

Bounds on eigenvalue decay rates and sensitivity of solutions to Lyapunov equations *

D. C. Sorensen and Y. Zhou

Computational and Applied Mathematics

e-mail: {sorensen,ykzhou}@rice.edu

<http://www.caam.rice.edu/~sorensen>

Rice University

Houston, Texas, 77005

June 3, 2002

Abstract

Balanced model reduction is a technique for producing a low dimensional approximation to a linear time invariant system. An important feature of balanced reduction is the existence of an error bound that is closely related to the decay rate of the eigenvalues of certain system Gramians. Rapidly decaying eigenvalues imply low dimensional reduced systems. New bounds are developed for the eigen-decay rate of the solution of Lyapunov equation $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T$. These bounds take into account the low rank right hand side structure of the Lyapunov equation. They are valid for any diagonalizable matrix \mathbf{A} . Numerical results are presented to illustrate the effectiveness of these bounds when the eigensystem of \mathbf{A} is moderately conditioned.

We also present a bound on the norm of the solution \mathbf{P} when \mathbf{A} is diagonalizable and derive bounds on the conditioning of the Lyapunov operator for general \mathbf{A} .

Key words: Lyapunov equation; Lyapunov operator; Low rank; Conditioning; Eigenvalue decay rate

*This work was supported in part by the NSF through Grants DMS-9972591 and CCR-9988393.

Contents

1	Introduction	2
2	Eigenvalue decay bounds for solutions of Lyapunov equations: the general case	3
2.1	Numerical results	8
3	Bounds on the solution norm and the conditioning of the Lyapunov operator	10
4	Conclusions	16
R	References	16
A	Derivation of the Stein equation: general complex shifts	18

1 Introduction

This paper shall discuss two related topics concerning the solution of the continuous-time Lyapunov equation

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times p}$, and \mathbf{A} is assumed to be stable. We are particularly concerned with problems where n is large. We shall derive bounds on decay rates of the eigenvalues of \mathbf{P} and we shall also present bounds on the norm of \mathbf{P} and on the conditioning of the continuous time Lyapunov operator.

In Section 2 we derive bounds on the eigenvalue decay rates for the case $p \ll n$. This case is most common for large scale linear time invariant (LTI) systems in engineering applications. When the integer p is much less than n , we call (1) a Lyapunov equation with low rank right hand side (RHS).

Lyapunov equations (1) with a low rank RHS appear frequently in engineering applications. One important property for a matrix equation with low rank RHS is that the solution \mathbf{P} is nearly a low rank matrix in most cases. Theoretically, we know that \mathbf{P} is positive definite whenever $\text{rank}([\mathbf{B}, \mathbf{A}\mathbf{B}, \dots, \mathbf{A}^{n-1}\mathbf{B}]) = n$ (i.e., the corresponding LTI system is controllable). However, even though \mathbf{P} is positive definite in theory, it is often the case that the eigenvalues $\lambda_j(\mathbf{P})$ tend to decay rapidly $\frac{\lambda_k(\mathbf{P})}{\lambda_1(\mathbf{P})} \ll 1$ for small k and hence \mathbf{P} is numerically low rank. The low rank solution property is also very important for iterative solution methods [7, 11, 14, 15, 19, 20, 23, 25, 24] that compute low rank approximants to \mathbf{P} . Iterative methods reduce the cost of solving Lyapunov equations from $O(n^3)$ flops and $O(n^2)$ storage required by a direct method to $O(nk^2)$ flops and $O(nk)$ storage, where k is an integer much less than n .

A fundamental question is: why and when does (1) have a nearly low rank solution? This question has not been answered satisfactorily in the existing literature. For large

scale Lyapunov equations, this question becomes very important. When n is large one needs to resort to iterative methods to avoid the work and storage costs of direct methods.

There are many papers concerning properties (including eigenvalue bounds) of the solutions of Lyapunov equations, but few of them directly address the above key question. Many papers consider Lyapunov equations of form

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T = -\mathbf{M} \quad (2)$$

where \mathbf{M} is positive definite (see [18] and references therein), or \mathbf{M} is positive semidefinite. Most of these do not exploit the low rank positive semidefinite structure $\mathbf{M} = \mathbf{B}\mathbf{B}^T$. Excellent theoretical results are obtained in [21] where $\mathbf{M} = \mathbf{B}\mathbf{B}^T$ structure is not neglected. The bounds they derive for the small eigenvalues of the solution \mathbf{P} are nontrivial. However, their results do not directly address the numerical rank of the solution of (1).

Since the solution \mathbf{P} of (1) is positive (semi)definite, its numerical rank can be revealed by the relative magnitude of its eigenvalues. Hence we study the eigen-decay rate $\frac{\lambda_i(\mathbf{P})}{\lambda_1(\mathbf{P})}$ where $\{\lambda_i(\mathbf{P})\}$ are the eigenvalues of \mathbf{P} ordered in non-increasing order.

Apparently, Penzl's work [22] is the first to directly address the numerical rank of the solution \mathbf{P} . In [22] the structure $\mathbf{M} = \mathbf{B}\mathbf{B}^T$ is exploited, an eigen-decay-rate bound for (1) is developed. However, Penzl's result requires \mathbf{A} to be a symmetric matrix. In a previous paper [2] we derived estimates for eigen-decay rate of the solution \mathbf{P} through a special Cauchy matrix defined by the eigenvalues of \mathbf{A} . Our estimated decay rates were valid for non-symmetric but diagonalizable \mathbf{A} and predicted the actual eigen-decay rates very well in computed examples. However, these estimates do not provide rigorous bounds. The results presented here in Section 2 will provide explicit eigen-decay-rate bounds that generalize the bounds given in [22] to the nonsymmetric case.

In Section 3, we derive an upper bound for the norm of the solution \mathbf{P} , and lower and upper bounds for the condition number of the Lyapunov operator. These bounds highlight the significance of the term $\{\min_i |Re(\lambda_i(\mathbf{A}))|\}$. The bounds on the condition number of the Lyapunov operator are easy to verify and they are able to give good indications of the conditioning of the related Lyapunov equation.

In this paper, capital bold face letters are used to denote matrices and lower case bold face letters shall denote vectors. The complex conjugate-transpose of a matrix \mathbf{A} is denoted by \mathbf{A}^* and this notation is also used to denote the complex conjugation of a scalar.

2 Eigenvalue decay bounds for solutions of Lyapunov equations: the general case

As mentioned in the introduction, we are most interested in the bounds for eigenvalue decay rate $\frac{\lambda_i(\mathbf{P})}{\lambda_1(\mathbf{P})}$, where $\{\lambda_i(\mathbf{P})\}$ are eigenvalues of the solution \mathbf{P} of the following Lyapunov

equation,

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0} \quad (3)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is stable, $\mathbf{B} \in \mathbb{R}^{n \times p}$, $p \ll n$. $\{\lambda_i(\mathbf{P})\}$ are ordered in non-increasing order. The low rank positive semidefinite structure of $\mathbf{B}\mathbf{B}^T$ plays a key role in the low rank property of the solution \mathbf{P} .

