LOW RANK ITERATIVE METHODS FOR PROJECTED GENERALIZED LYAPUNOV EQUATIONS *

TATJANA STYKEL†

Abstract. We generalize an alternating direction implicit method and the Smith method for large-scale projected generalized Lyapunov equations. Such equations arise in model reduction for descriptor systems. Low rank versions of these methods are also presented, that can be used to compute low rank approximations to the solution of projected generalized Lyapunov equations with low rank symmetric, positive semidefinite right-hand side. Numerical examples are presented.

Key words. projected generalized Lyapunov equations, alternating direction implicit method, Smith method, low rank approximation

AMS subject classifications. 65F10, 65F30, 15A22, 15A24, 93C05

1. Introduction.

Consider a linear time-invariant descriptor system

$$\begin{aligned} E\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{1.1}$$

where E, $A \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, $C \in \mathbb{R}^{p,n}$, $x(t) \in \mathbb{R}^n$ is the state vector, $u(t) \in \mathbb{R}^m$ is the control input and $y(t) \in \mathbb{R}^p$ is the output. The matrix E may be singular, but the pencil $\lambda E - A$ is assumed to be *regular*, i.e., $\det(\lambda E - A) \not\equiv 0$. Descriptor systems arise in many different applications including electrical circuit simulation, multibody dynamics and spatial discretization of partial differential equations, e.g. [7, 8, 9]. Stability analysis and some control problems for (1.1) are strongly related to the projected generalized continuous-time algebraic Lyapunov equations (GCALEs)

$$EXA^T + AXE^T = -P_l BB^T P_l^T, \qquad P_r X P_r^T = X, \tag{1.2}$$

$$E^T XA + A^T XE = -P_r^T C^T C P_r, \qquad P_l^T X P_l = X \tag{1.3}$$

and the projected generalized discrete-time algebraic Lyapunov equations (GDALEs)

$$AXA^T - EXE^T = (I - P_l)BB^T(I - P_l)^T, \quad P_r X P_r^T = 0, \tag{1.4}$$

$$A^T XA - E^T XE = (I - P_r)^T C^T C(I - P_r), \quad P_l^T X P_l = 0, \tag{1.5}$$

where P_l and P_r are the spectral projections onto the left and right deflating subspaces corresponding to the finite eigenvalues of the pencil $\lambda E - A$. It has been shown in [34] that if the pencil $\lambda E - A$ is *stable*, i.e., all its finite eigenvalues have negative real part, then the projected GCALEs (1.2) and (1.3) have the unique symmetric, positive semidefinite solutions which define the proper controllability and observability Gramians of the descriptor system (1.1). Furthermore, the projected GDALEs (1.4) and (1.5) have the unique symmetric, positive semidefinite solutions which are the improper controllability and observability Gramians of (1.1). The Gramians play a central role in analysis and control design problems for descriptor systems, such as the characterization of controllability and observability properties, computing the $\mathbb{H}_2$

*This work was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

†Institut für Mathematik, MA 3-3, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany (stykel@math.tu-berlin.de).

or Hankel norm, the minimal and balanced realizations as well as balanced truncation model order reduction [5, 34, 35, 36].

The numerical solution of standard Lyapunov equations with $E = I$ has been the topic of numerous publications [3, 4, 12, 13, 14, 15, 19, 20, 24, 26, 30]. A variety of direct and iterative methods has been proposed there for computing the solutions of such equations, their Cholesky factors or low rank approximations. The case of nonsingular E has been considered in [6, 10, 16, 17, 23]. Until now, only direct methods have been extended to projected Lyapunov equations [33]. The solutions of (1.2) – (1.5) can be computed by the generalized Schur–Bartels–Stewart or the generalized Schur–Hammarling methods that are based on the preliminary reduction of the pencil $\lambda E - A$ to the generalized Schur form, solution of the generalized Sylvester and Lyapunov equations and back transformation. Since these methods cost $O(n^3)$ operations and require $O(n^2)$ memory location, they can be used only for problems of small or medium size.

Due to the practical importance of the numerical solution of large-scale projected generalized Lyapunov equations that occur in balanced truncation model reduction for descriptor systems [36], the development of iterative methods for such equations is a challenging problem. In this paper we generalize an alternating direction implicit (ADI) method [19, 20, 24] and a Smith method [24, 30] to the projected generalized Lyapunov equations (1.2) and (1.4) with large sparse matrix coefficients. The dual equations (1.3) and (1.5) can be solved in a similar way. Low rank versions of the ADI and Smith methods are also presented that can be used to compute low rank approximations to the solutions of (1.2) and (1.4) with a low rank right-hand side. Such a problem arises, e.g., in model reduction. Note that the number m of columns of the matrix B in (1.2) and (1.4) relates to the number of inputs of the underlying descriptor system (1.1) and is usually very small compared to order n of the problem.

A major difficulty in the numerical solution of projected Lyapunov equations is that we need to compute the spectral projections P_l and P_r onto the left and right deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite eigenvalues. Fortunately, in many applications such as computational fluid dynamics, electrical circuit simulation and constrained structural mechanics, the matrices E and A have some special block structure. As the following examples show, this structure can be used to construct the projections P_l and P_r in explicit form.

Example 1.1. Consider the descriptor system (1.1), where E and A have the form

$$E = \begin{bmatrix} E_{11} & 0 \\ 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix}. \tag{1.6}$$

Such systems arise, for instance, in spatial discretization of the instationary incompressible Stokes equation [7, 8], and the convection equation [21]. If E_{11} is nonsingular and the matrices A_{12} and A_{21} have full rank, then the pencil $\lambda E - A$ as in (1.6) is of index 2, see, e.g., [8] for the definition of index. In this case the spectral projections P_l and P_r have the following form

$$P_l = \begin{bmatrix} \Pi_l & -\Pi_l A_{11} E_{11}^{-1} A_{12} (A_{21} E_{11}^{-1} A_{12})^{-1} \\ 0 & 0 \end{bmatrix},$$
$$P_r = \begin{bmatrix} \Pi_r & 0 \\ -(A_{21} E_{11}^{-1} A_{12})^{-1} A_{21} E_{11}^{-1} A_{11} \Pi_r & 0 \end{bmatrix},$$

where $\Pi_l = I - A_{12}(A_{21}E_{11}^{-1}A_{12})^{-1}A_{21}E_{11}^{-1}$ is a projection onto the nullspace of $A_{21}E_{11}^{-1}$ along the image of A_{12} and $\Pi_r = I - E_{11}^{-1}A_{12}(A_{21}E_{11}^{-1}A_{12})^{-1}A_{21} = E_{11}^{-1}\Pi_l E_{11}$.

