π and $k_{1,\max}$ and $k_{2,\max}$ by π/h, see for example [19,20,22,23].

A similar analysis as in Subsection 3.1 shows that the min-max problem (3.11) is solved by the solution $(\alpha_{opt},\theta_{opt})$ of the equi-oscillation equations

$$\lim_{k_2\to\infty}\rho(k_{1,\min},k_2,\nu,\alpha,\theta) = \lim_{k_1\to\infty}\rho(k_1,k_{2,\min},\nu,\alpha,\theta) = \rho(k_{1,\min},k_{2,\min},\nu,\alpha,\theta).$$

Solving these equations shows that $\theta_{opt} = 1/2$ and

$$\alpha_{opt} = \frac{\sqrt{2k_{1,\min}k_{2,\min}}\sqrt{k_{1,\min}k_{2,\min}T_1+(k_{1,\min}^2+k_{2,\min}^2)\sqrt{T_2}}}{(k_{1,\min}u_0+k_{2,\min}v_0)\sqrt{T_3)}} \tag{3.12}$$

with $T_1 = 2(k_{1,\min}u_0+k_{2,\min}v_0)^2-(k_{1,\min}^2+k_{2,\min}^2)^2\nu^2$, $T_2 = 4(k_{1,\min}u_0+k_{2,\min}v_0)^4+4(k_{1,\min}^4+k_{2,\min}^4)(k_{1,\min}u_0+k_{2,\min}v_0)^2\nu^2+k_{1,\min}^2k_{2,\min}^2(k_{1,\min}^2+k_{2,\min}^2)^2\nu^4$ and $T_3 = (k_{1,\min}u_0+k_{2,\min}v_0)^2+(k_{1,\min}^2+k_{2,\min}^2)^2\nu^2$, and the associated maximum of the convergence factor ρ is given by

$$\max_{(k_1,k_2)\in\mathbb{K}}\rho(k_1,k_2,\nu,\alpha_{opt},\theta_{opt}) = \left|\frac{\alpha_{opt}\nu-2}{\alpha_{opt}\nu+2}\right|.$$

This expression is of the same form as in the case of the Stokes equation, but with a different optimized relaxation parameter α_{opt}. Expanding α_{opt} in ν for ν small gives

$$\alpha_{opt} = C_1+C_2\nu^2+O(\nu^4),$$

where the constants C_1 and C_2 depend on $k_{1,\min}$, $k_{2,\min}$, u_0 and v_0 only. This reveals that for a small viscosity ν the Oseen problem is harder to solve with DSSR than the Stokes problem.

Remark 3.2 The relaxation parameter θ does not appear in the resulting maximum of the convergence factor, and one might wonder what it is useful for. To investigate this, we analyzed the DSSR iteration without the parameter θ, i.e. with $\hat{E}_1 = \texttt{diag}(0,1,1)$ and $\hat{E}_2 = \texttt{diag}(1,0,1)$. The optimized parameter α^* we then get is related to the earlier one including θ by the relation $\alpha^* = \frac{1}{2}\alpha_{opt}$, which leads to the maximum of the convergence factor $\max\rho = \left|\frac{\alpha^* v-1}{\alpha^* v+1}\right|$, the same value as in the case where $\theta_{opt} = 1/2$ is applied. The convergence factor is not improved, as expected, but the relation between α^* and $\frac{1}{2}\alpha_{opt}$ shows that including the relaxation parameter θ in the optimization increases the region of the relaxation parameter that leads to best performance. This also holds for the 2D Stokes problem.

4 Numerical experiments

In this section we test DSSR numerically. We discretize the Stokes and Oseen problems using the marker and cell (MAC) method invented by Harlow and Welch [24], see also [28] and the independent work by Lebedev [25]. All the experiments are run on a MacBook Pro with a 2.4 GHz Intel Core i5 CPU (dual core) and 8GB memory, and except for Subsection 4.4, backslash in Matlab is used to solve the inner linear systems. In our Fourier analysis, we did not consider usual boundary conditions for the Stokes problem, since we used an unbounded domain. To illustrate the accuracy of our analysis, we first solve the Stokes equations with periodic boundary conditions, and then study numerically the influence of Dirichlet boundary conditions on our optimization results. We then also test the performance of DSSR for steady Oseen problems and investigate the use of inexact inner solvers to further improve the performance for large scale linear systems. We also compare the DSSR method with the RDF preconditioner, which was shown to be better than many other state-of-the-art preconditioners [12], such as the pressure convection diffusion preconditioner PCD, the modified version mPCD and the least squares commutator preconditioner LSC, see [18].

4.1 Stokes equation with periodic boundary conditions

In this subsection, we consider the Stokes equation (3.1) on the unit square and also the 3D version on the unit cube with periodic boundary conditions. This problem is singular, also after the discretization with the MAC scheme. In 2D, there are three different eigenvectors associated with the zero eigenvalue, corresponding to constant values of u, v, and p. One way to deal with these singularities would be to use special variants of GMRES, see for example [13] and in particular [26]. Since the singular modes are however explicitly known, we just remove them from the operators by projection in our experiments when needed.

