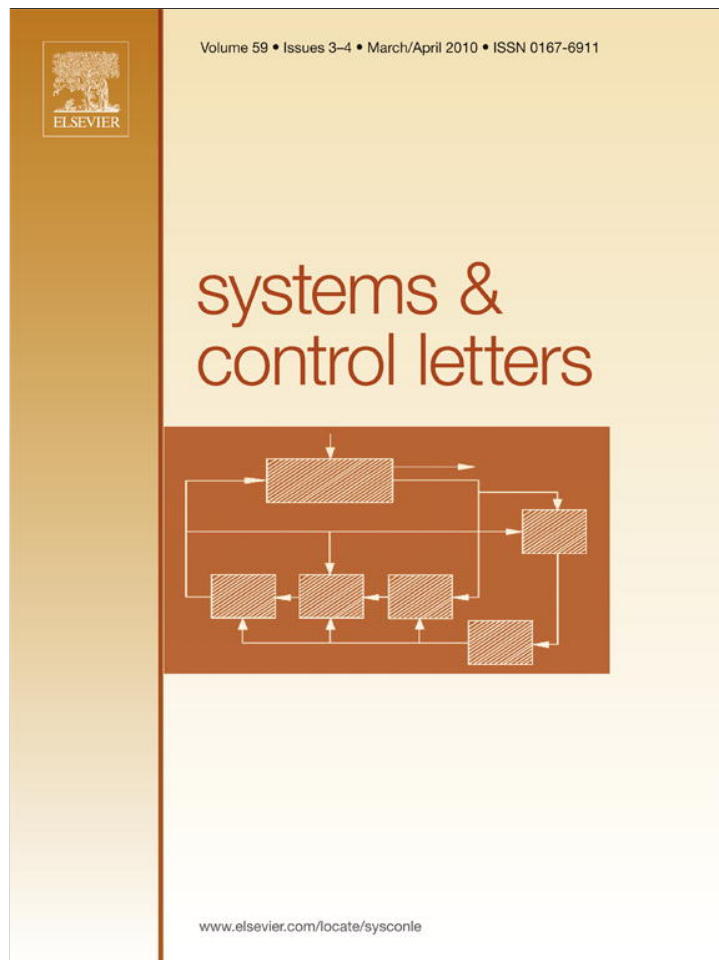


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

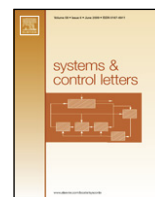
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Analysis of the solution of the Sylvester equation using low-rank ADI with exact shifts

Ninoslav Truhar^{a,*}, Zoran Tomljanović^a, Ren-Cang Li^b

^a Department of Mathematics, University J.J. Strossmayer, Trg Ljudevita Gaja 6, 31000 Osijek, Croatia

^b Department of Mathematics, University of Texas at Arlington, 411 South Nedderman Drive, Arlington, TX 76019, USA

ARTICLE INFO

Article history:

Received 10 June 2008

Received in revised form

2 December 2009

Accepted 10 February 2010

Available online 11 March 2010

Keywords:

Sylvester equation

Low-rank Alternating-Directional-Implicit (LR-ADI) method

Upper bounds

Perturbation bounds

ABSTRACT

The solution to a general Sylvester equation $AX - XB = GF^*$ with a low-rank right-hand side is analyzed quantitatively through the Low-rank Alternating-Directional-Implicit method (LR-ADI) with exact shifts. New bounds and perturbation bounds on X are obtained. A distinguished feature of these bounds is that they reflect the interplay between the eigenvalue decompositions of A and B and the right-hand side factors G and F . Numerical examples suggest that because of this inclusion of details, new perturbation bounds are much sharper than the existing ones.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In this paper we consider the properties of the solution of the $m \times n$ Sylvester equation

$$AX - XB = C, \quad (1.1)$$

where A , B , and C are $m \times m$, $n \times n$, and $m \times n$, respectively, and an unknown matrix X is $m \times n$. Eq. (1.1) has a unique solution if and only if A and B have no common eigenvalues, which will be assumed throughout this paper.

Sylvester equations appear frequently in many areas of applied mathematics. We refer the reader to the elegant survey by Bhatia and Rosenthal [1] and the references therein for a history of the equation and many interesting and important theoretical results. Sylvester equations are important in a number of applications such as matrix eigen-decompositions [2,3], control theory [4,5,3], model reduction [6–9], numerical solution of matrix differential Riccati equations [10–12], and many more.

There are several numerical algorithms for calculating the solution of Sylvester equations. The standard ones are the Bartels–Stewart algorithm [13] and the Hessenberg–Schur method first described by Enright [12], but more often attributed to Golub,

Nash, and Van Loan [14]. Another computationally efficient approach for the case that both A and $-B$ are (Hurwitz) stable, i.e., have all their eigenvalues in the open left half plane, is the sign function method [15], or the Smith method [16]. All these methods are efficient for dense matrices A and B .

However, recent interest is directed more towards large and sparse matrices A and B , and $C = GF^*$ with very low rank, where G and F have only a few columns. For dense A and B , an approach based on the sign function method is suggested in [17] that exploits the low-rank structure of C . This approach is further used in [18] in order to solve large-scale Sylvester equations with data-sparse A and B , i.e., dense matrices A and B that can be represented by $\mathcal{O}(n \log(n))$ data.

On the other hand, the problem for the sensitivity of the solution of the Sylvester equation is also a widely studied one. There are several books which contain results on this, e.g., [19–21].

Our main concern is how the solution will behave when C is a low-rank matrix. Our investigation is through applying Low-rank Alternating-Directional-Implicit method (LR-ADI) introduced in [22] (and in [23] with more details) with exact shifts, i.e., all or part of the eigenvalues of A and B . This method is a proper extension of the Cholesky factor ADI for Lyapunov equations in [24–27].

Our results show that the low-rank right-hand side of the Sylvester equation can sometimes greatly influence the norm of X and how it changes in the face of perturbations. The bound on the norm of X can be considered as a proper generalization of the results from [28].

* Corresponding author.

E-mail addresses: ntruhar@mathos.hr (N. Truhar), ztomljan@mathos.hr (Z. Tomljanović), rcli@uta.edu (R.-C. Li).

Notation. Throughout this paper, $\mathbb{C}^{n \times m}$ is the set of all $n \times m$ complex matrices, $\mathbb{C}^n = \mathbb{C}^{n \times 1}$, and $\mathbb{C} = \mathbb{C}^1$. Similarly define $\mathbb{R}^{n \times m}$, \mathbb{R}^n , and \mathbb{R} except replacing the word *complex* by *real*. I_n (or simply I if its dimension is clear from the context) is the $n \times n$ identity matrix, and e_j is its j th column. The superscript “ \cdot^* ” takes conjugate transpose while “ \cdot^T ” takes transpose only. We shall also adopt MATLAB-like convention to access the entries of vectors and matrices. $i : j$ is the set of integers from i to j inclusive and $i : i = \{i\}$. For a vector u and a matrix X , $u_{(j)}$ is u 's j th entry, $X_{(i,j)}$ is X 's (i, j) th entry; X 's submatrices $X_{(k:\ell, i:j)}$, $X_{(k:\ell, :)}$, and $X_{(:, i:j)}$ consist of intersections of row k to row ℓ and column i to column j , row k to row ℓ , and column i to column j , respectively. $\|\cdot\|$ and $\|\cdot\|_F$ stands for the spectral norm and the Frobenius norm of a matrix, respectively. $\kappa(A) = \|A\| \|A^{-1}\|$ is the spectral condition number of A .

2. Low-rank ADI for Sylvester equation

Our first result is a generalization of the main results from [28], where some new estimates for the eigenvalue decay rate of the Lyapunov equation $AX + XA^T = C$ with a low-rank right-hand side C has been derived. Our main result will be a bound on the norm of the solution of the following $m \times n$ Sylvester equation

$$AX - XB = GF^*, \quad (2.1)$$

where A, B, G and F are $m \times m, n \times n, m \times r$ and $n \times r$, respectively, and unknown matrix X is $m \times n$. It is assumed $r \ll \min\{m, n\}$.

But before we continue, we will briefly describe the Low-Rank Alternating-Directional-Implicit (LR-ADI) method for solving Sylvester equation (2.1) (more details can be found in [23] or [22]).

Given two sets of parameters $\{\alpha_i\}$ and $\{\beta_i\}$, the ADI iteration for iteratively solving (2.1) goes as follows: For $i = 0, 1, \dots$,

- (1) solve $(A - \beta_i I)X_{i+1/2} = X_i(B - \beta_i I) + C$ for $X_{i+1/2}$;
- (2) solve $X_{i+1}(B - \alpha_i I) = (A - \alpha_i I)X_{i+1/2} - C$ for X_{i+1} .

for given initial guess X_0 which is assumed to be 0 in this paper. A straightforward implementation for ADI can be given based on this.

Note that parameters $\{\alpha_i\}$ and $\{\beta_i\}$ in (LR-ADI) method for solving Sylvester equation (2.1) should be chosen such that $\alpha_i \neq \beta_i$, for all i . In the case of Lyapunov equation $AX + XA^T = C$ with the stable matrix A parameters $\{\alpha_i\}$ should be chosen such that $Re(\alpha_i) < 0$ for all i and $\beta_i = -\alpha_i$, where $Re(z)$ means real part of z .

