Published in *Linear Algebra and its Applications*, **349(1-3), 2002, pp. 155-185**

Numerical Solution and Perturbation Theory for Generalized Lyapunov Equations

Tatjana Stykel*

Abstract

We discuss the numerical solution and perturbation theory for the generalized continuous-time Lyapunov equation $E^*XA + A^*XE = -G$ with a singular matrix E. If this equation has a solution, it is not unique. We generalize a Bartels-Stewart method and a Hammarling method to compute a partial solution of the generalized Lyapunov equation with a special right-hand side. A spectral condition number is introduced and perturbation bounds for such an equation are presented. Numerical examples are given.

Key words. generalized Lyapunov equations, matrix pencils, deflating subspaces, spectral projections, perturbation theory, condition numbers.

AMS subject classification. 15A22, 15A24, 47A55, 65F35

1 Introduction

Consider the generalized continuous-time algebraic Lyapunov equation (GCALE)

$$E^*XA + A^*XE = -G \tag{1.1}$$

with given matrices E, A, G and unknown matrix X. Such equations play an important role in stability theory [13, 38], optimal control problems [33, 37] and balanced model reduction [36]. Equation (1.1) has a unique Hermitian, positive definite solution X for every Hermitian positive definite matrix G if and only if all eigenvalues of the pencil $\lambda E - A$ are finite and lie in the open left half-plane [39].

The classical numerical methods for the standard Lyapunov equations ($E = I$) are the Bartels-Stewart method [2], the Hammarling method [19] and the Hessenberg-Schur method [18]. An extension of these methods for the generalized Lyapunov equations with the non-singular matrix E is given in [9, 14, 15, 18, 39]. These methods are based on the preliminary reduction of the matrix (matrix pencil) to the (generalized) Schur form [17] or the Hessenberg-Schur form [18], calculation of the solution of a reduced system and back transformation.

An alternative approach to solve the (generalized) Lyapunov equations is the sign function method [6, 30, 35]. Comparison of the sign function method to the Bartels-Stewart and Hammarling methods with respect to accuracy and computational cost can be found in [6].

*Institut für Mathematik, MA 4-5, Technische Universität Berlin, Straße des 17. Juni 136, D-10623 Berlin, Germany, e-mail: stykel@math.tu-berlin.de. Supported by Deutsche Forschungsgemeinschaft, Research Grant ME 790/12-1.

The numerical solution of the generalized Lyapunov equations with a singular matrix E is more complicated. Such equations may not have solutions even if all finite eigenvalues of the pencil $\lambda E - A$ lie in the open left half-plane. Moreover, even if a solution exists, it is, in general, not unique. In this paper we consider the *projected generalized continuous-time algebraic Lyapunov equation*

$$\begin{aligned} &E^*XA + A^*XE = -P_r^*GP_r, \\ &X = XP_l, \end{aligned} \tag{1.2}$$

where G is Hermitian, positive (semi)definite, P_l and P_r are the spectral projections onto the left and right deflating subspaces of the pencil $\lambda E - A$ corresponding to the finite eigenvalues. Such an equation arises in stability theory and control problems for descriptor systems [4, 43]. The projected GCALE (1.2) has a unique Hermitian solution X that is positive definite on $\operatorname{im} P_l$ if and only if the pencil $\lambda E - A$ is *c-stable*, i.e., it is regular and all finite eigenvalues of $\lambda E - A$ lie in the open left half-plane, see [43] for details. Generalizations of the Bartels-Stewart and Hammarling methods to compute the solution of (1.2) are presented in Section 2.

In numerical problems it is very important to study the sensitivity of the solution to perturbations in the input data and to bound errors in the computed solution. There are several papers concerned with the perturbation theory and the backward error bounds for standard continuous-time Lyapunov equations, see [16, 20, 21] and references therein. The sensitivity analysis for generalized Lyapunov equations has been presented in [32], where only the case of nonsingular E was considered. In this paper we discuss the perturbation theory for the projected GCALE (1.2). In Section 3 we review condition numbers and Frobenius norm based condition estimators for the deflating subspaces of the pencil corresponding to the finite eigenvalues as well as the Lyapunov equations with nonsingular E. For the projected GCALE (1.2), we define a spectral norm based condition number which can be efficiently computed by solving (1.2) with $G = I$. Using this condition number we derive the perturbation bound for the solution of the projected GCALE (1.2) under perturbations that preserve the deflating subspaces of the pencil $\lambda E - A$ corresponding to the infinite eigenvalues. Section 4 contains some results of numerical experiments.

Throughout the paper $\mathbb{F}$ denotes the field of real ($\mathbb{F} = \mathbb{R}$) or complex ($\mathbb{F} = \mathbb{C}$) numbers, $\mathbb{F}^{n,m}$ is the space of $n \times m$-matrices over $\mathbb{F}$. The matrix $A^* = A^T$ denotes the transpose of the real matrix A, $A^* = A^H$ denotes the complex conjugate transpose of complex A and $A^{-*} = (A^{-1})^*$. We denote by $\|A\|_2$ the spectral norm of the matrix A and by $\|A\|_F$ the Frobenius norm of A. The vector formed by stacking the columns of the matrix A is denoted by $\operatorname{vec}(A)$, Π_{n^2} is the vec-permutation matrix of size $n^2 \times n^2$ such that $\operatorname{vec}(A^T) = \Pi_{n^2}\operatorname{vec}(A)$ and $A \otimes B = [a_{ij}B]$ is the Kronecker product of matrices A and B.

2 Numerical solution of projected generalized Lyapunov equations

The traditional methods to solve (generalized) Lyapunov equations are (generalized) Bartels-Stewart and Hammarling methods [2, 9, 14, 15, 19, 39] that are based on the preliminary reduction of the matrix (matrix pencil) to the (generalized) Schur form [17], calculation of the solution of a reduced quasi-triangular system and back transformation. In this section we extend these methods for the projected GCALE (1.2).

2.1 Generalizations of Schur and Bartels-Stewart methods

Let E and A be real square matrices (the complex case is similar). Assume that the pencil $\lambda E - A$ is regular, i.e., $\det(\lambda E - A) \neq 0$ for some $\lambda \in \mathbb{C}$. Then $\lambda E - A$ can be reduced to the GUPTRI form

$$E = V \begin{pmatrix} E_f & E_u \\ 0 & E_\infty \end{pmatrix} U^T, \qquad A = V \begin{pmatrix} A_f & A_u \\ 0 & A_\infty \end{pmatrix} U^T, \tag{2.1}$$

where matrices V and U are orthogonal, the pencil $\lambda E_f - A_f$ is quasi-triangular and has only finite eigenvalues, while the pencil $\lambda E_\infty - A_\infty$ is triangular and all its eigenvalues are infinite [11, 12]. Clearly, in this case the matrices E_f and A_∞ are nonsingular and E_∞ is nilpotent.

To compute the right and left deflating subspaces of $\lambda E - A$ corresponding to the finite eigenvalues we need to compute matrices Y and Z such that

$$\begin{pmatrix} I & -Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \lambda E_f - A_f & \lambda E_u - A_u \\ 0 & \lambda E_\infty - A_\infty \end{pmatrix} \begin{pmatrix} I & Y \\ 0 & I \end{pmatrix} = \begin{pmatrix} \lambda E_f - A_f & 0 \\ 0 & \lambda E_\infty - A_\infty \end{pmatrix}.$$

This leads to the generalized Sylvester equation

$$\begin{array}{rcl} E_f Y - Z E_\infty &=& -E_u, \\ A_f Y - Z A_\infty &=& -A_u. \end{array} \tag{2.2}$$

Since the pencils $\lambda E_f - A_f$ and $\lambda E_\infty - A_\infty$ have no common eigenvalues, equation (2.2) has a unique solution (Y, Z), e.g., [40]. Then the matrix pencil $\lambda E - A$ can be reduced by an equivalence transformation to the Weierstrass canonical form [41], i.e.,

$$\begin{aligned} \lambda E - A &= V \begin{pmatrix} I & Z \\ 0 & I \end{pmatrix} \begin{pmatrix} \lambda E_f - A_f & 0 \\ 0 & \lambda E_\infty - A_\infty \end{pmatrix} \begin{pmatrix} I & -Y \\ 0 & I \end{pmatrix} U^T \\ &= W \begin{pmatrix} \lambda E_f - A_f & 0 \\ 0 & \lambda E_\infty - A_\infty \end{pmatrix} T, \end{aligned}$$

where the matrices

$$W = V \begin{pmatrix} I & Z \\ 0 & I \end{pmatrix} \qquad \text{and} \qquad T = \begin{pmatrix} I & -Y \\ 0 & I \end{pmatrix} U^T$$

are nonsingular. In this case the spectral projections P_r and P_l onto the right and left deflating subspaces of $\lambda E - A$ corresponding to the finite eigenvalues have the form

$$P_r = T^{-1} \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} T = U \begin{pmatrix} I & -Y \\ 0 & 0 \end{pmatrix} U^T, \tag{2.3}$$

$$P_l = W \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} W^{-1} = V \begin{pmatrix} I & -Z \\ 0 & 0 \end{pmatrix} V^T. \tag{2.4}$$

Assume that the matrix pencil $\lambda E - A$ is c-stable. Setting

$$V^T X V = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \qquad \text{and} \qquad U^T G U = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix}, \tag{2.5}$$

we obtain from the GCALE in (1.2) the decoupled system of matrix equations

$$\begin{aligned}
E_f^T X_{11} A_f + A_f^T X_{11} E_f &= -G_{11}, && (2.6)\\
E_f^T X_{12} A_\infty + A_f^T X_{12} E_\infty &= G_{11} Y - E_f^T X_{11} A_u - A_f^T X_{11} E_u, && (2.7)\\
E_\infty^T X_{21} A_f + A_\infty^T X_{21} E_f &= Y^T G_{11} - E_u^T X_{11} A_f - A_u^T X_{11} E_f, && (2.8)\\
E_\infty^T X_{22} A_\infty + A_\infty^T X_{22} E_\infty &= -Y^T G_{11} Y - E_u^T X_{11} A_u - A_u^T X_{11} E_u - E_\infty^T X_{21} A_u \\
&\quad - A_\infty^T X_{21} E_u - E_u^T X_{12} A_\infty - A_u^T X_{12} E_\infty. && (2.9)
\end{aligned}$$

Since all eigenvalues of the pencil $\lambda E_f - A_f$ are finite and lie in the open left half-plane, the GCALE (2.6) has a unique solution X_{11} [9]. The pencils $\lambda E_f - A_f$ and $-\lambda E_\infty - A_\infty$ have no eigenvalues in common and, hence, the generalized Sylvester equations (2.7) and (2.8) are uniquely solvable [9]. To show that the matrix $X_{12} = -X_{11} Z$ satisfies equation (2.7), we substitute this matrix in (2.7). Taking into account equations (2.2) and (2.6), we obtain

$$\begin{aligned}
E_f^T X_{12} A_\infty + A_f^T X_{12} E_\infty &= -E_f^T X_{11}(A_f Y + A_u) - A_f^T X_{11}(E_f Y + E_u) \\
&= -(E_f^T X_{11} A_f + A_f^T X_{11} E_f) Y - E_f^T X_{11} A_u - A_f^T X_{11} E_u \\
&= G_{11} Y - E_f^T X_{11} A_u - A_f^T X_{11} E_u.
\end{aligned}$$

Similarly, it can be verified that the matrix $X_{21} = -Z^T X_{11}$ is the solution of (2.8).