We plot the spectrum of the iteration matrix M_α in Figure 4.1, both for the 2D and 3D case. We clearly see the value 1 in the spectrum, stemming from the singular modes. If we neglect this value and draw the spectral radius for the remaining eigen-

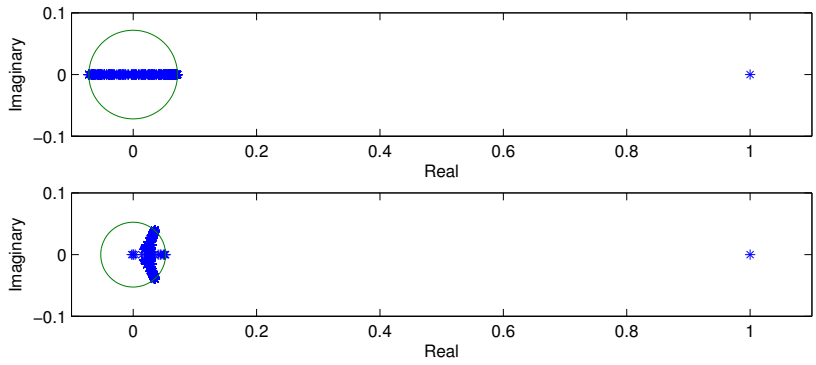

Fig. 4.1 The spectrum of the iteration matrix M_α for the periodic case with viscosity $\nu = 0.01$. Top: 2D Stokes with $\alpha = \alpha^*$ from Theorem 3.2 and mesh size $h = 1/40$. Bottom: 3D Stokes with $\alpha = \alpha_{opt}$ and mesh size $h = 1/10$.

ν	1	0.1	0.01	0.001	0.0001
2D spectral radius	0.0718	0.0718	0.0718	0.0718	0.0718
3D spectral radius	0.0525	0.0525	0.0525	0.0525	0.0525

Table 4.1 The spectral radius of the iteration matrix M_α versus the viscosity ν, for 2D Stokes with $\alpha = \alpha^*$ from Theorem 3.2 and mesh size $h = 1/40$, for 3D Stokes with $\alpha = \alpha_{opt}$ and mesh size $h = 1/10$.

values, we find in 2D 0.0718, as predicted in Theorem 3.2 by the formula $\frac{2-\sqrt{3}}{2+\sqrt{3}}$, and in 3D we get 0.0525, also as predicted.

To study how robust the DSSR method is with the optimized choice of the relaxation parameter, we show in Table 4.1 the spectral radii, by removing again the singular modes giving the eigenvalue 1, when ν is becoming smaller and smaller. We clearly see that the spectral radius does not depend on ν with the optimal choice of the relaxation parameter, and thus the method is robust[2].

We next show in Figure 4.2 how the different choice of α from Theorem 3.3 permits the formation of two clusters in the spectrum of the preconditioned operator. We see that the eigenvalues form two tight clusters when α becomes large, which can be advantageous for Krylov methods, and confirms our theoretical results based on Fourier analysis.

4.2 Stokes equation with Dirichlet boundary conditions

We now study numerically the more realistic case of the Stokes equation (3.1) with homogeneous Dirichlet boundary conditions. This operator is again singular, with the constant pressure mode being an eigenvector with zero eigenvalue, also after the discretization with the MAC scheme. We again remove this mode by projection when needed in the DSSR iteration.

We show the spectrum of the iteration matrix M_α on the left of Figure 4.3 with

[2] We thank an anonymous referee for pointing out that in the Stokes case, ν could actually be scaled out of the problem.

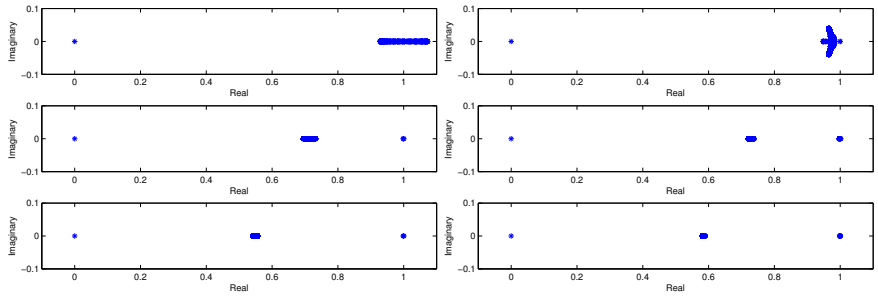

Fig. 4.2 Spectrum of the preconditioned matrix for the viscosity $\nu = 0.01$. The left column is for the 2D Stokes problem with mesh size $h = 1/40$: on the top for $\alpha = 100\sqrt{3}$ which is optimal for the stationary iterative method, in the middle for $\alpha = 200\sqrt{3}$ and at the bottom for $\alpha = 300\sqrt{3}$ which is better for the Krylov method. The right column is for the 3D Stokes problem with mesh size $h = 1/10$: on the top for $\alpha = \alpha_{opt}$, in the middle for $\alpha = 1.5\alpha_{opt}$ and at the bottom for $\alpha = 2\alpha_{opt}$.