Expressing X_{i+1} in terms of X_i , we have¹

$$X_{i+1} = (\beta_i - \alpha_i)(A - \beta_i I)^{-1}C(B - \alpha_i I)^{-1} + (A - \alpha_i I)(A - \beta_i I)^{-1}X_i(B - \beta_i I)(B - \alpha_i I)^{-1},$$

and the error equation

$$\begin{aligned} X_{i+1} - X &= (A - \alpha_i I)(A - \beta_i I)^{-1}(X_i - X)(B - \beta_i I)(B - \alpha_i I)^{-1}, \\ &= \left[\prod_{j=0}^i (A - \alpha_j I)(A - \beta_j I)^{-1} \right] (X_0 - X) \\ &\quad \times \left[\prod_{j=0}^i (B - \beta_j I)(B - \alpha_j I)^{-1} \right], \end{aligned} \quad (2.2)$$

where X denotes the exact solution. If convergence occurs much earlier in the sense that it takes much fewer than $\min\{m, n\}/r$

steps, then ADI in the factored form as below would be more economical. Let $X_i = Z_i D_i Y_i^*$. We have

$$\begin{aligned} X_{i+1} &= ((A - \beta_i I)^{-1}G \quad (A - \alpha_i I)(A - \beta_i I)^{-1}Z_i) \\ &\quad \times \begin{pmatrix} (\beta_i - \alpha_i)I & \\ & D_i \end{pmatrix} \times \begin{pmatrix} F^*(B - \alpha_i I)^{-1} \\ Y_i^*(B - \beta_i I)(B - \alpha_i I)^{-1} \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} Z_{i+1} &= ((A - \beta_i I)^{-1}G \quad (A - \alpha_i I)(A - \beta_i I)^{-1}Z_i), \\ D_{i+1} &= \begin{pmatrix} (\beta_i - \alpha_i)I & \\ & D_i \end{pmatrix}, \\ Y_{i+1}^* &= \begin{pmatrix} F^*(B - \alpha_i I)^{-1} \\ Y_i^*(B - \beta_i I)(B - \alpha_i I)^{-1} \end{pmatrix}. \end{aligned}$$

After renaming the parameters $\{\alpha_i\}$ and $\{\beta_i\}$ as in [23] or [22], and using the fact that $Z_0 = 0$ and $Y_0 = 0$, we can write

$$\begin{aligned} Z_k &= (Z^{(1)} \quad Z^{(2)} \quad \dots \quad Z^{(k)}), \\ \text{with } \begin{cases} Z^{(1)} = (A - \beta_1 I)^{-1}G, \\ Z^{(i+1)} = (A - \alpha_i I)(A - \beta_{i+1} I)^{-1}Z^{(i)} \\ \quad = Z^{(i)} + (\beta_{i+1} - \alpha_i)(A - \beta_{i+1} I)^{-1}Z^{(i)}, \end{cases} \end{aligned} \quad (2.3)$$

$$\text{and } Y_k = (Y^{(1)} \quad Y^{(2)} \quad \dots \quad Y^{(k)}),$$

$$\text{with } \begin{cases} Y^{(1)*} = F^*(B - \alpha_1 I)^{-1}, \\ Y^{(i+1)*} = Y^{(i)*}(B - \alpha_{i+1} I)^{-1}(B - \beta_i I) \\ \quad = Y^{(i)*} + (\alpha_{i+1} - \beta_i)Y^{(i)*}(B - \alpha_{i+1} I)^{-1}. \end{cases} \quad (2.4)$$

Combining these expressions yields

$$X_k = Z_k D_k Y_k^*, \quad D_k = \text{diag}((\beta_1 - \alpha_1)I, \dots, (\beta_k - \alpha_k)I),$$

or

$$X_k = \sum_{j=1}^k (\beta_j - \alpha_j) Z^{(j)} Y^{(j)*}. \quad (2.5)$$

Formulas (2.3)–(2.5) yields a new method for solving Sylvester equation (2.1) (see [23] or [22]) which is a natural extension of CF-ADI [24] and LR-ADI [25,26] for stable Lyapunov equations. Note that in the case of the low-rank matrices on the right-hand side of Sylvester equation (2.1), one can usually expect that the approximate solution (2.5) has a small rank too.

3. The diagonalizable case

In this section, we will present an upper bound on the norm of the solution of Sylvester equation (2.1) for diagonalizable matrices A and B , an error bound for the k th LR-ADI solution with exact shifts, and a first order perturbation bound when A, B, G , and F are perturbed slightly.

Suppose A and B are diagonalizable, i.e.,

$$A = SAS^{-1}, \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m), \quad (3.1)$$

$$B = T\Omega T^{-1}, \quad \Omega = \text{diag}(\mu_1, \dots, \mu_n). \quad (3.2)$$

Then the $(i + 1)$ st block of Z_k and Y_k^* can be written as

$$Z^{(i+1)} = S(A - \alpha_i I)(A - \beta_{i+1} I)^{-1}S^{-1}Z^{(i)}, \quad (3.3)$$

$$Y^{(i+1)*} = Y^{(i)*}T(\Omega - \alpha_{i+1} I)^{-1}(\Omega - \beta_i I)T^{-1}. \quad (3.4)$$

Define

$$\begin{aligned} \Delta_i &= (A - \alpha_i I)(A - \beta_i I)^{-1}, \quad (\Delta_i)_{(j,j)} = \frac{\lambda_j - \alpha_i}{\lambda_j - \beta_i}, \\ \Theta_i &= (\Omega - \beta_i I)(\Omega - \alpha_i I)^{-1}, \quad (\Theta_i)_{(j,j)} = \frac{\mu_j - \beta_i}{\mu_j - \alpha_i}. \end{aligned}$$

¹ It can be also thought of simply as an iterative method obtained from the identity

$$X = (\beta - \alpha)(A - \beta I)^{-1}C(B - \alpha I)^{-1} + (A - \beta I)^{-1}(A - \alpha I)X(B - \beta I)(B - \alpha I)^{-1}.$$

Eqs. (3.3) and (3.4) imply

$$\begin{aligned} Z^{(i+1)} &= S(\Lambda - \alpha_i I)(\Lambda - \beta_{i+1} I)^{-1}(\Lambda - \alpha_{i-1} I)(\Lambda - \beta_i I)^{-1} \cdots \\ &\quad \times (\Lambda - \alpha_1 I)(\Lambda - \beta_2 I)^{-1}(\Lambda - \beta_1 I)^{-1} S^{-1} G \\ &= S(\Lambda - \beta_{i+1} I)^{-1} \Delta_i \Delta_{i-1} \cdots \Delta_1 S^{-1} G, \\ Y^{(i+1)*} &= F^* T (\Omega - \alpha_1 I)^{-1} (\Omega - \alpha_2 I)^{-1} (\Omega - \beta_1 I) \cdots \\ &\quad \times (\Omega - \alpha_{i+1} I)^{-1} (\Omega - \beta_i I) T^{-1} \\ &= F^* T \Theta_1 \Theta_2 \cdots \Theta_i (\Omega - \alpha_{i+1} I)^{-1} T^{-1}. \end{aligned}$$

Notice $(\Lambda - \beta_{i+1} I)^{-1} \Delta_i \Delta_{i-1} \cdots \Delta_1$ is diagonal with its j th diagonal entry

$$((\Lambda - \beta_{i+1} I)^{-1} \Delta_i \Delta_{i-1} \cdots \Delta_1)_{(j,j)} = \prod_{\ell=1}^i \frac{\lambda_j - \alpha_\ell}{\lambda_j - \beta_\ell} \cdot \frac{1}{\lambda_j - \beta_{i+1}},$$

and similarly $\Theta_1 \Theta_2 \cdots \Theta_i (\Omega - \alpha_{i+1} I)^{-1}$ is also diagonal with its j th diagonal entry

$$(\Theta_1 \Theta_2 \cdots \Theta_i (\Omega - \alpha_{i+1} I)^{-1})_{(j,j)} = \prod_{\ell=1}^i \frac{\mu_j - \beta_\ell}{\mu_j - \alpha_\ell} \cdot \frac{1}{\mu_j - \alpha_{i+1}}.$$

Our first theorem gives an upper bound on the norm of the solution of Sylvester equation (2.1). But before we state the first theorem, we note that (2.2) implies that if $\{\alpha_j\}_{j=0}^i$ contains all of A 's eigenvalues (multiple eigenvalues counted as many times as their algebraic multiplicities) or if $\{\beta_j\}_{j=0}^i$ contains all of B 's eigenvalues, then $X_{i+1} - X \equiv 0$. This is because, by the Cayley–Hamilton theorem, $p(A) \equiv 0$ for A 's characteristic polynomial $p(\lambda) \stackrel{\text{def}}{=} \det(\lambda I - A)$ and $q(B) \equiv 0$ for $q(\lambda) \stackrel{\text{def}}{=} \det(\lambda I - B)$. Choose parameters such that

either $\alpha_i = \lambda_{p_i}$ for $i = 1, \dots, m$, or $\beta_j = \mu_{q_j}$ for $j = 1, \dots, n$, where $\{p_i\}$ and $\{q_j\}$ denote some permutations of indices $1, \dots, m$ and $1, \dots, n$, respectively. Then the solution X of Sylvester equation (2.1) can be written as

$$\begin{aligned} X \equiv X_{n_0} &= S \left(\sum_{j=1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) \Phi^{(j)} \Psi^{(j)} \right) T^{-1}, \\ n_0 &= \min\{m, n\}, \end{aligned} \quad (3.5)$$

where

$$\Phi^{(j)} = \text{diag} \left(\frac{\sigma(1, j-1)}{\lambda_1 - \mu_{q_j}}, \dots, \frac{\sigma(m, j-1)}{\lambda_m - \mu_{q_j}} \right) S^{-1} G, \quad (3.6)$$

$$\sigma(i, j-1) = \prod_{s=1}^{j-1} \frac{\lambda_i - \lambda_{p_s}}{\lambda_i - \mu_{q_s}}, \quad \sigma(i, 0) = 1, \quad i = 1, \dots, m, \quad (3.7)$$

and

$$\Psi^{(j)} = F^* T \text{diag} \left(\frac{\tau(1, j-1)}{\mu_1 - \lambda_{p_j}}, \dots, \frac{\tau(n, j-1)}{\mu_n - \lambda_{p_j}} \right), \quad (3.8)$$

$$\tau(i, j-1) = \prod_{s=1}^{j-1} \frac{\mu_i - \mu_{q_s}}{\mu_i - \lambda_{p_s}}, \quad \tau(i, 0) = 1, \quad i = 1, \dots, n. \quad (3.9)$$

Before we state our first result, we like to emphasize that Eq. (3.5) is a proper generalization of the similar equation for the solution of the Lyapunov equation from [28]. Eq. (3.5) is obtained as a by-product of the ADI method from [22] or [23] for Sylvester equations. Another commonly used expression for X , given eigen-decompositions (3.1) and (3.2), is

$$X = S[W \circ (S^{-1} G F^* T)] T^{-1}, \quad W = \left(\frac{1}{\lambda_i - \mu_j} \right) \in \mathbb{C}^{m \times n}, \quad (3.10)$$

where \circ is the entry-wise product of two matrices. It can be verified that the expressions for X in [9, Theorem 3.1] are essentially the same as (3.10). A simple consequence of (3.5) and (3.10) is

$$W \circ (S^{-1} G F^* T) = \sum_{j=1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) \Phi^{(j)} \Psi^{(j)}$$

which is not at all obvious. The above equality also implies another nontrivial equality:

$$\sum_{j=1}^{n_0} (\mu_j - \lambda_j) \frac{\sigma(i, j-1)}{\lambda_i - \mu_j} \frac{\tau(s, j-1)}{\mu_s - \lambda_{p_j}} = \frac{1}{\lambda_i - \mu_s}.$$

While our formula (3.5) is more complicated in form than the usual (3.10), it does allow us to separate $S^{-1} G$ and $F^* T$ as in Theorem 3.1 to illustrate the different effect (caused by the eigenvalue distributions of A and B) of their contribution to the norm of X . Such a separation appears hard to derive from (3.10). See also Remark 3.1 and Example 4.2 for a short discussion.

