Jan Šípek; Jan Zítko
Alternating iterative scheme for the solution of block-structured systems

In: Jan Chleboun and Petr Přikryl and Karel Segeth (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 6-11, 2004. Institute of Mathematics AS CR, Prague, 2004.
pp. 214–229.

Persistent URL: http://dml.cz/dmlcz/702799

ALTERNATING ITERATIVE SCHEME FOR THE SOLUTION OF BLOCK-STRUCTURED SYSTEMS *

Jan Šípek, Jan Zítko

Abstract

We consider the solution of linear system with a block-structured matrix of saddle point type. The solution technique is based on the idea of the classical alternating-direction implicit iterative method where symmetric-antisymmetric splitting of the coefficient matrix is used. To find an optimal parameter for solving the system with a symmetric matrix, the polynomial filters are considered. The CGW method is used for systems with skew-symmetric matrix. The numerical tests compare the results obtained by using alternating iteration and GMRES and point out advantages of alternative iterations for larger systems.

1. Introduction

We consider solution of linear system of equations with the following block structure

$$\underbrace{\begin{bmatrix} A & B^T \\ -B & C \end{bmatrix}}_{\mathcal{A}} \underbrace{\begin{bmatrix} u \\ p \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} f \\ -g \end{bmatrix}}_{b}, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{m\times m}$, $f \in \mathbb{R}^n$, $g \in \mathbb{R}^m$, and $m < n$. Moreover, we assume that $\operatorname{rank}(B) = m$ and $N(A) \cap N(B) = \{0\}$. Systems of the form (1) arise in a number of scientific and engineering applications. There are obtained by calculation of the minimum of a functional with m linear constrains or by the discretization of Navier-Stokes equations using the finite element or finite volume method; see References [1, 2, 14, 15]. In some special cases the system (1) has more interesting properties such as A is symmetric positive definite (SPD) or $C = 0$.

In our further investigation, we will assume that the matrix A is non-symmetric with a positive definite symmetric part $\frac{1}{2}(A + A^T)$. In such a case we refer matrix A as positive real. Define

$$\mathcal{G} := \frac{1}{2}(\mathcal{A} + \mathcal{A}^T), \quad \mathcal{S} := \frac{1}{2}(\mathcal{A} - \mathcal{A}^T) \tag{2}$$

and similarly

$$G = \frac{1}{2}(A + A^T), \quad S = \frac{1}{2}(A - A^T). \tag{3}$$

*This work was supported by Grant Agency of the Czech Republic under Grant No. 201/04/1503 and under Grant MSM113200007.

It was proved (see [7]) that if G is positive semidefinite, $\mathrm{rank}(B) = m$ and $N(G) \cap N(B) = \{0\}$, then $\mathcal{A}$ is nonsingular, semipositive real and positive semistable (i.e. the eigenvalues of $\mathcal{A}$ have nonnegative real part). A number of methods have been proposed for the solution of (1) in the literature. Let us mention [12], [7], [2], [16], [10], [3], [18], [19], [20]. Very interesting technique based on an idea of the classic alternating direction implicit method described in Reference [13] appears in the last years.

Let $\alpha \in \mathbb{R}$ be a positive number. Write

$$\mathcal{A} = (\mathcal{G} + \alpha\mathcal{I}) + (\mathcal{S} - \alpha\mathcal{I}) = (\mathcal{G} - \alpha\mathcal{I}) + (\mathcal{S} + \alpha\mathcal{I})$$

where $\mathcal{I}$ is identity matrix in $\mathbb{R}^{(n+m)\times(n+m)}$. Let us mention that

$$\mathcal{G} + \alpha\mathcal{I} = \begin{bmatrix} G + \alpha I_n & 0 \\ 0 & C + \alpha I_m \end{bmatrix}, \quad \mathcal{S} + \alpha\mathcal{I} = \begin{bmatrix} S + \alpha I_n & B^T \\ -B & \alpha I_m \end{bmatrix}. \tag{4}$$

According to the technique formulated for example in [13, 7], we write the following algorithm:

Algorithm 1: ADI method

Input: $x_0 = [u^0, p^0]^T$: an initial approximation.
Output: The solution x of (1)

For $k = 0$ **till convergence do**
$(\mathcal{G} + \alpha\mathcal{I})x^{k+\frac{1}{2}} = (\alpha\mathcal{I} - \mathcal{S})x^k + b,$ (G)
$(\mathcal{S} + \alpha\mathcal{I})x^{k+1} = (\alpha\mathcal{I} - \mathcal{G})x^{k+\frac{1}{2}} + b.$ (S)
end do

The matrix-vector multiplication on the right-hand side of (G) and (S) is evaluated only once for $k = 0$. The right-hand side of each following "half-iteration" yields the left-hand side of previously executed one, after an easy manipulation.