Example 1.2. The motion of multibody systems with holonomic constraints can be described by nonlinear differential-algebraic equations of the first order [9, 29]. Linearization of these equations around an equilibrium state leads to the descriptor system (1.1) with

$$E = \begin{bmatrix} I & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} 0 & I & 0 \\ K & D & -G^T \\ G & 0 & 0 \end{bmatrix}. \tag{1.7}$$

Here M is the symmetric, positive definite mass matrix, K is the stiffness matrix, D is the damping matrix and G is the matrix of constraints. If the matrix G is of full row rank, then the pencil $\lambda E - A$ is of index 3 and the spectral projections P_l and P_r can be computed as

$$P_l = \begin{bmatrix} \Pi & 0 & -\Pi M^{-1} D G_1 \\ -\Pi^T D(I-\Pi) & \Pi^T & -\Pi^T(K + D\Pi M^{-1} D) G_1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$P_r = \begin{bmatrix} \Pi & 0 & 0 \\ -\Pi M^{-1} D(I-\Pi) & \Pi & 0 \\ G_1^T(K\Pi - D\Pi M^{-1} D(I-\Pi)) & G_1^T D\Pi & 0 \end{bmatrix},$$

where $G_1 = M^{-1}G^T(GM^{-1}G^T)^{-1}$ and $\Pi = I - M^{-1}G^T(GM^{-1}G^T)^{-1}G = I - G_1 G$ is a projection onto the nullspace of G along the image of $M^{-1}G^T$, see [29] for details.

In the following we will assume that the projections P_l and P_r are given. Clearly, we do not compute these projections explicitly. Instead, we use matrix-vector multiplication and linear system solvers. In Section 2 we present a generalization of the ADI method and its low rank version for the projected GCALE (1.2). In Section 3 we discuss the numerical solution of the projected Lyapunov equations (1.2) and (1.4) via the (cyclic) Smith method. Section 4 contains some results of numerical experiments.

Throughout the paper the complex plane is denoted by $\mathbb{C}$ and the open left half-plane is denoted by $\mathbb{C}^-$. We will denote by $\mathbb{R}^{n,m}$ and $\mathbb{C}^{n,m}$ the spaces of $n \times m$ real and complex matrices, respectively. The real part of a complex number z is denoted by $\mathrm{Re}(z)$. The matrix A^T stands for the transpose of $A \in \mathbb{R}^{n,m}$, A^* denotes the complex conjugate and transpose of $A \in \mathbb{C}^{n,m}$, and $A^{-T} = (A^{-1})^T$. An identity matrix of order n is denoted by I_n or simply I.

2. Alternating direction implicit method. The ADI method was originally proposed for linear systems [22] and then used in [18, 19, 20, 24, 38] to solve standard continuous-time Lyapunov equations. The case of nonsingular E has been considered in [17]. In this section we present a generalization of the ADI method for the projected GCALE (1.2).

For any parameter $\tau \in \mathbb{C}$, the first equation in (1.2) can be rewritten as

$$(E + \tau A) X A^T + A X (E - \tau A)^T = -P_l B B^T P_l^T. \tag{2.1}$$

Then the generalized ADI iteration for the projected GCALE (1.2) is given by

$$\begin{aligned} (E + \tau_k A) X_{k-1/2} A^T &= -P_l B B^T P_l^T - A X_{k-1} (E - \tau_k A)^T, \\ (E + \overline{\tau}_k A) X_k^T A^T &= -P_l B B^T P_l^T - A X_{k-1/2}^T (E - \overline{\tau}_k A)^T \end{aligned} \tag{2.2}$$

with an initial matrix $X_0 = 0$ and the shift parameters $\tau_1, \ldots, \tau_k \in \mathbb{C}^-$. If the pencil $\lambda E - A$ is stable, then the matrices A and $E + \tau_k A$ are nonsingular. In this case the

matrices $(E+\tau_k A)^{-1}A$ and $A^{-1}(E-\overline{\tau}_k A)$ commute. From (2.2) we find that

$$\begin{aligned} X_k &= (E+\tau_k A)^{-1}(E-\overline{\tau}_k A)X_{k-1}(E-\tau_k A)^T(E+\overline{\tau}_k A)^{-T} \\ &\quad -2\mathrm{Re}(\tau_k)(E+\tau_k A)^{-1}P_l BB^T P_l^T(E+\overline{\tau}_k A)^{-T}. \end{aligned} \tag{2.3}$$

Moreover, it follows from

$$P_r(E+\tau_k A)^{-1} = (E+\tau_k A)^{-1}P_l, \qquad P_l(E-\overline{\tau}_k A) = (E-\overline{\tau}_k A)P_r$$

that $X_k = P_r X_k P_r^T$, i.e., the second equation in (1.2) is satisfied exactly.

