Bounds on the trace of a solution to the Lyapunov equation with a general stable matrix*

Ninoslav Truhar †and Krešimir Veselić‡

Abstract

Some new estimates for the eigenvalue decay rate of the Lyapunov equation $AX + XA^T = B$ with a low rank right-hand side B are derived. The new bounds show that the right-hand side B can greatly influence the eigenvalue decay rate of the solution. This suggests a new choice of the ADI-parameters for the iterative solution. The advantage of these new parameters is illustrated on second order damped systems with a low rank damping matrix.

Keywords: Lyapunov equation, eigenvalue decay rate, ADI-parameters.

1 Introduction

In this paper we consider two related topics concerning the solution of the continuous-time Lyapunov equation

$$AX + XA^T = -GG^T\,, \tag{1.1}$$

*part of this work was done while the author was visiting University of Kentucky, Department of Mathematics, Lexington, Kentucky, USA under the support of Postdoctoral Research Award from the National Foundation for Science, Higher Education and Technological Development of the Republic of Croatia.

†University of Osijek, Department of Mathematics, 31000 Osijek, Croatia, *email:ntruhar@mathos.hr*

‡Lehrgebiet Mathematische Physik, Fernuniversität, 58084 Hagen, Germany *kresimir.veselic@fernuni-hagen.de*

where $A \in \mathbb{R}^{m\times m}$ is assumed to be stable and $G \in \mathbb{R}^{m\times s}$ with $\operatorname{rank}(G) = s \ll m$.

Our first result contains a bound for the difference between the traces of the solution X of Lypunov equation (1.1) and its ADI approximation. From this bound it can be seen that the right-hand side of Lyapunov equation can sometimes greatly influence the eigenvalue decay rate of the solution. The second result contains a proposal for a new suboptimal choice of ADI parameters suggested by the aforementioned new error bound.

2 Bound for eigenvalue decay rate

We consider the following Lyapunov equation

$$AX + XA^T = -GG^T\,,$$

where $A \in \mathbb{R}^{m\times m}$ is stable, and $G \in \mathbb{R}^{m\times s}$ with $\operatorname{rank}(G) = s \ll m$.

The main result of this section is a new bound which generalizes bounds obtained by Antoulas, Sorensen and Zhou [1], Sorensen and Zhou [8] and Penzl [5] and [7].

We start with a simple generalization of the first step, from the analysis in [8] and [7] by dropping the assumption on diagonalizability. ADI iterates [14] $(X_l)_{l=0}^{\infty}$ generated by an initial matrix X_0 are

$$X_l = s_{p_l}(A)X_{l-1}s_{p_l}^*(A) - 2\operatorname{Re}(p_l)\,(A+\bar{p}_l I)^{-1}GG^T(A+\bar{p}_l I)^{-*}\,, \qquad (2.2)$$

where $s_{p_l}(A) = (A - p_l I)(A + \bar{p}_l I)^{-1}$. Since the solution X of (1.1) satisfies the Stein equation (for more details see [8])

$$X = s_{p_l}(A)Xs_{p_l}^*(A) - 2\operatorname{Re}(p_l)\,(A+\bar{p}_l I)^{-1}GG^T(A+\bar{p}_l I)^{-*}\,,$$

one can see that X is a stationary point of mapping (2.2). Hence

$$X - X_l = s_{p_l}(A)(X - X_{l-1})s_{p_l}^*(A)\,,$$

By choosing $X_0 = 0$, we obtain

$$X - X_l = s_{\{p_1,\ldots,p_l\}}(A)\,X\,s_{\{p_1,\ldots,p_l\}}^*(A)\,,$$

where

$$s_{\{p_1,\ldots,p_l\}}(A) = \prod_{k=1}^{l}(A - p_k I)(A + \bar{p}_k I)^{-1}\,.$$

If we set ADI parameters $\{p_1, \ldots, p_m\}$ to be equal to the eigenvalues of the matrix A, that is $\{p_1, \ldots, p_m\} = \sigma(A)$ (counted according to their multiplicities), then from Hamilton-Cayley theorem it follows that

$$X - X_m = 0\,. \tag{2.3}$$

The above fact will be used later in our investigations.

Since the right-hand side of the Lyapunov equation (1.1) has the rank $s \ll m$, for the approximation of the solution X of (1.1) we will use the Low Rank Cholesky Factor ADI (LRCF-ADI) algorithm (see [5] or [3]) which has the following form:

Algorithm 1 (Low rank Cholesky factor ADI (LRCF-ADI))

INPUT: A, G, $\{p_1, p_2, \ldots, p_l\}$

OUTPUT: $V = V_l \in \mathbb{C}^{m \times sl}$, such that $VV^* \approx X$.

1. $W_1 = \sqrt{-2\mathrm{Re}(p_1)}\,(A + p_1 I)^{-1} G$
2. $V = W_1$

FOR: $i = 2, 3, \ldots, l$

3. $W_i = \sqrt{\mathrm{Re}(p_i)/\mathrm{Re}(p_{i-1})}\;(W_{i-1} - (p_i + \bar{p}_{i-1})(A + p_i I)^{-1} W_{i-1})$
4. $V_i = [V_{i-1}\,, W_i]$

END

The l-th approximation X_l of the solution X can be written as

$$X_l = \sum_{j=1}^{l} W_j W_j^*\,. \tag{2.4}$$

From (2.4) it follows that the trace of the l-th approximation X_l is equal to

$$\mathrm{tr}(X_l) = \sum_{j=1}^{l} \mathrm{tr}(W_j W_j^*) = \sum_{j=1}^{l} \mathrm{tr}(W_j^* W_j) = \sum_{j=1}^{l} \sum_{i=1}^{s} \|W_j(:, i)\|^2\,, \tag{2.5}$$

where $\|\cdot\|$ represents a 2-norm.

From (2.5) and (2.3) it follows that the trace of the solution X can be written as (this holds for $\{p_1, \ldots, p_m\} = \sigma(A)$)

$$\mathrm{tr}(X) = \sum_{j=1}^{m} \sum_{i=1}^{s} \|W_j(:, i)\|^2\,. \tag{2.6}$$

Using

$$\|W_j\|_F^2 = \sum_{i=1}^{s} \|W_j(:,i)\|^2 \,,$$

(2.5) can be written as

$$\mathrm{tr}(X_l) = \sum_{j=1}^{l} \|W_j\|_F^2 \,, \tag{2.7}$$

and similarly, (2.6) can be written as

$$\mathrm{tr}(X) = \sum_{j=1}^{m} \|W_j\|_F^2 \,. \tag{2.8}$$

From (2.7) and (2.8) it follows that

$$\mathrm{tr}(X) - \mathrm{tr}(X_l) = \sum_{j=l+1}^{m} \|W_j\|_F^2 \,. \tag{2.9}$$

The main task in this section is to derive a bound for (2.9). Once we obtain the bound for (2.9) we will be able to bound the relative error for the solution of the Lyapunov equation (1.1) by simply using the inequality

$$\frac{\|X - X_l\|}{\|X\|} \leq \frac{\mathrm{tr}(X) - \mathrm{tr}(X_l)}{\mu_1} \,, \tag{2.10}$$

where $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_m$ denote eigenvalues of the matrix X.

The above inequality follows from the fact that $\|W_k W_k^*\| \leq \mathrm{tr}(W_k W_k^*)$, which implies $\|X - X_l\| \leq \mathrm{tr}(X) - \mathrm{tr}(X_l)$.

Also, as it has been shown in [8] or [5], the left-hand side in (2.10) is the upper bound for the eigenvalue decay, that is

$$\frac{\mu_{s\,l+1}}{\mu_1} \leq \frac{\|X - X_l\|}{\|X\|} \leq \frac{\mathrm{tr}(X) - \mathrm{tr}(X_l)}{\mu_1} \,, \tag{2.11}$$

where $s\,l < m$.