The result of Penzl [22] does take into account the low rank right hand side structure to analyze the numerical rank of \mathbf{P} directly. The bound he obtained is:

$$\frac{\lambda_{pk+1}(\mathbf{P})}{\lambda_1(\mathbf{P})} \leq \left(\prod_{j=0}^{k-1} \frac{\kappa(\mathbf{A})^{(2j+1)/(2k)} - 1}{\kappa(\mathbf{A})^{(2j+1)/(2k)} + 1} \right)^2. \quad (4)$$

This bound was established for symmetric \mathbf{A} . The advantages and disadvantages of using only $\kappa(\mathbf{A})$ to bound the eigen-decay rate are discussed in [2]. Here we generalize Penzl's result to include both symmetric and nonsymmetric \mathbf{A} .

We shall begin with a modification to the very first step in Penzl's analysis that will overcome the symmetry restriction.

Define the following rational functions which map the left half plane into the unit disk:

$$s_\tau(t) := \frac{\tau - t}{\tau^* + t}, \quad \text{and} \quad s_{\{\tau_1, \dots, \tau_k\}}(t) := \prod_{i=1}^k s_{\tau_i}(t) \quad (5)$$

where $Re(\tau) < 0$, $Re(\tau_i) < 0$, $i = 1, \dots, k$.

The Alternating Direction Implicit (ADI) method of Peaceman and Rachford [10] was adapted by Wachpress in [29] to iteratively solve Lyapunov equations and more recently this approach has been adapted to take advantage of the near low rank RHS structure in [19, 20, 23]. The ADI iterates $\{\mathbf{P}_i\}_{i=0}^\infty$ generated from an initial matrix \mathbf{P}_0 are:¹

$$\mathbf{P}_i = s_{\tau_i}(\mathbf{A}) \mathbf{P}_{i-1} s_{\tau_i}^*(\mathbf{A}) - 2 Re(\tau_i) (\mathbf{A} + \tau_i^* \mathbf{I})^{-1} \mathbf{B}\mathbf{B}^T (\mathbf{A} + \tau_i^* \mathbf{I})^{-*}. \quad (6)$$

Since the solution \mathbf{P} of (3) satisfies the following **Stein equation**:

$$\mathbf{P} = s_{\tau_i}(\mathbf{A}) \mathbf{P} s_{\tau_i}^*(\mathbf{A}) - 2 Re(\tau_i) (\mathbf{A} + \tau_i^* \mathbf{I})^{-1} \mathbf{B}\mathbf{B}^T (\mathbf{A} + \tau_i^* \mathbf{I})^{-*}, \quad (7)$$

we see \mathbf{P} is a stationary point of the mapping (6). Hence:

$$\mathbf{P} - \mathbf{P}_i = s_{\tau_i}(\mathbf{A}) (\mathbf{P} - \mathbf{P}_{i-1}) s_{\tau_i}^*(\mathbf{A}).$$

Choose $\mathbf{P}_0 = \mathbf{0}$, choose k s.t. $kp < n$, iterate recursively k steps, we get

$$\mathbf{P} - \mathbf{P}_k = s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A}) \mathbf{P} s_{\{\tau_1, \dots, \tau_k\}}^*(\mathbf{A}). \quad (8)$$

¹The complex case is more complicated than the real case, we derive (6) and (7) separately in the appendix with some discussions.

From (6) we see $\text{rank}(\mathbf{P}_{i-1}) \leq \text{rank}(\mathbf{P}_i) \leq \text{rank}(\mathbf{P}_{i-1}) + p$, so $\text{rank}(\mathbf{P}_k) \leq kp$. The low rank right hand side structure is used in this rank inequality.

By the Schmidt-Mirsky theorem [1, 5] we have:

$$\frac{\lambda_i(\mathbf{P})}{\lambda_1(\mathbf{P})} = \min_{\tilde{\mathbf{P}} \in \mathbb{R}^{n \times n}, \text{rank}(\tilde{\mathbf{P}}) \leq i-1} \frac{\|\mathbf{P} - \tilde{\mathbf{P}}\|_2}{\|\mathbf{P}\|_2}. \quad (9)$$

Combining (9) and (8) gives:

$$\begin{aligned} \frac{\lambda_{pk+1}(\mathbf{P})}{\lambda_1(\mathbf{P})} &\leq \frac{\|\mathbf{P} - \mathbf{P}_k\|_2}{\|\mathbf{P}\|_2} \leq \|s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A})\|_2 \|s_{\{\tau_1, \dots, \tau_k\}}^*(\mathbf{A})\|_2 \\ &= \|s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A})\|_2^2. \end{aligned} \quad (10)$$

Penzl's approach requires $s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A})$ to be Hermitian and this property is generally lost when \mathbf{A} is nonsymmetric. To overcome this difficulty, we note that $\|s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A})\|_2$ is the square root of the largest eigenvalue value of the following positive semidefinite matrix:

$$\begin{aligned} &s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A}) s_{\{\tau_1, \dots, \tau_k\}}^*(\mathbf{A}) \\ &= \prod_{i=1}^k (\tau_i \mathbf{I} - \mathbf{A})(\tau_i^* \mathbf{I} + \mathbf{A})^{-1} (\tau_i^* \mathbf{I} - \mathbf{A}^T)(\tau_i \mathbf{I} + \mathbf{A}^T)^{-1}. \end{aligned} \quad (11)$$

Observe that all the matrices in (11) commute with each other and that $\mathbf{A}^T = \mathbf{A}^*$ since it is real. We will establish an upper bound for $\|s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A})\|_2$ based on (11).

In the following, we assume \mathbf{A} is diagonalizable (this assumption is more general than symmetry). Let the eigendecomposition of \mathbf{A} be

$$\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda} \quad (12)$$

where the columns of \mathbf{X} are the right eigenvectors of \mathbf{A} , the diagonal matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues of \mathbf{A} , $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. (For notation simplicity, in the following part of this paper we use $\{\lambda_i\}$ to denote the eigenvalues of \mathbf{A} , i.e., $\lambda_i := \lambda_i(\mathbf{A})$).