Theorem 3.1. Assume A and B have eigen-decompositions (3.1) and (3.2). Let X be the solution of the Sylvester equation (2.1). Then

$$\begin{aligned} \|X\| &\leq \|S\| \|T^{-1}\| \sum_{j=1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \sum_{i=1}^m \frac{|\sigma(i, j-1)| \cdot \|\hat{g}_i\|}{|\lambda_i - \mu_{q_j}|} \\ &\quad \times \sum_{\ell=1}^n \frac{|\tau(\ell, j-1)| \cdot \|\hat{f}_\ell\|}{|\mu_\ell - \lambda_{p_j}|}, \end{aligned} \quad (3.11)$$

where $\sigma(\cdot, \cdot)$ and $\tau(\cdot, \cdot)$ are defined as in (3.7) and (3.9), respectively, and \hat{g}_i, \hat{f}_ℓ denote the i th row of $\hat{G} = S^{-1} G$ and the ℓ th column of $\hat{F}^* = F^* T$, respectively.

Proof. Inequality (3.11) is a consequence of taking the norms at both sides of (3.5). \square

Remark 3.1. Note that in the bound from Theorem 3.1 we left each of G and F^* together with the eigenvectors matrices of A and B . This approach has two advantages. First, if certain eigenvectors of S are nearly linearly dependent, then it is straightforward to see that for some G it can be $\|S^{-1} G\| \ll \|S^{-1}\| \|G\|$. On the other hand if for some i , numbers $\frac{\sigma(i, j-1)}{\lambda_i - \mu_{q_j}}$ are large and the corresponding $\|\hat{g}_i\|$ are small, then again it may lead to

$$\sum_{i=1}^m \frac{|\sigma(i, j-1)| \cdot \|\hat{g}_i\|}{|\lambda_i - \mu_{q_j}|} \ll \max_i \frac{|\sigma(i, j-1)|}{|\lambda_i - \mu_{q_j}|} \cdot \|S^{-1} G\|.$$

3.1. Error bounds

In this section we will present an error bound for on approximate solution of Sylvester equation (2.1) by (2.3)–(2.5).

Theorem 3.2. Assume A and B have eigen-decompositions (3.1) and (3.2). Let X_k be the k th approximation obtained by (2.3)–(2.5) with the set of ADI parameters corresponding to any subset of exact eigenvalues of the matrix A and B , i.e. $\{\alpha_1, \alpha_2, \dots, \alpha_k\} = \{\lambda_{p_1}, \lambda_{p_2}, \dots, \lambda_{p_k}\}$ and $\{\beta_1, \beta_2, \dots, \beta_k\} = \{\mu_{q_1}, \mu_{q_2}, \dots, \mu_{q_k}\}$. Then

$$\begin{aligned} \|X - X_k\| &\leq \|S\| \|T^{-1}\| \sum_{j=k+1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \sum_{i=1}^m \frac{|\sigma(i, j-1)| \cdot \|\hat{g}_i\|}{|\lambda_i - \mu_{q_j}|} \\ &\quad \times \sum_{\ell=1}^n \frac{|\tau(\ell, j-1)| \cdot \|\hat{f}_\ell\|}{|\mu_\ell - \lambda_{p_j}|}, \end{aligned} \quad (3.12)$$

where $\sigma(\cdot, \cdot)$ and $\tau(\cdot, \cdot)$ are defined as in (3.6) and (3.8), and \hat{g}_i and \hat{f}_ℓ are as in Theorem 3.1, and $\{\lambda_{p_j}, j > k\}$ and $\{\mu_{q_j}, j > k\}$ are the eigenvalue subsets of A and B complement to the ones already used as shifts for obtaining X_k .

Proof. It follows from (2.5) that

$$X_k = \sum_{j=1}^k (\mu_{q_j} - \lambda_{p_j}) Z^{(j)} Y^{(j)*}.$$

For $k = n_0 \stackrel{\text{def}}{=} \min\{m, n\}$, we have

$$X \equiv X_{n_0} = \sum_{j=1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) Z^{(j)} Y^{(j)*}.$$

Therefore

$$\begin{aligned} X - X_k &= \sum_{j=k+1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) Z^{(j)} Y^{(j)*} \\ &= S \left(\sum_{j=k+1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) \Phi^{(j)} \Psi^{(j)} \right) T^{-1}. \end{aligned} \quad (3.13)$$

Inequality (3.12) is a consequence of taking the norms at both sides of the above equality. \square

Remark 3.2. The bound (3.12) can be used as an upper bound for the decay of singular values of the solution X . One can find similar bounds in [29,30,28]. Especially, in the case of Lyapunov equations, it holds that

$$\|X - X_k\| \leq \text{trace}(X) - \text{trace}(X_k);$$

thus (3.12) may be thought of as the generalization of [28, Theorem 3.1].

Example 3.1. The following example illustrates the quality of the bound (3.11). For simplicity, we will use some of Matlab notation: $\text{ones}(m, n) \in \mathbb{R}^{m \times n}$ has all entry 1 and $\text{ones}(n) \equiv \text{ones}(n, n)$, and $\text{zeros}(m, n) \in \mathbb{R}^{m \times n}$ has all entry 0. Let

$$\begin{aligned} A &= S \Lambda S^{-1}, \quad \Lambda = 10 \text{diag}(1, 2, \dots, n), \\ S &= 100I_n + 0.1 \text{ones}(n), \\ B &= T \Omega T^{-1}, \quad \Omega = \Lambda + \text{diag}(\text{ones}(1, n)), \\ T &= 200I_n + 0.5 \text{ones}(n). \end{aligned}$$

We take $n = 100$,

$$\begin{aligned} G &= \text{ones}(n, 2) + 10 \cdot [I_2; \text{zeros}(n-2, 2)], \\ F &= 0.1 \cdot \text{ones}(n, 2) - 10 \cdot [I_2; \text{zeros}(n-2, 2)]. \end{aligned}$$

It can be computed that

$$\|X\| = 107.9026,$$

while the bound (3.11) gives

$$\|X\| \leq 319.8178.$$

The bound would be tighter if the eigenvector matrices S and T are better-conditioned. On the other hand, the bound can be attained, by, e.g., A and B are diagonal (thus $S = T = I$) and $F = G = (1, 0, \dots, 0)^T$. \diamond

3.2. Perturbation bound

We'll present a perturbation bound for the solution of Sylvester equation (2.1) perturbed to

$$(A + \delta A)(X + \delta X) - (X + \delta X)(B + \delta B) = (G + \delta G)(F + \delta F)^*. \quad (3.14)$$

Set

$$\epsilon = \max\{\|\delta A\|, \|\delta B\|, \|\delta G\|, \|\delta F\|\}, \quad (3.15)$$

and assume that it is sufficiently small. Neglecting the second order terms and subtracting the unperturbed Sylvester equation from the perturbed one yield

$$A \delta X - \delta X B \approx G \delta F^* + \delta G F^* - \delta A X + X \delta B. \quad (3.16)$$

δX can then be approximated by $\delta X \approx \delta X_1 + \delta X_2 + \delta X_3$, where

$$A \delta X_1 - \delta X_1 B = -\delta A X, \quad (3.17)$$

$$A \delta X_2 - \delta X_2 B = -X \delta B, \quad (3.18)$$

$$A \delta X_3 - \delta X_3 B = (G \delta F^* + \delta G F^*). \quad (3.19)$$

For (3.17), we again choose parameters

either $\alpha_i = \lambda_i$ for $i = 1, \dots, m$, or $\beta_j = \mu_j$ for $j = 1, \dots, n$,

to get

$$\delta X_1 = S \left(\sum_{j=1}^{n_0} (\mu_j - \lambda_j) \hat{\Phi}^{(j)} \hat{\Psi}^{(j)} \right) T^{-1}, \quad n_0 = \min\{m, n\}, \quad (3.20)$$

where

$$\hat{\Phi}^{(j)} = \text{diag} \left(\frac{\sigma(1, j-1)}{\lambda_1 - \mu_j}, \dots, \frac{\sigma(m, j-1)}{\lambda_m - \mu_j} \right) S^{-1} \delta A,$$

$$\sigma(i, j-1) = \prod_{s=1}^{j-1} \frac{\lambda_i - \lambda_s}{\lambda_i - \mu_s},$$

and

$$\hat{\Psi}^{(j)} = X T \text{diag} \left(\frac{\tau(1, j-1)}{\mu_1 - \lambda_j}, \dots, \frac{\tau(n, j-1)}{\mu_n - \lambda_j} \right),$$

$$\tau(i, j-1) = \prod_{s=1}^{j-1} \frac{\mu_i - \mu_s}{\mu_i - \lambda_s}.$$

Define

$$\Delta_{\hat{\Phi}^{(j)}} = \text{diag} \left(\frac{\sigma(1, j-1)}{\lambda_1 - \mu_j}, \dots, \frac{\sigma(m, j-1)}{\lambda_m - \mu_j} \right), \quad (3.21)$$

$$\Delta_{\hat{\Psi}^{(j)}} = \text{diag} \left(\frac{\tau(1, j-1)}{\mu_1 - \lambda_j}, \dots, \frac{\tau(n, j-1)}{\mu_n - \lambda_j} \right), \quad (3.22)$$

and

$$\sigma_{\max}^{(j)} = \|\Delta_{\hat{\Phi}^{(j)}}\| = \max_i \frac{|\sigma(i, j-1)|}{|\lambda_i - \mu_j|}, \quad (3.23)$$

$$\tau_{\max}^{(j)} = \|\Delta_{\hat{\Psi}^{(j)}}\| = \max_i \frac{|\tau(i, j-1)|}{|\mu_i - \lambda_j|}.$$