The first of our results concerns a diagonalizable matrix A, thus let

$$A = S\Lambda S^{-1}; \quad S \in \mathbb{R}^{m\times m} \,, \quad \Lambda = \mathrm{diag}\{\lambda_1, \ldots \lambda_m\} \tag{2.12}$$

be the eigenvalue decomposition of the matrix A.

If we write

$$\widehat{G} = S^{-1}G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1s} \\ g_{21} & g_{22} & \cdots & g_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ g_{m1} & g_{m2} & \cdots & g_{ms} \end{bmatrix} = \begin{bmatrix} \widehat{g}_1 \\ \widehat{g}_2 \\ \vdots \\ \widehat{g}_m \end{bmatrix} \tag{2.13}$$

($\widehat{g}_i$ denotes the i-th row of the matrix $\widehat{G}$), then the following theorem contains the bound for (2.9).

Theorem 2.1 *Let X_l be the l-th approximation obtained by* **Algorithm 1** *with the set of ADI parameters corresponding to any subset of the exact eigenvalues of the matrix A (i.e. $\{p_1, p_2, \ldots, p_l\} = \{\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_l}\}$). Then the following bound holds:*

$$\mathrm{tr}(X) - \mathrm{tr}(X_l) \leq \|S\|^2 \sum_{j=l+1}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=1}^{m} |\sigma(j,k)|^2 \cdot \|\widehat{g}_k\|^2 \,, \tag{2.14}$$

where

$$\sigma(1,k) = \frac{1}{\lambda_k + p_1} \,, \quad \textit{and} \quad \sigma(j,k) = \frac{1}{\lambda_k + p_j} \prod_{t=1}^{j-1} \frac{\lambda_k - \bar{p}_t}{\lambda_k + p_t} \quad \textit{for} \quad j > 1 \,. \tag{2.15}$$

Proof. From (2.9) it follows that we have to bound $\|W_j\|_F^2$ for $j = l+1, \ldots, m$. By **Algorithm 1** we can write:

$$\begin{aligned} W_j &= \sqrt{\mathrm{Re}(p_j)/\mathrm{Re}(p_{j-1})}\, \left(I - (p_j + \bar{p}_{j-1})(A + p_j I)^{-1}\right) W_{j-1} \,, \\ W_{j-1} &= \sqrt{\mathrm{Re}(p_{j-1})/\mathrm{Re}(p_{j-2})}\, \left(I - (p_{j-1} + \bar{p}_{j-2})(A + p_{j-1} I)^{-1}\right) W_{j-2} \\ &\cdots \\ W_2 &= \sqrt{\mathrm{Re}(p_2)/\mathrm{Re}(p_1)}\, \left(I - (p_2 + \bar{p}_1)(A + p_2 I)^{-1}\right) W_1 \,, \\ W_1 &= \sqrt{-2\mathrm{Re}(p_1)}\, (A + p_1 I)^{-1} G \,, \end{aligned}$$

where $p_j = \lambda_{k_j}$ is one of the exact eigenvalues of the matrix A. Together, the above equalities give:

$$\begin{aligned} W_j &= \sqrt{-2\,\mathrm{Re}(p_j)}\, \left(I - (p_j + \bar{p}_{j-1})(A + p_j I)^{-1}\right) \cdot \left(I - (p_{j-1} + \bar{p}_{j-2})(A + p_{j-1} I)^{-1}\right) \cdots \\ &\quad \cdots \left(I - (p_2 + \bar{p}_1)(A + p_2 I)^{-1}\right) (A + p_1 I)^{-1} G \,. \end{aligned} \tag{2.16}$$

Now using the eigenvalue decomposition (2.12) it is easy to show that (2.16) can be written as:

$$\begin{aligned} W_j &= \sqrt{-2\,\mathrm{Re}(p_j)}\,S\cdot\left(I-(p_j+\bar{p}_{j-1})(\Lambda+p_jI)^{-1}\right)\cdot\left(I-(p_{j-1}+\bar{p}_{j-2})(\Lambda+p_{j-1}I)^{-1}\right)\cdots \\ &\quad \cdots\left(I-(p_2+\bar{p}_1)(\Lambda+p_2I)^{-1}\right)(\Lambda+p_1I)^{-1}\widehat{G}\,, \end{aligned}$$

Note that the above equality contains $j-1$ diagonal matrices of the form

$$\left(I-(p_k+\bar{p}_{k-1})(\Lambda+p_kI)^{-1}\right)=\mathrm{diag}\left(\frac{\lambda_i-\bar{p}_{k-1}}{\lambda_i+p_k}\right)_i \quad i=1,\ldots,m\,,\quad k=2,\ldots,j\,.$$

Thus, the matrix W_j has the following form:

$$W_j=\sqrt{-2\,\mathrm{Re}(p_j)}\,S\cdot\begin{bmatrix}\sigma(j,1)\cdot\widehat{g}_1\\ \sigma(j,2)\cdot\widehat{g}_2\\ \vdots\\ \sigma(j,m)\cdot\widehat{g}_m\end{bmatrix},$$

where $\sigma(j,k)$ are defined as

$$\sigma(1,k)=\frac{1}{\lambda_k+p_1}\,,\quad\text{and}\quad \sigma(j,k)=\frac{1}{\lambda_k+p_1}\prod_{t=2}^{j}\frac{\lambda_k-\bar{p}_{t-1}}{\lambda_k+p_t}\quad\text{for}\quad j>1\,.$$

Now it is easy to show that $\sigma(j,k)$ defined above are the same as the ones defined in (2.15).

Using the inequality $\|AB\|_F\le\|A\|\|B\|_F$ we can write:

$$\|W_j\|_F^2 \;\le\; (-2\,\mathrm{Re}(p_j))\,\|S\|^2\cdot\sum_{k=1}^{m}|\sigma(j,k)|^2\cdot\|\widehat{g}_k\|^2\,, \tag{2.17}$$

here we have used the fact that $\|\widehat{g}_k\|_F^2=\|\widehat{g}_k\|^2$ since $\widehat{g}$ is a row-vector.

Now bound (2.14) is obtained simply by summing all terms from the right-hand side in (2.17) for $j=l+1,\ldots,m$. ∎

The above theorem can be generalized to any stable matrix A. For the sake of simplicity, we will consider the matrix A whose Jordan blocks are at most 2×2.

Let the Jordan canonical form of the matrix A be

$$A=SJS^{-1};\quad S\in\mathbb{C}^{m\times m}\,,\quad J=J_1\oplus\ldots\oplus J_{k_0}\,, \tag{2.18}$$

where $J_i \oplus J_k$ stands for a direct sum of J_i and J_k and each J_i, $i = 1, \ldots, k_0$ corresponds to subspaces associated with the eigenvalue λ_i, with the following structure

$$J_i = [\lambda_i] \qquad \text{for} \quad i = 1, \ldots, n_0\,,$$
$$J_i = \begin{bmatrix} \lambda_i & 1 \\ 0 & \lambda_i \end{bmatrix}, \quad \text{or} \quad J_i = \lambda_i I + N\,, \quad \text{for} \quad i = n_0 + 1, \ldots, k_0\,,$$

where I is 2×2 identity matrix and

$$N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \text{is nilpotent of order 2.}$$

Let matrix

$$\widehat{G} = S^{-1}G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1s} \\ g_{21} & g_{22} & \cdots & g_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ g_{k_0 1} & g_{k_0 2} & \cdots & g_{k_0 s} \end{bmatrix} = \begin{bmatrix} \widehat{g}_1 \\ \widehat{g}_2 \\ \vdots \\ \widehat{g}_{k_0} \end{bmatrix} \tag{2.19}$$

be partitioned according to the Jordan structure of the matrix A, that is for $i = 1, \ldots, n_0$, $\widehat{g}_i$ denotes the i-th $1 \times s$, and for $i = n_0 + 1, \ldots, k_0$, the i-th $2 \times s$, submatrix of the matrix $\widehat{G}$, respectively.

The following theorem contains the bound for the difference of traces (2.9).