Then (11) and (12) lead to

$$\begin{aligned} &s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A}) s_{\{\tau_1, \dots, \tau_k\}}^*(\mathbf{A}) \\ &= \left\{ \prod_{i=1}^k (\tau_i \mathbf{I} - \mathbf{A})(\tau_i^* \mathbf{I} + \mathbf{A})^{-1} \right\} \left\{ \prod_{j=1}^k (\tau_j^* \mathbf{I} - \mathbf{A}^T)(\tau_j \mathbf{I} + \mathbf{A}^T)^{-1} \right\} \\ &= \mathbf{X} \left\{ \prod_{i=1}^k (\tau_i \mathbf{I} - \mathbf{\Lambda})(\tau_i^* \mathbf{I} + \mathbf{\Lambda})^{-1} \right\} \mathbf{X}^{-1} \mathbf{X}^{-*} \left\{ \prod_{j=1}^k (\tau_j^* \mathbf{I} - \mathbf{\Lambda}^*)(\tau_j \mathbf{I} + \mathbf{\Lambda}^*)^{-1} \right\} \mathbf{X}^*. \end{aligned}$$

We finally obtain

$$\begin{aligned}
& \|s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A})\|_2^2 \\
&= \|s_{\{\tau_1, \dots, \tau_k\}}(\mathbf{A}) s_{\{\tau_1, \dots, \tau_k\}}^*(\mathbf{A})\|_2 \\
&\leq \kappa(\mathbf{X}) \kappa(\mathbf{X}^*) \left\{ \max_{1 \leq l \leq n} \prod_{i=1}^k \left| \frac{\tau_i - \lambda_l}{\tau_i^* + \lambda_l} \right| \right\} \left\{ \max_{1 \leq m \leq n} \prod_{j=1}^k \left| \frac{\tau_j^* - \lambda_m^*}{\tau_j + \lambda_m^*} \right| \right\} \\
&= \kappa^2(\mathbf{X}) \left\{ \max_{1 \leq l \leq n} \prod_{i=1}^k \left| \frac{\tau_i - \lambda_l}{\tau_i^* + \lambda_l} \right| \right\}^2. \tag{13}
\end{aligned}$$

This derivation is valid for any choice of $\{\tau_i, i = 1, \dots, k\} \subset \mathbb{C}_-$. Therefore, we might wish to make the bound in (13) as small as possible with an appropriate choice. This leads to a classical min-max problem:

$$\{\hat{\tau}_1, \hat{\tau}_2, \dots, \hat{\tau}_k\} = \arg \min_{\tau_1, \dots, \tau_k \in \mathbb{C}_-} \left\{ \max_{1 \leq l \leq n} \prod_{i=1}^k \left| \frac{\tau_i - \lambda_l}{\tau_i^* + \lambda_l} \right| \right\}. \tag{14}$$

The previous discussion establishes the following theorem:

Theorem 2.1 *For stable and diagonalizable \mathbf{A} , the eigenvalues of the solution \mathbf{P} of the Lyapunov equation (3) satisfy the following decay-rate bound:*

$$\frac{\lambda_{pk+1}(\mathbf{P})}{\lambda_1(\mathbf{P})} \leq \kappa^2(\mathbf{X}) \left\{ \max_{1 \leq l \leq n} \prod_{i=1}^k \left| \frac{\hat{\tau}_i - \lambda_l}{\hat{\tau}_i^* + \lambda_l} \right|^2 \right\}, \tag{15}$$

where $\{\hat{\tau}_i, i = 1, \dots, k\}$ are the solution of (14) and \mathbf{X} is the right eigenvector matrix of \mathbf{A} .

If $\kappa(\mathbf{X})$ is of modest size, this bound is a reasonable estimate of the actual decay rate. On the other hand when \mathbf{A} is highly nonnormal ($\kappa(\mathbf{X})$ is large) this bound is typically very pessimistic.

Obviously, in its present form, this bound is mainly of theoretical interest because it requires complete knowledge of the spectrum of \mathbf{A} and this generally be unavailable in the large scale setting. However, even with such information, a major practical difficulty with (15) is that the min-max problem (14) has not yet been solved for general \mathbf{A} (results for symmetric \mathbf{A} may be found in Wachspress [29]). We can apply other (suboptimal) shifts. One possibility is to use the eigenvalues of \mathbf{A} as the shifts. This choice leads to the following suboptimal bound:

$$\frac{\lambda_{pk+1}(\mathbf{P})}{\lambda_1(\mathbf{P})} \leq \kappa^2(\mathbf{X}) \left\{ \max_{1 \leq l \leq n} \prod_{i=1}^k \left| \frac{\lambda_i - \lambda_l}{\lambda_i^* + \lambda_l} \right|^2 \right\}. \tag{16}$$

Although this upper bound is not as tight as with optimal shifts, the numerical results of the next subsection indicate that (16) gives a reasonably good bound for the actual eigen-decay rate.

We shall make four remarks on Theorem 2.1:

1. Theorem 2.1 includes the symmetric case (Theorem 1 in Penzl's [22]) as its special case.
2. If all the eigenvalues of \mathbf{A} are real, then the derivation of Penzl's bound (4) still holds. For this case we immediately have the following bound:

$$\frac{\lambda_{pk+1}(\mathbf{P})}{\lambda_1(\mathbf{P})} \leq \kappa^2(\mathbf{X}) \left(\prod_{j=0}^{k-1} \frac{\kappa_0^{(2j+1)/(2k)} - 1}{\kappa_0^{(2j+1)/(2k)} + 1} \right)^2 \quad (17)$$

where $\kappa_0 := \frac{\lambda_{\min}(\mathbf{A})}{\lambda_{\max}(\mathbf{A})} = \frac{|\lambda_{\max}(\mathbf{A})|}{|\lambda_{\min}(\mathbf{A})|} \geq 1$. Note $\kappa_0 = \kappa(\mathbf{A})$ only if \mathbf{A} is normal.

3. Bound (15) has a striking resemblance to the eigen-decay rate we recently obtained in [2] even though the approaches that lead to the results are very different.

In [2] we first linked the solution \mathbf{P} of (3) to a special Cauchy matrix ($\mathcal{C} := \left[\frac{-1}{\lambda_i + \lambda_j^*} \right]$) defined by the spectrum of \mathbf{A} , then we applied ordered Cholesky decomposition to \mathcal{C} . We found the eigen-decay rate $\frac{\lambda_k(\mathbf{P})}{\lambda_1(\mathbf{P})}$ can be estimated quite well by $\frac{\delta_k}{\delta_1}$, where

$$\delta_k = \frac{-1}{2 \operatorname{Re}(\lambda_k)} \prod_{j=1}^{k-1} \left| \frac{\lambda_k - \lambda_j}{\lambda_k^* + \lambda_j} \right|^2, \quad (18)$$

$\{\lambda_i\}_{i=1}^n$ are ordered by what we called *Cholesky ordering*. This ordering leads to a non-increasing sequence $\{\delta_i\}$, with the largest number $\delta_1 = \frac{1}{2 \min\{\operatorname{Re}(-\lambda_i)\}}$. While in this paper, we derived the eigen-decay bound mainly through the Schmidt-Mirsky theorem.

Note that if we choose $\{\tau_i, i = 1, \dots, k\}$ to be the k eigenvalues of \mathbf{A} , as done in bound (16), then bound (16) is very similar to the approximate rate (18). Since we have the freedom in choosing other $\{\tau_i\}$, bound (15) is likely a better approximation than (16) and (18).

4. As shown in [2], (18) provides a good estimate of the actual eigen-decay rate when the matrix \mathbf{A} has a modestly conditioned eigen-system. From the similarity between (16) and (18), we see (16) should be a reasonably good bound. We also note that bound (15) is theoretically better than (16).