Let X be a solution of the projected GCALE (1.2). Then the error matrices $X - X_k$ can be computed from (2.1) and (2.3) recursively as $X - X_k = \mathcal{A}_k X \mathcal{A}_k^*$, where

$$\mathcal{A}_k = P_r(E+\tau_k A)^{-1}(E-\overline{\tau}_k A)\cdot\ldots\cdot(E+\tau_1 A)^{-1}(E-\overline{\tau}_1 A). \tag{2.4}$$

If the pencil $\lambda E - A$ is stable, then all the eigenvalues of $\mathcal{A}_k$ lie inside the unite circle, and, hence, X_k converges to the solution of the projected GCALE (1.2). The rate of convergence is determined by the spectral radius of the matrix $\mathcal{A}_k$ which depends strongly on the choice of the shift parameters. The minimization of this spectral radius with respect to the parameters $\tau_1, \ldots, \tau_k$ leads to the generalized ADI minimax problem

$$\{\tau_1, \ldots, \tau_k\} = \operatorname*{arg\,min}_{\{\tau_1,\ldots,\tau_k\}\in\mathbb{C}^-} \max_{t\in\mathrm{Sp}_f(E,A)} \frac{|(1-\overline{\tau}_1 t)\cdot\ldots\cdot(1-\overline{\tau}_k t)|}{|(1+\tau_1 t)\cdot\ldots\cdot(1+\tau_k t)|}, \tag{2.5}$$

where $\mathrm{Sp}_f(E,A)$ denotes the set of finite eigenvalues of the pencil $\lambda E - A$. The computation of the optimal shift parameters is a difficult problem, since the finite eigenvalues of the pencil $\lambda E - A$ (in particular, if it is large and sparse) are in general unknown and expensive to compute. This problem is solved for standard equations with $E = I$ and symmetric A, e.g., [39], while the case of complex eigenvalues is still not completely understood, see [20, 32, 39] for some contributions. To compute the suboptimal ADI shift parameters for the standard problem a heuristic algorithm has been proposed in [24]. This algorithm is based on Arnoldi iterations [27] applied to the matrices A and A^{-1}. It can also be extended to the generalized problem (2.5). Due to the nonsingularity of A this problem is rewritten as

$$\{\tau_1, \ldots, \tau_k\} = \operatorname*{arg\,min}_{\{\tau_1,\ldots,\tau_k\}\in\mathbb{C}^-} \max_{t\in\mathrm{Sp}(EA^{-1})\setminus\{0\}} \frac{|(t-\overline{\tau}_1)\cdot\ldots\cdot(t-\overline{\tau}_k)|}{|(t+\tau_1)\cdot\ldots\cdot(t+\tau_k)|}, \tag{2.6}$$

where $\mathrm{Sp}\,(EA^{-1})$ denotes the spectrum of the matrix EA^{-1}. Thus, the suboptimal ADI shift parameters $\tau_1, \ldots, \tau_k$ can be determined by the heuristic procedure [24, Algorithm 5.1] from a set of largest and smallest (in modulus) nonzero approximate eigenvalues of EA^{-1} computed by an Arnoldi process.

The computational cost of the generalized ADI method is, in general, $O(n^3)$ and the memory complexity is $O(n^2)$. In many cases the storage requirement rather than the computational cost is a limiting factor for feasibility of numerical methods for large-scale problems.

2.1. Low rank version of the generalized ADI method. Recently, an efficient modification of the ADI method has been proposed to compute low rank approximations to the solutions of standard Lyapunov equations with large-scale matrix

coefficients [18, 19, 24]. This is the low rank alternating direction implicit (LR-ADI) method. It was observed that the eigenvalues of the symmetric solutions of Lyapunov equations with low rank right-hand side generally decay very rapidly, and such solutions may be well approximated by low rank matrices [2, 11, 25, 31]. A similar result holds for projected generalized Lyapunov equations. In other words, it is possible to find a matrix Z with a small number of columns such that ZZ^T is an approximate solution of the projected GCALE (1.2). The matrix Z is referred to as the *low rank Cholesky factor* of the solution X of (1.2).

A low rank version of the generalized ADI iteration (2.2) can be derived analogously to the standard case [18, 19, 24]. First of all note that the matrix X_k in (2.3) is Hermitian, positive semidefinite and the Cholesky factor Z_k of $X_k = Z_k Z_k^*$ has the form

$$\begin{aligned} Z_k &= [\,\sqrt{-2\mathrm{Re}(\tau_k)}(E+\tau_k A)^{-1}P_l B, \quad (E+\tau_k A)^{-1}(E-\overline{\tau}_k A)Z_{k-1}\,] \\ &= [\,\alpha_k S_k P_l B, \quad \alpha_{k-1} S_k R_k S_{k-1} P_l B, \quad \ldots, \quad \alpha_1 S_k R_k \cdots R_2 S_1 P_l B\,], \end{aligned}$$

where $\alpha_j = \sqrt{-2\mathrm{Re}(\tau_j)}$, $S_j = (E+\tau_j A)^{-1}$ and $R_j = E - \overline{\tau}_j A$. Taking into account that

$$S_k A S_j = S_j A S_k, \qquad R_k A^{-1} R_j = R_j A^{-1} R_k, \qquad S_k R_j = A^{-1} R_j S_k A$$

for all k, $j = 1, 2, \ldots$, the matrix Z_k can be rewritten as

$$Z_k = [\,B_0, F_{k-1}B_0, \;\; F_{k-2}F_{k-1}B_0, \;\; \ldots, \;\; F_1 F_2 \cdots F_{k-1} B_0\,],$$

where $B_0 = \sqrt{-2\mathrm{Re}(\tau_k)}(E+\tau_k A)^{-1}P_l B = \sqrt{-2\mathrm{Re}(\tau_k)}P_r(E+\tau_k A)^{-1}B$ and

$$F_j = \frac{\sqrt{-\mathrm{Re}(\tau_j)}}{\sqrt{-\mathrm{Re}(\tau_{j+1})}} S_j R_{j+1} = \frac{\sqrt{-\mathrm{Re}(\tau_j)}}{\sqrt{-\mathrm{Re}(\tau_{j+1})}} \left(I - (\tau_j + \overline{\tau}_{j+1})(E+\tau_j A)^{-1}A\right).$$

If we reenumerate the shift parameters in reverse order, then we obtain the following algorithm to compute the low rank Cholesky factor of the solution of (1.2).