Theorem 2.2 *Let X_l be the l-th approximation obtained by* **Algorithm 1** *with the set of ADI parameters corresponding to any subset of exact eigenvalues of the matrix A (i.e. $\{p_1, p_2, \ldots, p_l\} = \{\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_l}\}$). Then the following bound holds:*

$$\mathrm{tr}(X) - \mathrm{tr}(X_l) \leq \|S\|^2 \sum_{j=l+1}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=1}^{k_0} \|\eta(j,k)\|^2 \cdot \|\widehat{g}_k\|_F^2\,, \tag{2.20}$$

where

$$\eta(j,k) = \sigma(j,k) \quad \textit{for} \quad k = 1, \ldots n_0\,, \tag{2.21}$$

$$\eta(j,k) = (\sigma(j,k)I - \mu(j,k)\,N)\,\left(I - (\lambda_k + p_1)^{-1}N\right) \in \mathbb{C}^{2\times 2} \tag{2.22}$$
$$\textit{for } k = n_0 + 1, \ldots k_0\,,$$

$$\sigma(1,k) = \frac{1}{\lambda_k + p_1}\,, \quad \textit{and} \quad \sigma(j,k) = \frac{1}{\lambda_k + p_j} \prod_{t=1}^{j-1} \frac{\lambda_k - \bar{p}_t}{\lambda_k + p_t} \quad \textit{for} \quad j > 1\,. \tag{2.23}$$

and for $k = n_0 + 1, \ldots k_0$

$$\mu(j,k) = \sum_{l=2}^{j} \frac{1}{\lambda_k + p_1} \prod_{\substack{t=2 \\ t\neq l}}^{j} \frac{\lambda_k - \bar{p}_{t-1}}{\lambda_k + p_t} \frac{p_l + \bar{p}_{l-1}}{(\lambda_k + p_l)^2}\,. \tag{2.24}$$

Proof. The first part of the proof is similar to the first part of the proof of Theorem 2.1. Thus, we continue from equality (2.16):

$$\begin{aligned} W_j &= \sqrt{-2\,\mathrm{Re}(p_j)}\,\left(I - (p_j + \bar{p}_{j-1})(A + p_j I)^{-1}\right) \cdot \left(I - (p_{j-1} + \bar{p}_{j-2})(A + p_{j-1} I)^{-1}\right) \cdots \\ &\quad \cdots \left(I - (p_2 + \bar{p}_1)(A + p_2 I)^{-1}\right) (A + p_1 I)^{-1} G\,. \end{aligned}$$

Now using the Jordan canonical form of the matrix A (2.18) it is easy to show that the above equality can be written as:

$$\begin{aligned} W_j &= \sqrt{-2\,\mathrm{Re}(p_j)}\, S \cdot \left(I - (p_j + \bar{p}_{j-1})(J + p_j I)^{-1}\right) \cdot \left(I - (p_{j-1} + \bar{p}_{j-2})(J + p_{j-1} I)^{-1}\right) \cdots \\ &\quad \cdots \left(I - (p_2 + \bar{p}_1)(J + p_2 I)^{-1}\right) (J + p_1 I)^{-1} \widehat{G}\,, \end{aligned} \tag{2.25}$$

Note that for n_0 leading 1×1 Jordan blocks the structure of the matrix W_j from (2.25) is equal to the structure of the matrix W_j from (2.16). Thus, we will consider only $n_0 + 1, \ldots, k_0$, 2×2 Jordan blocks.

Note that on the right-hand side of (2.25) we have a product of $j - 1$ block-diagonal matrices of the following form

$$I - (p_k + \bar{p}_{k-1})(J + p_k I)^{-1} \quad k = 2, \ldots j\,. \tag{2.26}$$

The i-th diagonal block is given by

$$I - (p_k + \bar{p}_{k-1})(J_i + p_k I)^{-1} = I - (p_k + \bar{p}_{k-1})\,(\lambda_i I + N + p_k I)^{-1} \quad i = n_0 + 1, \ldots, k_0\,.$$

Now using

$$(\lambda_i I + N + p_k I)^{-1} = \frac{1}{\lambda_i + p_k} \left(I - \frac{1}{\lambda_i + p_k} N\right)$$

one can see that the i-th diagonal block of the k-th matrix (2.26) can be written as

$$I - (p_k + \bar{p}_{k-1})\,(\lambda_i I + N + p_k I)^{-1} = \frac{\lambda_i - \bar{p}_{k-1}}{\lambda_i + p_k}\, I - \frac{p_k + \bar{p}_{k-1}}{(\lambda_i + p_k)^2}\, N\,. \quad (2.27)$$

Multiplying together all matrices (2.27) for $k = n_0 + 1, \dots k_0$ we get the block-diagonal matrix whose i-th block has the form:

$$\left(\frac{\lambda_i - \bar{p}_{j-1}}{\lambda_i + p_j}\, I - \frac{p_j + \bar{p}_{j-1}}{(\lambda_i + p_j)^2}\, N\right) \cdots \left(\frac{\lambda_i - \bar{p}_1}{\lambda_i + p_2}\, I - \frac{p_2 + \bar{p}_1}{(\lambda_i + p_2)^2}\, N\right) \cdot \left(\frac{1}{\lambda_i + p_1}\, I - \frac{1}{(\lambda_i + p_1)^2}\, N\right)$$

Using the fact that $N^2 = 0$, for $j \geq 2$ the above expression can be written as

$$\left(\frac{1}{\lambda_i + p_1} \prod_{t=2}^{j} \frac{\lambda_i - \bar{p}_{t-1}}{\lambda_i + p_t} - \sum_{l=2}^{j} \frac{1}{\lambda_i + p_1} \prod_{\substack{t=2 \\ t \neq l}}^{j} \frac{\lambda_i - \bar{p}_{t-1}}{\lambda_i + p_t}\, \frac{p_l + \bar{p}_{l-1}}{(\lambda_i + p_l)^2}\right) \cdot \left(I - \frac{1}{\lambda_i + p_1}\, N\right)\,.$$

Thus, one can see that the matrix W_j has the following form:

$$W_j = \sqrt{-2\,\mathrm{Re}(p_j)}\, S \cdot \begin{bmatrix} \eta(j,1) \cdot \widehat{g}_1 \\ \eta(j,2) \cdot \widehat{g}_2 \\ \vdots \\ \eta(j,k_0) \cdot \widehat{g}_{k_0} \end{bmatrix}\,,$$

where $\eta(j,k)$ is defined as in (2.21) and (2.24), and $\widehat{g}$ as in (2.19).

Similarly to the proof of Theorem 2.1 we have

$$\|W_j\|_F^2 \;\leq\; (-2\,\mathrm{Re}(p_j))\, \|S\|^2 \cdot \sum_{k=1}^{k_0} \|\eta(j,k) \cdot \widehat{g}_k\|_F^2\,. \quad (2.28)$$

Now bound (2.20) is obtained using the fact that $\|\eta(j,k) \cdot \widehat{g}_k\|_F \leq \|\eta(j,k)\| \cdot \|\widehat{g}_k\|_F$ and then summing all inequalities for $j = l+1, \dots, m$. ∎

Note that Theorem 2.2 is a proper generalization of Theorem 2.1, since the assumption on diagonalizability of A implies $n_0 = m$, that is all J_i are one-dimensional blocks, which insures that bound (2.20) is equal to bound (2.14).

We will now compare our bound with the bounds from [1] and [8]. For that purpose, note that if we assume a certain order for eigenvalues, $\lambda_1, \dots, \lambda_m$,

of the matrix A, and if we choose ADI parameters $p_1, \ldots, p_m$ to be their conjugate values in the same order, that is $p_i = \bar{\lambda}_i$, for $i = 1, \ldots m$, then

$$\sigma(j,k) = \frac{1}{\lambda_k + p_j} \prod_{t=1}^{j-1} \frac{\lambda_k - \bar{p}_t}{\lambda_k + p_t} = 0 \qquad \text{for} \qquad k < j\,, \quad j > 1.$$

Then bound (2.14) has the following form

$$\mathrm{tr}(X) - \mathrm{tr}(X_l) \leq \|S\|^2 \sum_{j=l+1}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=j}^{m} \sigma(j,k)^2 \cdot \|\widehat{g}_k\|^2\,, \tag{2.29}$$

where $\sigma(1,k)$ are given by (2.15).