Even though the min-max problem (14) is not well solved, the bound (15) is still able to shed light on the fundamental question about the eigen-decay rate of the solutions of Lyapunov equations. Two immediate consequences of (15) are:

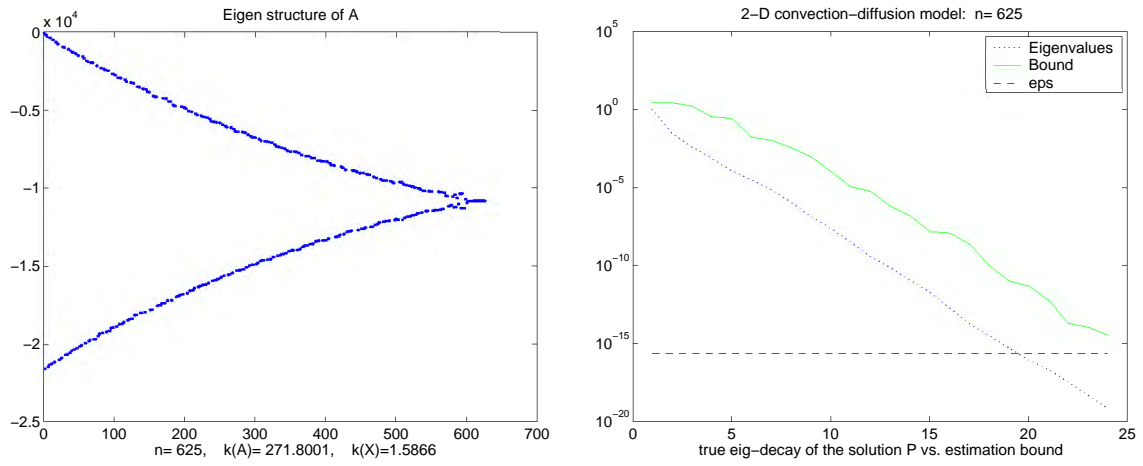


Figure 1: Actual Eigen-decay rate vs estimation bound: 2-D convection-diffusion model

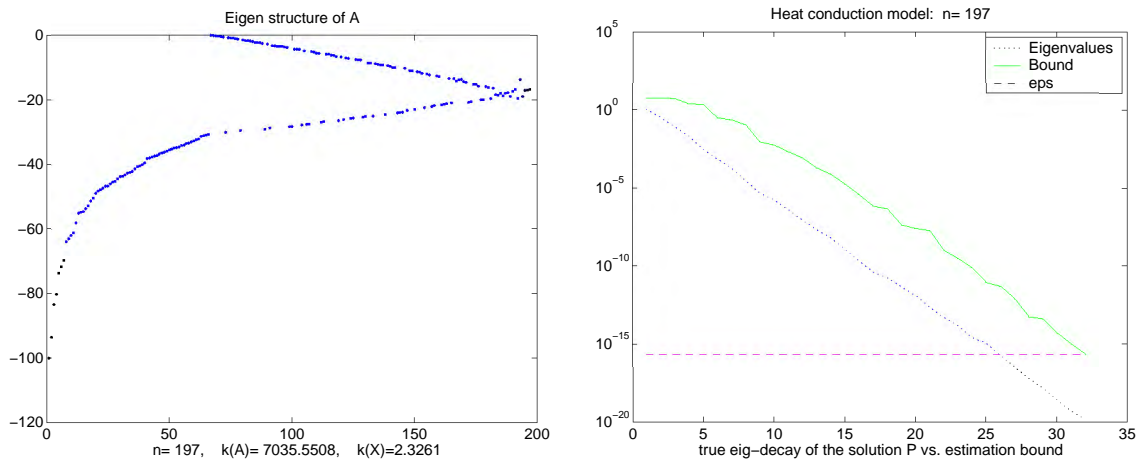


Figure 2: Actual Eigen-decay rate vs. estimation bound: Heat conduction model

of problems similar to those arising in practice. The last two are taken from actual engineering applications. The “CD player” is a model that describes the dynamics between the lens actuator and the radial arm position of a portable compact disc player [6]; the matrix \mathbf{A} for this model is anti-symmetric. The “heat conduction” model is a finite element model of heat conduction in a plate with boundary controls. The “clamped beam structure” model is a finite element model of a clamped beam with a control force applied at the free end. The “ISS 1r-c04” and “ISS 12a-c69” models are actual finite element discretization of the flex modes of the Zvezda Service Module of the International Space Station (ISS); both of the models have 3 inputs and 3 outputs, namely the roll, pitch, yaw jets, and the roll, pitch, yaw rate gyros readings, respectively; these models were provided to us by Draper Labs.

We see from Table 1 that even though \mathbf{A} sometimes can have large condition number, the condition number of the right eigenvector matrix \mathbf{X} is usually moderate. This is also the case in many other engineering applications where the models are PDEs, since the \mathbf{A} ’s obtained by discretization of the PDEs via finite element or finite difference method are usually symmetric positive definite or diagonally dominant. Hence for most of these systems, the $\kappa(\mathbf{X})^2$ term does not prevent the bounds from being fairly sharp.

Table 1: The condition number for some real models

<i>Model</i>	<i>n</i>	$\kappa(\mathbf{A})$	$\kappa(\mathbf{X})$
CD player	120	1.78 e+4	1.0
Heat conduction	197	7.04 e+3	2.3261
Clamped beam structure	384	6.87 e+6	358.09
ISS 1r-c04	270	9.68 e+3	61.341
ISS 12a-c69	1412	5.77 e+3	32.728

3 Bounds on the solution norm and the conditioning of the Lyapunov operator

In this section we present a series of propositions that will establish two results on the sensitivity of the Lyapunov equation. The first bounds the norm of the solution \mathbf{P} of $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}$ by linking \mathbf{P} to a special Cauchy matrix defined by the spectrum of \mathbf{A} and applying a result of Horn and Kittaneh [13]. The second establishes lower and upper bounds on the condition number of the Lyapunov operator and hence on the condition of the continuous-time Lyapunov equation.

As before, we assume \mathbf{A} is diagonalizable and let the eigen decomposition of \mathbf{A} be:

$$\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}, \quad \text{where } \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (20)$$

then $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}$ is equivalent to

$$\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}\mathbf{P} + \mathbf{P}\mathbf{X}^{-*}\mathbf{\Lambda}^*\mathbf{X}^* + \mathbf{B}\mathbf{B}^T = \mathbf{0}.$$

Let $\tilde{\mathbf{P}} = \mathbf{X}^{-1}\mathbf{P}\mathbf{X}^{-*}$, $\tilde{\mathbf{B}} = \mathbf{X}^{-1}\mathbf{B}$, we get

$$\mathbf{\Lambda}\tilde{\mathbf{P}} + \tilde{\mathbf{P}}\mathbf{\Lambda}^* + \tilde{\mathbf{B}}\tilde{\mathbf{B}}^* = \mathbf{0}.$$

It is straightforward to see that $\tilde{\mathbf{P}}$ may be expressed in terms of a *Hadarmard product* [12]:²

$$\tilde{\mathbf{P}} = \left[\frac{-1}{\lambda_i + \lambda_j^*} \right] \circ (\tilde{\mathbf{B}}\tilde{\mathbf{B}}^*), \quad (21)$$

where $\left[\frac{-1}{\lambda_i + \lambda_j^*} \right]$ is a Cauchy matrix defined by the eigenvalues of \mathbf{A} .