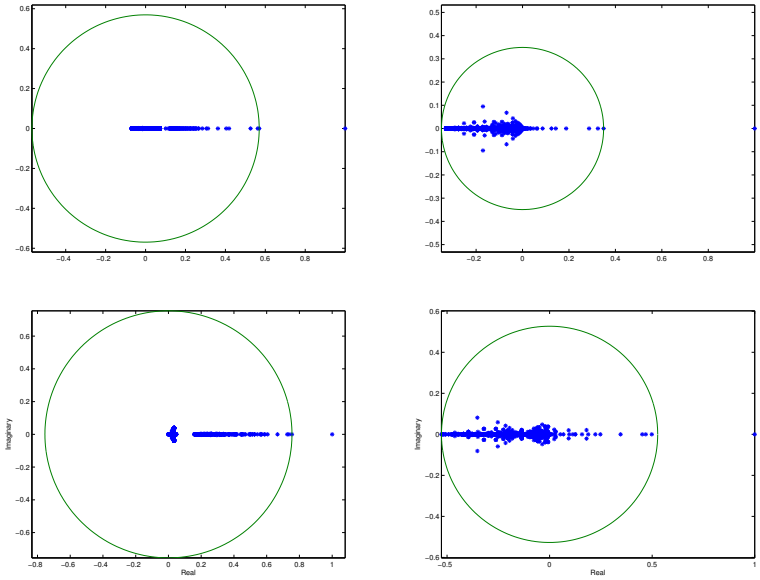

Fig. 4.3 Top: spectrum of the iteration matrix for 2D Stokes with homogeneous Dirichlet boundary conditions, mesh size $1/40$ and viscosity $\nu = 0.01$. On the left for $\alpha = \sqrt{3}/\nu$ and on the right for $\alpha = 1/\nu$. Bottom: spectrum of the iteration matrix for 3D Stokes with homogeneous Dirichlet boundary conditions, mesh size $1/10$ and viscosity $\nu = 0.01$. On the left for $\alpha = \alpha_{opt}$ and on the right $\alpha = \alpha_{opt}/\sqrt{5}$.

the optimized parameter from our analysis. We see that with the Dirichlet boundary conditions, the spectral radius of the preconditioned operator is substantially bigger, and our optimized choice for the periodic case does not give equi-oscillation any more. It is however just the constant in our estimate which has to be changed slightly, both in 2D and 3D, as we can see on the right in Figure 4.3, to get back to equi-oscillation. This is the best performance one can achieve with DSSR as a stationary iterative method, and we see that because of the Dirichlet boundary conditions, the

ν	1	0.1	0.01	0.001	0.0001
$\alpha = \sqrt{3}/\nu$	0.5694	0.5694	0.5694	0.5694	0.5694
$\alpha = 1/\nu$	0.3492	0.3492	0.3492	0.3492	0.3492

Table 4.2 The spectral radius of the iteration matrix M_α versus the viscosity ν for 2D Stokes and mesh size $h = 1/40$.

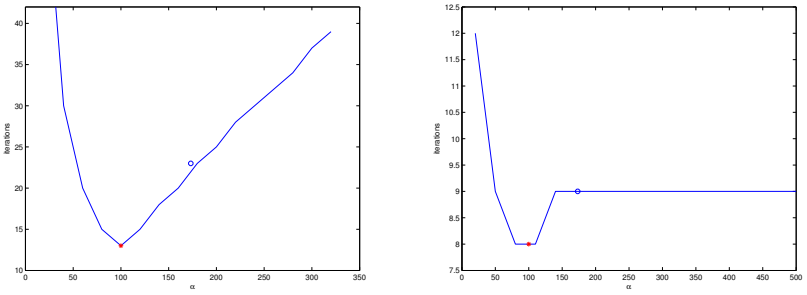

Fig. 4.4 Number of iterations for 2D Stokes required by the stationary iterative DSSR method (left) and using DSSR as a preconditioner for GMRES (right) with $\alpha = 1/\nu$ (the red star) compared to $\alpha = \sqrt{3}/\nu$ (the black circle), as well as other choices of the relaxation parameters. The mesh size is $h = 1/80$ and the viscosity $\nu = 0.01$.

performance will be slower than in the case of periodic boundary conditions or on an unbounded domain.

To investigate if the dependence on ν of the optimized parameter based on Fourier analysis remains correct also with Dirichlet boundary conditions, and if the DSSR method is robust in ν in that case, we show for the fixed mesh size $h = 1/40$ in Table 4.2 the spectral radii (removing again the pressure mode with eigenvalue 1) depending on a diminishing viscosity ν. We observe again that the spectral radius does not depend on ν for both the optimal choice of the relaxation parameter $\alpha^* = \sqrt{3}/\nu$ based on Fourier analysis, and the choice with the smaller constant chosen based on our numerical experiments, $\alpha = 1/\nu$. In both cases, DSSR is a robust solver when the viscosity becomes small, and similar results are obtained in 3D.

We next test if the numerically chosen constant $\alpha = 1/\nu$ is indeed optimal for Dirichlet boundary conditions. To this end, we vary the relaxation parameter α for 2D Stokes with 16 samples and calculate for each α the number of iterations required by the stationary iterative DSSR method, and using DSSR as preconditioner for GMRES, see Figure 4.4. This shows that with Dirichlet boundary conditions, the choice $\alpha = 1/\nu$ (the red star) is indeed optimal, and better than our prediction $\alpha = \sqrt{3}/\nu$ (the black circle) based on Fourier analysis, which could not take Dirichlet boundary conditions into account.