As it has been pointed out in [8], the suboptimal bound

$$\frac{\|X - X_l\|}{\|X\|} \leq \kappa^2(S) \left\{ \max_{1 \leq k \leq m} \prod_{i}^{l} \left| \frac{\lambda_i - \lambda_k}{\bar{\lambda}_i + \lambda_k} \right|^2 \right\} \tag{2.30}$$

is very similar to the approximate rate $\dfrac{\delta_k}{\delta_1}$, since it holds [1, Theorem 3.1.]:

$$\|X - X_l\| \leq (m-l)^2 \kappa^2(S) \delta_{k+1} \|G\|^2\,, \tag{2.31}$$

where

$$\delta_1 = \frac{-1}{2\,Re(\lambda_1)} \qquad \text{and} \qquad \delta_j = \frac{-1}{2\,Re(\lambda_j)} \prod_{i=1}^{j-1} \left| \frac{\lambda_j - \lambda_i}{\bar{\lambda}_j + \lambda_i} \right|^2\,, \tag{2.32}$$

and the eigenvalues of the matrix A are taken according to the so-called Cholesky ordering, that is, $\lambda_1 = \mathrm{argmax}\{-1/(2\,Re(\lambda)) : \lambda \in \lambda(A)\}$ and

$$\lambda_j = \mathrm{argmax} \left\{ \frac{-1}{2\,Re(\lambda)} \prod_{i=1}^{j-1} \left| \frac{\lambda - \lambda_i}{\bar{\lambda} + \lambda_i} \right|^2 \,:\, \lambda \in \lambda(A) \backslash \{\lambda_i : 1 \leq i \leq j-1\} \right\}. \tag{2.33}$$

For comparison we will assume that A is diagonalizable, which means that we will compare bounds (2.32) and (2.30) with our bound (2.14). The following remark will show that bound (2.14) is sharper than bounds (2.32) and (2.30).

REMARK **2.1** *If we assume that eigenvalues of the matrix A, $\lambda_1, \ldots, \lambda_m$, are sorted according to the Cholesky order (2.33), and if we choose the ADI shifts as conjugate eigenvalues, that is $p_i = \bar{\lambda}_i$, then we can write*

$$-2Re(p_1) \cdot |\sigma(1,k)|^2 \leq \delta_1 \quad \textit{for all } k = 1, \ldots, m$$

and

$$-2Re(p_j) \cdot |\sigma(j,k)|^2 \leq \delta_j \quad \textit{for all } k = 1, \ldots, m\,.$$

One can easily prove the above inequalities. Indeed,

$$-2Re(p_1)\cdot|\sigma(1,k)|^2 = -2Re(\lambda_1)\cdot\left(\frac{1}{|\lambda_k + p_1|}\right)^2 = \frac{-2Re(\lambda_1)}{|(\lambda_k + \bar{\lambda}_1)|^2} \leq \frac{-1}{2Re(\lambda_1)} \equiv \delta_1\,,$$

for all $k = 1, \ldots, m$. Further, note that for $p_i = \bar{\lambda}_i$, $i = 1, \ldots m$ expressions

$$-2Re(p_j) \cdot |\sigma(j,k)|^2 = \frac{-2Re(\lambda_j)}{|(\lambda_k + \bar{\lambda}_j)|^2} \cdot \prod_{t=1}^{j-1} \left|\frac{\lambda_k - \lambda_t}{\lambda_k + \bar{\lambda}_t}\right|^2$$

are equal to zero for $k = 1, \ldots, j-1$ (as we have pointed out earlier for such a choice of ADI parameters $\sigma(j,k) = 0$ for $k < j$). Thus, we can write

$$-2Re(p_j) \cdot |\sigma(j,k)|^2 \leq \frac{-1}{2Re(\lambda_j)} \cdot \prod_{t=1}^{j-1} \left|\frac{\lambda_k - \lambda_t}{\lambda_k + \bar{\lambda}_t}\right|^2 \leq \delta_j \text{ for } k = j, \ldots, m\,.$$

For this choice of eigenvalue ordering (2.33), it can be shown that $\delta_1 \geq \ldots \geq \delta_m > 0$ (for details see [1]), where δ_k is defined in (2.32).

Now, from (2.14) it follows

$$\begin{aligned}
\mathrm{tr}(X) - \mathrm{tr}(X_l) &\leq \|S\|^2 \sum_{j=l+1}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=1}^{m} |\sigma(j,k)|^2 \cdot \|\widehat{g}_k\|^2 \\
&\leq \|S\|^2 \|S^{-1}\|^2 \delta_{l+1} \sum_{j=l+1}^{m} \sum_{k=1}^{m} \|g_k\|^2 \\
&= (m-l)\,\kappa(S)^2\,\delta_{l+1}\,\|G\|_F^2 \\
&\leq m(m-l)\,\kappa(S)^2\,\delta_{l+1}\,\|G\|^2\,.
\end{aligned}$$

The last bound is similar to the bound from [1, Theorem 3.1.]. Note that bound (2.14) is sharper than (2.31).

In spite of the fact that our bound (2.20) is more general and sharper than existing bounds, this is not its main advantage. As we will see in the next section, the main advantage of the bounds (2.14) and (2.20) is the fact that these bounds include the influence of the right-hand side on the eigenvalue decay. In the next section we will consider this influence in some detail.

2.1 The influence of the right-hand side on the eigenvalue decay rate of the solution of Lyapunov equation

As it has been already pointed out, in this section we will consider the influence of the right-hand side on the eigenvalue decay rate of the solution of Lyapunov equation (1.1)

$$AX + XA^T = -GG^T\,.$$

Let $\mu_1, \ldots, \mu_m$ be the eigenvalues of the solution X and let $sl < m$. Then from (2.11) and (2.20) it follows

$$\frac{\mu_{s\,l+1}}{\mu_1} \leq \|S\|^2 \sum_{j=l+1}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=1}^{k_0} \frac{\eta(j,k)^2}{\mu_1} \cdot \|\widehat{g}_k\|_F^2\,, \tag{2.34}$$

where η is defined by (2.22) and $\widehat{g}_k$ by (2.19).

Now, one can easily see that the right-hand side of (2.34) strongly depends on $\|\widehat{g}_k\|_F^2$, $k = 1, \ldots, k_0$ (the structure of the matrix $\widehat{G}$ is important). For example, if $\widehat{G}$ has a structure such that

$$\|\widehat{g}_1\|_F \gg \|\widehat{g}_2\|_F \approx \|\widehat{g}_3\|_F \approx \cdots \approx \|\widehat{g}_{k_0}\|_F \approx \sqrt{\varepsilon}$$

then we can choose p_1 and p_2 such that $\eta(j,1) = 0$ for $j > 2$. If $\|S\|$, $\mathrm{Re}(p_j)$ and the rest of $\eta(j,k)$'s have modest magnitudes, then from (2.34) we have

$$\frac{\mu_{s\,l+1}}{\mu_1} \leq O(\varepsilon)\,.$$

Thus we can conclude that, although the right-hand side of bound (2.20) depends on the magnitude of numbers $\eta(j,k)$ defined in (2.22), it also strongly depends on the structure of the eigenvector matrix S defined in (2.18) and the norms of the rows of the matrix $\widehat{G}$ defined in (2.19).

To confirm our statements we will consider a well known example with no eigen-decay case considered in [1]. As it has been pointed out in [1], these no eigen-decay cases are related to all-pass systems. For example, if we take the Lyapunov equation

$$AX + XA^T = -bb^T\,, \tag{2.35}$$

where $A = T - bb^T/2$, $T = -T^T$, and b is a vector ensuring the stability of A, then it is obvious that the solution of the above Lyapunov equation is $X = I$, with no eigen-decay at all.