So the original solution \mathbf{P} is given by

$$\mathbf{P} = \mathbf{X} \left\{ \left[\frac{-1}{\lambda_i + \lambda_j^*} \right] \circ (\tilde{\mathbf{B}}\tilde{\mathbf{B}}^*) \right\} \mathbf{X}^*. \quad (22)$$

The following lemma will establish the desired bound on $\|\tilde{\mathbf{P}}\|$.

Lemma 3.1 [13], *Let $\{\alpha_i\}_{i=1}^m$, $\{\beta_i\}_{i=1}^n$ be given complex numbers with positive real parts, let $\|\cdot\|$ be a given norm on $\mathbb{C}^{m \times n}$. If $m \neq n$, assume $\|\cdot\|$ is unitarily invariant; if $m = n$, assume $\|\cdot\|$ is either unitarily invariant or induced by an absolute norm on \mathbb{C}^n , then $\forall \mathbf{M} \in \mathbb{C}^{m \times n}$*

$$\left\| \left[\frac{1}{\alpha_i + \beta_j} \right] \circ \mathbf{M} \right\| \leq \frac{1}{\min \text{Re}(\alpha_i) + \min \text{Re}(\beta_j)} \|\mathbf{M}\|. \quad (23)$$

Applying Lemma 3.1 to (21) we obtain

$$\|\tilde{\mathbf{P}}\| \leq \frac{1}{2 \min \text{Re}(-\lambda_i)} \|\tilde{\mathbf{B}}\tilde{\mathbf{B}}^*\| \leq \frac{1}{2 \min \text{Re}(-\lambda_i)} \|\mathbf{B}\mathbf{B}^T\| \|\mathbf{X}^{-1}\| \|\mathbf{X}^{-*}\|. \quad (24)$$

From this we obtain an upper bound for the norm of \mathbf{P} .

Proposition 3.1 *Let \mathbf{P} solve the Lyapunov equation $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}$ where $A \in \mathbb{R}^{n \times n}$ is stable, with eigenvalues $\mathbf{\Lambda} = \text{diag}(\{\lambda_i\}_{i=1}^n)$ and eigen decomposition $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}$. Then*

$$\|\mathbf{P}\| \leq \frac{1}{2 \min \text{Re}(-\lambda_i)} \|\mathbf{B}\mathbf{B}^T\| \kappa^2(\mathbf{X}), \quad (25)$$

where $\|\cdot\|$ is any unitarily invariant norm or any induced norm.

²Hadarmard product is defined as the elementwise product, $\mathbf{A} = \mathbf{B} \circ \mathbf{C} \Leftrightarrow a_{ij} = b_{ij}c_{ij}$

Proof : Combining (24) and (22) we get

$$\|\mathbf{P}\| \leq \frac{1}{2 \min_i \operatorname{Re}(-\lambda_i)} \|\mathbf{B}\mathbf{B}^T\| \|\mathbf{X}^{-1}\| \|\mathbf{X}^{-*}\| \|\mathbf{X}\| \|\mathbf{X}^*\|.$$

Note $\kappa(\mathbf{X}^*) = \kappa(\mathbf{X}) = \|\mathbf{X}^{-1}\| \|\mathbf{X}\|$ to establish (25). ■

From bound (25) we see that when \mathbf{A} has eigenvalues with real parts that are very small relative to $\|\mathbf{B}\|$, one might expect a $\|\mathbf{P}\|$ to be large.

It turns out that the situation just described is also closely related to predicting an ill-conditioned case of the Lyapunov equation. The conditioning of the Lyapunov equation

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{M} = \mathbf{0} \quad (26)$$

can be understood by studying the condition number of the following matrix

$$\mathcal{L} = \mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I}, \quad (27)$$

where \otimes is the Kronecker product [12]. \mathcal{L} is a matrix representation of the Lyapunov operator associated with (26).

It is well known that the eigenvalues of \mathcal{L} are

$$\lambda_l(\mathcal{L}) = \lambda_i(\mathbf{A}) + \lambda_j^*(\mathbf{A}),$$

where $l = i + (j - 1)n$; $i, j = 1, \dots, n$. Hence

$$\max_i |\lambda_l(\mathcal{L})| \geq 2 \max_i |\operatorname{Re}(\lambda_i(\mathbf{A}))|, \quad \min_i |\lambda_l(\mathcal{L})| \leq 2 \min_i |\operatorname{Re}(\lambda_i(\mathbf{A}))|. \quad (28)$$

It is also well known [5, 26, 9] that the conditioning of (26) is closely related to the separation between \mathbf{A} and $-\mathbf{A}$ defined as,

$$\operatorname{sep}(\mathbf{A}, -\mathbf{A}) = \min_{\mathbf{Y} \in \mathbb{R}^{n \times n}} \frac{\|\mathbf{A}\mathbf{Y} + \mathbf{Y}\mathbf{A}^T\|_F}{\|\mathbf{Y}\|_F}.$$

An easily seen fact is,

$$\sigma_{\min}(\mathcal{L}) = \min_{y \in \mathbb{R}^{2n}} \frac{\|\mathcal{L}y\|_2}{\|y\|_2} = \operatorname{sep}(\mathbf{A}, -\mathbf{A}) = \|\mathcal{L}^{-1}\|_2^{-1}. \quad (29)$$

In [16, 4] it is shown that for normal \mathbf{A} , \mathcal{L} is also normal (actually the converse is also true), and

$$\kappa_2(\mathcal{L}) = \|\mathcal{L}\|_2 \|\mathcal{L}^{-1}\|_2 = \frac{\max_i |\lambda_i(\mathbf{A}) + \lambda_j^*(\mathbf{A})|}{\min_i |\lambda_i(\mathbf{A}) + \lambda_j^*(\mathbf{A})|}.$$

We are not aware of any results for the nonnormal case. However, it is straightforward to get a lower bound of $k_2(\mathcal{L})$ if we note that,

$$\sigma_{\max}(\mathcal{L}) \geq \max_i \{|\lambda_i(\mathcal{L})|\}, \quad \text{and} \quad \sigma_{\min}(\mathcal{L}) \leq \min_i \{|\lambda_i(\mathcal{L})|\}.$$

Proposition 3.2 Assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ is stable. The condition number of \mathcal{L} can be bounded below by,

$$\kappa_2(\mathcal{L}) \geq \frac{\max_i \operatorname{Re}(-\lambda_i(\mathbf{A}))}{\min_i \operatorname{Re}(-\lambda_i(\mathbf{A}))}. \quad (30)$$

Proof : From the properties of singular values and (28) we get,

$$\begin{aligned} \sigma_{max}(\mathcal{L}) &= \max_{x \in \mathbb{R}^{n^2}} \frac{\|\mathcal{L}y\|_2}{\|y\|_2} \geq \max_i |\lambda_i(\mathcal{L})| \geq 2 \max_i |\operatorname{Re}(\lambda_i(\mathbf{A}))|, \\ \sigma_{min}(\mathcal{L}) &= \min_{x \in \mathbb{R}^{n^2}} \frac{\|\mathcal{L}y\|_2}{\|y\|_2} \leq \min_i |\lambda_i(\mathcal{L})| \leq 2 \min_i |\operatorname{Re}(\lambda_i(\mathbf{A}))|. \end{aligned}$$