Eq. (3.20) implies

$$\|\delta X_1\| \leq \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \|S \Delta_{\hat{\Phi}^{(j)}}\| \|S^{-1} \delta A\| \|X\| \|T \Delta_{\hat{\Psi}^{(j)}} T^{-1}\|,$$

and thus

$$\frac{\|\delta X_1\|}{\|X\|} \leq \|S^{-1} \delta A\| \|S\| \kappa(T) \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \tau_{\max}^{(j)} \sigma_{\max}^{(j)}. \quad (3.24)$$

Similarly for δX_2 , we have by (3.18)

$$\|\delta X_2\| \leq \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \|S \Delta_{\hat{\Phi}^{(j)}} S^{-1}\| \|\delta B T\| \|X\| \|\Delta_{\hat{\Psi}^{(j)}} T^{-1}\|,$$

and consequently

$$\frac{\|\delta X_2\|}{\|X\|} \leq \|\delta B T\| \|T^{-1}\| \kappa(S) \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \tau_{\max}^{(j)} \sigma_{\max}^{(j)}. \quad (3.25)$$

Lastly for δX_3 , we have by (3.19)

$$\|\delta X_3\| \leq \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \|S \Delta_{\phi^{(j)}}\| \|S^{-1}(G\delta F^* + \delta G F^*)T\| \|\Delta_{\psi^{(j)}} T^{-1}\|$$

which implies

$$\begin{aligned} \|\delta X_3\| &\leq \|S^{-1}(G\delta F^* + \delta G F^*)T\| \|S\| \|T^{-1}\| \\ &\times \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \tau_{\max}^{(j)} \sigma_{\max}^{(j)}. \end{aligned} \quad (3.26)$$

Finally use $\|\delta X\| \leq \|\delta X_1\| + \|\delta X_2\| + \|\delta X_3\| + \mathcal{O}(\epsilon^2)$ to arrive at a first order perturbation theorem below.

Theorem 3.3. Assume that A and B have the eigen-decompositions (3.1) and (3.2). Let X be the solution of the Sylvester equation (2.1) and let $X + \delta X$ be the solution of the perturbed Sylvester equation (3.14). For sufficiently small ϵ defined by (3.15), we have

$$\begin{aligned} \frac{\|\delta X\|}{\|X\|} &\leq \left(\kappa(T) \|S\| \|S^{-1} \delta A\| + \kappa(S) \|T^{-1}\| \|\delta B T\| \right. \\ &\left. + \|S\| \|T^{-1}\| \frac{\|S^{-1}(G\delta F^* + \delta G F^*)T\|}{\|X\|} \right) \gamma + \mathcal{O}(\epsilon^2), \end{aligned} \quad (3.27)$$

where $\gamma = \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \tau_{\max}^{(j)} \sigma_{\max}^{(j)}$ and $\sigma_{\max}^{(j)}, \tau_{\max}^{(j)}$ are defined as in (3.23).

Remark 3.3. From (3.20) and similar equalities for δX_2 and δX_3 , we can derive a different version of (3.27), too:

$$\begin{aligned} \frac{\|\delta X\|}{\|X\|} &\leq \sum_{j=1}^{n_0} |\mu_j - \lambda_j| \|S \Delta_{\phi^{(j)}} S^{-1}\| \|T \Delta_{\psi^{(j)}} T^{-1}\| \\ &\times \left(\|\delta A\| + \|\delta B\| + \frac{\|S^{-1}(G\delta F^* + \delta G F^*)T\|}{\|X\|} \right) + \mathcal{O}(\epsilon^2), \end{aligned}$$

where diagonal matrices $\Delta_{\phi^{(j)}}$ and $\Delta_{\psi^{(j)}}$ are defined as in (3.21) and (3.22).

Remark 3.4. Using $\|AB\|_F \leq \|A\| \|B\|_F$ and $\|AB\|_F \leq \|A\|_F \|B\|$ (see for example [21, Theorem II.3.9]), one can obtain the following bound for the Frobenius norm

$$\begin{aligned} \frac{\|\delta X\|_F}{\|X\|_F} &\leq \left(\kappa(T) \|S\| \|S^{-1} \delta A\| + \kappa(S) \|T^{-1}\| \|\delta B T\| \right. \\ &\left. + \|S\| \|T^{-1}\| \frac{\|S^{-1}(G\delta F^* + \delta G F^*)T\|_F}{\|X\|_F} \right) \gamma + \mathcal{O}(\epsilon^2). \end{aligned} \quad (3.28)$$

As an illustration of the quality of the perturbation bound (3.28), we will present a comparison between this bound and two bounds from [20]. The first one is

$$\frac{\|\delta X\|_F}{\|X\|_F} \leq \sqrt{3} \psi \epsilon + \mathcal{O}(\epsilon^2), \quad (3.29)$$

where

$$\begin{aligned} \psi &= \|P^{-1}[\alpha(X^T \otimes I_m) - \beta(I_n \otimes X) - \delta I_{mn}]\| / \|X\|_F, \\ P &= I_n \otimes A - B^T \otimes I_m \end{aligned}$$

and $\epsilon = \max \left\{ \frac{\|\delta A\|_F}{\alpha}, \frac{\|\delta B\|_F}{\beta}, \frac{\|\delta C\|_F}{\delta} \right\}$, while α, β and δ are scaling factors as in [20]. The second bound is a weaker version of (3.29):

$$\frac{\|\delta X\|_F}{\|X\|_F} \leq \sqrt{3} \phi \epsilon + \mathcal{O}(\epsilon^2), \quad (3.30)$$

where $\phi = \|P^{-1}\| \frac{(\alpha + \beta) \|X\|_F + \delta}{\|X\|_F}$.

As pointed out in [20], the perturbation bound (3.30) with $\alpha = \|A\|_F, \beta = \|B\|_F$ and $\delta = \|C\|_F$ is the one that is usually quoted in the literature. For example this bound can also be found in [4]. Note that both (3.29) and (3.30) do not require diagonalizability, and, moreover, they do not require any knowledge of the spectral decompositions.

The following example compares our (3.28) with (3.29) and (3.30).

Example 3.2. Let $A = S \Lambda S^{-1}$ with $\Lambda = \text{diag}(30, 40, 60, 70, 90)$ and

$$S = 5 I_5 + 0.5 \text{ones}(5).$$

Let $B = T \Omega T^{-1}$ with $\Omega = \text{diag}(100, 200, 300, 400, 450)$ and

$$T = \text{diag}([10^4, 10^3, 3, 2, 1]) \cdot (I_5 + 0.1 \text{ones}(5)).$$

Further,

$$G = \begin{pmatrix} -10 & -60 & 0 & 0 & 0 \\ -60 & 400 & 0 & 0 & 0 \end{pmatrix}^T, \quad F = G.$$

Perturbations are

$$\begin{aligned} \delta B &= 10^{-9} \cdot \text{diag}(10^{-5}, 10^{-5}, 1, 1, 1), \quad \delta A = 0, \\ \delta F &= \delta G = 0. \end{aligned}$$

It can be computed that

$$\frac{\|\delta X\|_F}{\|X\|_F} = 2.0878 \cdot 10^{-12}, \quad (3.31)$$

while the perturbation bound (3.29) and (3.30) give

$$\frac{\|\delta X\|_F}{\|X\|_F} \lesssim 1.188 \cdot 10^{-7}, \quad \frac{\|\delta X\|_F}{\|X\|_F} \lesssim 2.5997 \cdot 10^{-7}, \quad (3.32)$$

respectively. Here and in what follows, we use the approximate-less symbol “ \lesssim ” to indicate that it is a bound with the second term ignored. On the other hand the perturbation bound (3.28) gives

$$\frac{\|\delta X\|_F}{\|X\|_F} \lesssim 6.6112 \cdot 10^{-10} \quad (3.33)$$

which is much close to the exact (3.31) than (3.32) obtained by (3.29) and (3.30). This example illustrates that the structure of the matrices F and G from the right-hand side in Sylvester equation (2.1) sometimes can greatly influence the perturbation of the solution. \diamond

We shall now explain this influence and present a reason why the perturbation bounds (3.27) and (3.28) prevail the bounds (3.29) and (3.30). Recall that all these bounds can be considered as the perturbation bounds for the solution of the linear system

$$P x = c \quad \text{perturbed to} \quad (P + \delta P)(x + \delta x) = c + \delta c, \quad (3.34)$$

where $c = \text{vec}(C)$ is the column vector obtained by stacking the columns of C one after another with its first column on the top, followed by its second column and then third column and so on, and similarly $c + \delta c = \text{vec}(C + \delta C)$, and

$$\begin{aligned} P &= I_n \otimes A - B^T \otimes I_m, \\ P + \delta P &= I_n \otimes (A + \delta A) - (B + \delta B)^T \otimes I_m, \end{aligned}$$

and “ \otimes ” denotes the Kronecker product. The reason why (3.27) and (3.28) prevail over the bounds (3.29) and (3.30) lies in the fact that they include the range of P and $P + \delta P$ and the orientation of c with the respect to the range of P .

The last property (which illustrates the influence of the right-hand side) of the perturbation bounds can also be found in [31],

where Chan and Foulser obtained a result, when applied to (3.34), yielding

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\sigma_{N+1-k}}{\sigma_N} \left(\frac{\|\mathcal{P}_k c\|}{\|c\|} \right)^{-1} \frac{\|\delta c\|}{\|c\|},$$

$$k = 1, 2, \dots, N, \quad N = nm, \quad (3.35)$$

where $\sigma_1 \geq \dots \geq \sigma_N$ are singular values of the matrix P , and $P = U \Sigma V^T$, and $\mathcal{P}_k = U_k U_k^T$, $U_k = [u_{N+1-k}, \dots, u_N]$.