A slight change of the vector b may cause a "nice" eigen-decay of the solution of a new Lyapunov equation

$$AX + XA^T = -\widetilde{b}\widetilde{b}^T\,, \tag{2.36}$$

with the same matrix A.

To illustrate this, we have constructed a small example which shows dependence of the eigen-decay rate of the solution and the angle between vectors $\widetilde{b}$ and b from (2.36) and (2.35), respectively. Vectors $\widetilde{b}$ are constructed from b such that all components of $\widetilde{b}$ and b are equal excluding the last two. The last two components of $\widetilde{b}$ were set to zero. For the measure of the eigen-decay rate we use the ratio $\mathrm{tr(X)}/\|X\|$, which indicates that if $\mathrm{tr(X)}/\|X\|$ is close to 1, then we have a strong eigen-decay in contrast to the case when $\mathrm{tr(X)}/\|X\|$ is close to matrix dimension.

We have considered 16 randomly generated systems of dimension 400. Table 1 shows all 16 results. The results point out that although the angles between vectors $\widetilde{b}$ and b are small ($\sim 3.5^\circ$), eigen-decay takes place.

$\varphi = \angle(\widetilde{b}, b)$	0.0586	0.0715	0.0477	0.0529	0.0858	0.0540	0.0906	0.0748
$\mathrm{tr(X)}/\|X\|$	19.6502	2.3113	4.7841	22.7262	19.7142	4.8621	18.4565	6.0325
$\varphi = \angle(\widetilde{b}, b)$	0.0801	0.0773	0.0714	0.0896	0.0978	0.0199	0.0683	0.0849
$\mathrm{tr(X)}/\|X\|$	31.4756	10.5892	2.2508	6.6712	13.2327	30.8580	10.8526	2.4981

Table 1: The eigen-decay case

This example shows the importance of the structure of the right-hand side for eigen-decay rate phenomena for solutions of Lyapunov equations with small rank right-hand sides.

The following figure shows eigenvalues of X (with a logaritmic vertical axis) for the best and the worst case of the eigen-decay rate shown in Table 1.

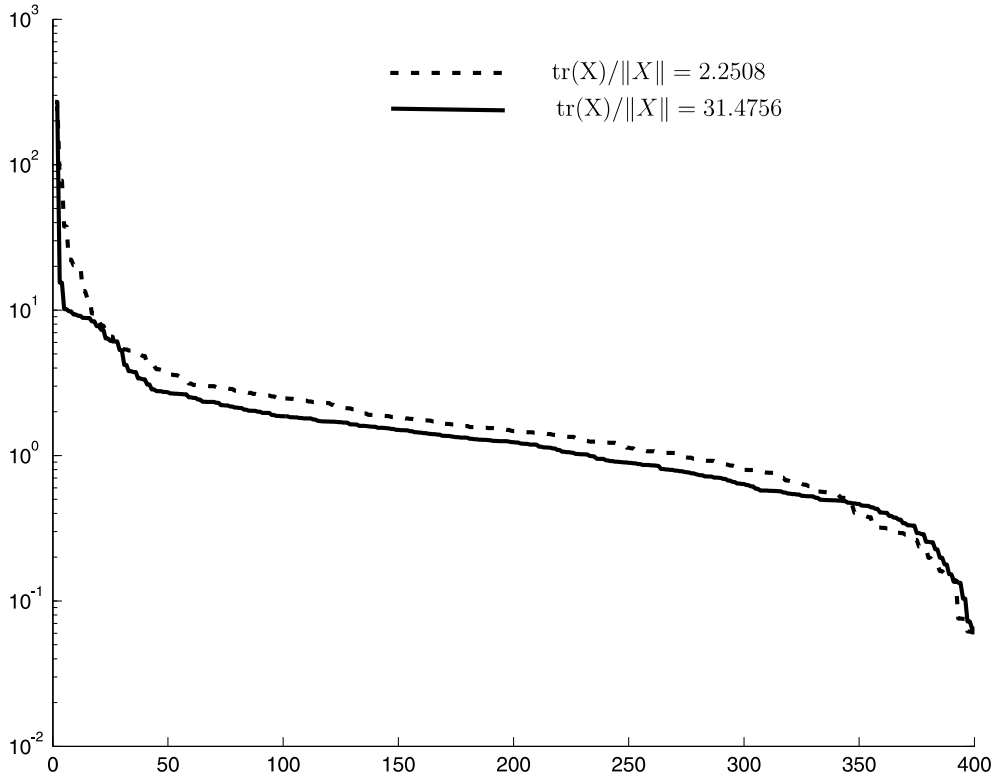

Figure 1: Eigenvalues of X

The next section contains our second result about a suboptimal choice of ADI parameters based on error bounds (2.14) and (2.20).

3 Suboptimal set of ADI parameters

In this section we will describe one possible application of Theorem 2.1 concerning a suboptimal choice of ADI parameters.

It is well known that the efficiency of the ADI method strongly depends on the selection of ADI parameters. Bounds (2.14) and (2.20) point out that

the structure of the right-hand side can be very important for the selection of a proper set of ADI parameters.

Our algorithm is inspired by the following example.

Example 3.1 *Consider Lyapunov equation*

$$AX + XA^T = -GG^T\,,$$

where A has a block structure

$$A = A_1 \oplus \ldots \oplus A_n\,, \quad A_i \in \mathbb{R}^{2\times 2}\,, \qquad \text{for} \quad i = 1, \ldots n.$$

Let, say $G = \begin{bmatrix} I_2 & 0_2 & \ldots & 0_2 \end{bmatrix}^T$, where I_2 and 0_2 denote, a 2×2 identity and a zero matrix, respectively.

Since A has the block structure, then the matrix of eigenvectors S has a similar block structure too. This further means that only the first two rows of the matrix $\widehat{G} = S^{-1}G$ will be non-zero rows while the rest of the rows will be zero rows (i.e. $\|\widehat{g}_k\| = 0$ for $k = 3, \ldots, m$). Now from (2.14) it follows

$$\frac{\|X - X_l\|}{\|X\|} \leq \frac{\|S\|^2}{\mu_1} \cdot \sum_{j=l+1}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=1}^{2} \sigma(j,k)^2 \, \|\widehat{g}_k\|^2\,,$$

where

$$\sigma(j,k) = \frac{1}{\lambda_k + p_j} \prod_{t=1}^{j-1} \frac{\lambda_k - \bar{p}_t}{\lambda_k + p_t}$$

Now, if we choose $p_1 = \bar{\lambda}_1$ and $p_2 = \bar{\lambda}_2$ (note that λ_1 and λ_2 are eigenvalues of the matrix A_1), then $\sigma(j,1) = \sigma(j,2) = 0$, for $j \geq 2$, which insures that the solution X_2 obtained by the LRCF-ADI algorithm with parameters p_1 and p_2 will be equal to the exact solution X, that is

$$\frac{\|X - X_2\|}{\|X\|} \leq \frac{\|S\|^2}{\mu_1} \cdot \sum_{j=3}^{m} (-2\mathrm{Re}(p_j)) \sum_{k=1}^{2} \sigma(j,k)^2 \, \|\widehat{g}_k\|^2 = 0\,.$$

As we have seen in the last example, which considers the so-called "modal damping" case, the optimal set of ADI parameters $\{p_1, \ldots, p_l\}$ is determined by the structure of the matrix GG^T, since A was given in a block-diagonal form. Thus for the determination of a suboptimal set of ADI parameters in a general case (with a general stable A), we propose the following algorithm:

Algorithm 2 (**ADI-parameters**)

1. Find the indices of $1'$s on the right-hand side (i.e. find positions of ones in the matrix GG^T).

2. Find the corresponding submatrix of A using this indexes.

3. Take a "little bit bigger block" A_{block} (which depends on a particular problem) which includes a submatrix chosen in the previous step.

4. The eigenvalues of the chosen matrix A_{block} are ADI parameters (that is, $p_1, \ldots, p_l \in \sigma(A_{block})$).

Figure 2. shows how we form the matrix $\mathbf{A}_{block}$.