The above together with the assumption that \mathbf{A} is stable leads to,

$$\kappa_2(\mathcal{L}) = \frac{\sigma_{max}(\mathcal{L})}{\sigma_{min}(\mathcal{L})} \geq \frac{\max_i |\operatorname{Re}(\lambda_i(\mathbf{A}))|}{\min_i |\operatorname{Re}(\lambda_i(\mathbf{A}))|} = \frac{\max_i \operatorname{Re}(-\lambda_i(\mathbf{A}))}{\min_i \operatorname{Re}(-\lambda_i(\mathbf{A}))}. \quad \blacksquare \quad (31)$$

Note that the first inequality in (31) holds for any square matrix (i.e., the stability of \mathbf{A} is not required). Clearly, the bound (30) may underestimate the ill conditioning of the Lyapunov equation (26), but it is possible to compute or estimate this bound with reasonable computational work.

An exact formula for the condition number of \mathcal{L} is:

$$\kappa_2(\mathcal{L}) = \frac{\max_{\|\mathbf{Y}\|_F=1} \|\mathbf{A}\mathbf{Y} + \mathbf{Y}\mathbf{A}^T\|_F}{\operatorname{sep}(\mathbf{A}, -\mathbf{A})}.$$

However, this formula is not convenient for practical use, while (30) immediately provides a computable sufficient condition for the stable continuous-time Lyapunov equation (26) to be ill conditioned.

We also obtain an upper bound for $\kappa(\mathcal{L})$ which links the conditioning to the norm and the logarithmic norm of \mathbf{A} . First, we recall the logarithmic norm of \mathbf{A} [27] is defined as,

$$\mu(\mathbf{A}) = \lim_{h \rightarrow 0^+} \frac{\|\mathbf{I} + h\mathbf{A}\| - 1}{h}, \quad (32)$$

where $\|\cdot\|$ is some *induced matrix norm*. Note that the logarithmic norm may assume negative value, so it is not really a norm in the traditional sense. One important property of $\mu(\cdot)$ is

$$\mu(\mathbf{A}) = \inf\{\alpha : \|\mathbf{e}^{\mathbf{A}t}\| \leq e^{\alpha t}, t \geq 0\}, \quad (33)$$

it is this property that leads to its name: *logarithmic norm*.

If the matrix 2-norm is used, we have the following relation

$$\mu_2(\mathbf{A}) = \frac{1}{2} \lambda_{max}(\mathbf{A} + \mathbf{A}^*).$$

Our upper bound can be stated as follows.

Proposition 3.3 Assume $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mu_2(\mathbf{A}) < 0$. The condition number of \mathcal{L} can be bounded above by

$$\kappa_2(\mathcal{L}) \leq -\frac{2\sigma_{\max}(\mathbf{A})}{\lambda_{\max}(\mathbf{A} + \mathbf{A}^T)}. \quad (34)$$

Proof : The upper bound of $\|\mathcal{L}\|_2$ is easily available,

$$\|\mathcal{L}\|_2 \leq \|\mathbf{I} \otimes \mathbf{A}\|_2 + \|\mathbf{A} \otimes \mathbf{I}\|_2 = 2\|\mathbf{A}\|_2.$$

Note that for the last equality, we used the property that the singular values of the Kronecker product of two matrices ($\sigma(\mathbf{M} \otimes \mathbf{N})$) are the product of the singular values of each matrix ($\sigma(\mathbf{M})\sigma(\mathbf{N})$).

We also need an upper bound for $\|\mathcal{L}^{-1}\|_2$. Consider the following Lyapunov equation with special right hand side,

$$\mathbf{A}\mathbf{H} + \mathbf{H}\mathbf{A}^T = -\mathbf{I}, \quad (35)$$

it is well known that \mathbf{H} can be expressed by the following integral,

$$\mathbf{H} = \int_0^\infty e^{\mathbf{A}t} e^{\mathbf{A}^T t} dt.$$

It is proved in [8, 3] that for any induced matrix norm $\|\cdot\|$, the following relations hold,

$$\|\mathcal{L}^{-1}\| = \|\mathcal{L}^{-1}(\mathbf{I})\| = \|\mathbf{H}\|. \quad (36)$$

So we only need to bound $\|\mathbf{H}\|_2$. By (33) and the condition that $\mu_2(\mathbf{A}) < 0$ we get,

$$\|\mathbf{H}\|_2 \leq \int_0^\infty e^{\mu_2(\mathbf{A})t} e^{\mu_2(\mathbf{A}^T)t} dt \leq -\frac{1}{2\mu_2(\mathbf{A})} \quad (37)$$

Noting that $\mu_2(\mathbf{A}) = \frac{1}{2}\lambda_{\max}(\mathbf{A} + \mathbf{A}^T)$ and $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$ we prove the desired bound,

$$\kappa_2(\mathcal{L}) = \|\mathcal{L}\|_2 \|\mathcal{L}^{-1}\|_2 \leq -\frac{2\sigma_{\max}(\mathbf{A})}{\lambda_{\max}(\mathbf{A} + \mathbf{A}^T)}. \quad \blacksquare$$

The above proof can easily be generalized to any induced matrix norm $\|\cdot\|$. The only difference is that $\mu(\mathbf{A}) = \mu(\mathbf{A}^T)$ may not hold. We actually have the following,

Proposition 3.4 Assume $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mu(\mathbf{A}) + \mu(\mathbf{A}^T) < 0$. Then for any induced matrix norm $\|\cdot\|$, the following upper bound for $\kappa(\mathcal{L}) = \|\mathcal{L}\| \|\mathcal{L}^{-1}\|$ holds,

$$\kappa(\mathcal{L}) \leq -\frac{2\|\mathbf{A}\|}{\mu(\mathbf{A}) + \mu(\mathbf{A}^T)}. \quad (38)$$

It is easy to show that $\max\{Re(\lambda_i(\mathbf{A}))\} \leq \mu_2(\mathbf{A})$. We see from Proposition 3.3 that if \mathbf{A} has a large norm and a small magnitude logarithmic norm (or some eigenvalues with tiny real part), then the associated Lyapunov equation is likely very ill conditioned.

We can get rid of the condition $\mu(\mathbf{A}) + \mu(\mathbf{A}^T) < 0$ in Proposition 3.4 since this condition is needed only for estimating $\|\mathbf{H}\|$ in (37). From the proof we see an obvious bound that directly uses $\|\mathbf{H}\|$ is: $\kappa(\mathcal{L}) \leq 2\|\mathbf{A}\|\|\mathbf{H}\|$, but this bound is not informative since one has no idea what $\|\mathbf{H}\|$ can be. One possibility is to apply Proposition 3.1 to equation (35) to obtain the following result.