Consider now equation (2.9). Substitute the matrices $X_{12} = -X_{11} Z$ and $X_{21} = -Z^T X_{11}$ in (2.9). Using (2.2) and (2.6) we obtain

$$\begin{aligned}
E_\infty^T X_{22} A_\infty + A_\infty^T X_{22} E_\infty &= Y^T E_f^T X_{11}(Z A_\infty - A_f Y) + Y^T A_f^T X_{11}(Z E_\infty - E_f Y) \\
&\quad + E_u^T X_{11} Z A_\infty + A_u^T X_{11} Z E_\infty - Y^T G_{11} Y \\
&= (E_f Y + E_u)^T X_{11} Z A_\infty + (A_f Y + A_u)^T X_{11} Z E_\infty \\
&= E_\infty^T Z^T X_{11} Z A_\infty + A_\infty^T Z^T X_{11} Z E_\infty.
\end{aligned}$$

Then

$$E_\infty^T (X_{22} - Z^T X_{11} Z) A_\infty + A_\infty^T (X_{22} - Z^T X_{11} Z) E_\infty = 0. \qquad (2.10)$$

Clearly, $X_{22} = Z^T X_{11} Z$ is the solution of (2.9). Moreover, we have

$$\begin{aligned}
X &= V \begin{pmatrix} X_{11} & -X_{11} Z \\ -Z^T X_{11} & Z^T X_{11} Z \end{pmatrix} V^T \\
&= V \begin{pmatrix} X_{11} & -X_{11} Z \\ -Z^T X_{11} & Z^T X_{11} Z \end{pmatrix} \begin{pmatrix} I & -Z \\ 0 & 0 \end{pmatrix} V^T = X P_l.
\end{aligned}$$

Thus, the matrix

$$X = V \begin{pmatrix} X_{11} & -X_{11} Z \\ -Z^T X_{11} & Z^T X_{11} Z \end{pmatrix} V^T \qquad (2.11)$$

is the solution of the projected GCALE (1.2).

In some applications we need the matrix $E^T X E$ rather that the solution X itself [42]. Using (2.1), (2.2) and (2.11) we obtain that

$$E^T X E = U \begin{pmatrix} E_f^T X_{11} E_f & -E_f^T X_{11} E_f Y \\ -Y^T E_f^T X_{11} E_f & Y^T E_f^T X_{11} E_f Y \end{pmatrix} U^T.$$

Remark 2.1 It follows from (2.10) that the general solution of the GCALE

$$E^T X A + A^T X E = -P_r^T G P_r \tag{2.12}$$

has the form

$$X = V \begin{pmatrix} X_{11} & -X_{11}Z \\ -Z^T X_{11} & X_\infty + Z^T X_{11} Z \end{pmatrix} V^T,$$

where X_∞ is the general solution of the homogeneous GCALE

$$E_\infty^T X_\infty A_\infty + A_\infty^T X_\infty E_\infty = 0.$$

If we require for the solution X of (2.12) to satisfy $X = XP_l$, then we obtain that $X_\infty = 0$.

In summary, we have the following algorithm for computing the solution X of the projected GCALE (1.2).

Algorithm 2.1 *Generalized Schur-Bartels-Stewart method for the projected GCALE (1.2).*
Input: *A real regular pencil $\lambda E - A$ and a real symmetric matrix G.*
Output: *A symmetric solution X of the projected GCALE (1.2).*
Step 1. *Use the GUPTRI algorithm [11, 12] to compute the orthogonal transformation matrices U and V such that*

$$V^T E U = \begin{pmatrix} E_f & E_u \\ 0 & E_\infty \end{pmatrix} \qquad \text{and} \qquad V^T A U = \begin{pmatrix} A_f & A_u \\ 0 & A_\infty \end{pmatrix}, \tag{2.13}$$

where E_f is upper triangular, nonsingular and E_∞ is upper triangular with zeros on the diagonal, A_f is upper quasi-triangular and A_∞ is upper triangular, nonsingular.
Step 2. *Use the generalized Schur method [26, 27] or the recursive blocked algorithm [22] to solve the generalized Sylvester equation*

$$\begin{aligned} E_f Y - Z E_\infty &= -E_u, \\ A_f Y - Z A_\infty &= -A_u. \end{aligned} \tag{2.14}$$

Step 3. *Compute the matrix*

$$U^T G U = \begin{pmatrix} G_{11} & G_{12} \\ G_{12}^T & G_{22} \end{pmatrix}. \tag{2.15}$$

Step 4. *Use the generalized Bartels-Stewart method [2, 39] or the recursive blocked algorithm [23] to solve the GCALE*

$$E_f^T X_{11} A_f + A_f^T X_{11} E_f = -G_{11}. \tag{2.16}$$

Step 5. *Compute the matrix*

$$X = V \begin{pmatrix} X_{11} & -X_{11}Z \\ -Z^T X_{11} & Z^T X_{11} Z \end{pmatrix} V^T. \tag{2.17}$$

2.2 Generalizations of Schur and Hammarling methods

In many applications it is necessary to have the Cholesky factor of the solution of the Lyapunov equation rather than the solution itself, e.g., [36]. An attractive algorithm for computing the Cholesky factor of the solution of the standard Lyapunov equation with a positive semidefinite right-hand side is the Hammarling method [19]. In [39] this method has been extended to the GCALE (1.1) with nonsingular E and positive semidefinite G. We will show that the Hammarling method can also be used to solve the projected GCALE

$$\begin{aligned} &E^T X A + A^T X E = -P_r^T C^T C P_r, \\ &X = X P_l \end{aligned} \tag{2.18}$$

with E, $A \in \mathbb{R}^{n,n}$, $C \in \mathbb{R}^{p,n}$. We may assume without loss of generality that C has full row rank, i.e., $\mathrm{rank}(C) = p \leq n$. If the pencil $\lambda E - A$ is c-stable, then the projected GCALE (2.18) has a unique symmetric, positive semidefinite solution X [43]. In fact, we can compute the full rank factorization [34] of the solution $X = R_X^T R_X$ without constructing X and the matrix product $C^T C$ explicitly.

Let $\lambda E - A$ be in the GUPTRI form (2.1) and let $CU = [\, C_1,\, C_2 \,]$ be partitioned in blocks conformally to E and A. Then the solution of the projected GCALE (2.18) has the form (2.11), where the symmetric, positive semidefinite matrix X_{11} satisfies the GCALE

$$E_f^T X_{11} A_f + A_f^T X_{11} E_f = -C_1^T C_1.$$

Let $U_{X_{11}}$ be a Cholesky factor of the solution $X_{11} = U_{X_{11}}^T U_{X_{11}}$. Compute the QR factorization

$$U_{X_{11}} = Q \begin{bmatrix} R_{X_{11}} \\ 0 \end{bmatrix},$$

where Q is orthogonal and $R_{X_{11}}$ has full row rank [17]. Then

$$\begin{aligned} X &= V \begin{bmatrix} U_{X_{11}}^T \\ -Z^T U_{X_{11}}^T \end{bmatrix} [\, U_{X_{11}},\ -U_{X_{11}} Z \,] V^T \\ &= V \begin{bmatrix} R_{X_{11}}^T \\ -Z^T R_{X_{11}}^T \end{bmatrix} [\, R_{X_{11}},\ -R_{X_{11}} Z \,] V^T = R_X^T R_X \end{aligned}$$

is the full rank factorization of X, where $R_X = [\, R_{X_{11}},\ -R_{X_{11}} Z \,] V^T$ has full row rank.

Thus, we have the following algorithm for computing the full row rank factor of the solution of the projected GCALE (2.18).

Algorithm 2.2 *Generalized Schur-Hammarling method for the projected GCALE (2.18).*
Input: *A real regular pencil $\lambda E - A$ and a real matrix C.*
Output: *A full row rank factor R_X of the solution $X = R_X^T R_X$ of (2.18).*
Step 1. *Use the GUPTRI algorithm [11, 12] to compute (2.1).*
Step 2. *Use the generalized Schur method [26, 27] or the recursive blocked algorithm [22] to compute the solution of the generalized Sylvester equation (2.2).*
Step 3. *Compute the matrix*

$$CU = [\, C_1,\ \ C_2 \,]. \tag{2.19}$$

Step 4. *Use the generalized Hammarling method [19, 39] to compute the Cholesky factor $U_{X_{11}}$ of the solution $X_{11} = U_{X_{11}}^T U_{X_{11}}$ of the GCALE*

$$E_f^T X_{11} A_f + A_f^T X_{11} E_f = -C_1^T C_1. \tag{2.20}$$

Step 5a. *Use Householder or Givens transformations [17] to compute the full row rank matrix $R_{X_{11}}$ from the QR-factorization*

$$U_{X_{11}} = Q \begin{bmatrix} R_{X_{11}} \\ 0 \end{bmatrix}.$$

Step 5b. *Compute the full row rank factor*

$$R_X = [\, R_{X_{11}},\ -R_{X_{11}} Z \,] \, V^T. \tag{2.21}$$

2.3 Numerical aspects

We will now discuss numerical aspects and computational cost for the algorithms described in the previous subsections in detail. We focus on Algorithm 2.1 and give a note about the differences to Algorithm 2.2.

Step 1. The numerical computation of the generalized Schur form of a matrix pencil has been intensively studied and various methods have been proposed, see [3, 11, 12, 17, 45] and the references therein. Comparison of the different algorithms can be found in [11].

To deflate the infinite eigenvalues of the matrix pencil $\lambda E - A$ and to reduce this pencil to the quasi-triangular form (2.13) we use the GUPTRI algorithm [11, 12]. This algorithm is based on the computation of the infinity-staircase form [44] of $\lambda E - A$ which exposes the Jordan structure of the infinite eigenvalues, and the QZ decomposition [17] of a subpencil which gives quasi-triangular blocks with the finite eigenvalues. The GUPTRI algorithm is numerically backwards stable and requires $O(n^3)$ operations [11].

Step 2. To solve the generalized Sylvester equation (2.14) we can use the generalized Schur method [26, 27]. Note that the pencils $\lambda E_f - A_f$ and $\lambda E_\infty - A_\infty$ are already in the generalized real Schur form [17], that is, the matrices E_f and E_∞ are upper triangular, whereas the matrices A_f and A_∞ are upper quasi-triangular. Since the infinite eigenvalues of $\lambda E_\infty - A_\infty$ correspond to the zero eigenvalues of the reciprocal pencil $E_\infty - \mu A_\infty$, we obtain that A_∞ is upper triangular. Let $A_f = [A_{ij}^f]_{i,j=1}^k$ and $A_\infty = [A_{ij}^\infty]_{i,j=1}^l$ be partitioned in blocks with diagonal blocks A_{jj}^f of size 1×1 or 2×2 and A_{jj}^∞ of size 1×1. Let $E_f = [E_{ij}^f]_{i,j=1}^k$, $E_\infty = [E_{ij}^\infty]_{i,j=1}^l$, $E_u = [E_{ij}^u]_{i,j=1}^{k,l}$, $A_u = [A_{ij}^u]_{i,j=1}^{k,l}$, $Y = [Y_{ij}]_{i,j=1}^{k,l}$ and $Z = [Z_{ij}]_{i,j=1}^{k,l}$ be partitioned in blocks conformally to A_f and A_∞. Then equation (2.14) is equivalent to the kl equations

$$E_{tt}^f Y_{tq} - Z_{tq} E_{qq}^\infty = -E_{tq} - \sum_{j=t+1}^{k} E_{tj}^f Y_{jq} + \sum_{j=1}^{q-1} Z_{tj} E_{jq}^\infty =: -\check{E}_{tq}, \tag{2.22}$$

$$A_{tt}^f Y_{tq} - Z_{tq} A_{qq}^\infty = -A_{tq} - \sum_{j=t+1}^{k} A_{tj}^f Y_{jq} + \sum_{j=1}^{q-1} Z_{tj} A_{jq}^\infty =: -\check{A}_{tq} \tag{2.23}$$

for $t = 1, \ldots, k$ and $q = 1, \ldots, l$. The matrices Y_{tq} and Z_{tq} can be computed successively in a row-wise order beginning with $t = k$ and $q = l$ from these equations. Since $E_{qq}^\infty = 0$, the 1×1 or 2×1 matrix Y_{tq} can be computed from the linear equation (2.22) of size 1×1 or 2×2 using Gaussian elimination with partial pivoting [17]. Then from (2.23) we obtain

$$Z_{tq} = (A_{tt}^f Y_{pq} + \check{A}_{tq})(A_{qq}^\infty)^{-1}.$$

The algorithm for solving the generalized Sylvester equation (2.14) via the generalized Schur method is available as the LAPACK subroutine _TGSYL [1] and costs $2m^2(n-m)+2m(n-m)^2$ flops [27].