We next investigate if it is also possible in the case of Dirichlet boundary conditions to form two clusters in the spectrum, see Figure 4.5. We observe again that the eigenvalues form two clusters, but with the Dirichlet conditions, it is not possible any more to make the second cluster as tight as we want, and thus the Krylov method will not become a direct solver in the presence of Dirichlet conditions. However, one

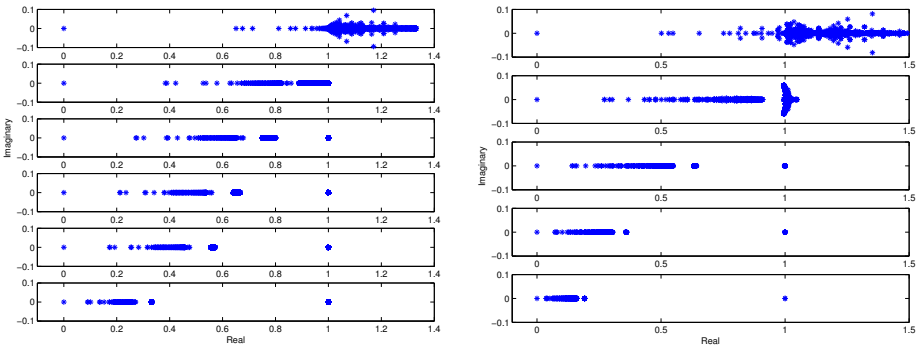

Fig. 4.5 Left: spectrum of the DSSR preconditioned matrix for 2D Stokes with homogeneous Dirichlet boundary conditions and mesh size $h = 1/40$ and viscosity $\nu = 0.01$. From the top to bottom $\alpha = 100, 200, 300, 400, 500, 1000$. Right: spectrum of the preconditioned matrix for 3D Stokes with homogeneous Dirichlet boundary condition for mesh size $h = 1/10$ and viscosity $\nu = 0.01$. From top to bottom $\alpha = \alpha_{no}, 2\alpha_{no}, 4\alpha_{no}, 8\alpha_{no}$ and $16\alpha_{no}$ with $\alpha_{no} = \alpha_{opt}/\sqrt{5}$.

2D Stokes equation: h	1/20	1/40	1/80	1/160	1/320
DSSR $\alpha = 1/\nu$	12	12	13	13	14
DSSR $\alpha = \sqrt{3}/\nu$	20	20	22	23	23
DSSRGMRES $\alpha = 1/\nu$	8	8	8	8	9
DSSRGMRES $\alpha = \sqrt{3}/\nu$	8	8	9	9	9
3D Stokes equation: h	1/20	1/30	1/40	1/50	1/60
DSSR $\alpha = \alpha_{opt}/\sqrt{5}$	22	23	23	24	24
DSSR $\alpha = \alpha_{opt}$	47	49	50	51	51
DSSRGMRES $\alpha = \alpha_{opt}/\sqrt{5}$	11	11	11	12	12
DSSRGMRES $\alpha = \alpha_{opt}$	12	12	12	13	13

Table 4.3 Number of iterations required by DSSR for homogeneous Dirichlet conditions for $\nu = 0.01$.

can still benefit from this observation: increasing α does not remarkably result in the increase of iteration counts for large α, see the right plot in Figure 4.4.

We finally show the number of iterations required by the stationary iterative DSSR method, as well as using DSSR as a preconditioner for GMRES in Table 4.3. We see that the iteration number does not depend on the mesh size, as predicted by the Fourier analysis.

We now compare DSSR with the RDF preconditioner [12]. In Figure 4.6 we see that for the 2D Stokes problem the spectrum of the DSSR preconditioned matrix with the numerically best parameter $1/\nu$ clustered the eigenvalues much farther from the zero eigenvalues than the RDF preconditioned one with experimentally optimal parameter, which is good for Krylov type solvers. The parameter α^* from the Fourier analysis for the DSSR preconditioner leads to similar result as the RDF preconditioner. In Table 4.4 we show the number of iterations required by the RDF preconditioner for 2D and 3D Stokes problems with homogeneous Dirichlet boundary conditions and the viscosity $\nu = 0.01$, where the relaxation parameter is numerically optimized. We see that the RDF preconditioner with the numerically optimized pa-

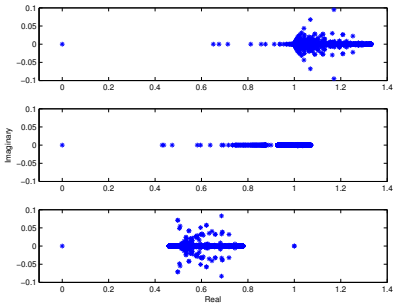

Fig. 4.6 Comparison of the spectrum for the DSSR preconditioned and RDF preconditioned matrix for the 2D Stokes equation with homogeneous Dirichlet boundary conditions for $\nu = 0.01$. On the top for the DSSR preconditioned matrix with numerically optimal parameter $1/\nu$, in the middle with the theoretically optimized parameter α^* and at bottom for the RDF preconditioned matrix with experimentally optimized parameter $\alpha = 28$.

2D Stokes h	1/20	1/40	1/80	1/160	1/320
RDFGMRES	8	9	9	9	9
3D Stokes h	1/20	1/30	1/40	1/50	1/60
RDFGMRES	10	11	11	11	11

Table 4.4 Number of iterations required by RDF preconditioned GMRES for the 2D and 3D Stokes equation with $\nu = 0.01$, with the experimentally best relaxation parameter.

rameter gives very similar iteration counts to the DSSR method for both 2D and 3D Stokes problems, see Table 4.3.

4.3 Steady Oseen equation

We consider the 2D steady Oseen problem with an exact solution of the form

$$u = \sin \pi x \sin \pi y, \quad v = x(1-x)y(1-y), \quad p = (x-\frac{1}{2})(y-\frac{1}{2}). \tag{4.1}$$

In Table 4.5, we show the iteration counts for DSSR preconditioned GMRES(20) stopped when the relative residual reaches $1e-6$. We see that the iteration counts are robust with respect to mesh refinement, but increase slowly when ν decreases.