Putting

$$f_1^k = \alpha u^k - Su^k + f - B^T p^k, \quad g_1^k = \alpha p^k - g + Bu^k,$$

we can see, that the first part of our algorithm involve the solution of two linear systems of the form

$$(G + \alpha I_n)u^{k+\frac{1}{2}} = f_1^k, \quad (C + \alpha I_m)p^{k+\frac{1}{2}} = g_1^k \tag{5}$$

with positive definite matrices. Let us remark that the problems with $C = 0$ are very often solved and in this case only one equation with positive definite matrix $G + \alpha I_n$ is solved. When solving the linear system (5) with conjugate gradient method, the smallest eigenvalues slow down the convergence. The same phenomenon has been observed by solving nonsymmetric systems using GMRES method and preconditioning

technique removing the small eigenvalues from the spectrum was studied by many authors. The situation in symmetric positive definite case is much more easier because the matrix is diagonalizable with real positive eigenvalues. This is the reason for splitting the coefficient matrix $\mathcal{A}$ into symmetric and skew-symmetric part and solve more systems of linear algebraic equations. For removing small eigenvalues, an invariant subspace associated with these ones is usually constructed. We have proposed the method of polynomial filters modified for symmetric systems. The authors [Giraud, Ruiz, Touhami, Arioli] proposed an algorithm which combine Chebyshev iteration with a block Lanczos procedure to accurately compute an orthonormal basis for the invariant subspace associated with the small eigenvalues of the matrix A. It uses Chebyshev polynomials to damp eigencomponents associated with the largest eigenvalues in the spectral decomposition of the considered vector. Both approaches for an efficient solving of symmetric system have the common idea to enforce conjugate gradient to work in the orthogonal complement of some invariant subspace associated with the smallest eigenvalues.

The second system in (S), rewritten according (4) as

$$\begin{aligned} lcl(\alpha I_n + S)u^{k+1} + B^T p^{k+1} &= f_2^k & (6)\\ -Bu^{k+1} + \alpha p^{k+1} &= g_2^k, & (7) \end{aligned}$$

where

$$f_2^k = (\alpha I_n - G)u^{k+\frac{1}{2}} + f, \quad g_2^k = (\alpha I_m - C)p^{k+\frac{1}{2}} - g,$$

can be solved in various ways. The CGW method is often recommended for the solution of (6); see Reference [22]. An alternative approach [7] is to eliminate u^{k+1} from the second equation using the first, which is useful for symmetric matrix $\mathcal{A}$ or to eliminate p^{k+1} from the first equation using the second, which seems to be more useful for non-symmetric case. The first reduction described, leads on solution of following equation

$$[B(I_n + \alpha S)^{-1}B^T + \alpha I_m]p^{k+1} = B(I_n + \alpha S)^{-1}f_2^k + g_2^k. \tag{8}$$

From (8) we get the solution vector p^{k+1}. Vector u^{k+1} is defined by

$$u^{k+1} = (\alpha I_n + S)^{-1}(f_2^k - B^T p^{k+1}).$$

If $S = 0$, then the system (8) degenerates to

$$(BB^T + \alpha^2 I_m)p^{k+1} = Bf_2^k + \alpha g_2^k \tag{9}$$

and $u^{k+1} = \frac{1}{\alpha}(f_2^k - B^T p^{k+1})$. Relation (8) is far to complicated in non-symmetric case and numerically too expensive. Hence, let us try to express p^{k+1}:

$$p^{k+1} = \frac{1}{\alpha}(Bu^{k+1} + g_2^k) \tag{10}$$

and substitute to the first equation, from which we get u_{k+1}. Therefore we have

$$((\alpha I_n + S) + \frac{1}{\alpha} B^T B) u^{k+1} = f_2^k - \frac{1}{\alpha} B^T g_2^k. \tag{11}$$

The aim of this paper is to compare several of these techniques in terms of numerical efficiency and to show the advantages of the alternating iterations.

The outline of this paper is as follows. In Section 2 the convergence and estimation of an optimal shift parameter is discussed. We devote Section 3 and 4 to application of polynomial filters for symmetric case. Various filtering techniques are shown for the construction of an invariant subspace. In Section 5 the CGW method is studied. In Section 6 the numerical results are shown. The comparison of alternating-iteration method with Krylov subspace methods is presented.

2. Convergence and estimation of optimal parameter α

We are looking for such α, that the method introduced, converge optimally. From (G) and (S) we become the following one step iterative method of the form

$$x_{k+1} = (\mathcal{S} + \alpha\mathcal{I})^{-1}(\mathcal{G} - \alpha\mathcal{I})(\mathcal{G} + \alpha\mathcal{I})^{-1}(\mathcal{S} - \alpha\mathcal{I})x_k + c(\alpha).$$

Using the relation $\mathcal{A} = \mathcal{G} + \mathcal{S}$ we define following

$$\begin{aligned} \mathcal{M} &= (2\alpha)^{-1}(\alpha\mathcal{I} + \mathcal{G})(\alpha\mathcal{I} + \mathcal{S}), \\ \mathcal{N} &= (2\alpha)^{-1}(\alpha\mathcal{I} - \mathcal{G})(\alpha\mathcal{I} - \mathcal{S}), \\ \mathcal{W} &= \mathcal{I} - \mathcal{A}\mathcal{M}^{-1} = \mathcal{N}\mathcal{M}^{-1} = (\alpha\mathcal{I} - \mathcal{G})(\alpha\mathcal{I} - \mathcal{S})(\alpha\mathcal{I} + \mathcal{S})^{-1}(\alpha\mathcal{I} + \mathcal{G})^{-1}. \end{aligned}$$

For the iteration $\{x_k\}$ defined by Algorithm 1 satisfies equation

$$\mathcal{M}x_{k+1} = \mathcal{N}x_k + b.$$

Residual vectors of x_{k+1} can be expressed as

$$r_{k+1} = (\mathcal{I} - \mathcal{A}\mathcal{M}^{-1})r_k = \mathcal{W}r_k$$

and

$$r_{k+1} \perp \Big(\mathcal{I} - \frac{r_k^T \mathcal{W} r_k}{r_k^T \mathcal{W}^T \mathcal{W} r_k}\mathcal{W}\Big) r_k \stackrel{\text{def}}{=} w_k.$$