To compute the solution of the quasi-triangular generalized Sylvester equation (2.14) we can also use the recursive blocked algorithm [22, Algorithm 3]. This algorithm consists in the recursive splitting equation (2.14) in smaller subproblems that can be solved using the high-performance kernel solvers. For comparison of the recursive blocked algorithm and the LAPACK subroutine, see [22].

Step 3 is a matrix multiplication. In fact, in Algorithm 2.1 only the $m \times m$ block G_{11} in (2.15) is needed. Let $U = [\,U_1,\, U_2\,]$, where the columns of the $n \times m$-matrix U_1 form the basis of the right finite deflating subspace of $\lambda E - A$. Exploiting the symmetry of G, the computation of $G_{11} = U_1^T G U_1$ requires $n^2m + 1/2nm^2$ flops. In Algorithm 2.2 we only need the $p \times m$ block C_1 in (2.19) which can be computed as $C_1 = CU_1$ in pmn flops.

Step 4. To solve the GCALE (2.16) with nonsingular E_f we can use the generalized Bartels-Stewart method [2, 39]. Let the matrices $X_{11} = [X'_{ij}]_{i,j=1}^{k}$ and $G_{11} = [G'_{ij}]_{i,j=1}^{k}$ be partitioned in blocks conformally to E_f and A_f. Then equation (2.16) is equivalent to k^2 equations

$$(E_{tt}^f)^T X'_{tq} A_{qq}^f + (A_{tt}^f)^T X'_{tq} E_{qq}^f = -\check{G}_{tq}, \qquad t, q = 1, \ldots k, \tag{2.24}$$

where

$$\begin{aligned}
\check{G}_{tq} &= G'_{tq} + \sum_{\substack{i=1,j=1 \\ (i,j)\neq(t,q)}}^{t,q} \left((E_{it}^f)^T X'_{ij} A_{jq}^f + (A_{it}^f)^T X'_{ij} E_{jq}^f \right) \\
&= G'_{tq} + \sum_{i=1}^{t} \left[(E_{it}^f)^T \left(\sum_{j=1}^{q-1} X'_{ij} A_{jq}^f \right) + (A_{it}^f)^T \left(\sum_{j=1}^{q-1} X'_{ij} E_{jq}^f \right) \right] \\
&\quad + \sum_{i=1}^{t-1} \left[(E_{it}^f)^T X'_{iq} A_{qq}^f + (A_{it}^f)^T X'_{iq} E_{qq}^f \right].
\end{aligned}$$

We compute the blocks X'_{tq} in a row-wise order beginning with $t = q = 1$. Using the column-wise vector representation of the matrices X'_{tq} and $\check{G}_{tq}$ we can rewrite the generalized Sylvester equation (2.24) as a linear system

$$\left((A_{qq}^f)^T \otimes (E_{tt}^f)^T + (E_{qq}^f)^T \otimes (A_{tt}^f)^T \right) \mathrm{vec}(X'_{tq}) = -\mathrm{vec}(\check{G}_{tq}) \tag{2.25}$$

of size 2×2, 4×4 or 8×8. The solution $\mathrm{vec}(X'_{tq})$ can be computed by solving (2.25) via Gaussian elimination with partial pivoting [17].

To compute the Cholesky factor of the solution of the GCALE (2.20) in Algorithm 2.2 we can use the generalized Hammarling method, see [19, 39] for details.

The solution of the GCALE (2.16) using the generalized Bartels-Stewart method requires $O(m^3)$ flops, while computing the Cholesky factor of the solution of the GCALE (2.20) by the generalized Hammarling method requires $O(m^3 + pm^2 + p^2m)$ flops [39].

The generalized Bartels-Stewart method and the generalized Hammarling method are implemented in LAPACK-style subroutines SG03AD and SG03BD, respectively, that are available in the SLICOT Library [5].

The quasi-triangular GCALE (2.16) can be also solved using the recursive blocked algorithm [23, Algorithm 3]. Comparison of this algorithm with the SLICOT subroutines can be found in [23].

Step 5. The matrix X in (2.17) is computed in $O(n^3 + m^2(n-m) + m(n-m)^2)$ flops. The computation of the full row rank factor R_X in (2.21) requires $O(m^3 + m(n-m)r + n^2 r)$ flops with $r = \text{rank}(X)$.

Thus, the total computational cost of the generalized Schur-Bartels-Stewart method as well as the generalized Schur-Hammarling method is estimated as $O(n^3)$.

3 Conditioning and condition estimators

In this section we discuss feasible condition numbers and condition estimators for the projected GCALE (1.2). A condition number for a problem is an important characteristic to measure the sensitivity of the solution of this problem to perturbations in the original data and to bound errors in the computed solution. If the condition number is large, then the problem is ill-conditioned in the sense that small perturbations in the data may lead to large variations in the solution.

The solution of the projected GCALE (1.2) is determined essentially in two steps that include first a computation of the deflating subspaces of a pencil corresponding to the finite and infinite eigenvalues due reduction to the GUPTRI form and solving the generalized Sylvester equation and then a calculation of the solution of the generalized Lyapunov equation. In such situation it may happen that although the projected GCALE (1.2) is well-conditioned, one of intermediate problems may be ill-conditioned. This may lead to large inaccuracy in the numerical solution of the original problem. In this case we may conclude that either the combined numerical method is unstable or the solution is ill-conditioned, since it is a composition of two mappings one of which is ill-conditioned. Therefore, along with the conditioning of the projected GCALE (1.2) we consider the perturbation theory for the deflating subspaces and the GCALE (1.1) with nonsingular E.

3.1 Conditioning of deflating subspaces and generalized Sylvester equations

The perturbation analysis for the deflating subspaces of a regular pencil corresponding to the specified eigenvalues and error bounds are presented in [10, 25, 26, 40, 41]. Here we briefly review the main results from there.

To compute the right and left deflating subspaces of $\lambda E - A$ corresponding to the finite eigenvalues we have to solve the generalized Sylvester equation (2.2). We define the *Sylvester operator* $\mathcal{S} : \mathbb{F}^{m,2(n-m)} \to \mathbb{F}^{m,2(n-m)}$ via

$$\mathcal{S}(Y, Z) := (E_f Y - Z E_\infty,\ A_f Y - Z A_\infty). \tag{3.1}$$

Using the column-wise vector representation for the matrices Y and Z we can rewrite (2.2) as a linear system

$$S \begin{bmatrix} \text{vec}(Y) \\ \text{vec}(Z) \end{bmatrix} = - \begin{bmatrix} \text{vec}(E_u) \\ \text{vec}(A_u) \end{bmatrix}, \tag{3.2}$$

where the $2m(n-m) \times 2m(n-m)$-matrix

$$S = \begin{bmatrix} I_{n-m} \otimes E_f & -E_\infty^T \otimes I_m \\ I_{n-m} \otimes A_f & -A_\infty^T \otimes I_m \end{bmatrix}$$

is the matrix representation of the Sylvester operator $\mathcal{S}$. The norm of $\mathcal{S}$ induced by the Frobenius matrix norm is given by

$$\|\mathcal{S}\|_F := \sup_{\|(Y,Z)\|_F=1} \|(E_fY - ZE_\infty, A_fY - ZA_\infty)\|_F = \|S\|_2.$$

We define the *separation* of two regular matrix pencils $\lambda E_f - A_f$ and $\lambda E_\infty - A_\infty$ as

$$\mathrm{Dif}_u \equiv \mathrm{Dif}_u(E_f, A_f; E_\infty, A_\infty) := \inf_{\|(Y,Z)\|_F=1} \|(E_fY - ZE_\infty, A_fY - ZA_\infty)\|_F = \sigma_{\min}(S),$$

where $\sigma_{\min}(S)$ is the smallest singular value of S [40]. Note that $\mathrm{Dif}_u(E_\infty, A_\infty; E_f, A_f)$ does not in general equal $\mathrm{Dif}_u(E_f, A_f; E_\infty, A_\infty)$. Therefore, we set

$$\mathrm{Dif}_l \equiv \mathrm{Dif}_l(E_f, A_f; E_\infty, A_\infty) := \mathrm{Dif}_u(E_\infty, A_\infty; E_f, A_f).$$

The values Dif_u and Dif_l measure how close the spectra of $\lambda E_f - A_f$ and $\lambda E_\infty - A_\infty$ are. In other words, if there is a small perturbation of $\lambda E_f - A_f$ and $\lambda E_\infty - A_\infty$ such that the perturbed pencils have a common eigenvalue, then either Dif_u or Dif_l is small. However, small separations do not imply that the corresponding deflating subspaces are ill-conditioned [41].

Important quantities that measure the sensitivity of the right and left deflating subspaces of the pencil $\lambda E - A$ to perturbations in E and A are the norms of the spectral projections P_r and P_l. If $\|P_r\|_2$ (or $\|P_l\|_2$) is large then the right (left) deflating subspace of $\lambda E - A$ corresponding to the finite eigenvalues is close to the right (left) deflating subspace corresponding to the infinite eigenvalues.

Let the pencil $\lambda E - A$ be in the GUPTRI form (2.1) and let the transformation matrices $U = [\,U_1,\, U_2\,]$ and $V = [\,V_1,\, V_2\,]$ be partitioned conformally to the blocks with the finite and infinite eigenvalues. In this case $\mathcal{U} = \mathrm{span}(U_1)$ and $\mathcal{V} = \mathrm{span}(V_1)$ are the right and left finite deflating subspaces of $\lambda E - A$, respectively, and they have dimension m. Consider a perturbed matrix pencil $\lambda\widetilde{E} - \widetilde{A} = \lambda(E + \Delta E) - (A + \Delta A)$. Let $\widetilde{\mathcal{U}}$ and $\widetilde{\mathcal{V}}$ be the right and left finite deflating subspaces of $\lambda\widetilde{E} - \widetilde{A}$, respectively, and suppose that they have the same dimensions as $\mathcal{U}$ and $\mathcal{V}$. A *distance between two subspaces* $\mathcal{U}$ and $\widetilde{\mathcal{U}}$ is given by

$$\theta_{\max}(\mathcal{U}, \widetilde{\mathcal{U}}) = \max_{u\in\mathcal{U}} \min_{\widetilde{u}\in\widetilde{\mathcal{U}}} \theta(u, \widetilde{u}),$$

where $\theta(u, \widetilde{u})$ is the acute angle between the vectors u and $\widetilde{u}$. Then one has the following perturbation bounds for the deflating subspaces of the regular pencil $\lambda E - A$.

Theorem 3.1 *[10] Suppose that the right and left finite deflating subspaces of a regular matrix pencil $\lambda E - A$ and a perturbed pencil $\lambda\widetilde{E} - \widetilde{A} = \lambda(E + \Delta E) - (A + \Delta A)$ corresponding to the finite eigenvalues have the same dimensions. If*

$$\|(\Delta E, \Delta A)\|_F < \frac{\min(\mathrm{Dif}_u, \mathrm{Dif}_l)}{\sqrt{\|P_l\|_2^2 + \|P_r\|_2^2 + \max(\|P_l\|_2, \|P_r\|_2)}} =: \rho,$$

then

$$\begin{aligned} \tan\theta_{\max}(\mathcal{U}, \widetilde{\mathcal{U}}) &\le \frac{\|(\Delta E, \Delta A)\|_F}{\rho\|P_r\|_2 - \|(\Delta E, \Delta A)\|_F\sqrt{\|P_r\|_2^2 - 1}} \\ &\le \|(\Delta E, \Delta A)\|_F \frac{\|P_r\|_2^2 + \sqrt{\|P_r\|_2^2 - 1}}{\rho} \end{aligned} \tag{3.3}$$

and

$$\begin{aligned}\tan\theta_{\max}(\mathcal{V},\widetilde{\mathcal{V}}) &\leq \frac{\|(\Delta E,\Delta A)\|_F}{\rho\|P_l\|_2-\|(\Delta E,\Delta A)\|_F\sqrt{\|P_l\|_2^2-1}} \\ &\leq \|(\Delta E,\Delta A)\|_F\frac{\|P_l\|_2^2+\sqrt{\|P_l\|_2^2-1}}{\rho}.\end{aligned} \tag{3.4}$$

Bounds (3.3) and (3.4) imply that for small enough $\|(\Delta E,\Delta A)\|_F$, the right and left deflating subspaces of the perturbed pencil $\lambda\widetilde{E}-\widetilde{A}$ corresponding to the finite eigenvalues are small perturbations of the corresponding right and left deflating subspaces of $\lambda E-A$. Perturbation $\|(\Delta E,\Delta A)\|_F$ is bounded by ρ which is small if the separations Dif_u and Dif_l are small or the norms $\|P_l\|_2$ and $\|P_r\|_2$ are large.