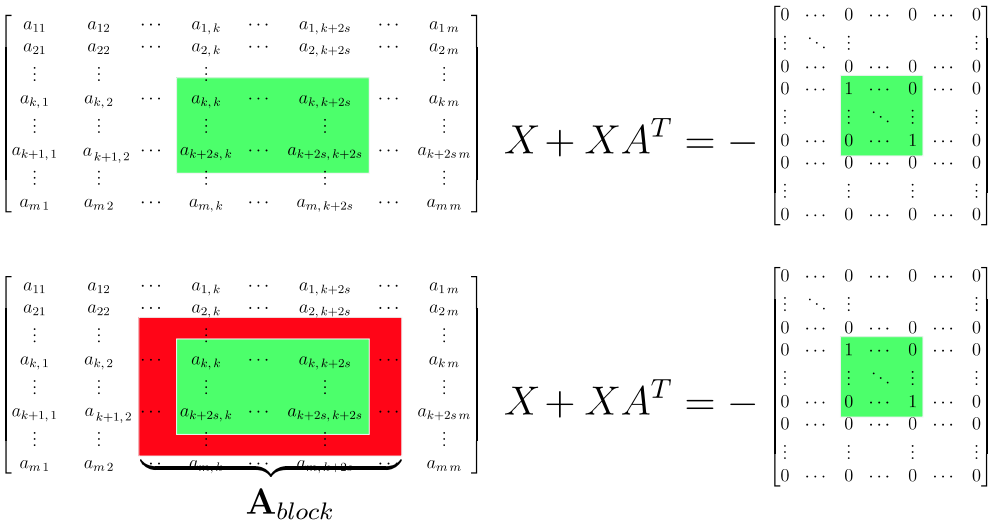

Figure 2: Choosing A_{block}

Here we have to emphasize two things:
1. when GG^T is not a projector (which is usual in our applications), one has to be able to predict positions of ones in the matrix GG^T without forming the matrix GG^T explicitly. For example, if G is close to some invariant subspace of A, then such a prediction will be possible.
2. generally it is not necessary for GG^T to have only ones and zeros, it is enough that one finds positions of the entries with the largest magnitude and then proceeds with steps 2.–4. of **Algorithm 2** (**ADI-parameters**).

Since we do not have a close theory which will insure that this choice of ADI parameters is (sub)optimal in any sense, we will compare the performance of the LRCF-ADI algorithm with a set of parameters generated by our algorithm with those generated by the algorithm proposed by Penzl in [5] and [7]. Penzl's algorithm for selection of ADI parameters is based on the following two ideas. First, we generate a discrete set, which "approximates" the spectrum, which is done by a pair of Arnoldi processes (we calculate the set of Ritz values). Then we choose a set of shift parameters which is a subset of the set of Ritz values by a heuristic that delivers a suboptimal set of ADI shifts.

In all the experiments we will use the following notation:

- Tr_{Penzl} denotes the trace of the solution of Lyapunov equation (1.1) obtained by LRCF-ADI generated by the set of ADI shifts generated by the algorithm proposed by T. Penzl [5].
- Tr_{new} denotes the trace of the solution of Lyapunov equation (1.1) obtained by LRCF-ADI generated by the set of ADI shifts generated by the new algorithm **ADI-parameters**
- trace(X) denotes the trace of the solution of Lyapunov equation (1.1) obtained by the Bartels–Stewart algorithm (implemented in MatLab as `lyap` function).

Application of Theorem 2.2 and the new algorithm for **ADI-parameters** shall be illustrated first on a mechanical system described in [13].

We consider the following Lyapunov equation

$$AX + XA^T = -GG^T\,, \tag{3.37}$$

where

$$A = A_0 - dd^T\,, \qquad A_0 = \Omega_1 \oplus \ldots \oplus \Omega_n\,, \qquad \Omega_i = \begin{bmatrix} 0 & \omega_i \\ -\omega_i & 0 \end{bmatrix}\,,$$

and d is an m dimensional vector. As it has been shown in [13], d can be chosen as to correspond to the system from (Fig. 3) with ω_i as undamped frequencies and $v > 0$ a damping constant of the damper applied to the mass m_1.

The system is described by the differential equation

$$M\ddot{x} + C\dot{x} + Kx = 0 \tag{3.38}$$

Figure 3: The n-mass oscillator

where M, C, K (called mass, damping, stiffness matrix, respectively) have the following form:

$$M = \operatorname{diag}(m_1, m_2, \ldots, m_n)\,,$$

$$K = \begin{bmatrix} k_1 & -k_1 & & & \\ -k_1 & k_1+k_2 & -k_2 & & \\ & \ddots & \ddots & \ddots & \\ & & -k_{n-2} & k_{n-2}+k_{n-1} & -k_{n-1} \\ & & & -k_{n-1} & k_{n-1}+k_n \end{bmatrix}, \quad C = \begin{bmatrix} v & 0 & \ldots & 0 \\ 0 & 0 & \ldots & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}.$$

In [13] it has been shown that for any choice of eigenvalues $\lambda_1, \ldots, \lambda_m$ in the open left half plane (symmetric with respect to the real axis) there exist unique matrices M, C and K of the above form, such that λ_i are eigenvalues of the corresponding quadratic eigenvalue problem

$$(\lambda^2 M + \lambda C + K)x = 0\,.$$

We first illustrate the application of Theorem 2.2 on the following simple example. Using the algorithm from [13] we will construct matrix A whose eigenvalues are $\lambda_1 = \lambda_2 = -1$ and $\lambda_3 = \lambda_4 = -2$. Let the matrix A_s be obtained from A by a perfect shuffle permutation and let $A_s = SJS^{-1}$, where

$$S = \begin{bmatrix} -0.3859 & -0.3939 & -1.4972 & -3.0098 \\ 0.68723 & 0.0141 & 5.3325 & 8.0535 \\ -0.5925 & 0.2097 & -6.1853 & -8.1836 \\ 0.1664 & -0.2252 & 3.4734 & 2.8589 \end{bmatrix}$$

and

$$J = J_1 \oplus J_2\,, J_1 = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}, \quad J_2 = \begin{bmatrix} -2 & 1 \\ 0 & -2 \end{bmatrix}.$$

We will consider Lyapunov equation (3.37) where the matrix G from the right-hand side has rank 1 and the 1 is set at the position $G(1,1) = 1$.

From (2.19) it follows that

$$\widehat{G} = S^{-1}G = \begin{bmatrix} 11.4354 & -6.9468 & -0.4505 & -0.6654 \end{bmatrix}^T .$$

Using (2.20) one can see that the best choice for ADI parameters will be the first two eigenvalues. Since, in general, we do not know the exact eigenvalues λ_1 and λ_2 we can apply the algorithm **ADI-parameters** for determination of ADI-parameters p_1 and p_2. From this algorithm it follows that p_1 and p_2 will be the eigenvalues of the matrix $A_{block} = A_s(1:2, 1:2)$ ($p_1 = -0.40859 + 0.38522\,i$, $p_2 = -0.40859 - 0.38522\,i$). If we assume that the damping constant is $v = 1$, then we have the following result:

$$\text{trace}(X) = 2.61022, \qquad Tr_{new} = 2.57693$$

(Tr_{new} is obtained by LRCF-ADI with p_1 and p_2 as shifts).

It should be pointed out, although our parameters are far away from the exact eigenvalues, the corresponding approximation is satisfactory. On the other hand, if we set the first two exact eigenvalues as ADI parameters $p_1 = -1$ and $p_2 = -1$, then the approximation which gives LRCF-ADI with these parameters is

$$Tr = 2.50127$$

which is obviously inferior to Tr_{new}.

A different choice of ADI parameters $p_1 = -1$ and $p_2 = -2$ (which will be the choice of Penzl's algorithm) will produce a less accurate result $Tr_{Penzl} = 2.24089$.

The above example was used as an illustration for a possible application of Theorem 2.2 and new algorithm **ADI-parameters** for deriving a suboptimal set of ADI parameters.