The integer k is to be chosen so that the right-hand side of (3.35) is the smallest.

The bounds (3.28) and (3.35) are based on the similar ideas, although (3.28) is more general.

Example 3.3. In Example 3.2, reset the perturbations to,

$$\delta B = 0, \quad \delta G = 10^{-6} \cdot \begin{pmatrix} 5 & 1 & 0 & 0 & 0 \\ 1 & 5 & 0 & 0 & 0 \end{pmatrix}^T, \quad \delta F = \delta G.$$

Then $\delta C = (G \delta F^* + \delta G F^*)$ and in (3.35) $\delta c = \text{vec}(\delta C)$. In this case (3.28) gives

$$\frac{\|\delta X\|_F}{\|X\|_F} \lesssim 3.942 \cdot 10^{-6},$$

while (3.35) gives

$$\frac{\|\delta x\|}{\|x\|} \leq 1.543 \cdot 10^{-6}.$$

So both bounds are of the same quality for the example.

On the other hand, Chan and Foulser also have another perturbation bound for $\delta c = 0$ only. This bound [31, Theorem 2] yields for Example 3.2

$$\frac{\|\delta x\|}{\|x\|} \leq 1.0114 \cdot 10^{-6},$$

which is obviously less sharp than (3.33). \diamond

Since in applications, often eigenvalues of A and B are close, interesting question arises: what will happen with the bound (3.11) in such cases? As was expected, the bound (3.11) are usually not sensitive to clustered eigenvalues, namely, if some eigenvalues of A are close to some eigenvalues of B the bound will usually increase proportionally to the norm of solution. On the other hand, the perturbation bound (3.27) is more sensitive to the clustered eigenvalues than the bound (3.11). But, for example, if the range of the matrices on the right-hand side is close to the rank- k dominant singular subspace of the eigenvector matrices, where the rank- k dominant singular subspaces is the subspace spanned by k singular vectors associated with the largest k singular values of the considered matrix, the bound (3.11) remains tight enough.

Remark 3.5. It is important to note that bounds (3.27) and (3.28) have more theoretical meaning than practical ones. In practice, when applying ADI, one does not necessarily choose the eigenvalues of A and B as the ADI shifts. One would use Ritz values (see for example [22] or [23] or [25, Section 5] in the case of Lyapunov equations) or some other approximate eigenvalues (see for example [28]) as opposed to a subset of the exact eigenvalues to extract information. One of such an example is given in Example 4.2 in the next section. A question naturally arises: *is it possible to obtain a version of the upper bound (3.27) that would use arbitrary shifts instead of the exact eigenvalues?* Unfortunately at this moment we do not have an answer to this question, and this remains as an open problem. We would like to emphasize that an attempt in the same direction for Lyapunov equations has been made in [32]. The bound obtained there is very complicated and hard to apply; thus from this point of view if one would like to derive a version of the upper bound (3.27), a different approach may be needed.

4. The non-diagonalizable case

This section accomplishes the same tasks as the previous section but for non-diagonalizable matrices A and B .

Recall that Eq. (2.1) has a unique solution if and only if A and B have no common eigenvalues, which has been assumed throughout the paper. For the sake of simplicity, we will consider matrices A and B whose Jordan blocks are at most 2×2 . The idea is easily extensible to more general cases, but more complicated. On the other hand, the case with only up to 2×2 Jordan blocks does happen in practice. For example a simple planar spacecraft model [33] has two 2×2 Jordan blocks. This makes our case more interesting for investigation.

Let the Jordan canonical form of A be

$$A = S J_A S^{-1}, \quad S \in \mathbb{C}^{m \times m}, \quad J_A = J_{A,1} \oplus \dots \oplus J_{A,k_A}, \quad (4.1)$$

where $J_{A,i} \oplus J_{A,j}$ stands for the direct sum of $J_{A,i}$ and $J_{A,j}$, and

$$J_{A,i} = \lambda_i \quad \text{for } i = 1, \dots, \ell_A,$$

$$J_{A,i} = \begin{pmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{pmatrix} \equiv \lambda_i I_2 + N_2, \quad \text{for } i = \ell_A + 1, \dots, k_A,$$

$2(k_A - \ell_A) + \ell_A = m$, and $N_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Similarly, let the Jordan canonical form of B be

$$B = T J_B T^{-1}, \quad T \in \mathbb{C}^{n \times n}, \quad J_B = J_{B,1} \oplus \dots \oplus J_{B,k_B}, \quad (4.2)$$

$$J_{B,i} = \mu_i \quad \text{for } i = 1, \dots, \ell_B,$$

$$J_{B,i} = \begin{pmatrix} \mu_i & 1 \\ 0 & \mu_i \end{pmatrix} \equiv \mu_i I_2 + N_2 \quad \text{for } i = \ell_B + 1, \dots, k_B,$$

where $2(k_B - \ell_B) + \ell_B = n$.

Theorem 4.1. Assume A and B have the Jordan canonical decompositions (4.1) and (4.2). Let X be the solution of the Sylvester equation (2.1), obtained by (2.3)–(2.5) (LR-ADI) with the set of ADI parameters

$$\{\alpha_1, \alpha_2, \dots, \alpha_m\} = \{\lambda_{p_1}, \lambda_{p_2}, \dots, \lambda_{p_m}\}$$

and

$$\{\beta_1, \beta_2, \dots, \beta_n\} = \{\mu_{q_1}, \mu_{q_2}, \dots, \mu_{q_n}\},$$

where each eigenvalue appears as many times as its algebraic multiplicity. Denote by \hat{g}_i the i th $1 \times r$ submatrix (for $i = 1, \dots, \ell_A$) and $2 \times r$ submatrix (for $i = \ell_A + 1, \dots, k_A$) of $\hat{G} = S^{-1}G$, respectively. Similarly, denote by \hat{f}_j the j th $1 \times r$ submatrix (for $j = 1, \dots, \ell_B$) and $2 \times r$ submatrix (for $j = \ell_B + 1, \dots, k_B$) of $\hat{F}^* = F^*T$, respectively. Then

$$\|X\| \leq \|S\| \|T^{-1}\| \sum_{j=1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \sum_{i=1}^{k_A} \frac{\|\eta(i, j-1)\| \cdot \|\hat{g}_i\|}{|\lambda_i - \mu_{q_j}|}$$

$$\times \sum_{s=1}^{k_B} \frac{\|\vartheta(s, j-1)\| \|\hat{f}_s\|}{|\mu_s - \lambda_{p_j}|}, \quad (4.3)$$

where $n_0 = \min\{m, n\}$,

$$\eta(i, j) = \sigma(i, j) \quad \text{for } i = 1, \dots, \ell_A, \quad (4.4)$$

$$\eta(i, j) = \left(I_2 - \frac{1}{\lambda_i - \mu_{q_{j+1}}} N_2 \right) [\sigma(i, j) I_2 - \mu(i, j) N_2]$$

$$\text{for } i = \ell_A + 1, \dots, k_A, \quad (4.5)$$

$$\sigma(i, j-1) = \prod_{s=1}^{j-1} \frac{\lambda_i - \lambda_{p_s}}{\lambda_i - \mu_{q_s}} \quad \text{and} \quad \sigma(i, 0) = 1, \quad \text{for all } i, \quad (4.6)$$

$$\mu(i, j) = \sum_{\ell=1}^j \prod_{\substack{t=1 \\ t \neq \ell}}^j \frac{\lambda_i - \lambda_{p_\ell}}{\lambda_i - \mu_{q_\ell}} \frac{\lambda_{p_t} - \mu_{q_t}}{(\lambda_i - \mu_{q_t})^2}$$

for $i = \ell_A + 1, \dots, k_A$, (4.7)

$$\vartheta(i, j) = \tau(i, j) \quad \text{for } i = 1, \dots, \ell_B, \quad (4.8)$$

$$\vartheta(i, j) = \left(I_2 - \frac{1}{\mu_i - \lambda_{p_{j+1}}} N_2 \right) [\tau(i, j) I_2 - \nu(i, j) N_2]$$

for $i = \ell_B + 1, \dots, k_B$, (4.9)

$$\tau(i, j-1) = \prod_{s=1}^{j-1} \frac{\mu_i - \mu_{q_s}}{\mu_i - \lambda_{p_s}} \quad \text{and} \quad \tau(i, 0) = 1, \quad \text{for all } i, \quad (4.10)$$

$$\nu(i, j) = \sum_{\ell=1}^j \prod_{\substack{t=1 \\ t \neq \ell}}^j \frac{\mu_i - \mu_{q_\ell}}{\mu_i - \lambda_{p_\ell}} \frac{\mu_{q_t} - \lambda_{p_t}}{(\mu_i - \lambda_{p_t})^2}$$

for $i = \ell_B + 1, \dots, k_B$. (4.11)