The goal is to find such α that fulfills the following

$$\arg\min_{\alpha>0} \|\mathcal{W}r_k\|/\|r_k\|.$$

In other words we are looking for such α that the angle between vectors r_k and w_k is minimal. However, such formulated task leads on solution of nonlinear problem. Hence a uniform estimate for the equation $\|\mathcal{W}r_k\|/\|r_k\|$ is applied, i.e. α^* (optimal α) is computed from the relation

$$\alpha^* = \arg\min_{\alpha>0} \|(\alpha\mathcal{I} - \mathcal{G})(\mathcal{G} + \alpha\mathcal{I})^{-1}\|. \tag{12}$$

Let us remark, that the matrix $(\alpha\mathcal{I}-\mathcal{S})(\alpha\mathcal{I}+\mathcal{S})^{-1}$ is orthonormal ($(\Leftarrow \mathcal{S}$ is skew-symmetric) and can be therefore omitted in the norm in (12).

We continue to explore the expression (12): Under the assumption that G is positive definite, C is symmetric positive semidefinite and B has maximum rank, it was shown in [7] that

$$\alpha^* = \underset{\alpha>0}{\arg\ \min} \max_{\substack{\lambda\in\{\lambda_1,\lambda_2,\ldots,\lambda_n\}\\ \lambda\in[\lambda_{\min},\lambda_{\max}]}} \left|\frac{\alpha-\lambda}{\alpha+\lambda}\right|, \tag{13}$$

where

$$\lambda_{\min} := \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n =: \lambda_{\max}$$

are eigenvalues of the matrix G. The solution (13) yields

$$\alpha* = \sqrt{\lambda_{\min}\lambda_{\max}}.$$

Moreover for the condition number of the matrix we have the estimate

$$\kappa(G+\alpha^* I) \leq 1+\sqrt{\kappa(G)}.$$

where G is the first diagonal block of matrix $\mathcal{G}$.

In the following will be shortly shown, how the relation (13) was developed and why the eigenvalues of matrix C disappeared.

Let

$$\mathcal{Z} = (\alpha\mathcal{I}-\mathcal{G})(\alpha\mathcal{I}+\mathcal{G})^{-1}(\alpha\mathcal{I}-\mathcal{S})(\alpha\mathcal{I}+\mathcal{S})^{-1}.$$

and

$$\nu_1 \leq \nu_2 \leq \ldots \leq \nu_m$$

be the eigenvalues of the matrix C. Because matrix $\mathcal{G}$ is symmetric, and matrix $(\alpha\mathcal{I}-\mathcal{S})(\alpha\mathcal{I}+\mathcal{S})^{-1}$ is orthonormal, there exists an orthonormal matrix $\mathcal{P}$ such, that

$$\mathcal{P}^T\mathcal{Z}\mathcal{P} = \underbrace{\mathcal{P}^T(\alpha\mathcal{I}-\mathcal{G})(\alpha\mathcal{I}+\mathcal{G})^{-1}\mathcal{P}}_{\begin{bmatrix} D_1 & 0\\ 0 & D_2\end{bmatrix} =: \mathcal{D}} \underbrace{\mathcal{Q}}_{\mathcal{P}^T(\alpha\mathcal{I}-\mathcal{S})(\alpha\mathcal{I}+\mathcal{S})^{-1}\mathcal{P}}, \tag{14}$$

where

$$\mathcal{Q} := \begin{bmatrix} Q_{11} & Q_{12}\\ Q_{21} & Q_{22}\end{bmatrix}$$

is orthonormal and

$$D_1 = \operatorname{diag}\left(\frac{\alpha-\lambda_1}{\alpha+\lambda_1},\ldots\frac{\alpha-\lambda_n}{\alpha+\lambda_n}\right), \quad \varrho(D_1)<1$$

and analogically with D_2, where we formally substitute ν_i for λ_i, $i=1,\ldots,m$. It holds $\varrho(D_2)\leq 1$. We try to show, that $\varrho(\mathcal{Q}\mathcal{D})<1$.

Let $(\lambda, [x_1, x_2]^T)$ be an eigenpair of $\mathcal{QD}$, that means

$$\begin{bmatrix} Q_{11}D_1 & Q_{12}D_2 \\ Q_{21}D_1 & Q_{22}D_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \overset{\text{if } x_1=0}{=} \begin{bmatrix} Q_{12}D_2x_2 \\ Q_{22}D_2x_2 \end{bmatrix} \overset{\text{if } x_1=0}{=} \begin{bmatrix} 0 \\ \lambda x_2 \end{bmatrix}.$$

It remains to show that for all eigenvalues λ the corresponding subvector x_1 is not a nullvector. From the last equation we immediately obtain

$$|\lambda|^2 = \|D_1x_1\|^2 + \|D_2x_2\|^2 \leq \varrho(D_1)^2\|x_1\|^2 + \varrho(D_1)^2\|x_2\|^2 < \|x\|^2 = 1. \tag{15}$$

In the rest of our proof we will show that the matrix Q_{12} has maximum rank. If we denote

$$(\alpha\mathcal{I} - \mathcal{S})(\alpha\mathcal{I} + \mathcal{S})^{-1} = \underbrace{\begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}}_{\mathcal{U}},$$

than

$$U_{12} = -2\alpha(\alpha I + S)B^T[\alpha I_m + B(\alpha I_m + S)^{-1}B^T]^{-1}. \tag{16}$$