Further, we will compare performance of the LRCF-ADI algorithm with two suboptimal sets of ADI parameters on the same mechanical system with larger dimension.

Let $n = 100$ be a simple dimension of system (3.38) and let $\omega_1 = \ldots = \omega_n = 1$. Let the vector d be constructed according to the algorithm from [13].

We will consider Lyapunov equation (3.37) where the matrix G from the right-hand side will have rank $2s = 10$ and the ones will be set at positions

$G(41:40+s,1:s)=I_s$ $G(n+41:n+40+s,s+1:2s)=I_s$. If we assume that the damping constant $v=1$, then we have the following result:

$$\text{trace}(X)=1.2320e{+}003, \qquad Tr_{Penzl}=1.4783e{+}002, \quad Tr_{new}=1.2075e{+}003$$

where Tr_{Penzl} is derived by LRCF-ADI generated by 70 ADI shifts obtained by the algorithm proposed by T. Penzl [5].

The trace Tr_{new} was derived by LRCF-ADI generated by 40 ADI shifts obtained by the new algorithm **ADI-parameters** for deriving a suboptimal set of ADI parameters. Note that after a perfect shuffle permutation the ones on the right-hand side will be at positions $81, 82, \ldots, 90$ (positions of 1s on the diagonal of P^TGG^TP, where P denotes the perfect shuffle permutation), thus the matrix A_{block} was chosen as:

$$A_{block}=A_s(71:110,71:110)\,.$$

Note that the above example is the worst possible case because all matrices have the Jordan structure, which excludes usage of results for diagoalizable matrices. On the other hand, although our theory does not cover all possible cases, the above example shows that our choice of ADI parameters can be optimal in a certain sense.

The next example considers a 2-D convection-diffusion model on a square region (for example see [8]) and in contrast to the mechanical system from the last example, the right-hand sides in Lyapunov equation do not fit in our theory (**Algorithm 2** (**ADI-parameters**)). Anyway the results obtained by the LRCF-ADI method generated by the set of ADI parameters obtained by the new algorithm **ADI-parameters** are quite satisfactory.

Example 3.2 *Again we try to find the trace of the solution of Lyapunov equation (3.37) where the matrix A has the following form:*

$$A=-1/h^2\begin{bmatrix} A_1 & -I & & & \\ -I & A_1 & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & A_1 & -I \\ & & & -I & A_1 \end{bmatrix}_{n\times n},\ \textit{where } A_1=\begin{bmatrix} 4 & 1-h & & & \\ 1+h & 4 & 1-h & & \\ & \ddots & \ddots & \ddots & \\ & & 1+h & 4 & 1-h \\ & & & 1+h & 4 \end{bmatrix}_{\sqrt{n}\times\sqrt{}},$$

and $h=a/(\sqrt{n}+1)$. *We will take* $n=256$, $a=10$ *and* $\text{rank}(G)=20$.

For the matrix on the right-hand side we will take two different cases

1. $G(1:10,1:10)=I_{10}$ *and* $G(129:138,11:20)=I_{10}$,

2. $G(41:50,1:10) = I_{10}$ *and* $G(169:178,11:20) = I_{10}$

and in both cases we have the following results

1. $\dfrac{|\operatorname{trace}(X) - Tr_{Penzl}|}{\operatorname{trace}(X)} = 1.8e-014\,, \quad \dfrac{|\operatorname{trace}(X) - Tr_{new}|}{\operatorname{trace}(X)} = 1.3e-004\,,$
2. $\dfrac{|\operatorname{trace}(X) - Tr_{Penzl}|}{\operatorname{trace}(X)} = 1.6e-014\,, \quad \dfrac{|\operatorname{trace}(X) - Tr_{new}|}{\operatorname{trace}(X)} = 3.4e-003\,.$

where Tr_{Penzl} *is derived by LRCF-ADI generated by* 40 *ADI shifts obtained by the algorithm proposed by T. Penzl [5].*

The trace Tr_{new} *is derived by LRCF-ADI generated by* 40 *ADI shifts obtained by a new algorithm where the matrix* A_{block} *was chosen as:*

1. $A_{block} = A_s(1:40,1:40)$
2. $A_{block} = A_s(71:110,71:110)$

respectively.

Note that in both cases the trace calculated by LRCF-ADI using the ADI parameters generated by the new algorithm ***ADI-parameters*** *is inferior to the one obtained by ADI parameters obtained by the algorithm proposed by Penzl [5].*

The reason for this lies in the structure of the matrix $\widehat{G}$*. Fig. 4 shows the row norms of the matrix* $\widehat{G}$*. As one can see from the figure, most of the rows of* $\widehat{G}$ *have a similar norm, which means that Theorem 2.1 does not ensure a fast decay. Anyhow, the obtained traces contain* 3-4 *exact digits.*

In the last example we will again consider mechanical system with three rows of masses connected with springs. The system is shown in Figure 5.

Example 3.3 *The mechanical system from Figure 5 is differential equation (3.38) where* M, C, K *now have the following form*

$$M = \operatorname{diag}(M_{11}, M_{22}, M_{33}, m_0) \quad M_{ii} = \operatorname{diag}(m_i, \ldots m_i)\,,$$

$$K = \begin{bmatrix} K_{11} & & & -\kappa_1 \\ & K_{22} & & -\kappa_2 \\ & & K_{33} & -\kappa_3 \\ -\kappa_1^T & -\kappa_2^T & -\kappa_3^T & k_1{+}k_2{+}k_3{+}k_0 \end{bmatrix}, \quad K_{ii} = k_i \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix},$$

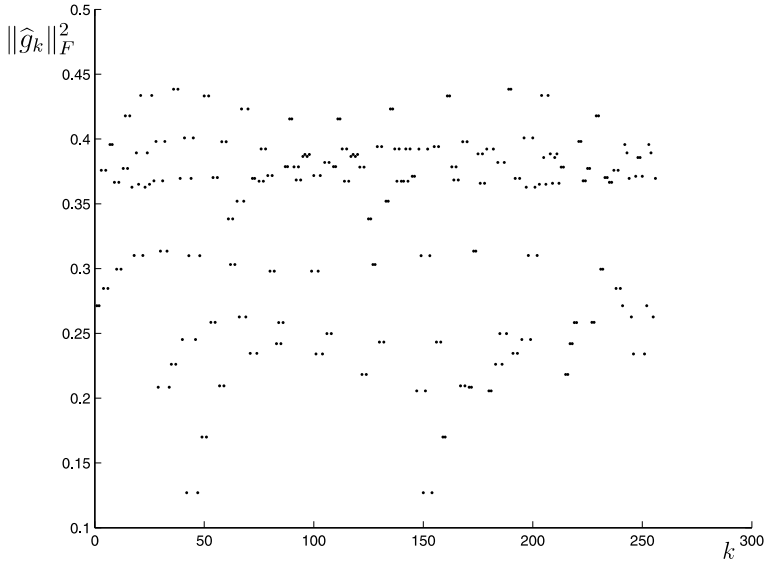

Figure 4: Row norms of $\widehat{G}$. No decay expected.

and $\kappa_i = \begin{bmatrix} 0 & \dots & 0 & k_i \end{bmatrix}^T$, $K_{ii} \in \mathbb{R}^{n\times n}$ *and* $\kappa_i \in \mathbb{R}^{n\times 1}$, *for* $i = 1, 2, 3$.

$$C \equiv C_u + C_v = C_u + v_1 e_1 e_1^T + v_2 e_n e_n^T + v_3 e_{2n+1} e_{2n+1}^T$$

Using the eigenvalue decomposition

$$\Phi^T K \Phi = \Omega^2\,, \quad \Phi^T M \Phi = I\,,$$

where $\Omega = \mathrm{diag}(\omega_1, \dots, \omega_n)$, $\omega_1 < \dots < \omega_n$ *and setting*

$$y_1 = \Omega\,\Phi^{-1} x \quad y_2 = \Phi^{-1}\dot{x}\,,$$

(3.38) can be written as

$$\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}\,, \tag{3.39}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\,, \quad \mathbf{A} = \begin{bmatrix} 0 & \Omega \\ -\Omega & -\Phi^T D \Phi \end{bmatrix}\,, \tag{3.40}$$

Note that M, C *and* K *are matrices of order* $3n + 1$. *We will set* $n = 50$, $C_u = 0.02\,\Omega$.