Thus, the quantities Dif_u, Dif_l, $\|P_l\|_2$ and $\|P_r\|_2$ can be used to characterize the conditioning of the right and left deflating subspaces of the pencil $\lambda E-A$ corresponding to the finite eigenvalues.

From representations (2.3) and (2.4) for the spectral projections P_r and P_l we have

$$\|P_r\|_2=\sqrt{1+\|Y\|_2^2},\qquad \|P_l\|_2=\sqrt{1+\|Z\|_2^2}, \tag{3.5}$$

where (Y,Z) is the solution of the generalized Sylvester equation (2.2). We see that the norms of Y and Z also characterize the sensitivity of the deflating subspaces. It follows from (3.2) that

$$\|(Y,Z)\|_F\leq \mathrm{Dif}_u^{-1}\|(E_u,A_u)\|_F. \tag{3.6}$$

This estimate gives a connection between the separation Dif_u and the norm of the solution of the generalized Sylvester equation (2.2).

The perturbation analysis, condition numbers and error bounds for the generalized Sylvester equation are presented in [24, 27]. Consider a perturbed generalized Sylvester equation

$$\begin{aligned}(E_f+\Delta E_f)\widetilde{Y}-\widetilde{Z}(E_\infty+\Delta E_\infty) &= -(E_u+\Delta E_u),\\ (A_f+\Delta A_f)\widetilde{Y}-\widetilde{Z}(A_\infty+\Delta A_\infty) &= -(A_u+\Delta E_u),\end{aligned} \tag{3.7}$$

where the perturbations are measured norm-wise by

$$\epsilon=\max\left\{\frac{\|(\Delta E_f,\Delta A_f)\|_F}{\alpha},\ \frac{\|(\Delta E_\infty,\Delta A_\infty)\|_F}{\beta},\ \frac{\|(\Delta E_u,\Delta A_u)\|_F}{\gamma}\right\} \tag{3.8}$$

with $\alpha=\|(E_f,A_f)\|_F$, $\beta=\|(E_\infty,A_\infty)\|_F$ and $\gamma=\|(E_u,A_u)\|_F$. Then one has the following first order relative perturbation bound for the solution of the generalized Sylvester equation (2.2).

Theorem 3.2 *[24] Let the perturbations in (3.7) satisfy (3.8). Assume that both the generalized Sylvester equations (2.2) and (3.7) are uniquely solvable. Then*

$$\frac{\|(\widetilde{Y},\widetilde{Z})-(Y,Z)\|_F}{\|(Y,Z)\|_F}\leq\sqrt{3}\,\epsilon\,\frac{\|S^{-1}M_S\|_2}{\|(Y,Z)\|_F}, \tag{3.9}$$

where the matrix M_S of size $2m(n-m)\times 2(n^2-nm+m^2)$ has the form

$$M_S=\begin{bmatrix} B_S & 0\\ 0 & B_S\end{bmatrix}$$

with $B_S=[\,\alpha(Y^T\otimes I_m),\ -\beta(I_{n-m}\otimes Z),\ \gamma I_{m(n-m)}\,]$.

The number

$$æ_{st} = \frac{\|S^{-1}M_S\|_2}{\|(Y,Z)\|_F}$$

is called the *structured condition number* for the generalized Sylvester equation (2.2). Bound (3.9) shows that the relative error in the solution of the perturbed equation (3.7) is small if $æ_{st}$ is not too large, i.e., if the problem is well-conditioned.

From (3.9) we obtain an other relative error bound

$$\frac{\|(\widetilde{Y},\widetilde{Z})-(Y,Z)\|_F}{\|(Y,Z)\|_F} \leq \sqrt{3}\,\epsilon\,\mathrm{Dif}_u^{-1}\,\frac{(\alpha+\beta)\,\|(Y,Z)\|_F+\gamma}{\|(Y,Z)\|_F}$$

that, in general, is worse than (3.9), since it does not take account of the special structure of perturbations in the generalized Sylvester equation [24].

Define the *condition number* for the generalized Sylvester equation (2.2) induced by the Frobenius norm as

$$æ_F := \Big(\|(E_f,A_f)\|_F^2+\|(E_\infty,A_\infty)\|_F^2\Big)^{1/2}\mathrm{Dif}_u^{-1}.$$

Then applying the standard linear system perturbation analysis [17] to (3.2) we have the following relative perturbation bounds.

Theorem 3.3 *[27] Suppose that the generalized Sylvester equation (2.2) has a unique solution (Y,Z). Let the perturbations in (3.7) satisfy (3.8). If $\epsilon\,æ_F<1$, then the perturbed generalized Sylvester equation (3.7) has a unique solution $(\widetilde{Y},\widetilde{Z})$ and*

$$\frac{\|(\widetilde{Y},\widetilde{Z})-(Y,Z)\|_F}{\|(Y,Z)\|_F} \leq \frac{\epsilon\,(æ_F\|(Y,Z)\|_F+\|(E_u,A_u)\|_F)}{(1-\epsilon\,æ_F)\|(Y,Z)\|_F} \leq \frac{2\,\epsilon\,æ_F}{1-\epsilon\,æ_F}. \tag{3.10}$$

Note that both the bounds in (3.10) may overestimate the true relative error in the solution, since they do not take into account the structured perturbations in the matrix S. Nevertheless, quantities Dif_u^{-1} and $æ_F$ are used in practice to characterize the conditioning of the generalized Sylvester equation (2.2).

The computation of $\mathrm{Dif}_u=\sigma_{\min}(S)$ is expensive even for modest m and $n-m$, since the cost of computing the smallest singular value of the matrix S is $O(m^3(n-m)^3)$ flops. It is more useful to compute lower bounds for Dif_u^{-1}, see [26, 27] for details. The Frobenius norm based Dif_u^{-1}-estimator can be computed by solving one generalized Sylvester equation in triangular form and costs $2m^2(n-m)+2m(n-m)^2$ flops. The one-norm based estimator is a factor 3 to 10 times more expensive and it does not differ more than a factor $\sqrt{2m(n-m)}$ from Dif_u^{-1} [26]. Computing both the Dif_u^{-1}-estimators is implemented in the LAPACK subroutine `_TGSEN` [1].

3.2 Condition numbers for the generalized Lyapunov equations

The perturbation theory and some useful condition numbers for the standard Lyapunov equations were presented in [16, 20, 21], see also the references therein. The case of the generalized Lyapunov equations with nonsingular E was considered in [31, 32]. In this subsection we review some results from there.

Consider the GCALE (1.1). Equation (1.1) is called *regular* if the matrix E is nonsingular and $\lambda_i + \overline{\lambda}_j \neq 0$ for all eigenvalues λ_i and λ_j of the pencil $\lambda E - A$. Clearly, the regular GCALE (1.1) has a unique solution X for every G, see [9].

Define the *continuous-time Lyapunov operator* $\mathcal{L} : \mathbb{F}^{n,n} \to \mathbb{F}^{n,n}$ via

$$\mathcal{L}(X) := E^*XA + A^*XE. \tag{3.11}$$

Then the GCALE (1.1) can be rewritten in the operator form $\mathcal{L}(X) = -G$ or as a linear system

$$Lx = -g, \tag{3.12}$$

where $x = \mathrm{vec}(X)$, $g = \mathrm{vec}(G)$ and the $n^2 \times n^2$-matrix

$$L = E^T \otimes A^* + A^T \otimes E^* \tag{3.13}$$

is the matrix representation of the Lyapunov operator $\mathcal{L}$.

The norm of $\mathcal{L}$ induced by the Frobenius matrix norm is computed via

$$\|\mathcal{L}\|_F := \sup_{\|X\|_F=1} \|E^*XA + A^*XE\|_F = \|L\|_2.$$

Analogously to the Sylvester equation, an important quantity in the sensitivity analysis for Lyapunov equations is a *separation* defined for the GCALE (1.1) by

$$\mathrm{Sep}(E, A) = \inf_{\|X\|_F=1} \|E^*XA + A^*XE\|_F = \sigma_{\min}(L),$$

where $\sigma_{\min}(L)$ is the smallest singular value of L, see [14]. If the GCALE (1.1) is regular, then the Lyapunov operator $\mathcal{L}$ is invertible and the matrix L is nonsingular. The norm of the inverse $\mathcal{L}^{-1}$ induced by the Frobenius norm can be computed as

$$\|\mathcal{L}^{-1}\|_F = \|L^{-1}\|_2 = \mathrm{Sep}^{-1}(E, A).$$

Consider now a perturbed GCALE

$$(E + \Delta E)^*\widetilde{X}(A + \Delta A) + (A + \Delta A)^*\widetilde{X}(E + \Delta E) = -(G + \Delta G), \tag{3.14}$$

where

$$\begin{array}{ll} \|\Delta E\|_F \leq \varepsilon_F, & \|\Delta A\|_F \leq \varepsilon_F, \\ \|\Delta G\|_F \leq \varepsilon_F, & (\Delta G)^* = \Delta G. \end{array} \tag{3.15}$$

Using the equivalent formulation (3.12) for the GCALE (1.1) we have the following norm-wise perturbation estimate for the solution of (1.1) in the real case, see [32] for the complex case.

Theorem 3.4 *[32] Let E, A, $G \in \mathbb{R}^{n,n}$ and let G be symmetric. Assume that the GCALE (1.1) is regular. Let the absolute perturbations in the GCALE (3.14) satisfy (3.15). If*

$$\varepsilon_F \left(l_E + l_A + 2\varepsilon_F Sep^{-1}(E, A)\right) < 1,$$

then the perturbed GCALE (3.14) is regular and the norm-wise absolute perturbation bound

$$\|\widetilde{X} - X\|_F \leq \frac{\sqrt{3}\,\varepsilon_F \|L^{-1}M_L\|_2 + 2\varepsilon_F^2 Sep^{-1}(E, A)\|X\|_2}{1 - \varepsilon_F \left(l_E + l_A + 2\varepsilon_F Sep^{-1}(E, A)\right)} \tag{3.16}$$

holds, where

$$\begin{aligned} M_L &= \Big[(I_{n^2} + \Pi_{n^2}) \left(I_n \otimes (A^T X) \right), \quad (I_{n^2} + \Pi_{n^2}) \left(I_n \otimes (E^T X) \right), \quad I_{n^2} \Big], \\ l_E &= \left\| L^{-1} (I_{n^2} + \Pi_{n^2}) \left(I_n \otimes A^T \right) \right\|_2, \\ l_A &= \left\| L^{-1} (I_{n^2} + \Pi_{n^2}) \left(I_n \otimes E^T \right) \right\|_2. \end{aligned}$$

The number

$$\kappa_{st}(E, A) = \frac{\|L^{-1} M_L\|_2}{\|X\|_F}$$

is called the *structured condition number* for the GCALE (1.1). Bound (3.16) shows that if $\kappa_{st}(E, A)$, $\mathrm{Sep}^{-1}(E, A)$, l_E and l_A are not too large, then the solution of the perturbed GCALE (3.14) is a small perturbation of the solution of (1.1). Note that bound (3.16) is asymptotically sharp.

We define the *condition number* for the GCALE (1.1) induced by the Frobenius norm as

$$\kappa_F(E, A) := 2\|E\|_2 \|A\|_2 \mathrm{Sep}^{-1}(E, A). \tag{3.17}$$

Applying the standard linear system perturbation analysis [17] to the linear system (3.12) and taking into account that $\|G\|_2 \leq 2\|E\|_2\|A\|_2\|X\|_F$, we obtain the following Frobenius norm based relative perturbation bounds.