Using (4) we evaluate $\mathcal{Q}$ from (14):

$$\mathcal{Q} = \underbrace{\begin{bmatrix} P_{11} & 0 \\ 0 & P_{22} \end{bmatrix}^T}_{\mathcal{P}^T} \underbrace{\begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}}_{\mathcal{U}} \underbrace{\begin{bmatrix} P_{11} & 0 \\ 0 & P_{22} \end{bmatrix}}_{\mathcal{P}} = \begin{bmatrix} P_{11}^TU_{11}P_{11} & P_{11}^TU_{12}P_{22} \\ P_{22}^TU_{21}P_{11} & P_{22}^TU_{22}P_{22} \end{bmatrix}.$$

Hence

$$Q_{12} = P_{11}^TU_{12}P_{22} = -2\alpha P_{11}^T(\alpha I + S)B^T \underbrace{[\alpha I_m + B(\alpha I_m + S)^{-1}B^T]^{-1}}_{E} P_{22}$$

where we have substituted from (16). Matrices P_{11} and P_{22} are orthonormal, the matrix E is nonsingular and B has maximal column rank, hence the matrix Q_{12} has maximum rank.

3. Polynomial filters

Assume that the eigenvalues of the matrix G are ordered according to

$$0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n.$$

Let $k << n$ be a positive integer. We want to find the eigenvalues included in the set $\sigma_w = \{\lambda_1, \ldots, \lambda_k\}$ and $\sigma_u = \{\lambda_{n-k+1}, \ldots, \lambda_n\}$ respectively.
Lanczos process starting with the vector v_1 gives the factorization

$$GV_m = V_mT_m + f_me_m^T, \tag{17}$$

where $V_m = \{v_1, v_2, \ldots, v_m\} \in \mathbb{R}^{n\times m}$, $f_m \in \mathbb{R}^n$, $T_m \in \mathbb{R}^{m\times m}$, $V_m^TV_m = I_m, V_m^Tf_m = 0$.

We will consider two ways for estimating the eigenvalues in σ_w or σ_u. Either for larger m to calculate the smallest and largest eigenvalues of the tridiagonal matrix T_m or for small m repeat the procedure based on the single-shift QR iterations to obtain more detail information about spectrum of the matrix G . The second procedure is analogous to the updating the Arnoldi factorization via QR-iteration (see References [24], [11])and we describe it shortly. Let μ_1, μ_2,..., μ_{m-k} be positive real numbers. From (17) we successively obtain

$$\begin{aligned}
(G-\mu_1 I)V_m - V_m(T_m-\mu_1 I) &= f_m e_m^T, &(18)\\
(G-\mu_1 I)V_m - V_m Q^{(1)}R^{(1)} &= f_m e_m^T, &(19)\\
(G-\mu_1 I)(V_m Q^{(1)}) - (V_m Q^{(1)})(R^{(1)}Q^{(1)}) &= f_m e_m^T Q^{(1)}, &(20)\\
G\underbrace{(V_m Q^{(1)})}_{V_m^{(1)}} - \underbrace{(V_m Q^{(1)})}_{V_m^{(1)}}\underbrace{(R^{(1)}Q^{(1)}+\mu_1 I)}_{T_m^{(1)}} &= f_m e_m^T Q^{(1)}. &(21)
\end{aligned}$$

Analogical application of the next shifts μ_2, μ_3, ..., μ_{m-k} gives

$$GV_m^{(m-k)} = V_m^{(m-k)}T_m^{(m-k)} + f_m e_m^T Q, \tag{22}$$

where

$$\begin{aligned}
&Q = Q^{(1)}Q^{(2)}\dots Q^{(m-k)}, \quad T_m^{(m-k)} = Q^T T_m Q \quad \text{and}\\
&V_m^{(m-k)}; = (v_1^{(m-k)}, v_2^{(m-k)}, \dots, v_m^{(m-k)}) = V_m Q,
\end{aligned}$$

The matrix $T_m^{(m-k)}$ has the form

$$T_m^{(m-k)} = \begin{pmatrix} T_k^{(m-k)} & t_k e_k e_1^T \\ t_k e_1 e_k^T & \widehat{T}_{m-k} \end{pmatrix},$$

and if we split $V_m^{(m-k)} = (V_k^{(m-k)}, \tilde{V}_{m-k}^{(m-k)})$ then

$$GV_k^{(m-k)} = V_k^{(m-k)}T_k^{(m-k)} + \underbrace{f_k^{(m-k)}}_{t_{k+1,k}v_{k+1}^{(m-k)}+f_m e_m^T Q e_k^T} e_k^T \tag{23}$$

where $V_k^T f_k^{(m-k)} = 0$ and the matrices $T_m^{(m-k)}$ and $T_k^{(m-k)}$ are tridiagonal positive definite because G is positive definite. The equalities (18)-(21) yield the formula

$$v_1^{m-k} = \nu\Pi_{j=1}^{m-k}(A-\mu_j I)v_1 \stackrel{\text{def}}{=} \psi(A)v_1 \tag{24}$$

The equation (23) gives Lanczos process with starting vector $V_k^{(m-k)}e_1$. Let

$$0 < \theta_1 \le \theta_2 \le \ldots \le \theta_m \tag{25}$$

be the eigenvalues of the matrix T_m. Now it is natural to put the following question. How to choose the shifts μ_j or alternatively a polynomial ψ such that $\|f_k^{m-k}\|$ is

"small", because if it would be in ideal case $\|f_k^{m-k}\| = 0$ then the columns of the matrix V_k^{m-k} form an invariant subspace of the matrix G. Let *tol* is a small positive number (for example $tol = 10^{-2}$). This will be discussed in the section concerning with numerical results. Let v_1 be the vector defined in (17). The polynomial ψ having the property that the Lanczos process starting with the initial vector $\psi(G)v_1$ gives $\|f_k^{m-k}\| \leq tol$ will be called **polynomial filter.** If the polynomial ψ is constructed iteratively then all polynomial approximations will be called filters.