Proposition 3.5 *Assume that $\mathbf{A} \in \mathbb{R}^{n \times n}$ is stable with eigen decomposition (20). Then for any induced matrix norm $\|\cdot\|$, the following upper bound for $\kappa(\mathcal{L}) = \|\mathcal{L}\|\|\mathcal{L}^{-1}\|$ holds,*

$$\kappa(\mathcal{L}) \leq \frac{\kappa^2(\mathbf{X})}{\min Re(-\lambda_i)} \|\mathbf{A}\|. \quad (39)$$

Proof : Applying Proposition 3.1 to the Lyapunov equation (35) we get:

$$\|\mathbf{H}\| \leq \frac{\kappa^2(\mathbf{X})}{2 \min Re(-\lambda_i)}.$$

Together with (36) and the definition of $\kappa(\mathcal{L})$ we proved (39). ■

From the definition of logarithmic norm we easily see that bound (39) is weaker than (38).

Remark: A nice discussion on properties and bounds for the Lyapunov operator corresponding to the discrete-time Lyapunov equation may be found in [28], where a contour integral formula is used to obtain an upper bound for \mathcal{L}^{-1} . It seems that the idea used in [28] (introducing the stability radius defined by the resolvent of \mathbf{A}) can not be applied directly here, because the contour integral (see [17]) for the solution \mathbf{H} of the continuous-time Lyapunov equation (35) is,

$$\mathbf{H} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (j\omega\mathbf{I} - \mathbf{A})^{-1} (-j\omega\mathbf{I} - \mathbf{A}^T)^{-1} d\omega,$$

the integral range is infinite; while for the discrete-time Lyapunov equation, the solution formula has a finite integral range. We avoided this problem by introducing the logarithmic norm. Moreover, the stability radius is generally hard to compute, while the logarithmic norm $\mu_2(\mathbf{A})$ can be computed efficiently by existing software since it is half of the largest eigenvalue of a symmetric matrix $\mathbf{A} + \mathbf{A}^T$. Hence our bounds are computationally more practical.

4 Conclusions

In this paper we have presented two qualitative results about solutions to Lyapunov equations. We established eigenvalue decay rate bounds for the solution of the Lyapunov equation $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T$. These bounds apply to diagonalizable \mathbf{A} and hence are more general than those of Penzl. Moreover, when \mathbf{A} has a moderately conditioned eigen-system, the bounds are fairly sharp. The phenomenon of rapidly decaying eigenvalues that is often observed in practice is both motivation and justification for the development of iterative methods that compute low rank approximate solutions in factored form.

An upper bound on the norm of the solution to the Lyapunov equation was derived in the case of diagonalizable \mathbf{A} . Then several new bounds for the condition number of the Lyapunov operator were also developed. These bounds are relatively easy to verify and seem to give a good estimates for the conditioning of the related Lyapunov equations.

References

- [1] A. C. Antoulas, *Lectures on the approximation of large-scale dynamical systems*, SIAM Press, to appear.
- [2] A. C. Antoulas, D. C. Sorensen, Y. Zhou *On the decay rate of Hankel singular values and related issues*, Systems and Control Letters, to appear.
- [3] R. Bhatia, *A note on Lyapunov equation*, Linear Algebra and Its Applications, 259 (1997) pp. 71-76.
- [4] A. R. Ghavimi and A. J. Laub, *An implicit deflation method for ill-conditioned Sylvester and Lyapunov equations*, Int. J. Control, 61 (1995) pp. 1119-1141.
- [5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, 3rd ed., 1996.
- [6] E. Grimme, D. C. Sorensen and P. Van Dooren, *Model reduction of state space systems via an implicitly restarted Lanczos method*, Numerical Algorithms, 12 (1996) pp. 1-32.
- [7] T. Gudmundsson, A. J. Laub, *Approximate Solution of Large Sparse Lyapunov Equations*, IEEE Trans on AC, AC-39 (1994) pp. 1110-1114.
- [8] G. Hewer, C. Kenney, *The sensitivity of the stable Lyapunov equation*, SIAM J. Control Optim., 26 (1988) pp. 321-344.
- [9] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, 1996.

- [10] D. W. Peaceman and H. H. Rachford, *The numerical solutions of parabolic and elliptic differential equations*, J. SIAM, {now: SIAM J. Appl. Math.}, v3 No.1 (1955) pp. 28-41.
- [11] A. Hodel, K. Poolla, B. Tenison, *Numerical solution of the Lyapunov equation by approximate power iteration*, Linear Algebra Appl., 236 (1996) pp. 205-230.
- [12] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge Univ. Press, 1991.
- [13] R. A. Horn, F. Kittaneh, *Two applications of a bound on the Hadamard Product with a Cauchy Matrix*, Electron. J. Linear Algebra 3, (1998) pp. 4-12.
- [14] D. Hu and L. Reichel, *Krylov-subspace methods for the Sylvester equation*, Linear Alg. and Appl., 172 (1992) pp. 283-313.
- [15] I. M. Jaimoukha and E. M. Kasenally, *Krylov Subspace methods for solving Large Lyapunov Equations*, SIAM J. Numer. Anal., 31 (1994) pp. 227-251.
- [16] A. J. Laub, *A Schur method for solving algebraic Riccati equations*, IEEE Trans. AC, AC-24 (1979) pp. 913-921.
- [17] P. Lancaster, *Explicit Solution of Linear Matrix Equations*, SIAM Review, 12 (1970) pp. 544-566.
- [18] C. Lee, S. Lee, *On the estimation of solution bounds of the generalized Lyapunov equations and the robust root clustering for the linear perturbed systems*, INT. J. Control, 74, (2001) pp. 996-1008.
- [19] J. Li, *Model Reduction of Large Linear Systems via Low Rank System Gramians*, PhD thesis, Dept. of Math, MIT, 2000.
- [20] J. Li, J. White, *Low rank solution of Lyapunov equations*, Leslie Fox Prize paper, 2001.
- [21] T. Mori, N. Fukuma, M. Kuwahara, *Explicit Solution and Eigenvalue Bounds in the Lyapunov Matrix Equation*, IEEE Trans on AC, AC-31 (1986) pp. 656-658.
- [22] T. Penzl, *Eigenvalue Decay Bounds for Solutions of Lyapunov Equations: The Symmetric Case*, Systems and Control Letters, 40 (2000) pp. 139-144.
- [23] T. Penzl, *A Cyclic Low-Rank Smith Method for Large Sparse Lyapunov Equations*, SIAM J. Sci. Comput., 21 (2000) pp. 1401-1418.