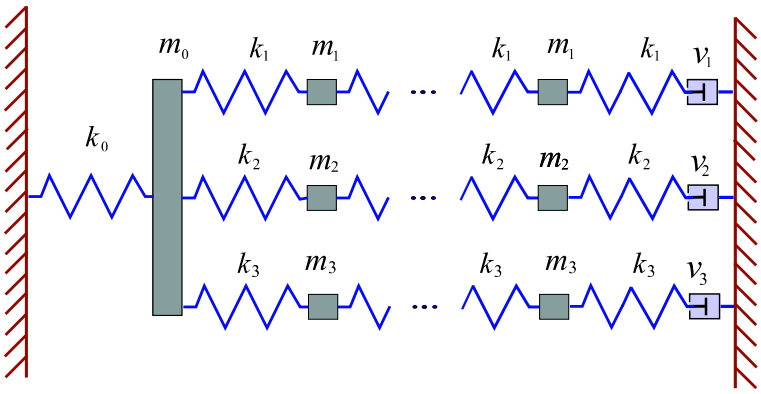

Figure 5: The mechanical system

There is a large number of papers which consider the problem of optimizing the damping matrix C. For example, in [12],[13] the one–dimensional damping has been considered, in [11] and [2] a general positive semidefinite damping has been considered, while [10] additionally assumes that all dampers have the same viscosity. In [4] and [2] among the other results the global minimum for positive definite damping has been presented.

In all mentioned papers one has to solve the following Lyapunov equation

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}^T = -\mathbf{G}\mathbf{G}^T \tag{3.41}$$

For this mechanical system we have performed a set of experiments using the following parameters. Let $m_2 = k_2 = 2$, $m_3 = k_3 = 4$ and $v_1 = 0.1$, $v_2 = 5$ and $v_3 = 0.01$ be fixed and let m_0, m_1, k_0, k_1 be chosen such that

$$m_0, k_0 \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2\} \text{ and } m_1, k_1 \in \{10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}\,.$$

This means that we have 1296 different configurations defined by different sets $\{m_0, m_1, m_2, m_3\}$ and $\{k_0, k_1, k_2, k_3\}$.

For each of these configurations we have derived the trace of the solution of Lyapunov equation (3.41) with two different matrices G on the right-hand side.

1. $G(1:10, 1:10) = I_{10}$, $G(152:161, 11:20) = I_{10}$, *and*
2. $G(51:60, 1:10) = I_{10}$, $G(202:211, 11:20) = I_{10}$.

In all cases we have used 50 *ADI shifts generated by the algorithm proposed by T. Penzl [5] for* Tr_{Penzl}*, while in the other case we have generated two different sets of* 50 *ADI parameters using a new algorithm, that is*

1. $p = eig(A_s(1:50, 1:50))$,
2. $p = eig(A_s(81:130, 81:130))$,

where A_s *denotes the matrix obtained from* A *by the perfect shuffle permutation.*

For all of 1296 different configurations in case 1. we have the following results:

$$\frac{|\operatorname{trace}(X) - Tr_{Penzl}|}{Tr_{Penzl}} \in \left(1.5 \cdot 10^{-6}, 0.02\right)$$
$$\frac{|\operatorname{trace}(X) - Tr_{new}|}{Tr_{new}} \in \left(8 \cdot 10^{-7}, 0.001\right)$$

while for all of 1296 different configurations in the case 2. the results are:

$$\frac{|\operatorname{trace}(X) - Tr_{Penzl}|}{Tr_{Penzl}} \in (32, 160)$$
$$\frac{|\operatorname{trace}(X) - Tr_{new}|}{Tr_{new}} \in (0.001, 0.01)$$

In both cases one can see that results obtained by the LRCF-ADI generated by the new set of ADI shifts are better than the one generated by shifts proposed by T. Penzl [5]. The reason for this lies again in the structure of the matrix $\widehat{G}$*. Fig. 6 shows the row norms of the matrix* $\widehat{G}$ *for one of the examples from case 2. As one can see from the figure, here Theorem 2.1 ensures a fast decay. As it is already shown, the results can be obtained efficiently.*

Acknowledgements. We would like to thank anonymous referees for a very careful reading of the manuscript and valuable comments.

References

[1] A. C. Antoulas, D. C. Sorensen and Y. Zhou, On the decay rate of Hankel singular values and related issues. *Systems and Control Letters*, **46**(1): 323-342; 2002.

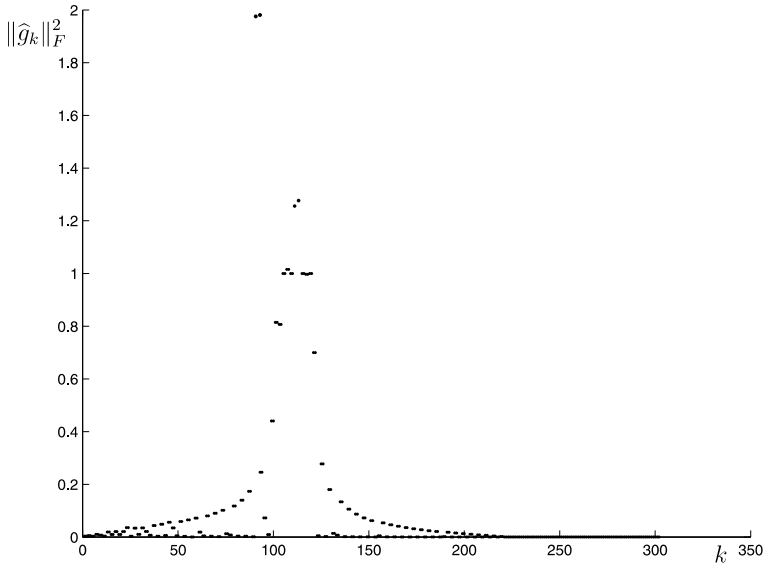

Figure 6: Row norms of $\widehat{G}$ for case *2.*

[2] S.J. Cox, I. Nakić and K. Veselić Minimization of energy of a damped system, submited for publication in Systems and Control Letters.

[3] J. Li and J. White, *Low rank solution of Lyapunov equation. SIAM Journal on Matrix Analysis and Appl.*, 24:1:260–280 (2002).

[4] I. Nakić, Optimal damping of vibrational systems. Ph. D. thesis Fernuniversität, Hagen, 2002.

[5] T. Penzl, *LYAPACK.* http://www.tu-chemnitz.de/sfb393/lyapack

[6] T. Penzl, *A cyclic low rank Smith method for large sparse Lyapunov equations. SIAM J. Sci. Comput.* 21 (2000) 1401-1418.

[7] T. Penzl, *Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case. Systems Control Lett.* 40 (2000) 139-144.

[8] D. C. Sorensen and Y. Zhou, Bounds on eigenvalue decay rates and sensitivity of solutions to Lyapunov equations. *Technical report*, 2002.

[9] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic press, Boston, 1990.

[10] N. Truhar, An efficient algorithm for damper optimization for linear vibrating systems using Lyapunov equation, *J. Comput. Appl. Math.* 172-1(2004), 169-182.

[11] K. Veselić, K. Brabender and K. Delinić, Passive control of linear systems. *Applied Mathematics and Computation, M. Rogina et al. Eds. Dept. of Math. Univ. Zagreb*, 2001, 39-68.

[12] K. Veselić, On linear vibrational systems with one dimensional damping. *Appl. Anal.* **29** (1988) 1-18.

[13] K. Veselić, On linear vibrational systems with one dimensional damping II. *Integral Eq. Operator Th.* **13** (1990) 883-897.

[14] E. L. Wachspress, Extended application of alternating direct implicit iteration model problem theory, Journal of SIAM, 11 (1963) pp. 994-1016.