We next compare our DSSR preconditioner with the RDF preconditioner for the Oseen problem with the exact solution in (4.1). To this end, we calculate only the fifth step of the Picard iteration. In Figure 4.7 for each integer α we show the number of iterations required by GMRES(20) preconditioned by DSSR (left column) and RDF (right column) as a function of the relaxation parameter α, for a mesh size $h = 1/80$ and the viscosity $\nu = 0.05$ (top row) and $\nu = 0.005$ (bottom row). We see that for the best relaxation parameters, both the DSSR and the RDF preconditioners require the same number of iterations for both $\nu = 0.05$ and $\nu = 0.005$. However, the region in which the DSSR preconditioner reaches its best performance is substantially wider than the one for the RDF preconditioner. This becomes even more apparent for small

	Picard Iteration	$h = 1/20$	$h = 1/40$	$h = 1/80$	$h = 1/160$
$\nu = 0.5$	1	10	10	10	10
	2	8	9	9	9
	3	5	5	5	5
	4	2	2	2	2
	5	0	0	0	0
$\nu = 0.05$	1	9	9	10	10
	2	11	11	11	11
	3	9	9	9	9
	4	7	7	7	7
	5	6	6	6	6
$\nu = 0.005$	1	9	9	10	10
	2	21	20	19	19
	3	20	19	18	18
	4	21	19	18	18
	5	19	16	16	16

Table 4.5 Iteration counts for the DSSR preconditioned Oseen problem with exact solution (4.1) for different values of the viscosity ν. For the first step we solve a steady Stokes problem with $\alpha = 1/\nu$ and in the following steps the Fourier analyzed relaxation parameter α_{opt} from (3.12) is applied.

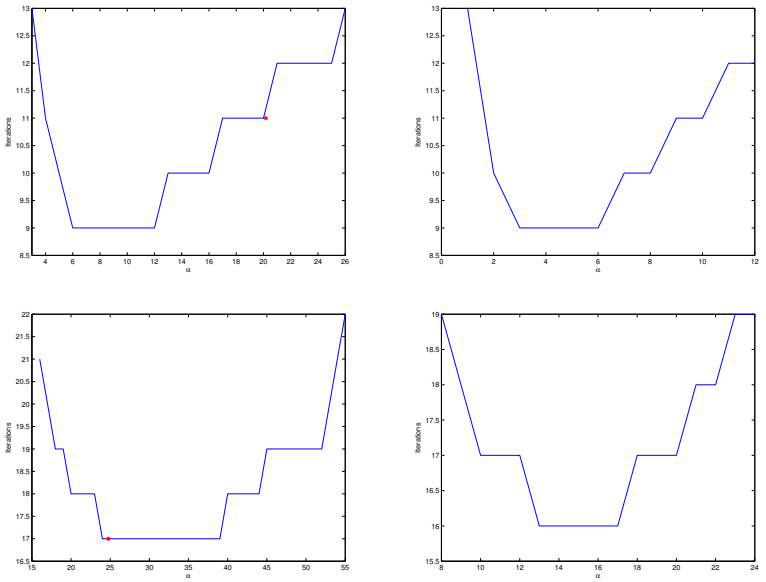

Fig. 4.7 Number of iterations required by preconditioned GMRES(20) for the fifth step of a Picard iteration for the steady Oseen problem with mesh size $h = 1/80$ as a function of the relaxation parameter. The left column is for the DSSR preconditioner, where the red star indicates the position of the Fourier analysis predicted parameter α_{opt} given by (3.12). The right column is for the RDF preconditioner. The top row corresponds to the viscosity $\nu = 0.05$ and the bottom row is for $\nu = 0.005$.

	h	1/20	1/30	1/40	1/50	1/60
EXACT	α_{opt}	12	12	12	13	13
	CPU	2.6347	20.2395	135.0461	620.8271	1786.7771
	$\alpha_{opt}/\sqrt{5}$	11	11	11	12	12
	CPU	2.3534	20.7897	110.8607	592.6696	1662.2182
ILU-GMRES	α_{opt}	17	18	18	17	17
	CPU	0.7925	3.1885	8.1035	15.7264	35.0197
	$\alpha_{opt}/\sqrt{5}$	15	15	16	16	17
	CPU	0.6836	2.6189	6.7926	14.5837	31.2215
ILU-PCG	α_{opt}	19	19	18	17	17
	CPU	0.7018	2.5171	6.4408	11.5749	21.2642
	$\alpha_{opt}/\sqrt{5}$	15	16	15	15	16
	CPU	0.6807	2.1094	5.1505	9.7508	18.3849

Table 4.6 GMRES iteration counts and timings for the 3D Stokes equation for $\nu = 0.01$.

viscosities. In addition, we see that our Fourier analysis predicted relaxation parameter α_{opt} (the red star in the left column of the Figure) is quite efficient, especially for small viscosities.

4.4 Inner and outer iterations

In the solution process, DSSR requires at each step to solve linear systems with $M_i = (\alpha E_i + H_i)$. For the 2D Stokes equations, we have the factorizations

$$M_1 = \begin{pmatrix} I & 0 & \frac{2}{\alpha}B_1^T \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} \hat{A}_1 & 0 & 0 \\ 0 & \alpha I & 0 \\ -B_1 & 0 & \frac{\alpha}{2}I \end{pmatrix}, \quad M_2 = \begin{pmatrix} \alpha I & 0 & 0 \\ 0 & \hat{A}_2 & B_2^T \\ 0 & 0 & \frac{\alpha}{2}I \end{pmatrix} \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\frac{2}{\alpha}B_2 & I \end{pmatrix},$$

with $\hat{A}_i = A_i + \frac{2}{\alpha}B_i^T B_i$. DSSR therefore requires to solve two linear systems at each step with coefficient matrices $\hat{A}_1$ and $\hat{A}_2$. The situation for 3D Stokes is similar. Solving inner systems with matrix $\hat{A}_i$ corresponds to solving elliptic partial differential equations, and for the Stokes problem, the matrices $\hat{A}_i$ are symmetric positive definite. Hence one can choose an efficient iterative solver for these inner problems, and thus improve the performance of DSSR further using inner and outer iterations.