Theorem 3.5 *Suppose that the GCALE (1.1) is regular. Let the perturbations in (3.14) satisfy $\|\Delta E\|_2 \leq \varepsilon\|E\|_2$, $\|\Delta A\|_2 \leq \varepsilon\|A\|_2$ and $\|\Delta G\|_2 \leq \varepsilon\|G\|_2$. If $\varepsilon\,(2+\varepsilon)\kappa_F(E, A) < 1$, then the perturbed GCALE (3.14) is regular and*

$$\begin{aligned} \frac{\|\widetilde{X} - X\|_F}{\|X\|_F} &\leq \frac{(2\varepsilon + \varepsilon^2)\kappa_F(E, A)\|X\|_F + \varepsilon\|G\|_2 Sep^{-1}(E, A)}{(1 - \varepsilon(2+\varepsilon)\kappa_F(E, A)\,)\|X\|_F} \\ &\leq \frac{\varepsilon(3+\varepsilon)\kappa_F(E, A)}{1 - \varepsilon(2+\varepsilon)\kappa_F(E, A)}. \end{aligned} \tag{3.18}$$

It should be noted that bounds (3.18) may overestimate the true relative error, since they do not take account of the specific structure of perturbations in (3.14). In the case of symmetric perturbations in G, sharp sensitivity estimates for general Lyapunov operators can be derived by using so-called Lyapunov singular values instead of standard singular values, see [31, 32] for details. Note that for the Lyapunov operator $\mathcal{L}$ as in (3.11), the Lyapunov singular values are equal to the standard singular values.

Let $\hat{X}$ be an approximate solution of the GCALE (1.1) and let

$$R := E^* \hat{X} A + A^* \hat{X} E + G \tag{3.19}$$

be a *residual* of (1.1) corresponding to $\hat{X}$. Then from Theorem 3.5 we obtain the following forward error bound

$$\frac{\|\hat{X} - X\|_F}{\|X\|_F} \leq \kappa_F(E, A)\, \frac{\|R\|_F}{2\|E\|_2\|A\|_2\|X\|_F} =: Est_F. \tag{3.20}$$

This bound shows that for well-conditioned problems, the small relative residual implies a small error in the approximate solution $\hat{X}$. However, if the condition number $\kappa_F(E, A)$ is large, then $\hat{X}$ may be inaccurate even for the small residual.

It follows from bounds (3.18) and (3.20) that $\kappa_F(E,A)$ and $\mathrm{Sep}(E,A) = \sigma_{\min}(L)$ can be used as a measure of the sensitivity of the solution of the regular GCALE (1.1). Since computing the smallest singular value of the $n^2 \times n^2$-matrix L is not acceptable for modest n, it is more useful to compute estimates for $\mathrm{Sep}^{-1}(E,A)$. A Sep^{-1}-estimator based on the one-norm differs from $\mathrm{Sep}^{-1}(E,A)$ at most by a factor n. Computing this estimator is implemented in the LAPACK subroutine `_LACON` [1] and costs $O(n^3)$ flops.

Unfortunately, if the matrix E is singular, then $\mathrm{Sep}(E,A) = 0$ and $\kappa_F(E,A) = \infty$. In this case we can not use (3.17) as the condition number for the projected GCALE (1.2).

In [16, 20] condition numbers based on the spectral norm have been used as a measure of sensitivity of the standard Lyapunov equations. In the following subsections we extend this idea to the projected GCALE (1.2).

3.3 Conditioning of the projected generalized Lyapunov equations

Assume that the pencil $\lambda E - A$ is c-stable. Let H be an Hermitian, positive semidefinite solution of the projected GCALE

$$\begin{aligned} &E^*HA + A^*HE = -P_r^*P_r, \\ &H = HP_l. \end{aligned} \tag{3.21}$$

The matrix H has the form

$$H = \frac{1}{2\pi}\int_{-\infty}^{\infty} (i\xi E - A)^{-*} P_r^* P_r (i\xi E - A)^{-1} d\xi.$$

Consider a linear operator $\mathcal{L}^- : \mathbb{F}^{n,n} \to \mathbb{F}^{n,n}$ defined as follows: for a matrix G, the image $X = -\mathcal{L}^-(G)$ is the unique solution of the projected GCALE (1.2). Note that the operator $\mathcal{L}^-$ is a (2)-pseudoinverse [8] of the Lyapunov operator L since it satisfies $\mathcal{L}^-\mathcal{L}\mathcal{L}^- = \mathcal{L}^-$.

Lemma 3.6 *Let $\lambda E - A$ be c-stable. Then $\|\mathcal{L}^-\|_2 = \|H\|_2$.*

Proof. Let u and v be the left and right singular vectors of unit length corresponding to the largest singular value of the solution X of the projected GCALE (1.2) with some matrix G. Then

$$\begin{aligned} \|\mathcal{L}^-(G)\|_2 &= \|X\|_2 = u^*Xv = \frac{1}{2\pi}\int_{-\infty}^{\infty} u^*(i\xi E - A)^{-*} P_r^* G P_r (i\xi E - A)^{-1} v \, d\xi \\ &\le \frac{1}{2\pi}\|G\|_2 \int_{-\infty}^{\infty} \left\|P_r(i\xi E - A)^{-1}u\right\|_2 \left\|P_r(i\xi E - A)^{-1}v\right\|_2 \, d\xi. \end{aligned}$$

Using the Cauchy-Schwarz inequality [28] we obtain

$$\begin{aligned} \|\mathcal{L}^-(G)\|_2 &\le \frac{1}{2\pi}\|G\|_2 \left(\int_{-\infty}^{\infty} \left\|P_r(i\xi E - A)^{-1}u\right\|_2^2 \, d\xi\right)^{\frac{1}{2}} \left(\int_{-\infty}^{\infty} \left\|P_r(i\xi E - A)^{-1}v\right\|_2^2 \, d\xi\right)^{\frac{1}{2}} \\ &\le \|G\|_2 \left\| \frac{1}{2\pi}\int_{-\infty}^{\infty} (i\xi E - A)^{-*} P_r^* P_r (i\xi E - A)^{-1} \, d\xi \right\|_2 = \|G\|_2 \|H\|_2. \end{aligned}$$

Hence, $\|\mathcal{L}^-\|_2 \le \|H\|_2$.

On the other hand, we have

$$\|\mathcal{L}^-\|_2 = \sup_{\|G\|_2=1} \|\mathcal{L}^-(-G)\|_2 \geq \|\mathcal{L}^-(-I)\|_2 = \|H\|_2.$$

Thus, $\|\mathcal{L}^-\|_2 = \|H\|_2$. □

Note that if E is nonsingular, then $\mathcal{L}^- = \mathcal{L}^{-1}$ is the inverse of the Lyapunov operator $\mathcal{L}$ and $\|\mathcal{L}^{-1}\|_2 = \|H\|_2$.

We define the *spectral condition number* for the projected GCALE (1.2) as

$$\kappa_2(E, A) := 2\|E\|_2\|A\|_2\|H\|_2. \tag{3.22}$$

We have

$$1 \leq \|P_r\|_2^2 = \|P_r^* P_r\|_2 = \|E^* H A + A^* H E\|_2 \leq 2\|E\|_2\|A\|_2\|H\|_2 = \kappa_2(E, A).$$

The matrix H and the parameter $\kappa_2(E, A)$ are closely related to the analysis of the asymptotic behavior of solutions of the differential-algebraic equation

$$E\dot{x}(t) = Ax(t). \tag{3.23}$$

It has been shown in [42] that

$$\|E^* H E\|_2 = \max_{\|P_r x_0\|=1} \int_0^\infty \|x(t)\|^2 dt,$$

i.e., the norm of the matrix E^*HE is the square of the maximum L_2-norm of the solution $x(t)$ of equation (3.23) with the initial condition $x(0) = P_r x_0$. Moreover, for this solution the pointwise estimate

$$\|x(t)\| \leq \sqrt{\kappa_2(E, A)\|E\|_2\|(EP_r + A(I - P_r))^{-1}\|_2}\, e^{-t\|A\|_2/(\|E\|_2\kappa_2(E,A))}\|P_r x_0\|$$

holds. These results are an extension of the known connection between the solution of the standard Lyapunov equation ($E = I$) and the asymptotic behavior of the dynamical system $\dot{x}(t) = Ax(t)$, see [16, 20].

Consider a perturbed pencil $\lambda\widetilde{E} - \widetilde{A} = \lambda(E + \Delta E) - (A + \Delta A)$ with $\|\Delta E\|_2 \leq \varepsilon\|E\|_2$ and $\|\Delta A\|_2 \leq \varepsilon\|A\|_2$. Assume that the right and left deflating subspaces of $\lambda E - A$ corresponding to the infinite eigenvalues are not changed under perturbations, i.e.,

$$\ker P_r = \ker \widetilde{P}_r, \qquad \ker P_l = \ker \widetilde{P}_l, \tag{3.24}$$

where $\widetilde{P}_r$ and $\widetilde{P}_l$ are the spectral projections onto the right and left finite deflating subspaces of the pencil $\lambda\widetilde{E} - \widetilde{A}$. In this case $\lambda E - A$ and $\lambda\widetilde{E} - \widetilde{A}$ have the right and left deflating subspaces corresponding to the finite eigenvalues of the same dimension. Such a restriction is motivated by applications, e.g., in the stability analysis for descriptor systems [7]. Moreover, we will assume for such allowable perturbations that we have an error bound $\|\widetilde{P}_r - P_r\|_2 \leq \varepsilon K$ with some constant K (for such estimate for the pencil $\lambda E - A$ of index one, see [42]). This estimate implies that the right deflating subspace of the perturbed pencil $\lambda\widetilde{E} - \widetilde{A}$ corresponding to the finite eigenvalues is close to the corresponding right deflating subspace of $\lambda E - A$.

Consider now the perturbed projected GCALE

$$\widetilde{E}^* \widetilde{X} \widetilde{A} + \widetilde{A}^* \widetilde{X} \widetilde{E} = -\widetilde{P}_r^* \widetilde{G} \widetilde{P}_r, \qquad \widetilde{X} = \widetilde{X}\widetilde{P}_l. \tag{3.25}$$

The following theorem gives a relative error bound for the solution of (1.2).