Proof. Similarly to (3.3), one gets

$$\begin{aligned} Z^{(j)} &= S(J_A - \lambda_{p_{j-1}} I)(J_A - \mu_{q_j} I)^{-1} S^{-1} Z^{(j-1)} \\ &= S(J_A - \lambda_{p_{j-1}} I)(J_A - \mu_{q_j} I)^{-1} (J_A - \lambda_{p_{j-2}} I)(J_A - \mu_{q_{j-1}} I)^{-1} \dots \\ &\quad \times (J_A - \lambda_{p_1} I)(J_A - \mu_{q_2} I)^{-1} (J_A - \mu_{q_1} I)^{-1} S^{-1} G \\ &= S(J_A - \mu_{q_j} I)^{-1} \Gamma_{j-1} \Gamma_{j-2} \dots \Gamma_1 S^{-1} G, \end{aligned} \quad (4.12)$$

where

$$\Gamma_j = (J_A - \lambda_{p_j} I)(J_A - \mu_{q_j} I)^{-1}.$$

Similarly to (3.4), one gets

$$\begin{aligned} Y^{(j)*} &= Y^{(j-1)*} T (J_B - \lambda_{p_j} I)^{-1} (J_B - \mu_{q_{j-1}} I) T^{-1} \\ &= F^* T (J_B - \lambda_{p_1} I)^{-1} (J_B - \lambda_{p_2} I)^{-1} (J_B - \mu_{q_1} I) \dots \\ &\quad \times (J_B - \lambda_{p_j} I)^{-1} (J_B - \mu_{q_{j-1}} I) T^{-1} \\ &= F^* T \mathcal{E}_1 \mathcal{E}_2 \dots \mathcal{E}_{j-1} (J_B - \lambda_{p_j} I)^{-1} T^{-1}, \end{aligned}$$

where

$$\mathcal{E}_j = (J_B - \mu_{q_j} I)(J_B - \lambda_{p_j} I)^{-1}.$$

Further, note that matrices Γ_j and \mathcal{E}_j are block diagonal matrices, and the i th diagonal block of Γ_j is

$$(\Gamma_j)_{(i,i)} = \frac{\lambda_i - \lambda_{p_j}}{\lambda_i - \mu_{q_j}} \quad \text{for } i = 1, \dots, \ell_A, \quad (4.13)$$

$$(\Gamma_j)_{(i,i)} = \frac{\lambda_i - \lambda_{p_j}}{\lambda_i - \mu_{q_j}} I_2 + \frac{\lambda_{p_j} - \mu_{q_j}}{(\lambda_i - \mu_{q_j})^2} N_2$$

for $i = \ell_A + 1, \dots, k_A$, (4.14)

while the i th diagonal block of \mathcal{E}_j is

$$(\mathcal{E}_j)_{(i,i)} = \frac{\mu_i - \mu_{q_j}}{\mu_i - \lambda_{p_j}} \quad \text{for } i = 1, \dots, \ell_B, \quad (4.15)$$

$$(\mathcal{E}_j)_{(i,i)} = \frac{\mu_i - \mu_{q_j}}{\mu_i - \lambda_{p_j}} I_2 + \frac{\mu_{q_j} - \lambda_{p_j}}{(\mu_i - \lambda_{p_j})^2} N_2$$

for $i = \ell_B + 1, \dots, k_B$. (4.16)

Now from (4.13) and (4.14), it follows that the i th diagonal block of $(J_A - \mu_{q_j} I)^{-1} \Gamma_{j-1} \Gamma_{j-2} \dots \Gamma_1$ is

$$(J_A - \mu_{q_j} I)^{-1} \Gamma_{j-1} \Gamma_{j-2} \dots \Gamma_1)_{(i,i)} = \frac{1}{\lambda_i - \mu_{q_j}} \prod_{s=1}^{j-1} \frac{\lambda_i - \lambda_{p_s}}{\lambda_i - \mu_{q_s}}$$

$$= \frac{\sigma(i, j-1)}{\lambda_i - \mu_{q_j}} \quad \text{for } i = 1, \dots, \ell_A, \quad (4.17)$$

and for $i = \ell_A + 1, \dots, k_A$ it is

$$\begin{aligned} &\left(I_2 - \frac{1}{\lambda_i - \mu_{q_j}} N_2 \right) \left(\frac{1}{\lambda_i - \mu_{q_j}} \prod_{s=1}^{j-1} \frac{\lambda_i - \lambda_{p_s}}{\lambda_i - \mu_{q_s}} I_2 + \frac{1}{\lambda_i - \mu_{q_j}} \right. \\ &\quad \times \left. \sum_{\ell=1}^{j-1} \prod_{\substack{t=1 \\ t \neq \ell}}^{j-1} \frac{\lambda_i - \lambda_{p_\ell}}{\lambda_i - \mu_{q_\ell}} \frac{\lambda_{p_t} - \mu_{q_t}}{(\lambda_i - \mu_{q_t})^2} N_2 \right) \\ &= \left(I_2 - \frac{1}{\lambda_i - \mu_{q_j}} N_2 \right) \left(\frac{\sigma(i, j-1)}{\lambda_i - \mu_{q_j}} I_2 + \frac{\mu(i, j-1)}{\lambda_i - \mu_{q_j}} N_2 \right), \end{aligned} \quad (4.18)$$

where $\sigma(i, j-1)$ and $\mu(i, j-1)$ are defined as in (4.6)–(4.7).

Similarly from (4.15) and (4.16), it follows that the i th diagonal block of $\mathcal{E}_1 \mathcal{E}_2 \dots \mathcal{E}_{j-1} (J_B - \lambda_{p_j} I)^{-1}$ is

$$\begin{aligned} (\mathcal{E}_1 \mathcal{E}_2 \dots \mathcal{E}_{j-1} (J_B - \lambda_{p_j} I)^{-1})_{(i,i)} &= \frac{1}{\mu_i - \lambda_{p_j}} \prod_{s=1}^{j-1} \frac{\mu_i - \mu_{q_s}}{\mu_i - \lambda_{p_s}} \\ &= \frac{\tau(i, j-1)}{\mu_i - \lambda_{p_j}} \quad \text{for } i = 1, \dots, \ell_B, \end{aligned} \quad (4.19)$$

and for $i = \ell_B + 1, \dots, k_B$ it is

$$\begin{aligned} &\left(I_2 - \frac{1}{\mu_i - \lambda_{p_j}} N_2 \right) \left(\frac{1}{\mu_i - \lambda_{p_j}} \prod_{s=1}^{j-1} \frac{\mu_i - \mu_{q_s}}{\mu_i - \lambda_{p_s}} I_2 + \frac{1}{\mu_i - \lambda_{p_j}} \right. \\ &\quad \times \left. \sum_{\ell=1}^{j-1} \prod_{\substack{t=1 \\ t \neq \ell}}^{j-1} \frac{\mu_i - \mu_{q_\ell}}{\mu_i - \lambda_{p_\ell}} \frac{\mu_{q_t} - \lambda_{p_t}}{(\mu_i - \lambda_{p_t})^2} N_2 \right) \\ &= \left(I_2 - \frac{1}{\mu_i - \lambda_{p_j}} N_2 \right) \left(\frac{\tau(i, j-1)}{\mu_i - \lambda_{p_j}} I_2 + \frac{\nu(i, j-1)}{\mu_i - \lambda_{p_j}} N_2 \right), \end{aligned} \quad (4.20)$$

where $\tau(i, j-1)$ and $\nu(i, j-1)$ are defined as in (4.10)–(4.11).

Now, from (2.5) it follows that

$$\begin{aligned} X &= \sum_{j=1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) Z^{(j)} Y^{(j)*} \\ &= S \left(\sum_{j=1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) (J_A - \mu_{q_j} I)^{-1} \Gamma_{j-1} \Gamma_{j-2} \dots \Gamma_1 \hat{G} \hat{F}^* \mathcal{E}_1 \mathcal{E}_2 \dots \right. \\ &\quad \times \left. \mathcal{E}_{j-1} (J_B - \lambda_{p_j} I)^{-1} \right) T^{-1}. \end{aligned} \quad (4.21)$$

Finally, (4.3) is a consequence of taking the norms at the both sides of the above equation and the definitions (4.4)–(4.11). \square

4.1. Error bound

Theorem 4.2. Assume that A and B admit the Jordan canonical decompositions (4.1) and (4.2). Let X_k be the k th approximation obtained by (2.3)–(2.5) with the set of ADI parameters corresponding to subsets of the exact eigenvalues of A and B , i.e., $\{\alpha_1, \alpha_2, \dots, \alpha_k\} = \{\lambda_{p_1}, \lambda_{p_2}, \dots, \lambda_{p_k}\}$ and $\{\beta_1, \beta_2, \dots, \beta_k\} = \{\mu_{q_1}, \mu_{q_2}, \dots, \mu_{q_k}\}$. Then the following inequality holds

$$\|X - X_k\| \leq \|S\| \|T^{-1}\| \sum_{j=k+1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \sum_{i=1}^{k_A} \frac{\|\eta(i, j-1)\| \cdot \|\hat{g}_i\|}{|\lambda_i - \mu_{q_j}|}$$

$$\times \sum_{\ell=1}^{k_B} \frac{\|\vartheta(\ell, j-1)\| \|\hat{f}_\ell\|}{|\mu_\ell - \lambda_{p_j}|}, \quad (4.22)$$

where $\eta(i, j)$ and $\vartheta(\ell, j)$ are defined as in (4.4)–(4.11), respectively, and \hat{g}_i, \hat{f}_ℓ are defined as in Theorem 4.1, and $\{\lambda_{p_j}, j > k\}$ and $\{\lambda_{q_j}, j > k\}$ are the subsets of eigenvalues of A and B complement to the ones already used as shifts for obtaining X_k .

Proof. An equation similar to (3.13) holds. Take norms and use the formulas from (4.12)–(4.20) to get (4.22). \square

4.2. Perturbation bound

We'll consider the same problem as we did in Section 3.2, except A and B are no longer diagonalizable but have the Jordan canonical forms as in (4.1) and (4.2), respectively. It can be seen that the development there up to (3.19) remains valid.

Choose parameters such that

either $\alpha_i = \lambda_{p_i}$ for $i = 1, \dots, m$, or $\beta_j = \mu_{q_j}$ for $j = 1, \dots, n$, where $p_i = i$ for $1 \leq i \leq \ell_A$ and $p_{\ell_A+2i-1} = p_{\ell_A+2i} = \ell_A + i$ for $1 \leq i \leq 2(k_A - \ell_A)$, and $q_i = i$ for $1 \leq i \leq \ell_B$ and $q_{\ell_B+2i-1} = q_{\ell_B+2i} = \ell_B + i$ for $1 \leq i \leq 2(k_B - \ell_B)$, recall that $\alpha_i \neq \beta_j$, for all i, j .