However we want to get eigenspace corresponding to the smallest or largest eigenvalues of G. Let us think the smallest eigenvalues. The algorithm for largest eigenvalues can be implemented equally. By analogy to the [24] paper let us consider the following iterative process.

Algorithm 2: Construction of an invariant subspace	
Input:	$G, m, k, tol, \max -$ maximal number of iterations
Output:	matrices V_k and a tridiagonal matrix T_k fulfilling the relations $\|GV_k - V_kT_k\| \leq tol$, $V_k^TV_k = I$
Step 1	Carry out the Lanczos process (17), $cycle := 1$
Step 2	Calculate the eigenvalues $\theta_j, j = 1, 2, \ldots, m$ of T_m
Step 3	Substitute $\mu_j := \theta_{m-j}$ for $j = 1, 2, \ldots, k$ and apply these shifts to obtain (22) and (23).
Step 4	**If** $\|f_k^{m-k}\| < tol$ **or** $cycle >$ max **then** STOP **else** perform the next $m - k$ steps of the the Lanczos process starting from (23). The resulting tridiagonal matrix is denoted again T_m. Put $cycle = cycle + 1$, go to Step 2.
end.	

The convergence is studied in [24]. Let us assume that the process converges, i.e. $\|f_k^{(m-k)}\| < tol$ after performing a finite number of steps.
We have

$$AV_k \doteq V_kT_k$$

and span$\{V_k\}$ approximate eigenspace corresponding to the k eigenvalues of λ_1, λ_2,... λ_k G. (See Reference [24].) If $\{\theta_j^{(k)}, y_j^{(k)}\}_{j=1}^k$ are eigenpairs of T_k, and

$$x_j = V_k y_j^{(k)}, \quad j = 1, 2, \ldots, k \quad \text{the Ritz vectors}$$

then the easy manipulation shows that the pairs $(\theta_j^{(k)}, x_j)$ approximate the eigenpairs of G and, moreover, the eigenpairs (λ_j, u_j), where the matrix $U = (u_1, u_2, \ldots, u_n)$ transforms G to the Jordan canonical form Λ =diag $(\lambda_1, \lambda_2, \ldots, \lambda_n)$. It is an interesting property proved by [24] that the k smallest Ritz eigenvalues (i.e. smallest eigenvalues of the matrix T_m) approximate successively $(\lambda_1, \lambda_2, \ldots, \lambda_k)$. (See Reference [24])

4. The Chebyshev filtering technique

Let

$$G = U\Lambda U^T = U_1\Lambda_1 U_1^T + U_2\Lambda_2 U_2^T \tag{26}$$

where $\Lambda_1 = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ and $\Lambda_2 = \mathrm{diag}(\lambda_{k+1}, \lambda_2, \ldots, \lambda_n$ be the eigendecomposition of the symmetric and positive definite matrix G. Let v_1 be a starting vector for Lanczos process and ψ an arbitrary polynomial. From (26)it follows that

$$\psi(G)v_1 = U_1\psi(\Lambda_1)U_1^T v_1 + U_2\psi(\Lambda_2)U_2^T v_1 \tag{27}$$

We want to find ψ which sets off the first term on the right hand-side (27) and damps the second one. The min-max property of Chebyshev polynomials leads us to the idea to seek a polynomial ψ between them. Generally, the Chebyshev polynomials T_m of degree m are defined for all complex numbers by the formula $T_m(z) = \cosh(m \times \cosh^{-1}(z))$ as an analytical function (see Reference [25]). But all eigenvalues of G are real and hence real version of these polynomials will be sufficient for our further investigations. For the interval $[-1, 1]$ the polynomials T_m reduce to very easy form $T_m(z) = \cos(m \times \cos^{-1}(z))$. The practical algorithms utilize often the recurrence formulas

$$T_0(z) = 1, \quad T_1(z) = z \quad \text{and} \quad T_{s+1}(z) = 2zT_s(z) - T_{s-1}(z) \quad \text{for all integer } s \geq 1.$$

We want theoretically "annihilate" the norm of the matrix $\psi(\Lambda_2)$. for the approximation of the smallest interval containing diagonal elements of Λ_2 we apply again Lanczos process and Ritz values. Let m and k has the same meaning as in the formulas (17) and (22). Let γ and δ be a positive numbers. The transformation

$$z \mapsto \vartheta(\lambda) := \frac{\delta + \gamma - 2\lambda}{\delta - \gamma} \tag{28}$$

maps the interval $[\gamma, \delta]$ onto $[-1, 1]$ and $\vartheta(0) = \frac{\delta+\gamma}{\delta-\gamma} =: q$. Hence assuming the matrix G possesses the unwanted (the largest or smallest) eigenvalues in the interval $[\gamma, \delta]$, the polynomial ψ,

$$\psi(\lambda) = \frac{T_{m-k}(\vartheta(\lambda))}{T_{m-k}(q)}$$

normed that $\|\psi(G)v_1\| = 1$ appears to be a good polynomial filter. Let us remark that an analogous technique was used in the papers [Arioli, Ruiz atd.] where the Chebyshev damping with the block Lanczos method is effectively used for solving large linear systems with a symmetric and positive definite matrix. Now the question appears how numerically perform this damping of eigenvalues lying in $[\gamma, \delta]$ and how to apply the product $\psi(A)v_1$. The process is very similar to the Algorithm 1 and we underline here only the differences. Denote $t = m - k$ and carry out the Lanczos process (17), i.e. the Step 1 in Algorithm 1 and we calculate all eigenvalues of T_m, see Step 2.