ALGORITHM 2.1. *The generalized LR-ADI method for the projected GCALE.*

INPUT: E, A, $P_l \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, *shift parameters* $\tau_1, \ldots, \tau_{k_{\max}} \in \mathbb{C}^-$.
OUTPUT: *A low rank Cholesky factor* Z_k *of the solution* $X \approx Z_k Z_k^*$ *of* (1.2).
1. $Z^{(1)} = \sqrt{-2\mathrm{Re}(\tau_1)}\,(E+\tau_1 A)^{-1}P_l B, \quad Z_1 = Z^{(1)}$;
2. FOR $k = 2, 3, \ldots$

   a. $Z^{(k)} = \dfrac{\sqrt{-\mathrm{Re}(\tau_k)}}{\sqrt{-\mathrm{Re}(\tau_{k-1})}} \left(I - (\overline{\tau}_{k-1} + \tau_k)(E+\tau_k A)^{-1}A\right) Z^{(k-1)}, \qquad (2.7)$

   b. $Z_k = [\,Z_{k-1}, \;\; Z^{(k)}\,]$.

   END FOR

Note that if at least one of the shift parameters is complex, then the low rank Cholesky factors Z_k may be complex although the solution X of (1.2) is real. As in the standard case [19, 24], this problem can be avoided if we take the complex

shift parameters in complex conjugate pairs $\{\tau_j, \tau_{j+1}\}$ with $\tau_{j+1} = \overline{\tau}_j$ and compute the iterate $Z_k = [\, Z_{k-1},\, Z^{(k)} \,]$ only if τ_k is real. If τ_k is complex and $\tau_{k+1} = \overline{\tau}_k$, then we do not form Z_k, but compute $Z^{(k+1)} = (I - 2\tau_{k+1}(E + \tau_{k+1}A)^{-1}A)Z^{(k)}$ and $Z_{k+1} = [\, Z_{k-1},\, Z_1^{(k)},\, Z_2^{(k)} \,]$, where

$$Z_1^{(k)} = \sqrt{2}\,|\tau_k|(E + \tau_{k+1}A)^{-1}AZ^{(k)}, \qquad Z_2^{(k)} = \sqrt{2}\,(E + \tau_{k+1}A)^{-1}EZ^{(k)}$$

and $Z^{(k)}$ is as in (2.7). In this case one can show that all the matrices Z_k are real. Therefore, in the following we will assume without loss of generality that the computed low rank Cholesky factors Z_k are real.

The ADI iteration can be stopped as soon as a *normalized residual norm* given by

$$\eta(Z_k) = \frac{\|EZ_kZ_k^TA^T + AZ_kZ_k^TE^T + P_lBB^TP_l^T\|_F}{\|P_lBB^TP_l^T\|_F} \tag{2.8}$$

satisfies the condition $\eta(Z_k) \le tol$ with a user-defined tolerance tol or a stagnation of normalized residual norms is observed. If the number of shift parameters is smaller than the number of iterations required to attain a prescribed tolerance, then we reuse these parameters in a cyclic manner. Note that computing the normalized residuals $\eta(Z_k)$ even via the efficient method proposed in [24] can still be quite expensive for large-scale problems. It should also be noted that for ill-conditioned problems, the small residual norm does not imply that the error in the computed solution is also small, see [33].

If the iterates $X_k = Z_kZ_k^T$ converge to the solution of (1.2), then

$$\lim_{k\to\infty} Z^{(k)}(Z^{(k)})^T = \lim_{k\to\infty} (X_k - X_{k-1}) = 0.$$

Therefore, as the stopping criterion in Algorithm 2.1 one can also use the condition $\|Z^{(k)}\| \le tol$ or $\|Z^{(k)}\|/\|Z_k\| \le tol$ with some matrix norm $\|\cdot\|$.

Note that the matrices $(E+\tau_kA)^{-1}$ in Algorithm 2.1 do not have to be computed explicitly. Instead, we solve linear systems of the form $(E + \tau_kA)x = P_lb$ either by computing (sparse) LU factorizations and forward/backward substitutions or by using iterative Krylov subspace methods [28]. In the latter case the generalized LR-ADI method has the memory complexity $O(k_{ADI}mn)$ and costs $O(k_{\rm ls}k_{ADI}mn)$ flops, where k_{ADI} is the number of outer ADI iterations and $k_{\rm ls}$ is the number of inner linear solver iterations. This method becomes efficient for large-scale sparse Lyapunov equations only if $k_{\rm ls}k_{ADI}m$ is much smaller than n.

Remark 2.1. In exact arithmetic the matrices Z_k satisfy $Z_k = P_rZ_k$ and, hence, the second equation in (1.2) is fulfilled for the low rank approximation $Z_kZ_k^T$. However, in finite precision arithmetic a drift-off effect may occur. In this case we need to project $Z^{(k)}$ onto the image of P_r by pre-multiplication with P_r. In order to limit the additional computation cost we can do this, for example, at every second or third iteration step.

Observe that if E is nonsingular, but ill-conditioned with respect to inversion, then Algorithm 2.1 may provide a better result than the classical LR-ADI method [17, 19, 24] applied to the matrices $E^{-1}A$ and $E^{-1}B$.