Theorem 3.7 *Let $\lambda E - A$ be c-stable and let X be a solution of the projected GCALE (1.2). Consider a perturbed pencil $\lambda\widetilde{E} - \widetilde{A} = \lambda(E + \Delta E) - (A + \Delta A)$ with $\|\Delta E\|_2 \leq \varepsilon\|E\|_2$ and $\|\Delta A\|_2 \leq \varepsilon\|A\|_2$. Assume that for the spectral projections $\widetilde{P}_r$ and $\widetilde{P}_l$ onto the right and left deflating subspaces corresponding to the finite eigenvalues of $\lambda\widetilde{E} - \widetilde{A}$, relations (3.24) are satisfied and a bound $\|\widetilde{P}_r - P_r\|_2 \leq \varepsilon K < 1$ holds with some constant K. Let $\widetilde{G}$ be a perturbation of G such that $\|\Delta G\|_2 \leq \varepsilon\|G\|_2$. If $\varepsilon(2+\varepsilon)\kappa_2(E,A) < 1$, then the perturbed projected GCALE (3.25) has a unique solution $\widetilde{X}$ and*

$$\frac{\|\widetilde{X} - X\|_2}{\|X\|_2} \leq \frac{\varepsilon\Big((\varepsilon K + \|P_r\|_2)(K + \|P_r\|_2)\|G\|_2 + 3\|E\|_2\|A\|_2\|X\|_2\Big)\kappa_2(E,A)}{\|E\|_2\|A\|_2\|X\|_2(1 - \varepsilon(2+\varepsilon)\kappa_2(E,A))}. \tag{3.26}$$

Proof. It follows from (3.24) that

$$\widetilde{P}_r P_r = \widetilde{P}_r, \qquad P_r\widetilde{P}_r = P_r, \qquad \widetilde{P}_l P_l = \widetilde{P}_l, \qquad P_l\widetilde{P}_l = P_l. \tag{3.27}$$

The perturbed GCALE in (3.25) can be rewritten as

$$E^*\widetilde{X}A + A^*\widetilde{X}E = -\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})\right),$$

where $\mathcal{K}(\widetilde{X}) = (\Delta E)^*\widetilde{X}\widetilde{A} + E^*\widetilde{X}\Delta A + (\Delta A)^*\widetilde{X}E + \widetilde{A}^*\widetilde{X}\Delta E$. Using (2.1) and (2.2) we can verify that $P_l E = P_l E P_r = E P_r$ and $P_l A = P_l A P_r = A P_r$. Analogous relations hold for the perturbed pencil $\lambda\widetilde{E} - \widetilde{A}$. Then by (3.27) we obtain that $\widetilde{X} = \widetilde{X}P_l = \widetilde{X}P_l\widetilde{P}_l = \widetilde{X}\widetilde{P}_l$ and

$$\begin{aligned}\widetilde{X}E &= \widetilde{X}P_lE = \widetilde{X}EP_r = \widetilde{X}P_lEP_r\widetilde{P}_r = \widetilde{X}E\widetilde{P}_r,\\ \widetilde{X}\widetilde{E} &= \widetilde{X}\widetilde{P}_l\widetilde{E} = \widetilde{X}\widetilde{E}\widetilde{P}_r = \widetilde{X}\widetilde{P}_l\widetilde{E}\widetilde{P}_rP_r = \widetilde{X}\widetilde{E}P_r.\end{aligned}$$

These relationships remain valid if we replace E by A and $\widetilde{E}$ by $\widetilde{A}$. In this case we obtain

$$\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X}) = P_r^*\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})\right)P_r = \widetilde{P}_r^*\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})\right)\widetilde{P}_r. \tag{3.28}$$

Then the perturbed projected GCALE (3.25) is equivalent to the projected GCALE

$$E^*\widetilde{X}A + A^*\widetilde{X}E = -P_r^*\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})\right)P_r, \qquad \widetilde{X} = \widetilde{X}P_l.$$

Since the pencil $\lambda E - A$ is c-stable, this equation has a unique solution given by

$$\widetilde{X} = \frac{1}{2\pi}\int_{-\infty}^{\infty}(i\xi E - A)^{-*}P_r^*\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})\right)P_r(i\xi E - A)^{-1}d\xi. \tag{3.29}$$

Thus, we have an integral equation $\widetilde{X} = \mathcal{I}(\widetilde{X})$ for the unknown matrix $\widetilde{X}$, where

$$\mathcal{I}(\widetilde{X}) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(i\xi E - A)^{-*}P_r^*\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})\right)P_r(i\xi E - A)^{-1}d\xi.$$

From

$$\|\mathcal{K}(\widetilde{X})\|_2 \leq 2(\|\Delta E\|_2\|\widetilde{A}\|_2 + \|\Delta A\|_2\|E\|_2)\|\widetilde{X}\|_2 \leq 2\varepsilon(2+\varepsilon)\|E\|_2\|A\|_2\|\widetilde{X}\|_2$$

we obtain for any matrices X_1 and X_2, that

$$\begin{aligned}\|\mathcal{I}(X_1)-\mathcal{I}(X_2)\|_2 &= \left\|\frac{1}{2\pi}\int_{-\infty}^{\infty}(i\xi E-A)^{-*}P_r^*\mathcal{K}(X_1-X_2)P_r(i\xi E-A)^{-1}\,d\xi\right\|_2\\ &\le \|\mathcal{K}(X_1-X_2)\|_2\|H\|_2 \le \varepsilon(2+\varepsilon)\kappa_2(E,A)\|X_1-X_2\|_2.\end{aligned}$$

Since $\varepsilon(2+\varepsilon)\kappa_2(E,A)<1$, the operator $\mathcal{I}(\widetilde{Z})$ is contractive. Then by the fixed point theorem [28] the equation $\widetilde{X}=\mathcal{I}(\widetilde{X})$ has a unique solution $\widetilde{X}$ and we can estimate the error

$$\begin{aligned}\|\widetilde{X}-X\|_2 &= \left\|\frac{1}{2\pi}\int_{-\infty}^{\infty}(i\xi E-A)^{-*}P_r^*\left(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r+\mathcal{K}(\widetilde{X})-P_r^*GP_r\right)P_r(i\xi E-A)^{-1}\,d\xi\right\|_2\\ &\le \left(\|\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r-P_r^*GP_r\|_2+\|\mathcal{K}(\widetilde{X})\|_2\right)\|H\|_2.\end{aligned}$$

Taking into account that

$$\begin{aligned}\|\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r-P_r^*GP_r\|_2 &\le \|\widetilde{P}_r-P_r\|_2(\|\widetilde{G}\|_2\|\widetilde{P}_r\|_2+\|P_r\|_2\|G\|_2)+\|P_r\|_2\|\widetilde{G}-G\|_2\|\widetilde{P}_r\|_2\\ &\le \varepsilon\left((\varepsilon K+\|P_r\|_2)((1+\varepsilon)K+\|P_r\|_2)+\varepsilon K\|P_r\|_2\right)\|G\|_2\\ &\le 2\varepsilon\left(\varepsilon K+\|P_r\|_2\right)\left(K+\|P_r\|_2\right)\|G\|_2\end{aligned}$$

and $\|\mathcal{K}(\widetilde{X})\|_2\le 2\varepsilon(2+\varepsilon)\|E\|_2\|A\|_2(\|X\|_2+\|\widetilde{X}-X\|_2)$ we obtain the relative perturbation bound (3.26). □

Bound (3.26) shows that if $\kappa_2(E,A)$, K and $\|P_r\|_2$ are not too large, then the solution of the perturbed projected GCALE (3.25) is a small perturbation of the solution of the projected GCALE (1.2).

From Theorem 3.7 we can obtain some useful consequences.

Corollary 3.8 *Under the assumptions of Theorem 3.7 we have that if G is Hermitian, positive definite and if*

$$2\varepsilon\left(2(1+2\varepsilon)(\varepsilon K+\|P_r\|_2)^2+1\right)\kappa_2(E,A)\|G\|_2<\lambda_{\min}(G), \tag{3.30}$$

where $\lambda_{\min}(G)$ is the smallest eigenvalue of the matrix G, then the perturbed pencil $\lambda\widetilde{E}-\widetilde{A}$ is c-stable and the following relative perturbation bound

$$\frac{|\kappa_2(\widetilde{E},\widetilde{A})-\kappa_2(E,A)|}{\kappa_2(E,A)}\le\frac{3\varepsilon\ (K(K+2\|P_r\|_2)+\kappa_2(E,A)+1)}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)} \tag{3.31}$$

holds.

Proof. First we will show that the matrix $\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r+\mathcal{K}(\widetilde{X})$ is positive definite on the subspace $\operatorname{im}P_r$. For all nonzero $z\in\operatorname{im}P_r$, we have

$$\begin{aligned}(\,(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r+\mathcal{K}(\widetilde{X}))z,z\,) &= (\,(\widetilde{P}_r^*(G+\Delta G)\widetilde{P}_r+\widetilde{P}_r^*\mathcal{K}(\widetilde{X})\widetilde{P}_r)z,z\,)\\ &\ge \left(\lambda_{\min}(G)-\|\mathcal{K}(\widetilde{X})\|_2-\|\Delta G\|_2\right)\|\widetilde{P}_rz\|^2.\end{aligned} \tag{3.32}$$

Suppose now that $\widetilde{P}_rz=0$. Then we obtain from (3.27) that $z\in\ker P_r$, but $z\in\operatorname{im}P_r$ and $z\ne0$. Hence, $\widetilde{P}_rz\ne0$. From (3.29) it follows that

$$\|\widetilde{X}\|_2\le\frac{\|\widetilde{P}_r\|_2^2\|\widetilde{G}\|_2\|H\|_2}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)}\le\frac{(1+\varepsilon)(\varepsilon K+\|P_r\|_2)^2\|G\|_2\|H\|_2}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)}. \tag{3.33}$$

Then taking into account estimate (3.30) we get

$$\|\mathcal{K}(\widetilde{X})\|_2 + \|\Delta G\|_2 \leq \frac{\varepsilon\left(2(1+2\varepsilon)(\varepsilon K + \|P_r\|_2)^2 + 1\right)\kappa_2(E,A)\|G\|_2}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)} < \lambda_{\min}(G).$$

Thus, $(\,(\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X}))z, z\,) > 0$ for all nonzero $z \in \operatorname{im} P_r$, i.e., the matrix $\widetilde{P}_r^*\widetilde{G}\widetilde{P}_r + \mathcal{K}(\widetilde{X})$ is positive definite on the subspace $\operatorname{im} P_r$. Consequently, the matrix $\widetilde{X}$ is positive definite on $\operatorname{im} P_l$ and positive semidefinite. Moreover, it follows from (3.30) that the matrix $\widetilde{G}$ is positive definite. We have that the positive semidefinite matrix $\widetilde{X}$ satisfies the projected GCALE (3.25) with positive definite $\widetilde{G}$. In this case the pencil $\lambda\widetilde{E} - \widetilde{A}$ is c-stable, see [43].

From the proof of Theorem 3.7 with $\widetilde{G} = G = I$ we have that

$$\|\widetilde{H} - H\|_2 \leq \frac{\varepsilon\left(K(\varepsilon K + 2\|P_r\|_2) + (2+\varepsilon)\kappa_2(E,A)\right)\|H\|_2}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)},$$

where $\widetilde{H}$ is the solution of the perturbed projected GCALE (3.25) with $\widetilde{G} = I$. Then

$$\begin{aligned} |\kappa_2(\widetilde{E},\widetilde{A}) \;-\; \kappa_2(E,A)| &= 2\left|\,\|\widetilde{E}\|_2\|\widetilde{A}\|_2\|\widetilde{H}\|_2 - \|E\|_2\|A\|_2\|H\|_2\,\right| \\ &\leq 2\left(\|\widetilde{E}\|_2\|\widetilde{A}\|_2\|\widetilde{H} - H\|_2 + \|\widetilde{E} - E\|\|\widetilde{A}\|_2\|H\|_2 + \|E\|_2\|\widetilde{A} - A\|_2\|H\|_2\right) \\ &\leq \frac{3\varepsilon\kappa_2(E,A)\;(K(K + 2\|P_r\|_2) + \kappa_2(E,A) + 1)}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)}. \end{aligned}$$

□

Furthermore, from the proof of Theorem 3.7 for $\widetilde{P}_r = P_r = I$ we obtain the following perturbation bound for the solution of the regular GCALE (1.1).

Corollary 3.9 *Let G be Hermitian and positive definite. Assume that the GCALE (1.1) is regular. Let ΔE, ΔA be perturbations of $\lambda E - A$ such that $\|\Delta E\|_2 \leq \varepsilon\|E\|_2$, $\|\Delta A\|_2 \leq \varepsilon\|A\|_2$ and let ΔG be a perturbation of G with $\|\Delta G\|_2 \leq \varepsilon\|G\|_2$. If $\varepsilon\,(2+\varepsilon)\,\kappa_2(E,A) < 1$, then the perturbed GCALE (3.14) is regular and the relative error bound*

$$\frac{\|\widetilde{X} - X\|_2}{\|X\|_2} \leq \frac{\varepsilon(3+\varepsilon)\kappa_2(E,A)}{1-\varepsilon(2+\varepsilon)\kappa_2(E,A)} \tag{3.34}$$

holds.

Note that bound (3.34) can be also obtained directly by applying the linear operator perturbation theory [29] to the regular GCALE (1.1) in the operator form $\mathcal{L}(X) = -G$.