We put $\gamma \in (0, \theta_1]$ and $\delta \in [\theta_k \theta_{k+1)})$ if small eigenvalues are damped, i.e. the invariant subspace for t largest eigenvalues of G is calculated or we substitute $\gamma \in (\theta_{k-1}\theta_k]$ and $\delta \in [\theta_m, \chi)$ for some $\chi > \theta_m$ in the opposite case. Let us still remark that the linear mapping

$$\lambda \mapsto \eta(\theta) := \delta + (1-\theta)(\gamma - \delta)/2 \tag{29}$$

transforms the interval $[-1, 1]$ onto $[\gamma, \delta]$ and to the roots of T_m in $[-1, 1]$ correspond the numbers $\mu_1, \mu_2, \ldots \mu_r$, where $r = k$ or $r = t$ dependent if the invariant subspace for t largest eigenvalues of G is calculated or conversely. Hence for the shifts we take the numbers μ_j

$$\mu_j = \eta(\cos\left(\frac{2j-1}{2t}\pi\right)) \quad \text{for} \quad j = 1, 2, \ldots r.$$

The Step 4 stays only instead of $m - k$ must be substituted k if largest eigenvalues are calculated.

5. Solution of systems with skew-symmetric matrix

In this section we deal with systems of the form

$$\begin{aligned} (\alpha\mathcal{I} + \mathcal{S})x^{k+1} &= (\alpha\mathcal{I} - \mathcal{G})x^{k+\frac{1}{2}} + b, \quad \text{and} \\ (\underbrace{(\alpha I_n + \frac{1}{\alpha}B^T B)}_{Q} &+ S)u^{k+1} = f_2^k - \frac{1}{\alpha}B^T g_2^k. \end{aligned}$$

The matrices of both systems have the form: "a SPD matrix + an skew-symmetric one" and therefore without any loss of generality we will consider only the second equation, we leave out the upper index k and denote the-right hand side by f_r, i.e. we mean the system

$$(Q + S)u = f_r. \tag{30}$$

Let us mention that Q is symmetric and positive definite and S is skew-symmetric matrix. The acceleration procedure for the CGW iterative method, formulated by Concus, Golub and Widlund, is recommended (see References [22], [7]). It accelerates the iterative process

$$\underbrace{u^{i+1} = \underbrace{(I - Q^{-1}(Q+S))}_{-Q^{-1}S=T}u^i + Q^{-1}b}_{u^{i+1}=Tu^i+c \quad \text{where} \quad c=Q^{-1}b.}$$

Let us shortly present following iteration formulas:

$$\begin{aligned} u^{i+1} &= \omega_{i+1}(Tu^i + c) + (1 - \omega_{i+1})u^{i-1}, \\ \omega_{i+1} &= \left(1 - \frac{(Q\delta^i)^T\delta^i}{(Q\delta^{i-1})^T\delta^{i-1}}\,\frac{1}{\omega_1}\right)^{-1} \quad \omega_1 = 1. \end{aligned}$$

where

$$\delta^i = Tu^i + c - u^i.$$

Let us remark, that the last formulas may be fast obtained by rewriting of the three-term recurrence formula for the preconditioned CG applied in symmetric case.

6. Numerical results

In this section, we compare some Krylov subspace methods with the alternating iteration method described in this paper. We show, that this method is faster for larger systems of the considered structure.

As it was shown, one of our main goals is to find an optimal parameter α^* to gain fast convergence of Algorithm 1. Therefore we have to calculate the smallest and largest eigenvalue of matrix G. We test the Lanczos method and the method described in Algorithm 2. In the second case we obtain more than one of the smallest or largest eigenvalues and this fact gives an information, which is used by the construction of a good preconditioner for the system (G).

For the solution of the symmetric part in Algorithm 1 we use preconditioned CG method. The preconditioning matrix M was constructed according to the algorithm presented in [23], where M was constructed analogously to the paper [11]. This preconditioner is very efficient in the case of matrices having very small eigenvalues. The first example will be academic.

Example. Let $A \in \mathbb{R}^{100\times 100}$ be symmetric positive definite matrix with eigenvalues 0.009887, 0.01803, 0.03207. The other ones lie in interval $[3.254, 100.7]$. Hence our matrix A has three very small eigenvalues. We use the Algorithm 2 with parameters $s = 8$, $t = 3$ and after 29 iterations we get the three smallest eigenvalues of A with precision 10^{-14} and the corresponding eigenvectors. If we solve linear system $Az = e$ using preconditioned CG method (for the construction of preconditioner were used polynomial filters). The residual norm is less then 10^{-4} after 140 iterations. Classical CG-method or CG-method preconditioned with incomplete Cholesky stagnates.

6.1. Test cases

Our concern is to solve large linear systems occurring in fluid dynamics, mostly constructed by finite element or finite volume method.