Then if we apply formula (4.21) to the Sylvester equation

$$A\delta X_1 - \delta X_1 B = -\delta A X,$$

we have

$$\delta X_1 = -S \left(\sum_{j=1}^{n_0} (\mu_{q_j} - \lambda_{p_j}) (J_A - \mu_{q_j} I)^{-1} \Gamma_{j-1} \dots \Gamma_1 S^{-1} \delta A X T \mathcal{E}_1 \right. \\ \left. \times \mathcal{E}_2 \dots \mathcal{E}_{j-1} (J_B - \lambda_{p_j} I)^{-1} \right) T^{-1},$$

from which we have

$$\|\delta X_1\| \leq \|S\| \kappa(T) \|S^{-1} \delta A\| \|X\| \left(\sum_{j=1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \eta_{\max}^{(j)} \vartheta_{\max}^{(j)} \right), \quad (4.23)$$

where $\eta_{\max}^{(j)} = \|(J_A - \mu_{q_j} I)^{-1} \Gamma_{j-1} \Gamma_{j-2} \dots \Gamma_1\|$ and $\vartheta_{\max}^{(j)} = \|\mathcal{E}_1 \mathcal{E}_2 \dots \mathcal{E}_{j-1} (J_B - \lambda_{p_j} I)^{-1}\|$. Further using (4.17), (4.18), (4.4) and (4.5), we obtain

$$\eta_{\max}^{(j)} = \left\| \text{diag} \left(\frac{\eta(1, j-1)}{\lambda_1 - \mu_{q_j}}, \frac{\eta(2, j-1)}{\lambda_2 - \mu_{q_j}}, \dots, \frac{\eta(k_A, j-1)}{\lambda_{k_A} - \mu_{q_j}} \right) \right\| \\ = \max_k \frac{\|\eta(k, j-1)\|}{|\lambda_k - \mu_{q_j}|}. \quad (4.24)$$

Similarly from (4.19), (4.20), (4.8) and (4.9), we have

$$\vartheta_{\max}^{(j)} = \left\| \text{diag} \left(\frac{\vartheta(1, j-1)}{\mu_1 - \lambda_{p_j}}, \frac{\vartheta(2, j-1)}{\mu_2 - \lambda_{p_j}}, \dots, \frac{\vartheta(k_B, j-1)}{\mu_{k_B} - \lambda_{p_j}} \right) \right\| \\ = \max_k \frac{\|\vartheta(k, j-1)\|}{|\mu_k - \lambda_{p_j}|}. \quad (4.25)$$

For the solutions of the Sylvester equation (3.18) and (3.19), one can obtain the following bounds

$$\|\delta X_2\| \leq \kappa(S) \|T^{-1}\| \|\delta B T\| \|X\| \left(\sum_{j=1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \eta_{\max}^{(j)} \vartheta_{\max}^{(j)} \right), \quad (4.26)$$

$$\|\delta X_3\| \leq \|S\| \|T^{-1}\| \|S^{-1} (G\delta F^* + \delta G F^*) T\| \\ \times \left(\sum_{j=1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \eta_{\max}^{(j)} \vartheta_{\max}^{(j)} \right). \quad (4.27)$$

The following theorem contains a first order perturbation bound for the solution of the Sylvester equation (2.1) perturbed as in (3.14), where A and B are no longer diagonalizable but have the Jordan canonical forms as in (4.1) and (4.2), respectively.

Theorem 4.3. Assume that A and B admits the Jordan canonical decompositions (4.1) and (4.2). Let X be the solution of the Sylvester equation (2.1) and let $X + \delta X$ be the solution of the perturbed Sylvester equation (3.14). If ϵ defined by (3.15) is sufficiently small, then

$$\frac{\|\delta X\|}{\|X\|} \leq \left(\kappa(T) \|S\| \|S^{-1} \delta A\| + \kappa(S) \|T^{-1}\| \|\delta B T\| \right. \\ \left. + \|S\| \|T^{-1}\| \frac{\|S^{-1} (G\delta F^* + \delta G F^*) T\|}{\|X\|} \right) \gamma + \mathcal{O}(\epsilon^2), \quad (4.28)$$

where $\gamma = \sum_{j=1}^{n_0} |\mu_{q_j} - \lambda_{p_j}| \vartheta_{\max}^{(j)} \eta_{\max}^{(j)}$, and $\eta_{\max}^{(j)}$ and $\vartheta_{\max}^{(j)}$ are defined as in (4.24) and (4.25).

Proof. From $\|\delta X\| \leq \|\delta X_1\| + \|\delta X_2\| + \|\delta X_3\| + \mathcal{O}(\epsilon^2)$, together with (4.23), (4.26) and (4.27) for $\|\delta X_1\|$, $\|\delta X_2\|$ and $\|\delta X_3\|$, respectively, we obtain (4.28). \square

Remark 4.1. Again, using $\|AB\|_F \leq \|A\| \|B\|_F$ and $\|AB\|_F \leq \|A\|_F \|B\|$ [21, Theorem II. 3.9.], we have for the Frobenius norm

$$\frac{\|\delta X\|_F}{\|X\|_F} \leq \left(\kappa(T) \|S\| \|S^{-1} \delta A\| + \kappa(S) \|T^{-1}\| \|\delta B T\| \right. \\ \left. + \|S\| \|T^{-1}\| \frac{\|S^{-1} (G\delta F^* + \delta G F^*) T\|_F}{\|X\|_F} \right) \gamma + \mathcal{O}(\epsilon^2). \quad (4.29)$$

The bounds (4.28) and (4.29) share the same properties as the bounds for the diagonalizable case. In order to compare our new bound (4.29) with bounds (3.29) and (3.30), we consider following example.

Example 4.1. Let $A = S J_A S^{-1}$ with $J = 10 \oplus 20 \oplus (40I_2 + N_2) \oplus (60I_2 + N_2)$ and

$$S = \text{diag}([10, 10, 0.1, 0.1, 0.1, 0.1]) \cdot (5I_6 + 0.5 \text{ones}(6)).$$

Let $B = T J_B T^{-1}$ with $J_B = 1.1 \oplus 2.2 \oplus 3.3 \oplus 4.4 \oplus (5.5I_2 + N_2)$ and

$$T = 10I_6 + 0.5 \text{ones}(6).$$

Further,

$$G = \begin{pmatrix} 20 & 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T, \quad F = G.$$

Perturbations are

$$\delta A = 10^{-10} \cdot \text{diag}(0.2, 0.2, 0.0002, 0.0002, 0.0002, 0.0002),$$

$$\delta B = 5 \delta A,$$

$$\delta G = 10^{-6} \cdot \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}^T, \quad \delta F = \delta G.$$

It can be computed that

$$\frac{\|\delta X\|_F}{\|X\|_F} = 1.0763 \cdot 10^{-7},$$

while the perturbation bounds (3.29) and (3.30) give

$$\frac{\|\delta X\|_F}{\|X\|_F} \lesssim 2.3228 \cdot 10^{-4}, \quad \frac{\|\delta X\|_F}{\|X\|_F} \lesssim 5.4664 \cdot 10^{-4},$$

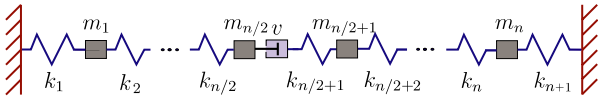


Fig. 1. The n -mass oscillator with one damper.

respectively. On the other hand, the perturbation bound (4.29) gives

$$\frac{\|\delta X\|_F}{\|X\|_F} \lesssim 4.0402 \cdot 10^{-7}.$$

As in the diagonalizable case, the above example shows that the structure of the matrices F and G from the Sylvester equation (2.1) can sometimes greatly influence the perturbation of the solution. \diamond

So far all examples were artificially designed for the purpose of illustration. The following example, however, is a more realistic one, and also a more complex one. We will consider the single-input and single-output system (SISO) which corresponds to the n -mass oscillator (or oscillator ladder) as in Fig. 1. As the input function $u(t)$, we will apply the force $f(t)$ to the k th mass, while the output function $y(t)$ will be velocity of the k th mass.

This example is to explain how the bound (3.11) can be used to guide us to determine very effective ADI parameters for efficiently computing the cross gramian (see [7]), in this case the solution to a Sylvester equation. This is similar to the idea presented in [28].

The similar approach can be used in gramian-based reduction methods [34,7] or balanced truncation model reduction [35,9,36], that is for the corresponding projection subspaces one can use so-called ADI subspace (for more details see [23,22]).

Example 4.2. Consider an n -mass oscillator (or oscillator ladder) as in Fig. 1:

$$M = \text{diag}(m_1, m_2, \dots, m_n), \quad (4.30)$$

$$K = \begin{pmatrix} k_1 + k_2 & -k_2 & & & & \\ -k_2 & k_2 + k_3 & -k_3 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & -k_{n-1} & k_{n-1} + k_n & -k_n \\ & & & & -k_n & k_n + k_{n+1} \end{pmatrix}. \quad (4.31)$$

Let $n = 2000$ be the dimension of the oscillator ladder, where $m_i = i$ for $i = 1, \dots, n - 31$, $m_i = 10 \cdot i$ ($i = 1960$) for $i = n - 30, \dots, 2000$ and $k_i = 1$, for $i = 1, \dots, n$. For the position of input and output function we will take $p_o = n$. Furthermore, we will apply the external damping on the mass m_{1000} (mass at the position 1000).

The cross gramian X is the solution of the Sylvester equation (see [7])

$$AX + XA + bc^T = 0, \quad (4.32)$$

where A is the matrix from the corresponding SISO system given in modal coordinates:

$$A = \Omega_1 \oplus \Omega_2 \oplus \dots \oplus \Omega_n + D,$$

$$\Omega_i = \begin{pmatrix} 0 & \omega_i \\ -\omega_i & -\alpha\omega_i \end{pmatrix}, \quad i = 1, \dots, n,$$

$$\Phi^T K \Phi = \Omega^2 = \text{diag}(\omega_1^2, \dots, \omega_n^2) \quad \text{and} \quad \Phi^T M \Phi = I,$$

$$D = \begin{pmatrix} D_{11} & D_{12} & \dots & D_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ D_{1n} & D_{2n} & \dots & D_{nn} \end{pmatrix},$$

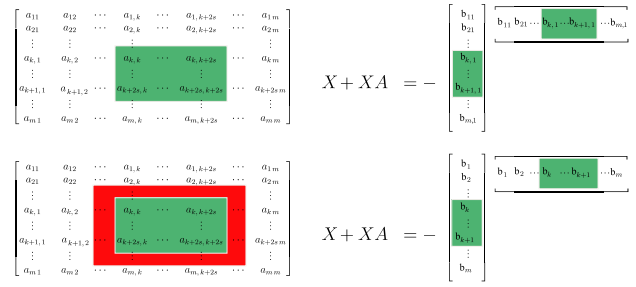


Fig. 2. ADI shift determination.

$$D_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & -v \cdot \Phi_{(1000,i)} \Phi_{(1000,j)} \end{pmatrix},$$

$$b = c = \Phi_{(k_o, \cdot)}^T.$$

Here α is the internal damping constant, and D presents the external damping and corresponds to the damper with viscosity v applied to the mass m_{1000} .