- [24] A. Ruhe, D. Skoogh, *Rational Krylov algorithm for eigenvalue computation and model reduction*, Applied Parallel Computing (PARA in Umeå, 1998), pp. 491-502, Springer Verlag, 1998.
- [25] Y. Saad, *Numerical Solution of Large Lyapunov Equations*, Proceedings of the Int. symposium MTNS-89, Vol. III, M. Kaashoek, J.V. Schuppen, A. Ran, eds., Birkhäuser, Boston, MA, (1990) pp. 503-511.
- [26] G. W. Stewart, *Error and Perturbation Bounds for Subspaces Associated with Certain Eigenvalue Problems* SIAM review, 15 (1973) pp. 727-764.
- [27] T. Ström, *On logarithmic norms*, SIAM J. Num. Anal., 12 (1975) pp. 741-753.
- [28] M. Tippett, S. Cohn, R. Todling, D. Marchesin, *Conditioning of the stable discrete-time Lyapunov operator* SIAM J. Matrix Anna. Appl., 22 (2000) pp. 56-65.
- [29] E. L. Wachspress, *Extended Application of Alternating Direction Implicit Iteration Model Problem Theory*, Journal of SIAM, 11 (1963) pp. 994-1016.

A Derivation of the Stein equation: general complex shifts

Consider general continuous-time Lyapunov Equation (CLE):

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{M} = \mathbf{0}, \quad (40)$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ be stable, $\mathbf{M} \in \mathbb{R}^{n \times n}$ is positive semidefinite.

Equation (40) is equivalent to the following two shifted equations: $\forall \tau \in \mathbb{C}_-$,

$$(\mathbf{A} - \tau\mathbf{I})\mathbf{P} + \mathbf{P}(\mathbf{A} + \tau^*\mathbf{I})^* + \mathbf{M} = \mathbf{0}, \quad (41)$$

$$(\mathbf{A} + \tau^*\mathbf{I})\mathbf{P} + \mathbf{P}(\mathbf{A} - \tau\mathbf{I})^* + \mathbf{M} = \mathbf{0}. \quad (42)$$

Performing $(\mathbf{A} + \tau^*\mathbf{I}) \times (41) - (\mathbf{A} - \tau\mathbf{I}) \times (42)$ we get:

$$\begin{aligned} & (\mathbf{A} + \tau^*\mathbf{I})(\mathbf{A} - \tau\mathbf{I})\mathbf{P} + (\mathbf{A} + \tau^*\mathbf{I})\mathbf{P}(\mathbf{A} + \tau^*\mathbf{I})^* + (\mathbf{A} + \tau^*\mathbf{I})\mathbf{M} \\ & - (\mathbf{A} - \tau\mathbf{I})(\mathbf{A} + \tau^*\mathbf{I})\mathbf{P} - (\mathbf{A} - \tau\mathbf{I})\mathbf{P}(\mathbf{A} - \tau\mathbf{I})^* - (\mathbf{A} - \tau\mathbf{I})\mathbf{M} = \mathbf{0}. \end{aligned}$$

Hence

$$(\mathbf{A} + \tau^*\mathbf{I})\mathbf{P}(\mathbf{A} + \tau^*\mathbf{I})^* - (\mathbf{A} - \tau\mathbf{I})\mathbf{P}(\mathbf{A} - \tau\mathbf{I})^* + 2 \operatorname{Re}(\tau) \mathbf{M} = \mathbf{0}. \quad (43)$$

By introducing the Cayley operator

$$s_\tau(\mathbf{A}) := (\mathbf{A} + \tau^*\mathbf{I})^{-1}(\mathbf{A} - \tau\mathbf{I}),$$

we arrive at the Stein Equation (with complex shift τ):

$$\mathbf{P} - s_\tau(\mathbf{A})\mathbf{P}s_\tau^*(\mathbf{A}) + 2 \operatorname{Re}(\tau) (\mathbf{A} + \tau^*\mathbf{I})^{-1}\mathbf{M}(\mathbf{A} + \tau^*\mathbf{I})^{-*} = \mathbf{0}. \quad (44)$$

In [23, 19] the following Cayley transform was used for the complex shifts case,

$$s_\tau(\mathbf{A}) := (\mathbf{A} + \tau\mathbf{I})^{-1}(\mathbf{A} - \tau\mathbf{I}). \quad (45)$$

To guarantee the necessary condition for convergence of the Smith or ADI iteration, namely $\|s_\tau(\mathbf{A})\| < 1$, the shifts can only be real or complex conjugate pairs. This choice is based on computational efficiency consideration. While for general complex shifts, we need to use formula (44). Another feature of the above derivation is that it is closely related to the ADI method, the only thing one needs to modify is to add subindex to the unknown \mathbf{X} and the shift τ ; the derivation proceeds as follows:

$$(\mathbf{A} - \tau_i\mathbf{I})\mathbf{P}_{i+1/2} + \mathbf{P}_{i+1}(\mathbf{A} + \tau_i^*\mathbf{I})^* + \mathbf{M} = \mathbf{0}, \quad (46)$$

$$(\mathbf{A} + \tau_i^*\mathbf{I})\mathbf{P}_{i+1/2} + \mathbf{P}_i(\mathbf{A} - \tau_i\mathbf{I})^* + \mathbf{M} = \mathbf{0}. \quad (47)$$

Then $(\mathbf{A} + \tau_i^*\mathbf{I}) \times (46) - (\mathbf{A} - \tau_i\mathbf{I}) \times (47)$ leads to:

$$\begin{aligned} & (\mathbf{A} + \tau_i^*\mathbf{I})(\mathbf{A} - \tau_i\mathbf{I})\mathbf{P}_{i+1/2} + (\mathbf{A} + \tau_i^*\mathbf{I})\mathbf{P}_{i+1}(\mathbf{A} + \tau_i^*\mathbf{I})^* + (\mathbf{A} + \tau_i^*\mathbf{I})\mathbf{M} \\ & - (\mathbf{A} - \tau_i\mathbf{I})(\mathbf{A} + \tau_i^*\mathbf{I})\mathbf{P}_{i+1/2} - (\mathbf{A} - \tau_i\mathbf{I})\mathbf{P}_i(\mathbf{A} - \tau_i\mathbf{I})^* - (\mathbf{A} - \tau_i\mathbf{I})\mathbf{M} = \mathbf{0}. \end{aligned}$$

Hence

$$(\mathbf{A} + \tau_i^*\mathbf{I})\mathbf{P}_{i+1}(\mathbf{A} + \tau_i^*\mathbf{I})^* - (\mathbf{A} - \tau_i\mathbf{I})\mathbf{P}_i(\mathbf{A} - \tau_i\mathbf{I})^* + 2 \operatorname{Re}(\tau_i) \mathbf{M} = \mathbf{0}. \quad (48)$$

Let

$$s_{\tau_i}(\mathbf{A}) := (\mathbf{A} + \tau_i^*\mathbf{I})^{-1}(\mathbf{A} - \tau_i\mathbf{I}),$$

we arrive at the Smith iteration (with complex shifts τ_i):

$$\mathbf{P}_{i+1} - s_{\tau_i}(\mathbf{A})\mathbf{P}_i s_{\tau_i}^*(\mathbf{A}) + 2 \operatorname{Re}(\tau_i) (\mathbf{A} + \tau_i^*\mathbf{I})^{-1}\mathbf{M}(\mathbf{A} + \tau_i^*\mathbf{I})^{-*} = \mathbf{0}. \quad (49)$$

So the two step ADI actually performs the one step Smith iteration. We used formula (44) (49) in Section 2.