Let us consider a flow in the unit square domain Ω described by the following equations

$$\begin{aligned} -\mu \triangle u_1 + \{v_1.(u_1)_x + v_2.(u_1)_y\} + p_x &= 0 \quad \text{in } \Omega \\ -\mu \triangle u_2 + \{v_1.(u_2)_x + v_2(u_2)_y\} + p_y &= 0 \quad \text{in } \Omega \\ (u_1)_x + (u_2)_y &= 0 \quad \text{in } \Omega \\ u_1 &= 1 \quad \text{on } \Gamma_1 \\ u_1 &= 0 \quad \text{on } \Gamma \setminus \Gamma_1 \\ u_2 &= 0 \quad \text{on } \Gamma \end{aligned}$$

where Γ is the boundary of the unit square and Γ_1 his part with $y = 1$. The functions u_1 and u_2 are the velocity components in x and y directions and p is the

Name	Size		Parameters
	Nr. of unknowns	Nr. of non-zero el.	
Stokes 1	759	3764	$h = 0.1$, $\nu = 10^{-1}$
Stokes 2	3119	15924	$h = 0.05$, $\nu = 10^{-1}$
Stokes 3	7079	36484	$h = 0.033$, $\nu = 10^{-1}$
Stokes 4	1263	65444	$h = 0.025$, $\nu = 10^{-1}$
Stokes 5	19799	102804	$h = 0.02$, $\nu = 10^{-1}$
Stokes 6	28559	148564	$h = 0.0167$, $\nu = 10^{-1}$
Stokes 7	38919	202724	$h = 0.0143$, $\nu = 10^{-1}$
Stokes 8	50879	265284	$h = 0.0125$, $\nu = 10^{-1}$
Stokes 9	64439	336244	$h = 0.0111$, $\nu = 10^{-1}$
Stokes 10	79599	415604	$h = 0.01$, $\nu = 10^{-1}$

Tab. 1: *Set of test matrices.*

pressure. The resulting linear system for the discrete solution, putting $u = [u_1, u_2]^T$, has the form (1) with $C = 0$ and A symmetric and positive definite.

6.2. Estimation of optimal parameter

We want to find an optimal parameter $\alpha^* = \sqrt{\lambda_{\min}\lambda_{\max}}$ for the matrices defined in Table 1. We compare the Lanczos method with the Algorithm 2. In Table 2 the following notation is used:

- $\varepsilon_{\min} = |\lambda_{\min}^{(comp)}(G) - \lambda_{\min}(G)|$, where $\lambda_{\min}^{(comp)}$ is computed and $\lambda_{\min}$ is the exact smallest eigenvalue of G.
- $\varepsilon_{\max} = |\lambda_{\max}^{(comp)}(G) - \lambda_{\max}(G)|$, where $\lambda_{\max}^{(comp)}$ is computed and $\lambda_{\max}$ is the exact largest eigenvalue of G.
- fl. = number of floating-point operations.
- s = number of Lanczos steps.
- s_{opt} = number of steps necessary to to obtain results comparable with Algorithm 2.
- m, k = constants defined in Algorithm 2.

We show the results obtained using Laczos method for $s = 10$ and s_{opt} in the first two columns. We choose constant s, such that the computation costs are similar to Algorithm 2. The last two columns represents the Algorithm 2. In column 3 and 4 different variants of shifts are used according to the text in Section 3.

Table 2 illustrates, that for smaller problems the Lanczos method gives similar results as Algorithm 2. But when the size of the problem grows Algorithm 2 appears to be more efficient.

Matrix	Lanczos		Chebyshev	Sorensen
	$s = 10$	$s = opt$	$m = 5, k = 3$	$m = 5, k = 3$
Stokes 1	$\varepsilon_{\min} = 10^{-3}$ $\varepsilon_{\max} = 10^{-2}$ fl.= 186742	$s_{opt} = 60$ fl.= 1164243	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-4}$ fl.= 152433	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 156476
Stokes 2	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 0.8×10^6	$s_{opt} = 65$ fl.= 1.9×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 8.86×10^6	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 8.87×10^6
Stokes 3	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 1.8×10^6	$s_{opt} = 68$ fl.= 2.4×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-4}$ fl.= 8.91×10^6	$\varepsilon_{\min} = 10^{-6}$ $\varepsilon_{\max} = 10^{-6}$ fl.= 8.93×10^6
Stokes 4	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 3.2×10^6	$s_{opt} = 74$ fl.= 7.4×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 9.67×10^6	$\varepsilon_{\min} = 10^{-7}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 9.98×10^6
Stokes 5	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 5.0×10^6	$s_{opt} = 74$ fl.= 8.4×10^7	$\varepsilon_{\min} = 10^{-6}$ $\varepsilon_{\max} = 10^{-6}$ fl.= 1.02×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.02×10^7
Stokes 6	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 7.2×10^6	$s_{opt} = 76$ fl.= 1.2×10^8	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.11×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.12×10^7
Stokes 7	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 9.8×10^6	$s_{opt} = 78$ fl.= 1.6×10^8	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.32×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.35×10^7
Stokes 8	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 1.3×10^7	$s_{opt} = 76$ fl.= 2.2×10^8	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.43×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.44×10^7
Stokes 9	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 1.6×10^7	$s_{opt} = 84$ fl.= 2.6×10^8	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.56×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.57×10^7
Stokes 10	$\varepsilon_{\min} = 10^{-1}$ $\varepsilon_{\max} = 10^{-1}$ fl.= 1.9×10^7	$s_{opt} = 89$ fl.= 2.9×10^8	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.65×10^7	$\varepsilon_{\min} = 10^{-5}$ $\varepsilon_{\max} = 10^{-5}$ fl.= 1.66×10^7

Tab. 2: *Estimation of optimal parameter* α^*.

6.3. Methods comparison

In this subsection we compare GMRES, GMRES(20), MINRES and ADI methods for the solution of our test cases. In the following table we compare the floating-point operations necessary to get the solution with residual norm less than 10^{-6}. The parameter α used in algorithm 1 was computed by Algorithm 2. The symbol $*$ means, that no results has been obtained in real time.