3. Smith method. For any parameter $\tau \in \mathbb{C}^-$, the projected GCALE (1.2) is equivalent to the projected discrete-time algebraic Lyapunov equation (DALE)

$$\mathcal{A}X\mathcal{A}^* - X = -P_r\,\mathcal{B}\mathcal{B}^*P_r^T, \qquad P_rXP_r^T = X, \tag{3.1}$$

where $\mathcal{A} = (E + \tau A)^{-1}(E - \overline{\tau}A)$ and $\mathcal{B} = \sqrt{-2\mathrm{Re}(\tau)}\,(E + \tau A)^{-1}B$. Note that if the pencil $\lambda E - A$ is stable, then P_r is the spectral projection onto the invariant subspace of the matrix $\mathcal{A}$ corresponding to the eigenvalues inside the unit circle. In this case the *Smith iteration*

$$X_0 = P_r\,\mathcal{B}\mathcal{B}^* P_r^T, \qquad X_k = P_r\,\mathcal{B}\mathcal{B}^* P_r^T + \mathcal{A} X_{k-1}\mathcal{A}^*$$

converges linearly to the solution X of (3.1), see [30]. The quadratic convergence can be achieved by using the *squared Smith method* [30] based on the iteration

$$X_0 = P_r\,\mathcal{B}\mathcal{B}^* P_r^T, \qquad X_k = \mathcal{A}^{2^{k-1}} X_{k-1} \left(\mathcal{A}^{2^{k-1}}\right)^*.$$

The number of iterations required for a desired accuracy in the approximate solution X_k of the projected DALE (3.1) depends on the parameter τ. Note that the Smith method is, in fact, the generalized ADI iteration with a single parameter. Therefore, an optimal value $\tau = \tau_1 = \ldots = \tau_k$ from (2.5) can be used to increase the convergence.

The Smith method costs $O(n^3)$ flops and has the memory complexity $O(n^2)$, since the solution X is computed explicitly and it is dense even if the matrices E and A are sparse.

A modification of the Smith method has been proposed in [1, 24] to compute a low rank Cholesky factor of the solution of standard Lyapunov equations with a low rank right-hand side. This version of the Smith method is based on the LR-ADI iteration with ℓ shift parameters applied in a cyclic manner and referred to as the *low rank cyclic Smith* (LR-Smith(ℓ)) *method*. It consists of two stages: first one computes the ℓ-th iterate of the LR-ADI method with the shift parameters $\tau_1, \ldots, \tau_\ell$ and then solve the discrete-time Lyapunov equation

$$\mathcal{A}_\ell X \mathcal{A}_\ell^* - X = -Z_\ell Z_\ell^*,$$

where $\mathcal{A}_\ell$ is as in (2.4) with $k = \ell$. We have the following algorithm to compute the low rank Cholesky factor of the solution of the projected GCALE (1.2).

ALGORITHM 3.1. *The generalized LR-Smith(ℓ) method for the projected GCALE.*

INPUT: *E, A, $P_l, P_r \in \mathbb{R}^{n,n}$, $B \in \mathbb{R}^{n,m}$, shift parameters $\tau_1, \ldots, \tau_\ell \in \mathbb{C}^-$.*
OUTPUT: *A low rank Cholesky factor $Z_{k\ell}$ of the solution $X \approx Z_{k\ell} Z_{k\ell}^*$ of* (1.2).
1. *Compute Z_ℓ using Algorithm* 2.1 *and $Z^{(\ell)} = Z_\ell$;*
2. FOR $k = 2, 3, \ldots$
   a. $Z^{(k\ell)} = \mathcal{A}_\ell Z^{((k-1)\ell)}$ *with* $\mathcal{A}_\ell = P_r \prod_{k=1}^{\ell} (E + \tau_k A)^{-1}(E - \overline{\tau}_k A)$, (3.2)
   b. $Z_{k\ell} = [\, Z_{(k-1)\ell},\ Z^{(k\ell)} \,]$. (3.3)

   END FOR

It follows from (2.3), (3.2) and (3.3) that the generalized LR-Smith(l) method is mathematically equivalent to the generalized LR-ADI iteration with the cyclically repeated shift parameters $\tau_1, \ldots, \tau_\ell$. If Z_ℓ is real and the complex shift parameters appear in complex conjugate pairs, then the matrices $Z_{k\ell}$ are also real.

Remark 3.1. Note that at every iteration step in Algorithms 2.1 and 3.1 the number of columns of the approximate solution factors Z_k and $Z_{k\ell}$ is increased by m and ℓm, respectively. To keep the low rank structure in the Cholesky factors in case of large m and slow convergence, we should replace the iterate by its low rank approximation computed via the updated singular value decomposition, see [1] for details.

Consider now the projected GDALE (1.4). If the pencil $\lambda E - A$ is stable, then the matrix A is nonsingular and equation (1.4) is equivalent to the projected DALE

$$X - \mathcal{E}X\mathcal{E}^T = (I - P_r)\hat{\mathcal{B}}\hat{\mathcal{B}}^T(I - P_r)^T, \qquad P_r X P_r^T = 0, \tag{3.4}$$

where $\mathcal{E} = A^{-1}E$ and $\hat{\mathcal{B}} = A^{-1}(I - P_l)B = (I - P_r)A^{-1}B$. Note that $I - P_r$ is the spectral projection onto the invariant subspace of the matrix $\mathcal{E}$ corresponding to the zero eigenvalues. In this case $(I - P_r)\mathcal{E} = \mathcal{E}(I - P_r)$ is nilpotent with the index of nilpotency ν that is equal to the index of the pencil $\lambda E - A$. The unique solution of the projected DALE (3.4) is given by

$$X = \sum_{k=0}^{\infty} \mathcal{E}^k \hat{\mathcal{B}}\hat{\mathcal{B}}^T(\mathcal{E}^T)^k = \sum_{k=0}^{\nu-1}((I - P_r)\mathcal{E})^k \hat{\mathcal{B}}\hat{\mathcal{B}}^T(\mathcal{E}^T(I - P_r)^T)^k.$$

Thus, the Cholesky factor Y of the solution $X = YY^T$ of (3.4) and also of the projected GDALE (1.4) has the form

$$Y = [\,\hat{\mathcal{B}},\ \mathcal{E}\hat{\mathcal{B}},\ \ldots,\ \mathcal{E}^{\nu-1}\hat{\mathcal{B}}\,].$$

It can be computed by the following algorithm that is a generalization of the Smith method for the projected GDALE (1.4).