In this example we will take $v = 1$ and $\alpha = 0.01$ (recall that $p_o = 2000$).

We use the LR-ADI method to solve Sylvester equation (4.32). For this we will need a set of ADI parameters. One possible choice of ADI parameters is given in [22], which is an easily implementable extension of Penzl's suboptimal choice [25,26] for Lyapunov equations. On the other hand we will propose a different choice of ADI parameters inspired by the similar one from [28] for Lyapunov equations.

We note that the problem of choosing ADI parameters is a widely open problem. A detailed discussion on the issue is out of scope of this paper. The interested reader is referred to, e.g., [37,28,23,22].

Our choice of ADI parameters is based on the application of the bound (3.11). Note that the vector b has structure such that $b_i = 0$ for $i = 1, \dots, n$, that is the first n components of the vector b are equal to zero. Further, if some part of the vector b has components smaller than a given tolerance, that is if index fin exists ($n/2 < fin$), such that $|b_i|$ are smaller than the given tolerance, for $i = fin + 1, \dots, 2n$, then we can choose the shifts λ_{p_i} such that $\sigma(j, k) = 0$ for $n/2 < j \leq fin$, $k = n + 1, \dots, fin$. Recall $\sigma(j, k)$ is defined as

$$\sigma(i, j - 1) = \prod_{s=1}^{j-1} \frac{\lambda_i - \lambda_{p_s}}{\lambda_i - \mu_{q_s}}, \quad \sigma(i, 0) = 1, \quad i = 1, \dots, 2n,$$

thus $\sigma(i, j - 1) = 0$ if $p_i = i$, for all i and j . Thus, the error bound (3.11) will be almost minimal if we choose ADI parameters $\lambda_{p_i} = \lambda_i$ for $n/2 < i \leq fin$. The problem here is that one does not know the exact eigenvalues. Thus similar to as in [28], we propose the following choice of ADI parameters, which will approximate appropriate part of the spectrum.

- (1) Find the indices of elements with magnitude greater than the given tolerance in the vector b .
- (2) Find the corresponding submatrix of A using this indices.
- (3) Take a "little bit bigger block" A_{block} which include the upper submatrix.
- (4) The eigenvalues of the chosen matrix A_{block} are the ADI parameters (see Fig. 2).

In this example we get the following residuals:

$$\|AY + YA + bc^T\| = 1.2498 \cdot 10^{-6},$$

$$\|AX + XA + bc^T\| = 1.3152 \cdot 10^{-13}.$$

Here Y is obtained using LR-ADI method (2.5) with 40 shifts generated by the algorithm from [22], while X is obtained using LR-ADI method (2.5) with 40 shifts generated by the proposed algorithm above. \diamond

5. Concluding remarks

We have analyzed the solution to a general Sylvester equation $AX - XB = GF^*$ with a low-rank right-hand side. LR-ADI with the exact shifts provides us the tool to do so. Our new results contain considerably more detailed information on the eigen-properties of A and B and the right-hand side GF^* as opposed to the existing ones. Because of this, our new bounds are sharper and provides better understanding of the solution structure, but are messier as a tradeoff.

Although we tackled the general case by considering when A and B have Jordan blocks of orders only up to 2, the technique is readily applicable to Jordan blocks of orders higher than 2 with little changes.

Acknowledgements

The authors would like to thank the referees for their remarks which helped to clarify some of our statements and arguments and improve the quality of the paper.

References

- [1] R. Bhatia, P. Rosenthal, How and why to solve the operator equation $AX - XB = Y$, *Bull. London Math. Soc.* 29 (1997) 1–21.
- [2] G.H. Golub, C.F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, Maryland, 1996.
- [3] V. Sima, *Algorithms for Linear-Quadratic Optimization*, in: *Pure and Applied Mathematics*, vol. 200, Marcel Dekker, Inc, New York, NY, 1996.
- [4] B. Datta, *Numerical Methods for Linear Control Systems*, Elsevier Academic Press, 2004.
- [5] A. Locatelli, *Optimal Control: An Introduction*, Birkhäuser, Basel, Boston, Berlin, 2001.
- [6] R. Aldhaheri, Model order reduction via real Schur-form decomposition, *Internat. J. Control* 53 (3) (1991) 709–716.
- [7] A.C. Antoulas, Approximation of large-scale dynamical systems, in: *Advances in Design and Control*, SIAM, Philadelphia, PA, 2005.
- [8] U. Baur, P. Benner, Cross-gramian based model reduction for data-sparse systems, *Electron. Trans. Numer. Anal. (ETNA)* 31 (2008) 256–270.
- [9] D. Sorensen, A. Antoulas, The Sylvester equation and approximate balanced reduction, *Linear Algebra Appl.* 351/352 (2002) 671–700.
- [10] C.H. Choi, A.J. Laub, Efficient matrix-valued algorithm for solving stiff Riccati differential equations, *IEEE Trans. Automat. Control* 35 (1990) 770–776.
- [11] L. Dieci, Numerical integration of the differential Riccati equation and some related issues, *SIAM J. Numer. Anal.* 29 (1992) 781–815.
- [12] W. Enright, Improving the efficiency of matrix operations in the numerical solution of stiff ordinary differential equations, *ACM Trans. Math. Software* 4 (1978) 127–136.
- [13] R.H. Bartels, G.W. Stewart, Algorithm 432: The solution of the matrix equation $AX - BX = C$, *Commun. ACM* 8 (1972) 820–826.
- [14] G.H. Golub, S. Nash, C.F. Van Loan, Hessenberg–Schur method for the problem $AX + XB = C$, *IEEE Trans. Automat. Control* AC-24 (1979) 909–913.
- [15] J. Roberts, Linear model reduction and solution of the algebraic Riccati equation by use of the sign function, *Internat. J. Control* 32 (1980) 677–687. Reprint of Technical Report No. TR-13, CUED/B-Control, Cambridge University, Engineering Department, 1971.
- [16] R.A. Smith, Matrix Equation $XA + BX = C$, *SIAM J. Appl. Math.* 16 (1) (1968) 198–201.
- [17] P. Benner, Factorized solution of Sylvester equations with applications in control, in: *Proc. Intl. Symp. Math. Theory Networks and Syst. MTNS 2004 (CD-ROM)*, Leuven, Belgium, 2004 (10 pages).
- [18] U. Baur, Low rank solution of data-sparse Sylvester equations, in: *Numer. Lin. Alg. Appl. Special Issue on Large-Scale Matrix Equations of Special Type*, *Numer. Linear Algebra Appl.* 15 (9) (2008) 837–851.
- [19] M. Konstantinov, D.-W. Gu, V. Mehrmann, P. Petkov, *Perturbation Theory for Matrix Equations*, Elsevier Science B.V, 2003.
- [20] Nicholas J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 1996.
- [21] G.W. Stewart, Ji-guang Sun, *Matrix Perturbation Theory*, Academic Press, Harcourt Brace Jovanovich, 1990.
- [22] P. Benner, R.-C. Li, N. Truhar, On ADI method for Sylvester equations, *J. Comput. Appl. Math.* 233 (2009) 1035–1045.
- [23] N. Truhar, R.-C. Li, On ADI method for Sylvester equations, Technical Report 2008-02, Department of Mathematics, University of Texas at Arlington, TX, USA, 2008.
- [24] J.-R. Li, J. White, Low-rank solution of Lyapunov equations, *SIAM J. Matrix Anal. Appl.* 24 (2002) 260–280.
- [25] T. Penzl, A cyclic low-rank Smith method for large sparse Lyapunov equations, *SIAM J. Sci. Comput.* 21 (2000) 1401–1418.
- [26] T. Penzl, LYAPACK: A MATLAB toolbox for large Lyapunov and Riccati equations, model reduction problems, and linear-quadratic optimal control problems, users' guide (ver. 1.0). 2000. Available at www.tu-chemnitz.de/sfb393/lyapack/.
- [27] E. Wachspress, Trail to a Lyapunov equation solver, *Comput. Math. Appl.* 55 (2008) 1653–1659.
- [28] N. Truhar, K. Veselić, Bounds on the trace of solution to the Lyapunov equation with a general stable matrix, *Systems Control Lett.* 56 (2007) 493–503.
- [29] A.C. Antoulas, D.C. Sorensen, Y. Zhou, On the decay rate of hankel singular values and related issues, *Systems Control Lett.* 46 (2002) 323–342.
- [30] T. Penzl, Eigenvalue decay bounds for solutions of Lyapunov equations: The symmetric case, *Systems Control Lett.* 40 (2000) 139–144.
- [31] T.F. Chan, D.E. Foulser, Effectively well-conditioned linear systems, *SIAM J. Sci. Stat. Comput.* 9 (6) (1988) 963–969.
- [32] N. Truhar, The perturbation bound for the solution of the Lyapunov equation, *Math. Commun.* 12 (1) (2007) 83–94.
- [33] G. Strang, *Linear Algebra and Its Applications*, third ed., Harcourt Brace Jovanovich, Philadelphia, PA, 1988.
- [34] F.D. Freitas, J. Rommes, N. Martins, Gramian-based reduction method applied to large sparse power system descriptor models, *IEEE Trans. Power Syst.* 23 (3) (2008) 1258–1270.
- [35] M. Heinkenschloss, D.C. Sorensen, K. Sun, Balanced truncation model reduction for a class of descriptor systems with application to the Oseen equations, *SIAM J. Sci. Comput.* 30 (2) (2008) 1038–1063.
- [36] S. Gugercin, D. Sorensen, A. Antoulas, A modified low-rank Smith method for large-scale Lyapunov equations, *Numer. Algorithms* 32 (2003) 27–55.
- [37] E.L. Wachspress, Iterative solution of the Lyapunov matrix equation, *Appl. Math. Lett.* 1 (1988) 87–90.