If $\hat{X}$ is an approximate solution of the GCALE (1.1) and if R is a residual given by (3.19), then from Corollary 3.9 with $\Delta E = 0$, $\Delta A = 0$ and $\Delta G = R$ we obtain the following forward error bound

$$\frac{\|\hat{X} - X\|_2}{\|X\|_2} \leq \kappa_2(E,A)\,\frac{\|R\|_2}{2\|E\|_2\|A\|_2\|X\|_2} =: Est_2. \tag{3.35}$$

Bounds (3.34) and (3.35) show that $\kappa_2(E,A)$ just as $\kappa_F(E,A)$ may also be used to measure the sensitivity of the solution of the regular GCALE (1.1). From the relationship

$$\frac{1}{\sqrt{n}}\|\mathcal{L}^{-1}\|_2 \leq \|\mathcal{L}^{-1}\|_F \leq \sqrt{n}\|\mathcal{L}^{-1}\|_2$$

we obtain that the Frobenius norm based condition number $\kappa_F(E,A)$ does not differ more than a factor $\sqrt{n}$ from the spectral condition number $\kappa_2(E,A)$. Thus, $\kappa_2(E,A)$ may be used as an estimator of $\kappa_F(E,A)$. Note that to compute one-norm or Frobenius norm based estimators of $\kappa_F(E,A)$ we need to solve around five generalized Lyapunov equations $E^*XA+A^*XE=-G$ and $EXA^*+AXE^*=-G$, see [1, 21], whereas the computation of $\kappa_2(E,A)$ requires solving only one additional generalized Lyapunov equation $E^*XA+A^*XE=-I$.

4 Numerical experiments

In this section we present the results of several numerical experiments. Computations were carried out on IBM RS 6000 44P Modell 270 with relative machine precision $\texttt{EPS} \approx 2.22 \cdot 10^{-16}$.

Example 4.1 [39] The matrices E and A are defined as

$$\begin{aligned} E &= I_n + 2^{-t}U_n, \\ A &= (2^{-t}-1)I_n + \mathrm{diag}(1,2,\ldots,n) + U_n^T, \end{aligned}$$

where U_n is the $n \times n$ matrix with unit entries below the diagonal and all other entries zero. Note that E is nonsingular. The matrix G is defined so that a true solution X of (1.1) is the matrix of all ones.

In Figure 1(a) we compare the spectral condition number $\kappa_2(E,A)$ and the Frobenius norm based condition number $\kappa_F(E,A)$. We see that $\kappa_2(E,A)$ is a factor 2-8 smaller than $\kappa_F(E,A)$ and the problem becomes ill-conditioned as the parameter t increases. Figure 1(b) shows the relative errors in the spectral and Frobenius norms

$$\texttt{RERR2} = \frac{\|\hat{X}-X\|_2}{\|X\|_2}, \qquad \texttt{RERRF} = \frac{\|\hat{X}-X\|_F}{\|X\|_F},$$

where $\hat{X}$ is an approximate solution of (1.1) computed by the generalized Bartels-Stewart method. As expected from the perturbation theory, the accuracy of $\hat{X}$ may get worse as the condition numbers are large, while the relative residuals

$$\texttt{RRES2} = \frac{\|E^*\hat{X}A + A^*\hat{X}E + G\|_2}{2\|E\|_2\|A\|_2\|X\|_2} \quad \text{and} \quad \texttt{RRESF} = \frac{\|E^*\hat{X}A + A^*\hat{X}E + G\|_F}{2\|E\|_2\|A\|_2\|X\|_F},$$

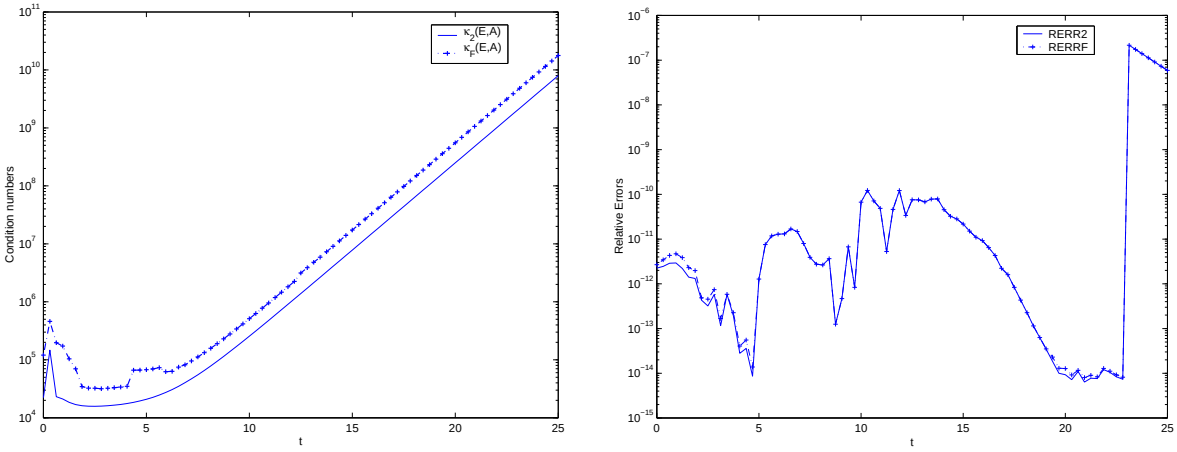
shown in Figure 2(a), remain small.

Figure 2(b) shows the ratios $\texttt{RERR2}/Est_2$ and $\texttt{RERRF}/Est_F$ between the relative errors and the computed residual based error estimates given by (3.20) and (3.35). One can see that the estimate in the spectral norm is sharper than the estimate in the Frobenius norm.

Example 4.2 Consider a family of projected GCALEs with

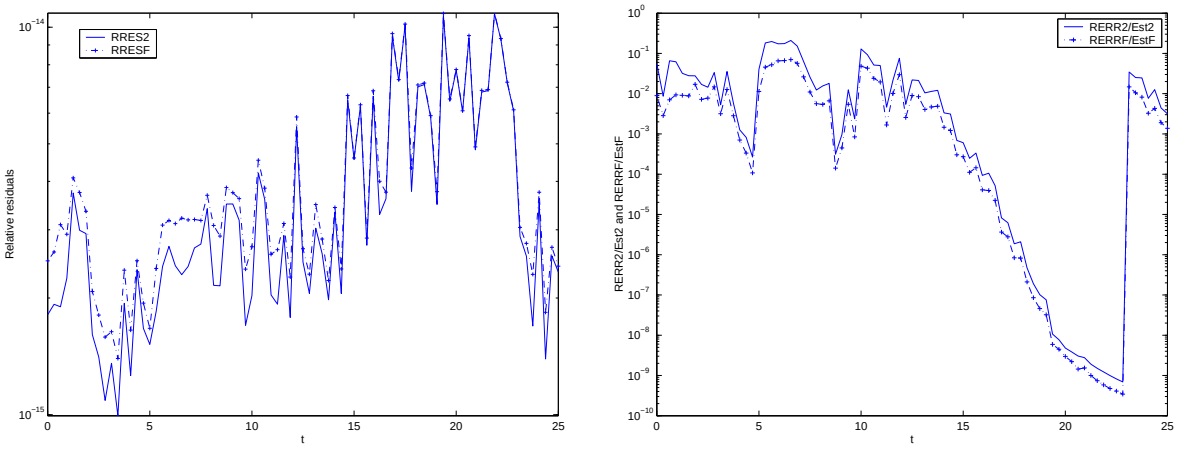
$$E = V \begin{pmatrix} I_3 & D(N_3 - I_3) \\ 0 & N_3 \end{pmatrix} U^T, \qquad A = V \begin{pmatrix} J & (I_3 - J)D \\ 0 & I_3 \end{pmatrix} U^T,$$

$$G = U \begin{pmatrix} G_{11} & -G_{11}D \\ -DG_{11} & DG_{11}D \end{pmatrix} U^T,$$



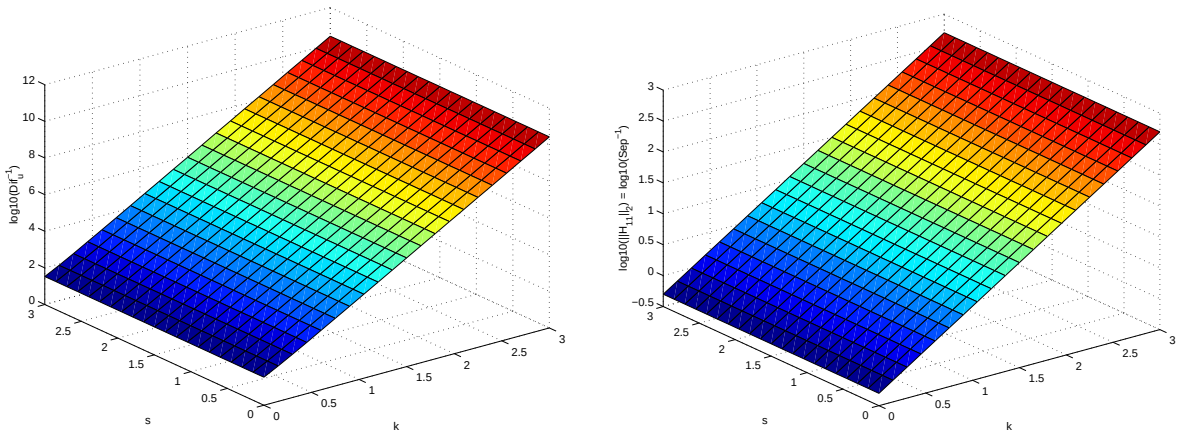
(a) (b)

Figure 1: Spectral norm and Frobenius norm condition numbers (a) and the relative errors in the solution (b)



(a) (b)

Figure 2: Relative residuals (a) and ratios between the relative errors and the error estimates in the spectral and Frobenius norms (b)



(a) (b)

Figure 3: Conditioning of the generalized Sylvester equation (a) and the regular generalized Lyapunov equation (b)

where N_3 is the nilpotent Jordan block of order 3,

$$\begin{aligned} J &= \mathrm{diag}(-10^{-k},\, -2,\, -3\times 10^{k}), && k \geq 0,\\ D &= \mathrm{diag}(10^{-s},\, 1,\, 10^{s}), && s \geq 0,\\ G_{11} &= \mathrm{diag}(2,\, 4,\, 6). \end{aligned}$$

The transformation matrices V and U are elementary reflections chosen as

$$\begin{aligned} V &= I_6 - \frac{1}{3}ee^T, \qquad e = (1,\, 1,\, 1,\, 1,\, 1,\, 1)^T,\\ U &= I_6 - \frac{1}{3}ff^T, \qquad f = (1, -1,\, 1, -1,\, 1, -1)^T. \end{aligned}$$

The exact solution of the projected GCALE (1.1) is given by

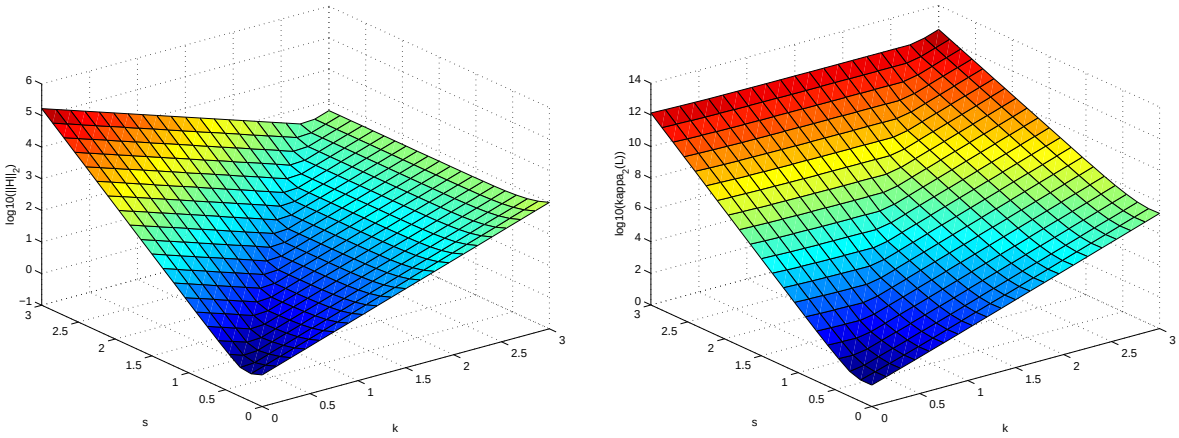
$$X = V \begin{pmatrix} X_{11} & -X_{11}D \\ -DX_{11} & DX_{11}D \end{pmatrix} V^T$$

with $X_{11} = \mathrm{diag}(10^{k},\, 1,\, 10^{-k})$. The problem becomes ill-conditioned when k and s increase.