We thus replace now the direct inner solver used so far by either GMRES(20) or PCG, both preconditioned by ILU(0) with a relative residual tolerance of $1e-1$. In these experiments, we had to use flexible GMRES (FGMRES [27]) as the outer solver, instead of standard GMRES and we solve the right preconditioned linear system. We show for the 3D Stokes equation the iteration count and CPU time corresponding to the relaxation parameter α_{opt}, as well as the scaled one $\alpha_{opt}/\sqrt{5}$ in Table 4.6 for $\nu = 0.01$. Comparing with the results for solving the inner linear systems exactly (see Table 4.6), we see that the use of inner and outer iterations can dramatically save CPU time. For example with ILU-PCG as inner solver and FGMRES as outer solver and our optimized relaxation parameter $\alpha_{opt}/\sqrt{5}$, we can reach the relative residual $1e-6$ in 56 steps costing 265.7382 seconds for a large system with mesh size $h = 1/100$, which corresponds to 3970000 unknowns.

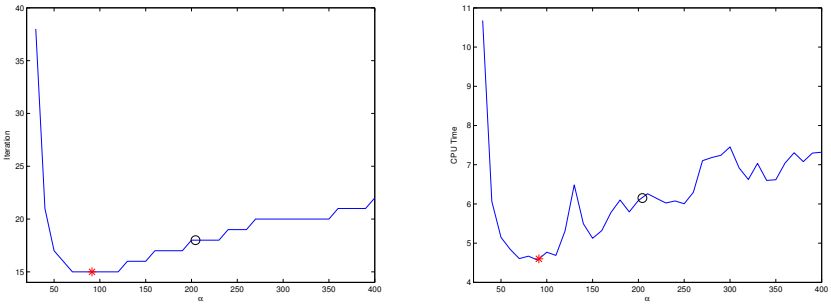

Fig. 4.8 The number of iterations and the CPU time in seconds as a function of the relaxation parameter α for 3D Stokes problems with mesh size $h = 1/40$ and viscosity $\nu = 0.01$. The black circle indicates the position of the Fourier analysis predicted parameter α_{opt} and the red star the one scaled by $1/\sqrt{5}$.

We next investigate how our Fourier analysis predicted relaxation parameter performs in the case of using inner and outer iterations. We use as the inner solver ILU-PCG, $\nu = 0.01$, $h = 1/40$ and vary the relaxation parameter α from 30 to 400 with 38 equidistant samples and record for each sample the number of iterations and the CPU time required. The results are shown in Figure 4.8. We can see that our Fourier analysis predicted parameter α_{opt} is fairly good both for iteration counts and CPU time, and the scaled parameter $\alpha_{opt}/\sqrt{5}$ is close to best possible.

We see as well from Figure 4.8 that increasing the parameter α, the number of iterations and the corresponding CPU time increases slowly, as it was indicated by our analysis of the clustering property of DSSR.

4.5 A lid-driven cavity flow

In this last subsection, we consider a lid-driven cavity flow: we solve the Stokes equation with homogeneous Dirichlet boundary conditions, except on top where we impose an inhomogeneous boundary condition $u = 1$. With this model problem, we compare our DSSR method with the PHSS method [6,4], where in the PHSS method the matrix C is taken as $B_1\hat{A}_1^{-1}B_1^T + B_2\hat{A}_2^{-1}B_2^T$, with the matrices $\hat{A}_1$ and $\hat{A}_2$ being the block diagonal part of A_1 and A_2, see [6] for more details. In Table 4.7 we show the number of iterations as well as the CPU time required by both the DSSR and PHSS methods to reach a relative residual of $1e-6$ when used as stationary iterative solvers and preconditioners of GMRES(20). For the PHSS method, we used the numerically optimized relaxation parameter, i.e. the one that worked best for the method in the experiment. We observe that even though the iteration counts are higher, the DSSR methods are much faster than the PHSS method, especially when the mesh size is small, where the inverse of a Schur complement like matrix becomes very expensive in PHSS. The solution for mesh size $h = 1/30$ is shown in Figure 4.9, on the left the velocity field, and on the right, using a log scaling, the level sets of the absolute value of the velocity.

	h	1/20	1/40	1/80	1/160
DSSR	$\alpha = \sqrt{3}/\nu$	40	42	43	44
	CPU	0.1179	0.4771	2.7185	11.7824
DSSRGMRES	$\alpha = \sqrt{3}/\nu$	8	8	8	9
	CPU	0.0828	0.2047	0.7873	3.6072
DSSR	$\alpha = 1/\nu$	24	25	26	26
	CPU	0.0683	0.3466	1.9072	7.9265
DSSRGMRES	$\alpha = 1/\nu$	8	8	8	8
	CPU	0.0858	0.1983	0.7755	3.3757
PHSS	$\alpha = \alpha_{exp}$	14	14	15	/
	CPU	1.6004	43.5979	2061.7826	/
PHSSGMRES	$\alpha = \alpha_{exp}$	2	2	2	/
	CPU	1.0081	20.0527	775.7383	/

Table 4.7 Number of iterations and CPU time (in seconds) required by DSSR compared to PHSS for the lid-driven cavity flow described in Subsection 4.5, used as both stationary iterations and preconditioners for GMRES(20) with $\nu = 0.01$. For PHSS, we used numerically optimized relaxation parameters, and for $h = 1/160$, the calculation exceeded the capacity of the computer used.