Matrix	GMRES	GMRES(20)	MINRES	ADI
Stokes 1	1.1×10^6	0.5×10^6	0.7×10^6	2.1×10^6
Stokes 2	4.2×10^6	1.2×10^6	1.1×10^6	3.2×10^6
Stokes 3	8.4×10^6	2.1×10^6	1.9×10^6	4.4×10^6
Stokes 4	*	3.3×10^6	3.5×10^6	5.1×10^6
Stokes 5	*	6.1×10^6	5.6×10^6	7.3×10^6
Stokes 6	*	7.4×10^6	6.3×10^6	7.6×10^6
Stokes 7	*	8.2×10^6	*	8.3×10^6
Stokes 8	*	9.3×10^6	*	8.9×10^6
Stokes 9	*	1.6×10^7	*	1.2×10^6
Stokes 10	*	2.8×10^7	*	2.1×10^6

Tab. 3: *Methods comparison.*

We can see from Table 3, that the ADI method is more efficient if the size of the solved problem grows. Beginning with Stokes 7 the ADI method is better than GMRES(20).

References

[1] M. Feistauer: *Mathematical methods in fluid dynamics.* Longman Group UK Limited 1993.

[2] H.C. Elman, D.J. Silvester, A.J. Wathen: *Iterative methods for problems in computational fluid dynamics.* Winter School on Iterative Methods in Scientific Computing Applications, July 1996.

[3] H.C. Elman, D.J. Silvester, A.J. Wathen: *Performance and analysis of saddle point preconditioners for the discrete steady-state Navier-Stokes equations.* Numer. Math. **90**, 2002, 665–668.

[4] P. Knobloch, L. Tobiska: *The P_1^{mod} element: A new nonconforming finite element for convection-diffusion problems.* Otto–von–Guericke–Universitt Magdeburg, Preprint Nr. 28, 1999.

[5] N.J. Higham: *Accuracy and stability of numerical algorithms.* Siam 1996.

[6] M. Fiedler: *Speciální matice a jejich použití v numerické matematice.* SNTL 1981.

[7] M. Benzi, G. H. Golub: *An iterative method for generalized saddle point problems.* Emory University, Draft 25 October 2002.

[8] M. Eirmann, O.G. Ernst, O. Schneider: *Analysis of acceleration strategies for restarted minimal residual methods.* Journal of Computational and Applied Mathematics **123**, 2000, 261–292.

[9] O.G. Ernst: *Residual–minimizing Krylov subspace methods for stabilized discretizations of convection–diffusion equations.* Siam J. Matrix Anal. Appl. **21**, No. 4, 1079–1101.

[10] D. Loghin, A.J. Wathen: *Schur complement preconditioners for the Navier–Stokes equations.* Oxford University Numerical Analysis Group.

[11] J. Baglama, D. Calvetti, G.H. Golub, L. Reichel: *Adaptively preconditioned GMRES algorithms.* SIAM J. Sci. Comput. **10**, No.1, 243–269.

[12] Zhong-Zhi Bai, G.H. Golub, M.K. NG: *Hermitian and Skew-Hermitian splitting methods for non-Hermitian positive definite linear systems.* SIAM J. Matrix Anal. Appl. **24**, No. 3, 603–626.

[13] R.S. Varga: *Matrix iterative analysis.* Prentice-Hall. Englewood Cliffs, NJ, 1962.

[14] M. Arioli, D. Ruiz: *A Chebyshev-based two-stage iterative method as an alternative to the direct solution of linear systems.* Atlas Centre, Rutherford Appleton Laboratory, Technical Report, Nr. RAL-TR-2002-021.

[15] L. Giraud, D. Ruiz, A. Touhami: *A comparative study of iterative solvers exploiting spectral information for SPD systems.* CERFACS, Technical Report, Nr. TR/PA/04/40.

[16] M. Benzi, J.M. Gander, G.H. Golub: *Optimization of the Hermitian and Skew-Hermitian splitting iteration for saddle-point problems.* BIT **43**, 2002, No. 1, 001–013.

[17] Saad Y.: *Iterative methods for sparse linear systems.* PWS Publishing Company, 1996.

[18] W. Zulenher: *Analysis of iterative methods for saddle point problems: a unified approach.* Math. Comp. **71**, 2001, 479–505.

[19] I. Perugia, V. Simoncini: *Block-diagonal and indefinite symmetric preconditioners for mixed finite element formulations.* Numer. Linear Algebra Appl. **7**, 2000, 585–616.

[20] L. Lukan: *Indefinitely preconditioned inexact Newton method for large sparse equality constrained non-linear programming problems.* Numer. Linear Algebra Appl. **5**, 1998, 219–247.

[21] M. Huhtanen: *A Hermitian Lanczos method for normal matrices.* Siam J. Matrix Anal. Appl. **23**, 2002, 1092–1108.

[22] L.A. Hageman, D.M. Young: *Applied iterative methods.* Academic Press 1981.

[23] G.H. Golub, C.F. Van Loan: *Matrix computation*. The Johns Hopkins University Press 1985.

[24] D.C. Sorensen: *Implicit application of polynomial filters in a k-step Arnoldi method.* SIAM J. on Matrix Anal. and Appl. **13**, 1992, Issue 1, 357–385.

[25] T.A. Manteuffel: *The Tchebychev iteration for nonsymmetric linear systems.* Numer. Math. **28**, 1977, 307–327.