In Figure 3 we show the values of Dif_u^{-1} and $\|H_{11}\|_2$ as functions of k and s. Here H_{11} is the solution of the regular GCALE $E_f^T H_{11} A_f + A_f^T H_{11} E_f = -I_{n_f}$. Note that in this example $\|H_{11}\|_2 = \mathrm{Sep}^{-1}(E_f, A_f)$. We see that the condition numbers of the generalized Sylvester equation (2.14) and the regular GCALE (2.16) are independent of s and increase for magnifying k.

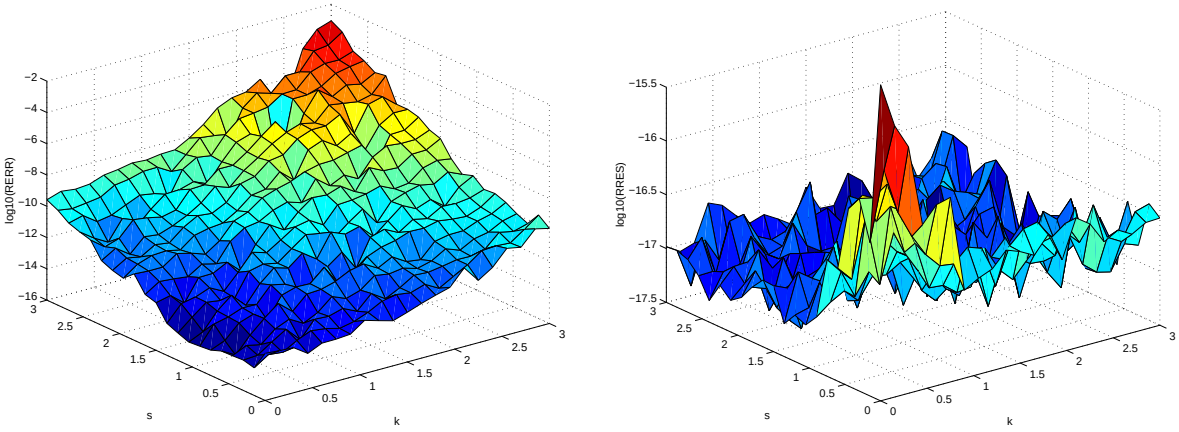
In Figure 4 we show the values of $\|H\|_2$ and the condition number $\kappa_2(E, A)$ of the projected GCALE (1.1) for the same values of k and s. When k and s are increased then the condition number $\kappa_2(E, A)$ increases more quickly than $\|H\|_2$. Finally, Figure 5(a) shows the relative error $\texttt{RERR} = \|\hat{X} - X\|_2/\|X\|_2$, where $\hat{X}$ is the computed solution, and Figure 5(b) shows the relative residual

$$\texttt{RRES} = \frac{\|E^T\hat{X}A + A^T\hat{X}E + \hat{P}_r^T G \hat{P}_r\|_2}{2\|E\|_2\|A\|_2\|X\|_2},$$



(a) (b)

Figure 4: Conditioning of the projected generalized Lyapunov equation



(a) (b)

Figure 5: Relative error (a) and relative residual (a).

where $\hat{P}_r$ is the computed projection onto the right deflating subspace of the pencil $\lambda E - A$ corresponding to the finite eigenvalues. We see that the relative residual is small even for the ill-conditioned problem. However, this does not imply that the relative error in the computed solution remains close to zero when the condition number $\kappa_2(E, A)$ is large. The relative error in $\hat{X}$ increases as $\kappa_2(E, A)$ grows. Moreover, the computed solution may be inaccurate, if one of intermediate problems is ill-conditioned.

Acknowledgement: The author would like to thank V. Mehrmann for helpful discussions and also B. Kågström for providing the `GUPTRI` routine written by J.W. Demmel and himself.

References

[1] E. Anderson, Z. Bai, C.H. Bischof, S. Blackford, J.M. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide,*

Third Edition. SIAM, Philadelphia, PA, 1999.

[2] R.H. Bartels and G.W. Stewart. Solution of the equation $AX + XB = C$. *Comm. ACM*, 15(9):820–826, 1972.

[3] T. Beelen and P. Van Dooren. An improved algorithm for the computation of Kronecker's canonical form of a singular pencil. *Linear Algebra Appl.*, 105:9–65, 1988.

[4] D.J. Bender. Lyapunov-like equations and reachability/observability Gramians for descriptor systems. *IEEE Trans. Automat. Control*, 32(4):343–348, 1987.

[5] P. Benner, V. Mehrmann, V. Sima, S. Van Huffel, and A. Varga. SLICOT - A subroutine library in systems and control theory. *Appl. Comput. Control Signals Circuits*, 1:499–539, 1999.

[6] P. Benner and E.S. Quintana-Ortí. Solving stable generalized Lyapunov equations with the matrix sign function. *Numerical Algorithms*, 20(1):75–100, 1999.

[7] R. Byers and N.K. Nichols. On the stability radius of a generalized state-space system. *Linear Algebra Appl.*, 188/189:113–134, 1993.

[8] S.L. Campbell and C.D. Meyer. *Generalized Inverses of Linear Transformations*. Dover Publications, New York, 1979.

[9] K.E. Chu. The solution of the matrix equations $AXB-CXD = E$ and $(YA-DZ, YC-BZ) = (E, F)$. *Linear Algebra Appl.*, 93:93–105, 1987.

[10] J.W. Demmel and B. Kågström. Computing stable eigendecompositions of matrix pencils. *Linear Algebra Appl.*, 88/89:139–186, 1987.

[11] J.W. Demmel and B. Kågström. The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: Robust software with error bounds and applications. Part I: Theory and algorithms. *ACM Trans. Math. Software*, 19(2):160–174, 1993.

[12] J.W. Demmel and B. Kågström. The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: Robust software with error bounds and applications. Part II: Software and applications. *ACM Trans. Math. Software*, 19(2):175–201, 1993.

[13] F.R. Gantmacher. *Theory of Matrices*. Chelsea, New York, 1959.

[14] 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(2):223–231, 1992.

[15] J.D. Gardiner, M.R.Wette, A.J. Laub, J.J. Amato, and C.B. Moler. Algorithm 705: A Fortran-77 software package for solving the Sylvester matrix equation $AXB^T+CXD^T = E$. *ACM Trans. Math. Software*, 18(2):232–238, 1992.

[16] S.K. Godunov. *Ordinary Differential Equations with Constant Coefficients*. Translations of Mathematical Monographs, 169. American Mathematical Society, Providence, RI, 1997.

[17] G.H. Golub and C.F. Van Loan. *Matrix Computations. 3rd ed.* The Johns Hopkins University Press, Baltimore, London, 1996.

[18] G.H. Golub, S. Nash, and C. Van Loan. A Hessenberg-Schur method for the problem $AX + XB = C$. *IEEE Trans. Automat. Control*, AC-24:909–913, 1979.

[19] S.J. Hammarling. Numerical solution of the stable non-negative definite Lyapunov equation. *IMA J. Numer. Anal.*, 2:303–323, 1982.

[20] G. Hewer and C. Kenney. The sensitivity of the stable Lyapunov equation. *SIAM J. Cont. Optim.*, 26(2):321–344, 1988.

[21] N.J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, PA, 1996.

[22] I. Jonsson and B. Kågström. Recursive bloked algorithms for solving triangular matrix equations – Part I: One-sided and coupled Sylvester-type equations. SLICOT Working Note 2001-4, 2001. Available from `ftp://wgs.esat.kuleuven.ac.be/pub/WGS/REPORTS/SLWN2001-4.ps.Z` Submitted to *ACM Trans. Math. Software*.

[23] I. Jonsson and B. Kågström. Recursive bloked algorithms for solving triangular matrix equations – Part II: Two-sided and generalized Sylvester and Lyapunov equations. SLICOT Working Note 2001-5, 2001. Available from `ftp://wgs.esat.kuleuven.ac.be/pub/WGS/REPORTS/SLWN2001-5.ps.Z` Submitted to *ACM Trans. Math. Software*.

[24] B. Kågström. A perturbation analysis of the generalized Sylvester equation $(AR - LB, DR - LE) = (C, F)$. *SIAM J. Matrix Anal. Appl.*, 15(4):1045–1060, 1994.

[25] B. Kågström and P. Poromaa. Computing eigenspaces with specified eigenvalues of a regular matrix pencil (A, B) and condition estimation: Theory, algorithms and software. *Numerical Algorithms*, 12:369–407, 1996.

[26] B. Kågström and P. Poromaa. LAPACK-Style algorithms and software for solving the generalized Sylvester equation and estimating the separation between regular matrix pairs. *ACM Trans. Math. Software*, 22(1):78–103, 1996.

[27] B. Kågström and L. Westin. Generalized Schur methods with condition estimators for solving the generalized Sylvester equation. *IEEE Trans. Automat. Control*, 34:745–751, 1989.

[28] L.V. Kantorovich and G.P. Akilov. *Functional Analysis*. Pergamon Press, Oxford, 1982.

[29] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, 1966.

[30] C.S. Kenney and A.J. Laub. The matrix sign function. *IEEE Trans. Automat. Control*, 40(8):1330–1348, 1995.

[31] M.M. Konstantinov, V. Mehrmann, and P. Petkov. On properties of Sylvester and Lyapunov operators. *Linear Algebra Appl.*, 312:35–71, 2000.

[32] M.M. Konstantinov, P.Hr. Petkov, D.W. Gu, and V. Mehrmann. Sensitivity of general Lyapunov equations. Technical Report 98-15, Depart. of Engineering, Leicester University, Leicester LE1 7RH, UK, 1998.

[33] P. Lancaster and L. Rodman. *The Algebraic Riccati Equation*. Oxford University Press, Oxford, 1995.

[34] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, Orlando, FL, 2nd edition, 1985.

[35] V.B. Larin and F.A. Aliev. Generalized Lyapunov equation and factorization of matrix polynomials. *Systems Control Lett.*, 21(6):485–491, 1993.

[36] A.J. Laub, M.T. Heath, C.C. Paige, and R.C. Ward. Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms. *IEEE Trans. Automat. Control*, AC-32(2):115–122, 1987.

[37] V. Mehrmann. *The Autonomous Linear Quadratic Control Problem, Theory and Numerical Solution*. Lecture Notes in Control and Information Sciences, 163. Springer-Verlag, Heidelberg, 1991.

[38] P.C. Müller. Stability of linear mechanical systems with holonomic constraints. *Appl. Mech. Rev.*, 46(11):160–164, 1993.

[39] T. Penzl. Numerical solution of generalized Lyapunov equations. *Adv. Comput. Math.*, 8(1-2):33–48, 1998.

[40] G.W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.*, 15:727–764, 1973.

[41] G.W. Stewart and J.-G. Sun. *Matrix Perturbation Theory*. Academic Press, New York, 1990.

[42] T. Stykel. On criteria for asymptotic stability of differential-algebraic equations. *Z. Angew. Math. Mech.*, 82(3):147–158, 2002.

[43] T. Stykel. Generalized Lyapunov equations for descriptor systems: stability and inertia theorems. Preprint SFB393/00-38, Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany, October 2000. Available from `http://www.tu-chemnitz.de/sfb393/sfb00pr.html`.

[44] P. Van Dooren. The computation of the Kronecker's canonical form of a singular pencil. *Linear Algebra Appl.*, 27:103–140, 1979.

[45] D.S. Watkins. Performance of the QZ algorithm in the presence of infinite eigenvalues. *SIAM J. Matrix Anal. Appl.*, 22(2):364–375, 2000.