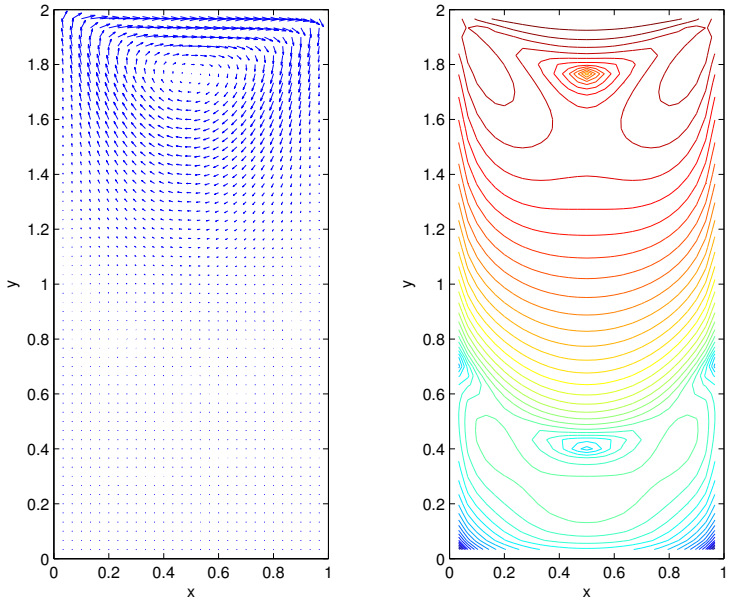

Fig. 4.9 Plot of the lid-driven cavity flow, on the left the velocity field, and on the right the level sets of the absolute value of the velocity using a log scale.

5 Conclusion

We proposed a new iterative method called dimension-wise splitting with selective relaxation (DSSR) for saddle point problems arising from the discretization of the Navier-Stokes equation. We used Fourier analysis to study the performance of the

DSSR iterative method applied to the Stokes and Oseen problems, and obtained optimized relaxation parameters, which capture correctly the dependence on the viscosity ν also in cases where Fourier analysis is not applicable. The numerical observation that for the Stokes problem with Dirichlet conditions, the optimized relaxation parameter should be scaled by $1/\sqrt{3}$ in 2D and $1/\sqrt{5}$ in 3D is intriguing, and merits further investigation.

With the asymptotically optimal relaxation parameter, DSSR is robust in the mesh size h and in ν as well. We have also observed numerically that the performance of DSSR is comparable to the performance of RDF, but DSSR is more robust in the relaxation parameter α, and we have an analytical formula for it. The numerical observation that PHSS preconditioned GMRES converges in two steps in our experiments deserves further investigation.

A Optimizing θ in DSSR for the Stokes problem

In this appendix we show that $\theta = 1/2$ is the best parameter value for DSSR when applied to the 2D Stokes problem. Using Fourier analysis as in Subsection 3.1 for the Stokes equation (3.1) with

$$\hat{E}_1 = \mathtt{diag}(0,1,\theta), \quad \hat{E}_2 = \mathtt{diag}(1,0,1-\theta),$$

one can compute the eigenvalues of the iteration matrix M_α, and we get $\lambda_{1,2}(k_1,k_2) = 0$ and

$$\lambda_3(k_1,k_2,\theta) = \frac{(\alpha\nu\theta(k_1^2+k_2^2)-k_2^2)(\alpha\nu(1-\theta)(k_1^2+k_2^2)-k_1^2)}{(\alpha\nu\theta(k_1^2+k_2^2)+k_1^2)(\alpha\nu(1-\theta)(k_1^2+k_2^2)+k_2^2)}.$$

We then find

$$\lim_{k_2\to 0}|\lambda_3(k_1,k_2,\theta)| = \left|\frac{\alpha\nu(\theta-1)\theta-\theta+1}{\alpha\nu(\theta-1)\theta-\theta}\right| =: \rho_{10}(\alpha,\nu,\theta),$$

$$\lim_{k_2\to 0}|\lambda_3(k_1,k_2,\theta)| = \left|\frac{\alpha\nu(\theta-1)\theta+\theta}{\alpha\nu(\theta-1)\theta+\theta-1}\right| =: \rho_{20}(\alpha,\nu,\theta).$$

It is easy to verify that ρ_{10} decreases monotonically in θ, ρ_{20} increases monotonically in θ for $\theta \in (0,1)$ and $\theta = 1/2$ solves $\rho_{10} = \rho_{20}$. Thus, if $\theta \neq 1/2$ we would find a spectral radius of $\hat{M}_\alpha$ with $\max_{k_1,k_2}\rho(k_1,k_2,\nu,\alpha) \geq \max\{\rho_{10},\rho_{20}\} > \rho_{10}(\alpha,\nu,\frac{1}{2})$, which shows that $\theta = 1/2$ is the best value for the relaxation parameter θ introduced in Section 2.

Acknowledgment

The authors would like to thank the organizing committee for the wonderful conference NASC2014, where the authors met each other and started their collaboration on this interesting topic. They are also very thankful for the constructive comments of the anonymous referees, which substantially enhanced the content and structure of this manuscript.