ALGORITHM 3.2. *The generalized Smith method for the projected GDALE.*

INPUT: *E, A, $P_r \in \mathbb{R}^{n,n}$ and $B \in \mathbb{R}^{n,m}$.*
OUTPUT: *A Cholesky factor Y_ν of the solution $X = Y_\nu Y_\nu^T$ of* (1.4).
1. $Y^{(1)} = (I - P_r)A^{-1}B, \quad Y_1 = Y^{(1)}$;
2. FOR $k = 2, 3, \ldots, \nu$
   $Y^{(k)} = A^{-1}EY^{(k-1)}, \quad Y_k = [\,Y_{k-1},\ Y^{(k)}\,]$.
   END FOR

Note that if the index ν of the pencil $\lambda E - A$ is unknown, then the iteration in Algorithm 3.2 can be stopped as soon as $\|Y^{(k)}\| \le tol$ or $\|Y^{(k)}\|/\|Y_k\| \le tol$ with some matrix norm $\|\cdot\|$ and a tolerance *tol*. If we want to compute the solution of (1.4) as accurate as possible, we should set *tol* to the machine precision. To avoid the drift-off of the columns of $Y^{(k)}$ from the image of $I - P_r$, the matrix $Y^{(k)}$ should be pre-multiplied with $I - P_r$ after some iteration steps.

It has been shown in [40] that the solution of the projected GDALE (1.4) also satisfies the projected DALE

$$X - FEXEF^T = FBB^TF^T, \qquad P_r X P_r^T = 0 \tag{3.5}$$

with $F = (EP_r - A(I - P_r))^{-1}P_r = P_l(P_lE - (I - P_l)A)^{-1}$. Therefore, the Cholesky factor Y_ν of the solution $X = Y_\nu Y_\nu^T$ of (1.4) can also be computed by applying the

generalized Smith method to equation (3.5). Similar to the projections P_r and P_l, the matrix F can be constructed in explicit form using the special structure of the matrix coefficients E and A.

4. Numerical examples. In this section we present some results of numerical experiments. Computations were done on IBM RS 6000 44P Modell 270 with machine precision $\varepsilon \approx 2.22 \times 10^{-16}$ using MATLAB 6.5.

Example 4.1. Consider the 2D instationary Stokes equation that describes the flow of an incompressible fluid in a domain. The spatial discretization of this equation by the finite difference method on a uniform staggered grid leads to the descriptor system (1.1) with the matrices E and A as in (1.6), where $E_{11} = I$, $A_{11} = A_{11}^T$ and $A_{21} = A_{12}^T$. In our experiments the order of the problem is $n = 7700$ and the matrix $B \in \mathbb{R}^{n,1}$ is chosen at random. Note that E and A are symmetric and E is positive semidefinite. In this case the finite eigenvalues of $\lambda E - A$ are real and we use only real ADI shift parameters.

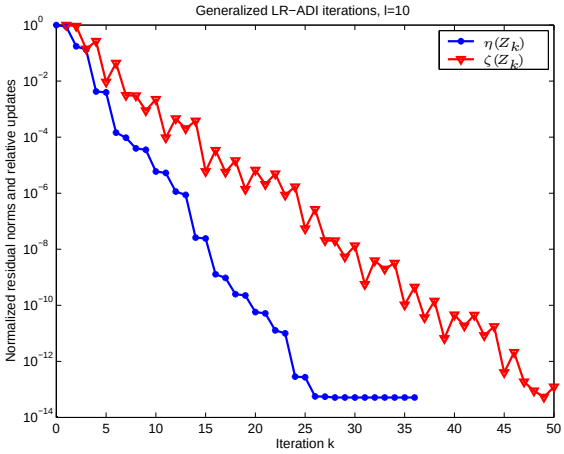

FIG. 4.1. *Example 4.1: convergence history for the generalized LR-ADI method.*

In Fig. 4.1 we present the normalized residual norm $\eta(Z_k)$ as in (2.8) and the ratio $\zeta(Z_k) = \|Z^{(k)}\|_F / \|Z_k\|_F$ for the generalized LR-ADI method with $\ell = 10$ real ADI shift parameters. One can see that the generalized LR-ADI method converges fast and the solution of the projected GCALE (1.2) of order 7700 can be approximated quite accurately by a matrix of rank 30. The normalized residual norm $\eta(Z_k)$ stagnates on a relatively small level, which is caused by round-off errors. Note that $\zeta(Z_k)$ does not decrease monotonically and more iteration steps are required to achieve $\zeta(Z_k) \leq tol$ than $\eta(Z_k) \leq tol$.

Figure 4.2 shows the convergence history of the normalized residual norm for the generalized LR-ADI method and the generalized LR-Smith(10) method versus the number of columns of Z_k. One can see that both methods give similar results.

Furthermore, we computed the full rank Cholesky factor $Y_2 \in \mathbb{R}^{n,2}$ of the solution $X = Y_2 Y_2^T$ of the projected GDALE (1.4) using Algorithm 3.2. The Frobenius norms of the update matrices are $\|Y^{(1)}\|_F = 1.13824 \times 10^5$, $\|Y^{(2)}\|_F = 60.6731$ and $\|Y^{(3)}\|_F \leq \varepsilon$. This is not surprising because the pencil $\lambda E - A$ is of index 2.

Example 4.2. Consider the descriptor system (1.1) arising from the finite element discretization of a convection problem with a boundary control, see [21] for details.

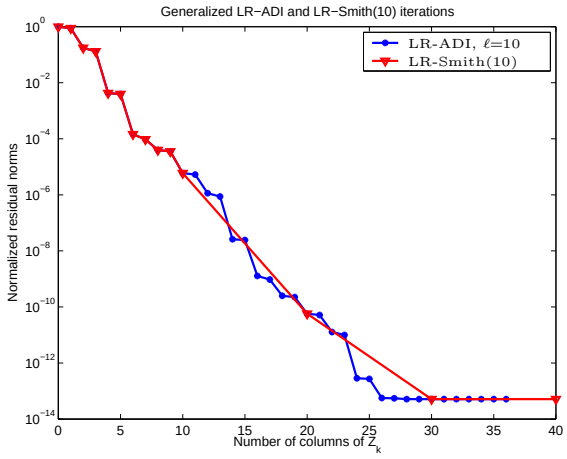

FIG. 4.2. *Example 4.1: normalized residual norms for the generalized LR-ADI and LR-Smith(10) methods.*

The matrices E and A as in (1.6) are non-symmetric. The problem is of order $n = 2909$ and the matrix $B = [\,0,\, B_2^T\,]^T \in \mathbb{R}^{n,1}$ results from the boundary control.