References

1. Bai, Z.-Z.: Optimal parameters in the HSS-like methods for saddle-point problems. Numer. Linear Algebra Appl. **16**(6), 447–479 (2009)
2. Bai, Z.-Z., Golub, G.H.: Accelerated Hermitian and skew-Hermitian splitting iteration methods for saddle-point problems. IMA J. Numer. Anal. **27**(1), 1–23 (2007)
3. Bai, Z.-Z., Golub, G.H., Li, C.-K.: Optimal parameter in Hermitian and skew-Hermitian splitting method for certain two-by-two block matrices. SIAM J. Sci. Comput. **28**(2), 583–603 (2006)

4. Bai, Z.-Z., Golub, G.H., Li, C.-K.: Convergence properties of preconditioned Hermitian and skew-Hermitian splitting methods for non-Hermitian positive semidefinite matrices. Math. Comp. **76**(257),287–298(2007)
5. Bai, Z.-Z., Golub, G.H., Ng, M.K.: Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems. SIAM J. Matrix Anal. Appl. **24**(3), 603–626 (2003)
6. Bai, Z.-Z., Golub, G.H., Pan, J.-Y.: Preconditioned Hermitian and skew-Hermitian splitting methods for non-Hermitian positive semidefinite linear systems. Numer. Math. **98**(1), 1–32(2004)
7. Bennequin, D., Gander, M.J., Halpern, L.: A homographic best approximation problem with application to optimized Schwarz waveform relaxation. Math. Comp. **78**(265), 185–223 (2009)
8. Bennequin, D., Gander, M.J., Gouarin, L., Halpern, L.: Optimized Schwarz waveform relaxation for advection reaction diffusion equations in two dimensions. Numer. Math. (to appear 2016)
9. Benzi, M., Gander, M.J., Golub, G.H.: Optimization of the Hermitian and skew-Hermitian splitting iteration for saddle-point problems. BIT **43**(5), 881–900 (2003)
10. Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. Acta Numer. **14**(1), 1–137 (2005)
11. Benzi, M., Guo, X.-P.: A dimensional split preconditioner for Stokes and linearized Navier-Stokes equations. Appl. Numer. Math. **61**(1), 66–76 (2011)
12. Benzi, M., Ng, M.K., Niu, Q., Wang, Z.: A relaxed dimensional factorization preconditioner for the incompressible Navier-Stokes equations. J. Comput. Phys. **230**(16), 6185–6202 (2011)
13. Brown, P.N., Walker, H.F.: GMRES on (nearly) singular systems. SIAM J. Matrix Anal. Appl. **18**(1), 37–51 (1997)
14. Cao, Y., Du, J., Niu, Q.: Shift-splitting preconditioners for saddle point problems. J. Comput. Appl. Math. **272**, 239–250 (2014)
15. Elman, H.C., Howle, V.E., Shadid, J., Shuttleworth, R., Tuminaro, R.S.: Block preconditioners based on approximate commutators. SIAM J. Sci. Comput. **27**(5), 1651–1668 (2006)
16. Elman, H.C., Howle, V.E., Shadid, J., Silvester, D.J., Tuminaro, R.S.: Least squares preconditioners for stabilized discretizations of the Navier-Stokes equations. SIAM J. Sci. Comput. **30**(1), 290–311 (2007)
17. Elman, H.C., Ramage, A., Silvester, D.J.: Algorithm 866: IFISS, a Matlab toolbox for modelling incompressible flow, ACM Trans. Math. Soft. **33**(2), 14(2007)
18. Elman, H.C., Tuminaro, R.S.: Boundary conditions in approximate commutator preconditioners for the Navier-Stokes equations. Electron. Trans. Numer. Anal. **35**, 257–280 (2009)
19. Gander, M.J.: Optimized Schwarz methods. SIAM J. Numer. Anal. **44**(2), 699–731 (2006)
20. Gander, M.J.: Schwarz methods over the course of time. ETNA **31**, 228–255 (2008)
21. Gander, M.J., Halpern, L.: Optimized Schwarz waveform relaxation for advection reaction diffusion problems. SIAM J. Numer. Anal. **45**(2), 666–697 (2007)
22. Gander, M.J., Xu, Y.: Optimized Schwarz methods for circular domain decompositions with overlap. SIAM J. Numer. Anal. **52**(4), 1981–2004 (2014)
23. Gander, M.J., Xu, Y.: Optimized Schwarz methods with nonoverlapping circular domain decomposition. Math. Comp. (to appear 2016)
24. Harlow, F.H., Welch, J.E.: Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. Physics of fluids **8**(12), 2182–2189 (1965)
25. Lebedev, V.I.: Difference analogues of orthogonal decompositions, basic differential operators and some boundary problems of mathematical physics. I. USSR Computational Mathematics and Mathematical Physics **4**(3), 69–92 (1964)
26. Reichel, L., Ye, Q.: Breakdown-free GMRES for singular systems. SIAM J. Matrix Anal. Appl. **26**(4), 1001–1021 (2005)
27. Saad, Y.: Iterative Methods for Sparse Linear Systems, SIAM, 2003.
28. Welch, J.E., Harlow, F.H., Shannon, J.P., Daly, B.J.: The MAC method-a computing technique for solving viscous, incompressible, transient fluid-flow problems involving free surfaces. Tech. Rep., Los Alamos Scientific Lab., Univ. of California, N. Mex. (1965)