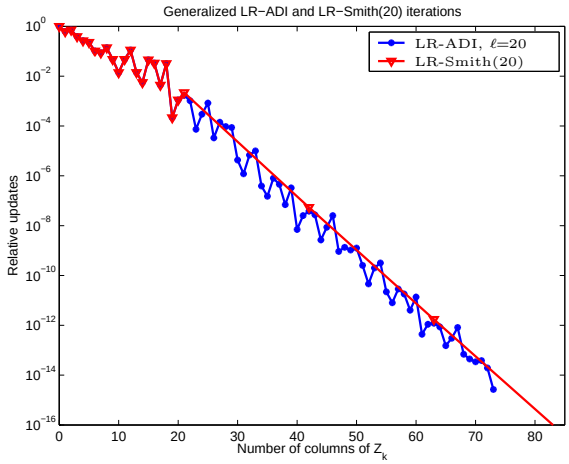

FIG. 4.3. *Example 4.2: relative updates for the generalized LR-ADI and LR-Smith(20) methods.*

In Fig. 4.3 we present the convergence history in terms of the relative updates $\zeta(Z_k) = \|Z^{(k)}\|_F / \|Z_k\|_F$ for the generalized LR-ADI method and the generalized LR-Smith(20) method. One can see that the solution of the projected GCALE (1.2) can be approximated by a matrix of rank 70.

The solution of the the projected GDALE (1.4) has been computed in factored form $X = Y_2 Y_2^T$ with the full rank Cholesky factor $Y_2 \in \mathbb{R}^{n,2}$. The Frobenius norms of the update matrices are $\|Y^{(1)}\|_F = 2.397$, $\|Y^{(2)}\|_F = 1.357 \times 10^{-3}$ and $\|Y^{(3)}\|_F \le \varepsilon$.

Example 4.3. Consider a damped mass-spring system with g masses, see [37, Section 3.9]. The ith mass is connected to the $(i+1)$st mass by a spring and a damper and also to the ground by another spring and damper. Additionally, we assume that

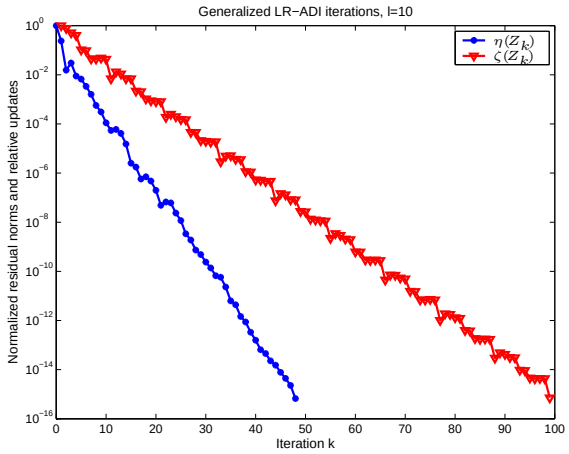

FIG. 4.4. *Example 4.3: convergence history for the generalized LR-ADI method.*

the first mass is connected to the last one by a rigid bar and it can be influenced by a control. The vibration of this system is described by a descriptor system (1.1) with the matrices E and A as in (1.7). For $g = 5000$, we obtain a problem of order $n = 10001$ with $B \in \mathbb{R}^{n,1}$. Figure 4.4 shows the normalized residual norms $\eta(Z_k)$ and the relative updates $\zeta(Z_k)$ for the generalized LR-ADI method with $\ell = 10$ complex ADI shift parameters. We see that the low rank Cholesky factor Z_k of the solution of the projected GCALE (1.2) computed with the stopping criterion $\zeta(Z_k) \leq tol$ has about twice more columns than those computed with the stopping criterion $\eta(Z_k) \leq tol$.

5. Conclusion. In this paper we have discussed the numerical solution of large-scale projected generalized Lyapunov equations that arise, for example, in model reduction for descriptor systems. We have presented the generalized low rank alternating direction implicit method and the generalized low rank cyclic Smith method for computing low rank approximations to the solutions of these equations. The efficiency of these methods has been demonstrated by numerical experiments.

Besides the ADI and Smith methods, there are several alternative approaches for solving large-scale standard Lyapunov equations. These are Krylov subspace methods [14, 15, 26], the sign function method [6, 16] and hierarchical matrix arithmetic based methods [4, 12]. The extension of these methods to projected generalized Lyapunov equations is a topic for future work.

REFERENCES

[1] A. C. ANTOULAS, D. C. SORENSEN, AND S. GUGERCIN, *A modified low-rank Smith method for large-scale Lyapunov equations*, Numerical Algorithms, 32 (2003), pp. 27–55.

[2] A. C. ANTOULAS, D. C. SORENSEN, AND Y. ZHOU, *On the decay rate of the Hankel singular values and related issues*, Systems Control Lett., 46 (2002), pp. 323–342.

[3] R. H. BARTELS AND G. W. STEWART, *Solution of the equation $AX + XB = C$*, Comm. ACM, 15 (1972), pp. 820–826.

[4] U. BAUR AND P. BENNER, *Factorized solution of Lyapunov equations based on hierarchical matrix arithmetic*, Preprint 161, DFG Research Center MATHEON, Technische Universität Berlin, Germany, 2004.

[5] D. J. BENDER, *Lyapunov-like equations and reachability/observability Gramians for descriptor systems*, IEEE Trans. Automat. Control, 32 (1987), pp. 343–348.

[6] P. Benner and E. S. Quintana-Ortí, *Solving stable generalized Lyapunov equations with the matrix sign function*, Numerical Algorithms, 20 (1999), pp. 75–100.
[7] K. Bernert, *Differenzenverfahren zur Lösung der Navier-Stokes-Gleichungen über Orthogonalen Netzen*, Wissenschaftliche Schriftenreihe 10/1990, Technische Universität Chemnitz, 1990. [German]
[8] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *The Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, Elsevier, North-Holland, New York, 1989.
[9] E. Eich-Soellner and C. Führer, *Numerical methods in multibody dynamics*, B.G. Teubner, Stuttgart, 1998.
[10] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, *Solution of the Sylvester matrix equation* $AXB^T + CXD^T = E$, ACM Trans. Math. Software, 18 (1992), pp. 223–231.
[11] L. Grasedyck, *Existence of a low rank of H-matrix approximation to the solution of the Sylvester equation*, Numer. Linear Algebra Appl., 11 (2004), pp. 371–389.
[12] L. Grasedyck, W. Hackbusch, and B. Khoromskij, *Solution of large scale algebraic matrix Riccati equations by use of hierarchical matrices*, Computing, 70 (2003), pp. 121–165.
[13] S. J. Hammarling, *Numerical solution of the stable non-negative definite Lyapunov equation*, IMA J. Numer. Anal., 2 (1982), pp. 303–323.
[14] D. Y. Hu and L. Reichel, *Krylov-subspace methods for the Sylvester equation*, Linear Algebra Appl., 172 (1992), pp. 283–313.
[15] I. M. Jaimoukha and E. M. Kasenally, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Numer. Anal., 31 (1994), pp. 227–251.
[16] V. B. Larin and F. A. Aliev, *Generalized Lyapunov equation and factorization of matrix polynomials*, Systems Control Lett., 21 (1993), pp. 485–491.
[17] J.-R. Li and T. Penzl, *Square root method via Cholesky factor ADI*, in Proceedings of the 14th International Symposium of Mathematical Theory of Networks and Systems, Perpignan, France, 2000.
[18] J.-R. Li, F. Wang, and J. White, *An efficient Lyapunov equation-based approach for generating reduced-order models of interconnect*, in Proceedings of the 36th Design Automation Conference, New Orleans, 1999, pp. 1–6.
[19] J.-R. Li and J. White, *Low rank solution of Lyapunov equations*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 260–280.
[20] A. Lu and E. Wachspress, *Solution of Lyapunov equations by alternating direction implicit iteration*, Comput. Math. Appl., 21 (1991), pp. 43–58.
[21] K. H. A. Olsson, *Model Order Reduction in FEMLAB by Dual Rational Arnoldi*, Master thesis, Chalmers University of Technology and Göteborg University, Göteborg, Sweden, 2002.
[22] D. W. Peaceman and H. H. Rachford, *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math., 3 (1955), pp. 28–41.
[23] T. Penzl, *Numerical solution of generalized Lyapunov equations*, Adv. Comput. Math., 8 (1998), pp. 33–48.
[24] ———, *A cyclic low-rank Smith method for large sparse Lyapunov equations*, SIAM J. Sci. Comput., 21 (1999/00), pp. 1401–1418.
[25] ———, *Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case*, Systems Control Lett., 40 (2000), pp. 139–144.
[26] Y. Saad, *Numerical solution of large Lyapunov equations*, in Signal Processing, Scattering, Operator Theory, and Numerical Methods, M. A. Kaashoek, J. H. Van Schuppen, and A. C. M. Ran, eds., Birkhäuser, Boston, 1990, pp. 503–511.
[27] ———, *Numerical Methods for Large Eigenvalue Problems*, Algorithms and Architectures for Advanced Scientific Computing, Manchester University Press, Manchester, 1992.
[28] ———, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, 1996.
[29] R. Schüpphaus, *Regelungstechnische Analyse und Synthese von Mehrkörpersystemen in Deskriptorform*, Ph.D. thesis, Bergische Universität-Gesamthochschule Wuppertal, Fortschritt-Berichte VDI, Reihe 8, Nr. 478, VDI Verlag, Düsseldorf, 1995. [German]
[30] R. A. Smith, *Matrix equation* $XA + BX = C$, SIAM J. Appl. Math., 16 (1968), pp. 198–201.
[31] D. C. Sorensen and Y. Zhou, *Bounds on eigenvalue decay rates and sensitivity of solutions of Lyapunov equations*, Technical Report TR02-07, Rice University, Houston, 2002.
[32] G. Starke, *Optimal alternating direction implicit parameters for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 28 (1991), pp. 1431–1445.
[33] T. Stykel, *Numerical solution and perturbation theory for generalized Lyapunov equations*, Linear Algebra Appl., 349 (2002), pp. 155–185.
[34] ———, *Stability and inertia theorems for generalized Lyapunov equations*, Linear Algebra Appl., 355 (2002) pp. 297–314.

[35] ———, *Input-output invariants for descriptor systems*, Preprint PIMS-03-1, The Pacific Institute for the Mathematical Sciences, Canada, 2003.
[36] ———, *Gramian-based model reduction for descriptor systems*, Math. Control Signals Systems, 16 (2004), pp. 297–319.
[37] F. Tisseur and K. Meerbergen, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.
[38] E. Wachspress, *Iterative solution of the Lyapunov matrix equation*, Appl. Math. Lett., 1 (1988), pp. 87–90.
[39] ———, *The ADI minimax problem for complex spectra*, in Iterative Methods for Large Linear Systems, D. Kincaid and L. Hayes, eds., Academic Press, 1990, pp. 251–271.
[40] L. Zhang, J. Lam, and Q. Zhang, *Lyapunov and Riccati equations for discrete-time descriptor systems*, IEEE Trans. Automat. Control, 44 (1999), pp. 2134–2139.